Journal on Interactive Systems, 2024, 15:1, doi: 10.5753/jis.2024.4350

A Systematic Mapping Study about Technologies for Hedonic Aspects Evaluation in Text-based Chatbots

Pamella A. de L. Mariano [**Federal University of Paraná** | ***pamella@ufpr.br***]
Ana Carolina R. de Souza [**Federal University of Paraná** | ***anacrossidesouza@gmail.com***]
Guilherme C. Guerino [**State University of Paraná** | ***guilherme.guerinosi@gmail.com***]
Ana Paula Chaves [**Northern Arizona University** | ***ana.chaves@nau.edu***]
Natasha M. C. Valentim [**Federal University of Paraná** | ***natasha@inf.ufpr.br***]

Department of Informatics, Federal University of Paraná, R. Cel. Francisco H. dos Santos, 100 - Centro Politécnico, Jardim das Américas, Curitiba - PR, 81531-980, Brazil

Received: 04 April 2024 • **Accepted:** 02 August 2024 • **Published:** 18 August 2024

Abstract: Many studies present and evaluate daily-use systems ranging from information to conversational systems. Chatbots, either text-based or voice-based, have attracted the attention of researchers. In particular, User eXperience (UX) has been pointed out as one of the chatbot's leading aspects of evaluation involving pragmatic and hedonic aspects. Pragmatic aspects deal with the usability and efficiency of the system, while hedonic aspects consider aspects related to the originality, innovation, beauty of the system, and the user's psychological well-being. Even with existing research on usability evaluation and human-computer interaction within conversational systems, there is a clear shortfall in studies specifically addressing the hedonic aspects of user experience in chatbots. Therefore, this paper presents a Systematic Mapping Study that investigates various UX evaluation technologies (questionnaires, methods, techniques, and models, among others), focusing on the hedonic aspect of chatbots. We focused on studies with chatbots that are activated by text, although they may be able to display click interactions, videos, and images in addition to the text modality. We discovered 69 technologies to evaluate hedonic aspects of UX in chatbots, and the most frequent aspect found is the General UX . Our study provides relevant data on the research topic, addressing the specific characteristics of human-chatbot interaction, such as identity and social interaction. Moreover, we highlight gaps in the hedonic aspect evaluation in chatbots, such as a few works investigating the assessment of user emotional state.

Keywords: Chatbots, User Experience, Systematic Mapping Study

1 Introduction

Chatbots mimic the unique human action of conversation [Ruane *et al.*, 2021] and are defined as online conversational systems where humans and computers interact using natural language [Jia and Jyou, 2021] by text or voice [Veglis *et al.*, 2019]. Unlike voice-activated smart assistants, chatbots are often text-activated with additional interactional resources such as point-and-click interactions, images, and videos [Candello and Pinhanez, 2016]. Additionally, chatbots can be powered with Artificial Intelligence (AI) to serve various purposes, including imitating human chat or performing various tasks. One example of a task-oriented chatbot is the study by Mudofi and Yuspin [2022] that addresses using chatbots in financial institutions to perform credit analysis and customer service.

Almost all of the chatbot developers in Brazil work with text-based chatbots (97%), while a smaller number of developers work with voice chatbots (68%) [mobiletime, 2022]. Nearly 40% of internet users worldwide prefer interacting with chatbots than virtual agents, and with major industries including retail and healthcare turning to digital technology, chatbots will likely increase in popularity moving forward [Yuen, 2022]. There are many advantages to using chatbots to perform services, as they help to reduce costs and increase efficiency in processes [Telner, 2021]. Chatbots are quick to implement and allow customization [Mudofi and Yuspin, 2022], which justifies their increasing popularity and stresses the need for well-designed, high-quality agents. Moreover, the development of Large Language Models like chatGPT has revolutionized chatbot technology, enhancing their conversational abilities and driving an increase in usage and popularity among users looking for more engaging and personalized interactions [Brown *et al.*, 2020].

However, chatbots may suffer from quality issues, such as the lack of conversational skills and social intelligence, which may impair the User eXperience (UX). Lack of quality concerns can decrease the user's interest in interacting with the chatbot. Additionally, social intelligence is crucial to engage the user in interesting and relevant conversations [Skjuve *et al.*, 2019].

One way to improve the quality of chatbots is by providing an appropriate UX. According to ISO [2019], UX is the user's response and perception when using a system, product, or service. These perceptions may include emotions, beliefs, preferences, perceptions, amenities, behaviors, and achievements that may occur before, during, and after use.

In line with the model for attractive software systems with good UX, proposed by Hassenzahl *et al.* [2000], software is described using different quality dimensions divided into two groups: pragmatic and hedonic quality. The first deals with the usability and efficiency of the system, while the second considers aspects related to originality, innovation, and beauty. The hedonic aspects of UX are also related to the user's psychological well-being [Hassenzahl, 2004]. In this paper, we only considered the hedonic aspect of UX, as it is the aspect that addresses the user's emotional well-being, mostly because products that go beyond the user's hedonic needs increase pleasure and the loyalty of the customer, more than satisfaction alone does [Chitturi *et al.*, 2008].

Current literature presents studies on technologies (questionnaires, methods, techniques, and models, among others [Santos *et al.*, 2012]) that assess the quality of conversational systems. Guerino and Valentim [2020] mapped the usability and UX assessment technologies to evaluate conversational systems that specifically use voice. Rapp *et al.* [2021] investigated human-computer interaction and chatbots, such as whether and why people accept and use this technology. Ren *et al.* [2019] investigated technologies for evaluating chatbots focusing specifically on the usability criteria. However, there is a gap in what assessment technologies can be used to evaluate the hedonic aspects of UX for text-based chatbots.

Therefore, this research is guided by the following question: "What technologies are used to evaluate hedonic aspects of UX in text-based chatbots?". To answer this question, we performed a Systematic Mapping Study (SMS) to investigate the existing literature on the subject.

An SMS is needed to unveil and connect practices and outcomes related to a certain research topic. Initially, we performed an SMS to identify and characterize technologies that evaluate hedonic aspects of UX for text-based chatbots. We found 26 papers published between 2017 and 2021 containing 29 different technologies. The User Experience Questionnaire (UEQ) [Laugwitz *et al.*, 2008] emerged as the predominant technology, while trust was the most frequently evaluated hedonic aspect. These results were published at IHC 2023 [De Souza *et al.*, 2024].

In recent years, the evolution of chatbots has been remarkable, driven by advances in artificial intelligence (AI) technologies such as natural language processing (NLP) and natural language understanding (NLU). These advancements have allowed chatbots to become more sophisticated and capable of meaningfully interacting with users. Furthermore, the rapid evolution of chatbots can be attributed to consumers' growing adoption of platforms that favor conversational interaction. Companies have recognized the potential of chatbots in several areas, leading to an increase in interest and research in the area [Følstad *et al.*, 2021].

Because the topic has evolved very quickly in recent years, we extended the previous SMS by analyzing papers published between 2021 and 2023. A total of 26 new papers were identified, revealing the presence of 40 new technologies about this research theme. The Chatbot Usability Questionnaire (CUQ) [Holmes *et al.*, 2019] was the most prevalent technology and the predominant hedonic aspect of these technologies was the general UX.

Our study provides relevant data on the research topic, addressing the specific characteristics of human-chatbot interaction, such as identity and social interaction. In this extension of SMS, we will trace the evolution of research into UX assessment technologies used to evaluate text chatbots. Additionally, we will discuss possible reasons for this evolution, highlighting emerging trends, identifying research gaps, and providing an analysis of the motivations behind the continued growth and diversification of UX evaluation approaches for text chatbots.

Moreover, we highlight gaps in the hedonic aspect evaluation in chatbots, such as a few works investigating the assessment of user emotional state. The SMS extension allows us to identify that the trends in SMS Part 1, in general, are being maintained in SMS Part 2. This reinforces the gaps identified in the first part, such as: the literature lacks empirical studies to evaluate the reliability and consistency of technologies to evaluate the hedonic aspects of UX for text chatbots. This has important implications for the validity of the results obtained by these technologies. There is also a lack of technologies that address the specific characteristics of human-chatbot interaction, indicating that particular aspects of chatbots, such as identity and social interaction, are not adequately considered when determining user experience. These findings can guide both researchers and professionals in the field (and can help in choosing appropriate evaluation methods and developing more effective and attractive text chatbots for users).

Section 2 presents the theoretical background and the related work. Section 3 draws the research methodology used to conduct this SMS. In Section 4, we present our quantitative and qualitative results, which are discussed in light of current literature in Section 5. Finally, Sections 6 and 7 present the limitations, conclusions, and future work.

2 Background and Related Work

Chatbots can be referred to by different terms such as *dialog system* and *chatterbot* [Shawar and Atwell, 2007]. Chatbots combine conversation with visual elements [Höhn and Bongard-Blanchy, 2021] and these systems are designed to simulate intelligent communication via text or speech [Dahiya, 2017]. Currently, chatbots facilitate various business processes, such as situations related to customer service and personalization, due to their accessibility, low cost, and ease of use for the end consumer [Przegalinska *et al.*, 2019].

Recently, there has been an increase in investment in the development of chatbots, virtual agents, and personal assistants. This growth has also attracted the interest of scholars in such systems and their various applications, such as Ashktorab *et al.* [2019] which presented a chatbot that helps in customer service in the help-desk service. Therefore, assessing the quality of chatbots becomes crucial.

Identifying which technologies are appropriate for chatbot assessment remains a challenge. Some studies aim to support this task by systematically mapping the assessment technologies. For example, Guerino and Valentim [2020] investigate conversational systems that use the human voice to perform actions. The study reports 31 assessment technologies to evaluate chatbot usability and UX. The study has searched the following virtual libraries: Scopus, IEEEXplore, ACM Digital Library, and Engineering Village. The results found that the assessment technologies are mainly created for a particular study without empirical evaluation. Most of the identified chatbots focused on assisting users in daily tasks.

Mafra *et al.* [2024] created the U2Chatbot inspection checklist — developed through a systematic literature review process that helped to identify relevant quality attributes from previous studies — is a tool designed for the evaluation and identification of defects in text-based chatbots. Composed of 107 items that cover various quality attributes related to usability and user experience, the checklist was created to be more comprehensive than existing tools, ensuring that crucial aspects affecting chatbot performance are not overlooked.

Ren *et al.* [2019] conducted an SMS to identify the use of chatbots and their application as a human-computer interaction technique focusing on evaluating chatbot usability. The study has searched the main scientific databases (Scopus, ACM Digital Library, IEEE Xplorer, SpringerLink, and Science Direct) and found that most technologies to evaluate chatbot usability elect a group of users to use the system freely or perform certain tasks and then measure satisfaction through the System Usability Scale (SUS) [Brooke *et al.*, 1996] questionnaire.

Rapp *et al.* [2021] evaluated 83 studies to investigate how users interact with text-based chatbots. The findings reveal that trust, engagement, and satisfaction are important aspects of user experience, and Wizard of Oz (WoZ) and fully developed prototypes are the most common tools to explore user experiences, attitudes, and behaviors.

Tubin *et al.* [2022] analyzed how to evaluate the experience with conversational agents to offer a more realistic and natural user experience. They focused on identifying how the user experience is measured when interacting with agents. For the authors, evaluating the user experience at different moments and applying combined methods to understand aspects such as the participant's feelings and behaviors is necessary.

The studies discussed above have their particularities, such as the different types of *chatbots*, the relationship with the HCI, and the methods, techniques, and technologies to evaluate conversational systems, whether through usability or user experience. In Guerino and Valentim [2020], the mapping performed considers both usability and UX, however, it only evaluates conversational voice systems and does not distinguish between the hedonic and pragmatic aspects of UX. In the study of Rapp *et al.* [2021], the HCI is considered as a whole, without making cuts for the user experience when using a chatbot. Although Mafra *et al.* [2024] have developed an inspection checklist for text-based chatbots, their main focus is on usability and user experience (UX) in general, without exclusively addressing the hedonic aspects of UX. Ren *et al.* [2019] present a general SMS without delimiting the scope of the nature of chatbots or regarding the aspect of quality considered in the mapping. Finally, Tubin *et al.* [2022], despite considering the UX when using conversational agents, do not distinguish the method used for data entry in the evaluated conversational systems. In this paper, we aim to fill the gaps in the literature by shedding light on text-based chatbots and examining UX technologies that focus on hedonic quality. Moreover, we investigated whether these UX evaluations consider aspects of the user's mental and emotional health.

3 Systematic Mapping Study

The methodology used in this paper is based on a secondary study, which reviews all primary studies related to a specific research question and aims at integrating evidence related to a specific research question [Kitchenham and Charters, 2007]. One of the types of secondary study is SMS.

An SMS aims to ascertain, qualify, and relate relevant research on a defined subject [Kitchenham and Charters, 2007]. Grounded on Kitchenham and Charters [2007] methodological steps, this SMS structure is divided into three phases:

- Planning: in this phase, we defined the mapping protocol, research questions, data sources, search string, and the paper selection's inclusion and exclusion criteria;
- Execution: we carried out the searches in the data sources, selected and extracted the primary studies, and conducted the data analysis;
- Reporting: as the last step, we presented the quantitative and qualitative results obtained from the analysis.

These phases are detailed in the following subsections.

3.1 Phase 1: Planning

The goal of this SMS was defined based on the *Goal-Question-Metric* (GQM) [Basili and Rombach, 1988] paradigm (see Table 1). The main research question is: "What technologies are used to evaluate hedonic aspects of UX in text-based chatbots?". We answered this question by defining the subquestions in Table 2.

Table 1. Purpose of the SMS.

Analyse	scientific publications
For the purpose of	characterize
Regarding	UX assessment technologies focusing on hedonic aspects of text-based chatbots
From the point of view of	HCI researchers
In the context of	primary sources available on ACM Digital Library and IEEE Xplore

This research was carried out from the ACM[1] and IEEEXplore[2] virtual libraries through an advanced search engine. These libraries provide a competent search engine, allow the use of similar terms in the string, and provide several papers in the HCI area.

We used the PICOC method [Kitchenham and Charters, 2007] to define the search string, presented in Table 3. The acronym refers to Population (P), Intervention (I), Comparison (C), Outcome (O), and Context (C); however, we will focus on the PIO since the remainder of the concepts are used when the SMS compares the results among each other. Therefore, PIO was established as: (P)opulation: Chatbots; (I)ntervention: Technologies to evaluate the hedonic aspects of UX in text-based chatbots; and (O)utcome: UX evaluation.

[1] https://dl.acm.org/

[2] https://ieeexplore.ieee.org/

Table 2. Sub-questions and possible answers.

Subquestions	Possible answers
SQ1. What hedonic aspects of UX does the technology assess?	Vary from paper to paper. Examples can be immersion, fatigue, and pleasure, among others.
SQ2. Is the assessment technology specific to chatbots?	**Specific:** UX assessment technology specific to chatbots. **Generic:** technology is not restricted to specific types of software.
SQ3. Was the assessment technology created for the study?	**Existing:** evaluation uses existing technology. **Created:** technology was created for the study and described in the paper.
SQ4. How were the participant's responses collected?	Vary from paper to paper. Verify how the user's feedback was captured, for example, using a Likert scale or checklist, etc.
SQ5. What is the composition of the assessment technology?	Vary from paper to paper. Extract attributes of the technologies, such as the questions and whether the technology is a questionnaire or interview, etc.
SQ6. Does the assessment technology extract quantitative or qualitative data?	**Quantitative:** the analysis uses quantitative methods. **Qualitative:** the analysis uses qualitative methods. **Mixed:** the analysis uses both qualitative and quantitative methods.
SQ7. What is the chatbot application?	The answers are subjective and identified during the readings. Examples: health, or education, among others.
SQ8. Was the chatbot created for a specific group? Which one?	**Yes**, it assists a specific group such as blind, deaf, and elderly users, etc. **No**, it is not intended for a specific group.
SQ9. Is the chatbot of a specific type? Which one?	**Yes, task-oriented:** helps users perform a task or solve a problem. **Yes, conversation-oriented:** holds a conversation with humans or establishes a relationship with them. **Yes, both task and conversation-oriented**. **No**, it has an undefined purpose.
SQ10. How was the chatbot evaluated?	The answers are subjective and were identified during the readings. Examples are experiment and observation.
SQ11. Has the assessment technology been empirically evaluated?	**Yes**, the paper carried out an empirical evaluation of the assessment technology. **No**, the technology was not empirically evaluated.
SQ12. Does the UX assessment consider aspects of the user's emotional state?	**Yes**, it considers aspects of the user's mental state. **No**, these aspects are not considered.
SQ13. How was the user's mental state assessed?	The answers are subjective and vary from paper to paper. Examples are interviews or questionnaires, among others.

The inclusion criteria are (I1) publications presenting technologies to evaluate hedonic aspects of UX for text-based chatbots and (I2) publications describing experimental studies about the evaluation of hedonic aspects of UX for text-based chatbots. The exclusion criteria are (E1) publications that do not meet the inclusion criteria; (E2) publications in languages other than the ones understood by the authors (English and Portuguese); (E3) publications for which the full text was not available to the authors; (E4) publications that are part of the gray literature, such as technical reports and work in progress; and (E5) duplicated publications.

3.2 Phase 2: Execution

The initial search was conducted in October 2021 (Part 1 - pt1) , with three researchers participating in defining the protocol, executing the filters, and extracting information. In June 2023, we conducted an extension of the SMS (Part 2 - pt2), with two researchers participating in executing the filters and extracting information. Having two or more researchers involved in the SMS process is necessary to preserve the research consistency and to reduce biases [Kitchenham and Charters, 2007].

Both parts of the SMS followed the same selection procedures: after submitting the search string to the search engines, the researchers filtered the studies by reading the title and abstract and evaluating them against the inclusion and exclusion criteria separately (1st filter). If there was a divergence in the process of inclusion and exclusion, we attempted to find a term of agreement. If agreement could not be reached, we adopted a conservative approach and escalated the paper to the second filter. Whenever a paper was rejected, we recorded a justification.

For the 2nd filter, the first author read and classified the papers and extracted the relevant information. Then, the other researchers reviewed the excluded papers and their justifications, the included papers, and their data extraction outcomes. To follow the same procedure, we recorded a justification for the excluded papers. We used the collaborative tool Porifera (https://porifera.app.br/ [Campos *et al.*, 2022]) to support this process. In SMS Part 1, for the 1st filter, the Fleiss Kappa among the researchers was 0.501. This value is considered moderate, according to Altman [1990]. In the 2nd filter, the Fleiss Kappa was 0.7991, which is considered good. In SMS Part 2, for the 1st filter, the Cohen Kappa was moderate (0.4611), according to Altman [1990] while the same index was considered good for the second filter (0.6633).

As presented in Table 4, the initial SMS started the filtering process with 630 papers gathered from the search engines. Out of them, 91 were selected after the 1st filter, and 26 pa-

Table 3. Terms and elements of the Search String

Population	(“chatbot*” OR “conversational agent*” OR “chatterbot” OR “artificial conversational entity” OR “ conversational interface” OR “conversational system” OR “conversation system” OR “dialogue system” OR “conversational user interface” OR “conversational UI”)	AND
Intervention	(“tool” OR “framework” OR “technique” OR “method” OR “guideline” OR “pattern” OR “metric” OR “approach” OR “inspection” OR “heuristic” OR “methodology”)	AND
Outcome	(“user experience” OR “UX”) AND (”evaluation” OR ”assessment” OR ”measure” OR ”measurement”)	AND

pers (cited in Table 6, 7 and 8) were selected after the 2nd filter.

In the SMS extension (see Table 5), we started the filtering process with 769 papers gathered from the search engines. Out of them, 104 were selected after the 1st filter, and 26 papers (cited in Table 6, 7 and 8) were selected after the 2nd filter.

The data extraction focused on obtaining answers for each research sub-questions (Table 2). By doing so, we ensure that we apply the same criteria for data extraction to all the selected papers. We also collected the study’s metadata, such as the year and place of publication. The outcomes for this step are presented in Subsection 3.3.

Table 4. Paper Selection

Source	# papers returned	# selected after 1st filter	# selected after 2nd filter
IEEExplore	8	3	1
ACM Digital Library	622	88	25
Total	630	91	26

Table 5. Paper Selection - SMS Extension

Source	#papers returned	# selected after 1st filter	# selected after 2nd filter
IEEExplore	383	31	11
ACM Digital Library	386	73	15
Total	769	104	26

The data analysis was based on descriptive statistics and data visualization. We used a Google Sheets document to support the data visualization and interpretation presented in the following sections.

3.3 Phase 3: Reporting

Our analysis shows that assessing the hedonic aspects of UX in text-based chatbots is a recent research topic. The year of publication of the selected papers ranges from 2017 to 2023. Additionally, as Figure 1 depicts, the number of published papers addressing this topic has increased. The year 2017 has the fewest number of publications, while 2021 has the highest number (19 studies).

Figure 2 presents the publications conferences found in this SMS. The conference with the highest number of publications is the ACM Conference on Human Factors in Computing Systems (CHI), with ten publications, followed by ACM Designing Interactive Systems (DIS) with 3 publications. International Conference on Mobile Human-Computer Interaction (Mobile HCI), International ACM Conference on Conversational User Interfaces (CUI), Nordic Conference on Human-Computer Interaction (NordiCHI), International Conference on Mobile and Ubiquitous Multimedia (MUM), and the ACM International Conference on Intelligent User Interface (IUI) all with two publications. Other eighteen conferences appear in the results with only one publication each.

Regarding journals, only one appears three times (ACM on Human-Computer PACMHCI) and ACM on Human-Computer Interaction (HCI) appears two times. Other six journals appear in the results with only one publication each (Figure 3). The full list of conferences and journals can be found in the technical report [Souza *et al.*, 2023] and technical report Part 2 [Mariano *et al.*, 2024].

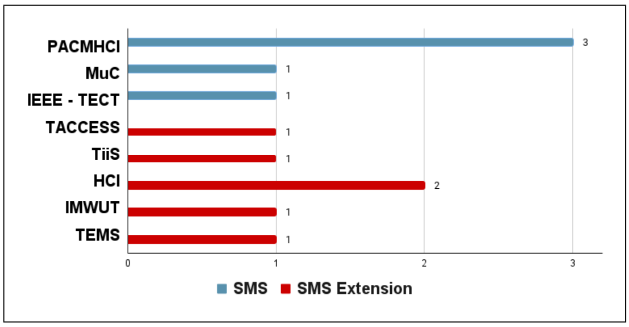


Figure 3. Journals

4 Results

Our main research question is “What technologies are being used to evaluate hedonic aspects of UX in text-based chatbots?”. In the initial SMS, we found 29 different technologies for this purpose. The most applied technology was the UEQ [Laugwitz *et al.*, 2008] (3 papers), followed by User Experience Questionnaire - Short (UEQ-S) [Schrepp *et al.*, 2017] (2 papers).

In the SMS Extension (Part 2), we found 40 different technologies used for this purpose. The most applied technology was the Chatbot Usability Questionnaire (CUQ) [Holmes *et al.*, 2019] (5 papers), followed by UEQ [Laugwitz *et al.*, 2008] (3 papers) and Technology Acceptance Model - TAM (2 papers).

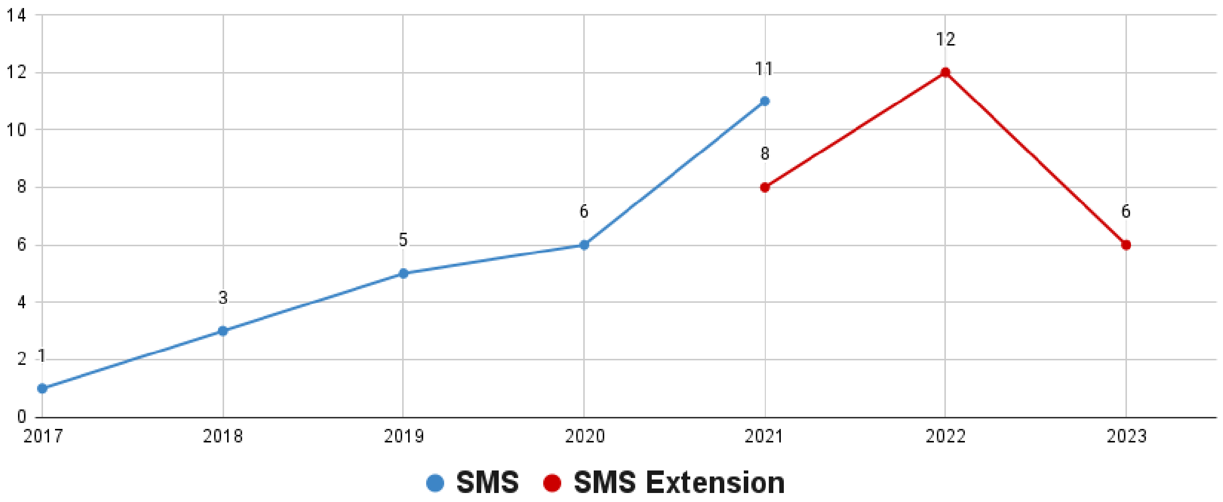


Figure 1. Publication years

Tables 6, 7 and 8 present a list of the identified UX evaluation technologies and the associated publications, besides identified hedonic aspects of UX. Moreover, a summary of quantitative results of sub-questions SQ2, SQ3, SQ6, SQ8, SQ9, SQ11 and SQ13 is presented in Table 9. The sub-questions SQ1, SQ4, SQ5, SQ7, SQ10, and SQ12 are qualitative or have many response options. Therefore, their results are only presented in the subsections below. All the results of sub-questions can be found in the technical report [Souza *et al.*, 2023] and technical report Part 2 [Mariano *et al.*, 2024].

4.1 SQ1. Hedonic aspects of UX that technologies assess

In the 1st part of this SMS , we identified 66 hedonic aspects of UX, listed below along with the respective number of studies: Trust (6); Enjoyment, Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, Novelty, and Engagement (5 each); Interest (3); Likeable, General UX, Easy to Report, Intention to Reuse, Fun, Frustration, Anxiety, Social Presence, Humanity, and Privacy (2 each); Pressure, Tension, Effort, Motivation, Attitude, Enjoyable, Privacy Intrusive, Diversity, Control, Feedback, Understanding, Difficulty, Expectation, Intimacy, Self-reflection, Self-awareness, Impressions, Psychological Well-being, Attention, Intention to Use, Adaptability, Sociability, Social Influence, Interpretation, Psychological Impact, Perceptions of Social Disclosure, Revealing Emotional Expression, Usefulness of Emotional Expression, Naturalness, Affection, Happiness, Sadness, Anger, Surprise, Tranquility, Vigor, Discomfort, Well-being, Empathy, Appreciation, Emotional Support, Emotion Perception, Expression of Emotion, Social Support, Commitment, and Unmet Expectations (1 each).

The most frequent aspect is trust. For example, Fadhil *et al.* [2018] apply a self-designed questionnaire to evaluate mental and physical well-being based on pleasure, attitude, and trust. Examples of the questions are “I enjoyed chatting with the conversational agent during the interaction” (pleasure), “I found the dialog with the conversational agent to be realistic” (estimate for attitude), and “The agent asked very personal questions” (trust).

In the SMS Extension (Part 2), we identified 132 hedonic aspects of UX, listed below along with the respective number of studies: General UX (15); Satisfaction, Perceived Usefulness, self-awareness, Intent to use, mental wellbeing (3 each); Novelty, user acceptance, perceived ease of use, Focused Attention, Perceived Usability, Aesthetic Appeal, Reward Factor, Overall (2 each); Subjective experience, usefulness, satisfaction rate, Behavior intention, Aspectos Hedônicos da UX, Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Habit, Technology Threat Avoidance Theory, Perceived Recommendation Quality, Perceived Conversational interaction, Perceived Efort, openness to experience, conscientiousness, extroversion, agreeableness, neuroticism, resistance to the bot’s answer, a little unpleasant, numbness, identify with the bot’s efforts, satisfaction with the answer, Interactional Enjoyability, Perceived Social Presence, Self-Disclosure, Attractiveness, perspicuity, efficiency, dependability, stimulation, Trust General, Task-specific Trust, Trusting Belief Reliability, Perceived Anthropomorphism, Social Presence, Positive and Negative Aspects of Personality, stimulation level, ease of use, frequency of use, Propensity to Trust, Intention detection, Identity recognition, Learning record, Emotional expression, Knowledge guide, Funniness, Appropriateness, Use Again, Damage Control, Thoroughness, Manners, Moral Agency, Emotional Intelligence, Recommendation Accuracy, Explanation, Interaction Adequacy, CUI Attentiveness, CUI Understanding, CUI Response Quality, User Control, Transparency, CUI Rapport, CUI Engagingness, CUI Humanness, Trust, Confidence, Clear, Fluent, Related, Useful, Helpful for Administrative procedures, Helpful for the characteristics of the department, future career planning, CUI Adaptability, Trust in Automation dimensions, disability disclosure, use of assistive technologies, alternative formats you may like to use in your module material and your preferences about tutors, tutorials, communication

Table 6. Identified Evaluation Technologies

Ref	ID	Technology	Hedonic Aspects of UX
[Ceha *et al.*, 2021], [Daniel *et al.*, 2022]	1	Intrinsic motivation inventory - IMI	Interest, Enjoyment, Pressure, Tension, Effort
[Ceha *et al.*, 2021]	2	Academic motivation scale - AMS	Motivation
[El Kamali *et al.*, 2020]	3	User Experience Questionnaire - Short -UEQ-S	Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, Novelty
[Fadhil *et al.*, 2018]	4	Questionnaire Created for the Study of Fadhil *et al.* [2018]	Enjoyment, Attitude, Trust
[Fahn and Riener, 2021]	3	User Experience Questionnaire - Short -UEQ-S	Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, Novelty
[Xiao *et al.*, 2019]	5	Interview Created for the Study of Xiao *et al.* [2019]	Enjoyable, Likeable
[Elsholz *et al.*, 2019]	6	Existing Questionnaire used in Elsholz *et al.* [2019]	General UX
[Kim *et al.*, 2019]	7	Usefulness, Satisfaction, and Ease of use - USE	Enjoyment
[Chen *et al.*, 2021]	8	Existing Questionnaire used in Chen *et al.* [2021]	Privacy Intrusive
[Fiore *et al.*, 2019], [Flohr *et al.*, 2021], [Denecke *et al.*, 2020], [Torkamaan, 2023] , [Kernan Freire *et al.*, 2023] , [Sharma *et al.*, 2021]	9	User Experience Questionnaire - UEQ	Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, Novelty
[Jin *et al.*, 2019]	10	Questionnaire Created for the Study of Jin *et al.* [2019]	Interest, Trust, Diversity, Easy to Report, Feedback, Understanding, Difficulty, Expectation, Intention to Use
[Lee *et al.*, 2021]	11	Existing Questionnaire used in Lee *et al.* [2021]	Trust, Engagement, Intimacy, Self-reflection,m Self-awareness
[Lee *et al.*, 2021]	12	Existing Interview used in Lee *et al.* [2021]	Engagement, Impressions
[Völkel and Kaya, 2021]	13	Big Five Inventory-2	Likeable
[Jain *et al.*, 2018]	14	Questionnaire Created for the Study of Jain et al.	Fun, Frustration
[Wald *et al.*, 2021]	15	Existing Questionnaire used in Wald *et al.* [2021]	Trust
[Park *et al.*, 2021]	16	Existing Questionnaire used in Park *et al.* [2021]	Enjoyment, Trust, Engagement, Psychological Well-being, Anxiety, Attention, Intention to Use, Adaptability, Sociability, Social Influence, Social Presence, Interpretation, Psychological Impact
[Park *et al.*, 2021]	17	Existing Interview used in Park *et al.* [2021]	General UX, Perceptions of Social Disclosure
[Yun *et al.*, 2020]	18	Questionnaire Created for the Study of Yun *et al.* [2020]	Revealing Emotional Expression, Usefulness of Emotional Expression, Naturalness
[Benke *et al.*, 2020]	19	Affective Benefits and Costs of Communication Technology - ABCCT	Emotion Perception
[Benke *et al.*, 2020]	20	Interview Created for the Study of Benke *et al.* [2020]	Engagement, Social Presence, Privacy, Expression of Emotion, Social Support, Commitment, Unmet Expectations
[De Nieva *et al.*, 2020]	21	Questionnaire Created for the Study of De Nieva *et al.* [2020]	Humanity, Affection
[Wambsganss *et al.*, 2021]	22	Existing Questionnaire used in Wambsganss *et al.* [2021]	Enjoyment
[Bawa *et al.*, 2020]	23	Questionnaire Created for the Study of Bawa *et al.* [2020]	Humanity
[Portela and Granell-Canut, 2017]	24	Visual Analogue Scale - VAS	Anxiety, Happiness, Sadness, Anger, Surprise, Tranquility, Vigor
[Portela and Granell-Canut, 2017]	25	Multidimensional Integrative Model - MIM	Interest, Frustration, Discomfort, Well-being
[Portela and Granell-Canut, 2017]	26	Interpersonal Reactivity Index - IRI	Empathy

Table 7. Identified UX Evaluation Technologies: Continuation

Ref	ID	Technology	Hedonic Aspects of UX
[Liu *et al.*, 2020]	27	Interview Created for the Study of Liu *et al.* [2020]	Engagement
[Kattenbeck *et al.*, 2018]	28	Questionnaire Created for the Study of Kattenbeck *et al.* [2018]	Easy to Report, Intention to Use, Fun
[Bae Brandtzæg *et al.*, 2021]	29	Interview Created for the Study of Bae Brandtzæg *et al.* [2021]	Trust, Privacy, Appreciation, Emotional Support
[Iniesto *et al.*, 2023]	30	Interaction with VA Created for the Study of Iniesto *et al.* [2023]	General UX
[Iniesto *et al.*, 2023]	31	Observation Created for the Study of Iniesto *et al.* [2023]	Disability disclosure, Use of Assistive Technologies, Alternative Formats you may Like to Use in your Module Material and your Preferences About Tutors, Tutorials, Communication Preferences
[Iniesto *et al.*, 2023]	32	Open-ended Experience questionnaire Created for the Study of Iniesto *et al.* [2023]	General UX
[Iniesto *et al.*, 2023]	33	Interview Created for the Study of Iniesto *et al.* [2023]	Experience, Language and Voice, Conversation, Summary, Relationship with the Disability Support Form, General
[Iniesto *et al.*, 2023] , [Cai *et al.*, 2023]	34	Technology Acceptance Model - TAM	Perceived Usefulness, Attitude, Intent to Use
[Iniesto *et al.*, 2023]	35	Conversational User Interface Accessibility Questionnaire - CUIAQ	Made Sense, Easy to Navigate, Able to Predict, Compatible, Accessibility Preferences, Not Excessively Demanding, Enough time, Well-defined Options, Communicate, Communicating
[Iniesto *et al.*, 2023]	36	Feedback questionnaire Created for the Study of Iniesto *et al.* [2023]	General UX
[Iniesto *et al.*, 2023]	37	Speech User Interface Service Quality Reduced - SUISQ-R	User Goal Orientation (UGO), Customer Service Behaviour (CSB), Verbosity (V)
[Cai *et al.*, 2023]	38	Proactive Guidance - PG	Self-awareness, User Acceptance, Mental Wellbeing
[Cai *et al.*, 2023]	39	Social Information - SI	Self-awareness, User Acceptance, Mental Wellbeing
[Cai *et al.*, 2023]	40	User study/test Created for the Study of Cai *et al.* [2023]	Emotional Resonance (times), Expression Length (words), Expression Depth Music, Rating Engagement Duration (seconds)
[Cai *et al.*, 2023]	41	Questionnaire to Measure Users' Perceived Need Satisfaction and User Acceptance Created for the Study of Cai *et al.* [2023]	Autonomy, Competence, Relatedness
[Cai *et al.*, 2023]	42	Warwick-Edinburgh Mental Well-being Scale - WEMWBS	Mental Wellbeing
[Cai *et al.*, 2023]	43	Open questions Created for the Study of Cai *et al.* [2023]	Self-awareness
[Zorrilla and Torres, 2022] , [Chen, 2022] , [Sharma *et al.*, 2021], [Gambetta *et al.*, 2021], [Daniel *et al.*, 2022]	44	Chatbot Usability Questionnaire - CUQ	General UX
[Zorrilla and Torres, 2022]	45	Hedonic Feelings Questionnaire - HFQ	Hedonic Aspects of UX
[Schmitt *et al.*, 2022]	46	User study/test Created for the Study of Schmitt *et al.* [2022]	Generally Accurate, Exciting, Enjoy, Users' Perceived Social, Sense of Sociability
[Jung *et al.*, 2022]	47	Intrinsic Motivation Inventory -IMI(Partial)	Interest-Enjoyment (INT-ENJ), Perceived Competence
[Jung *et al.*, 2022]	48	User Engagement Scale - UES-SF	Engagement, Focused Attention, Perceived Usability, Aesthetic Appeal, Reward Factor, Overall
[Jung *et al.*, 2022]	49	Trust in Automation - TiA	Propensity to Trust, Trust in Automation Dimensions
[Moilanen *et al.*, 2022]	50	User Engagement Scale survey in Short Form - UES-SF	Focused Attention, Perceived Usability, Aesthetic Appeal, Reward Factor
[Moilanen *et al.*, 2022]	51	Interview Created for the Study of Moilanen *et al.* [2022]	Positive and Negative Aspects of Personality

Table 8. Identified UX Evaluation Technologies: Continuation

Ref	ID	Technology	Hedonic Aspects of UX
[Jin *et al.*, 2021]	52	Conversational recommender system - User Experience - CRS-UX	Recommendation Accuracy, Explanation, Novelty, Interaction Adequacy, CUI Attentiveness, CUI Understanding, CUI Response Quality, User Control, Transparency, CUI Rapport, CUI Engagingness, CUI Humanness, Trust, Confidence, Satisfaction, CUI Adaptability, Perceived Usefulness, Intent to Use, Perceived Ease of Use, Overall
[Flandrin *et al.*, 2022]	53	Look-alike Method of Instruction	General UX
[Flandrin *et al.*, 2022]	54	UX curve	Stimulation level, Ease of Use, Frequency of Use
[Torkamaan, 2023]	55	Open questions Created for the Study of Torkamaan [2023]	General UX
[Wambsganss *et al.*, 2022]	56	User study/test Created for the Study of Wambsganss *et al.* [2022]	Interactional Enjoyability, Perceived Social Presence, Self-Disclosure
[Liu *et al.*, 2022]	57	User study/test Created for the Study of Liu *et al.* [2022]	Funniness, Appropriateness, Use Again, Damage Control, Thoroughness, Manners, Moral Agency, Emotional Intelligence, Satisfaction
[Law *et al.*, 2022]	58	Existing Questionnaire Used in Law *et al.* [2022]	Trust General, Task-specific Trust , Trusting Belief Reliability, Perceived Anthropomorphism, Social Presence
[Cai *et al.*, 2022]	59	Ten Item Personality Inventory (TIPI)	Openness to Experience, Conscientiousness, Extroversion, Agreeableness, Neuroticism
[Cai *et al.*, 2022]	60	Trust Measurement	Perceived Recommendation Quality, Perceived Conversational Interaction, Perceived Efort
[Essop *et al.*, 2023]	61	UTAUT2 Framework	Behavior Intention
[El Hefny *et al.*, 2021]	62	Acceptance scale	Usefulness, Satisfaction Rate
[Zhang *et al.*, 2022]	63	Questionnaire Created for the Study of Zhang *et al.* [2022]	Resistance to the Bot's Answer, a Little Unpleasant, Numbness, Identify With the Bot's Efforts, Satisfaction with the Answer
[Zhang *et al.*, 2022]	64	Importance Analysis of Persuasiveness and Self-efficacy	Intention Detection, Identity Recognition, Learning Record, Emotional Expression, Knowledge Guide
[Day and Shaw, 2021]	65	Existing Questionnaire Used in Day and Shaw [2021]	Clear, Fluent, Related, Useful, Helpful for Administrative Procedures, Helpful for the Characteristics of the Department, Future Career Planning
[Dopler and Göschlberger, 2022]	66	Questionnaire Created for the Study of Dopler and Göschlberger [2022]	General UX
[Al-Emran *et al.*, 2024]	67	Integrated chatbot acceptance-avoidance model - ICAAM	Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Habit, Technology Threat Avoidance Theory
[Alazraki *et al.*, 2021]	68	User study/test Created for the Study of Alazraki *et al.* [2021]	General UX
[Yu *et al.*, 2021]	69	User study/test Created for the Study of Yu *et al.* [2021]	General UX

preferences, Experience, Language and voice, Conversation, Summary, Relationship with the Disability Support Form, General, Attitude, made sense, easy to navigate, able to predict, compatible, accessibility preferences, not excessively demanding, enough time, well-defined options, communicate, communicating, User goal orientation (UGO), Customer service behaviour (CSB), Verbosity (V), Emotional Resonance (times), Expression Length (words), Expression Depth Music, Rating Engagement Duration (seconds), autonomy, competence, relatedness, generally accurate, exciting, enjoy, users' perceived social, sense of sociability, Interest-Enjoyment (INT-ENJ), Perceived Competence, Engagement (1 each).

The most frequent hedonic aspect found on the SMS extension is the general UX. For example, in Alazraki *et al.* [2021], the authors conducted a user study that included a questionnaire containing multiple-choice questions to evaluate various aspects of the user experience. Participants were asked to rate: "(a) the chatbot's ability to demonstrate empathy; (b) each user's level of engagement; (c) the usefulness of the platform; (d) the chatbot's ability to identify emotions." Additionally, other papers such as Zorrilla and Torres [2022] and Chen [2022], which used the CUQ (Chatbot Usability Questionnaire), were also identified as examples of general UX.

After a detailed analysis of the SMS extension, we identified 198 hedonic aspects. We verified that 10 of the 132 hedonic aspects (SMS Part 2) had already been identified in SMS Part 1, resulting in a total of 188 different hedonic aspects in the complete set (SMS Parts 1 and 2). Overall, the hedonic

Table 9. SMS results for each of the sub-questions

Subquestions	Possible anserws	Results SMS		Results Extension		Total	
		Technologies	%	Technologies	%	Technologies	%
SQ2	Specific	8	27.59%	16	32.65%	24	30.77%
	Generic	21	72.41%	33	67.35%	54	69.23%
SQ3	Existing	18	62.07%	31	62.27%	49	62.82%
	Created	11	37.93%	18	36.73%	29	37.18%
SQ6	Quantitative	23	79.31%	37	75.51%	60	76.92%
	Qualitative	5	17.24%	8	16.33%	13	16.67%
	Mixed	1	3.45%	4	8.16%	5	6.41%
		Chatbots	%	Chatbots	%	Chatbots	%
SQ8	Yes	5	19.23%	13	52%	18	35.30%
	No	21	80.77%	12	48%	33	64.70%
SQ9	Task oriented	0	0%	1	4%	1	1.95%
	Conversation-oriented	18	69.23%	14	56%	32	62.75%
	Conversation and task oriented	8	30.77%	10	40%	18	35.30%
	No	0	0%	0	0%	0	0%
		Assessments	%	Assessments	%	Assessments	%
SQ11	Yes	0	0%	1	3.85%	1	1.92%
	No	26	100%	25	96.15%	51	98.08%
SQ12	Yes	6	23.08%	3	11.54%	9	17.31%
	No	20	76.92%	23	88.46%	43	82.69%

aspects identified in Part 1 and 2 of SMS encompass General UX, Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, Novelty, Social Presence, Humanity, and Privacy.

4.2 SQ2. Specificity of evaluation technology

The analysis for this sub-question indicates a lack of chatbot-specific UX assessment technologies. Most of the identified technologies are designed to evaluate any software (72.41%, N = 21) as depicted in Table 9. One example is the User Experience Questionnaire (UEQ), applied by Fiore *et al.* [2019] to evaluate the experience with using a chatbot for IT support.

Regarding the chatbot-specific technologies, Jin *et al.* [2019] applied a questionnaire with 14 questions specifically developed to evaluate the chatbot Musicbot. One of the questions asked is “I felt in control of modifying my taste using MusicBot” (aspect: control). Despite being chatbot-specific, this technology cannot be reused in evaluating other conversational systems since it is context-dependent. In contrast, the questionnaire developed by Fadhil *et al.* [2018] uses more generic questions, such as “The more I interacted with the agent, the more I liked the experience” (aspect: enjoyment), so it can be used to evaluate chatbots regardless of function and application.

The SMS extension (Part 2) emphasizes the lack of chatbot-specific UX assessment technologies. Most of the identified technologies are designed to evaluate any software (67.35%, N = 33). One example is the Technology Acceptance Model (TAM), applied by Cai *et al.* [2023] in a chatbot that guides users to be self-aware and express their feelings when listening to music.

Regarding the chatbot-specific technologies, Iniesto *et al.* [2023] applied the Conversational User Interface Accessibility Questionnaire (CUIAQ) to explore the potential of CUI to improve the experience of disclosing disabilities and accessing support in the context of higher education. The questionnaire has 10 sentences that can be answered with a 7-point Likert scale. An example of the sentence is ”The sequence of the conversation made sense”.

Analyzing the total technologies identified in this SMS (Part 1 and Part 2), we found that 30.77% (N=24) of the technologies were specific to chatbots, and 69.23% (N=54) were generic.

Our results reveal that information on chatbot-specific experience is not being investigated in depth, which may impose barriers to performing a complete analysis of the chatbot’s UX. Chatbots have unique characteristics such as natu-

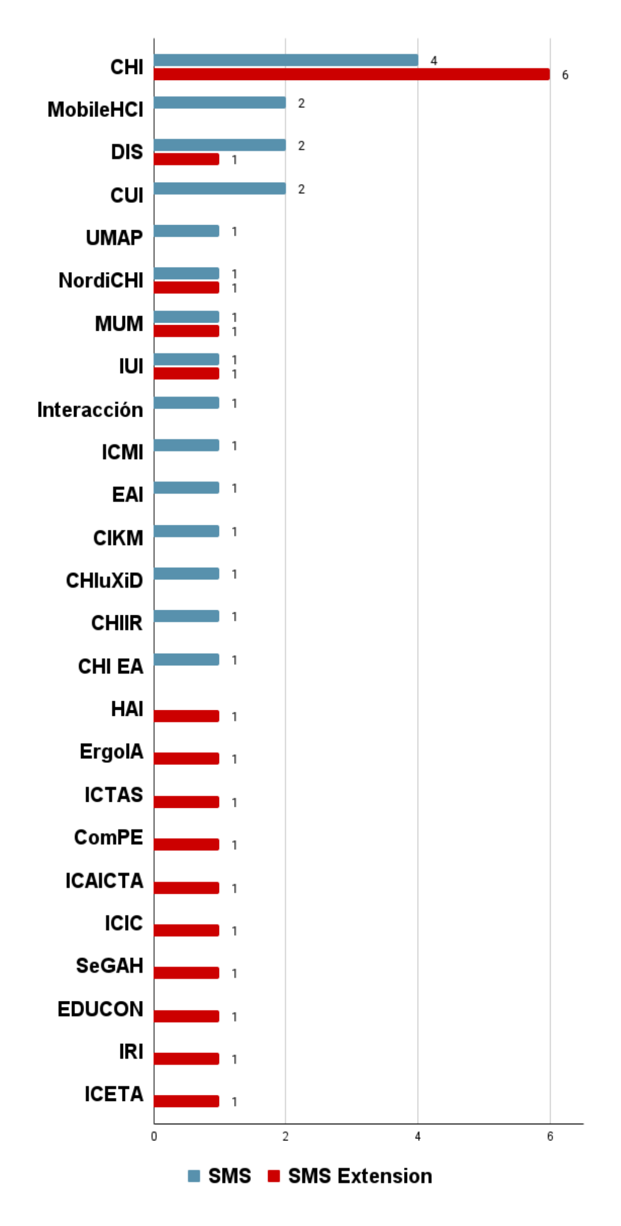


Figure 2. Conferences

ral language interaction and personification. The lack of instruments that address these characteristics may prevent possible quality improvements in UX.

4.3 SQ3. Basis of evaluation technology

In Part 1 of SMS, our analysis shows that creating a chatbot-specific UX assessment is a less common practice than using existing technologies. Most of our primary studies applied existing technologies. Only 37.93% (N = 11) of the identified technologies were created specifically for the study (Table 9). For example, Chen *et al.* [2021] applied an existing questionnaire based on the study by Xu *et al.* [2008] to evaluate UX in terms of intrusive privacy. In contrast, Kattenbeck *et al.* [2018] developed a questionnaire to determine the participants' experience with Airbot. This result suggests that not many technologies were created to carry out chatbot UX assessments and indicates that using existing technologies to carry out the assessments is more common.

When analyzing the 2nd part of this SMS (2021-2023), it becomes evident that most of the identified technologies are already existing, representing 63.27% (N=31) of the total (Table 9). For example, in the study conducted by Law *et al.* [2022], a questionnaire based on the work of Lankton *et al.* [2015] was used to evaluate a customer service chatbot. In contrast, Dopler and Göschlberger [2022]'s work adopted a previously designed questionnaire to evaluate an educational chatbot with 10 sentences. One example of a sentence is "The bot has motivated me to continue".

Analyzing the total technologies identified in this SMS (Part 1 and Part 2), we found that only 37.18% (N=29) of the technologies were created for the study and 62.82% (N=49) are existing technologies (Table 9) . It is important to note that there is no standard for building or applying technologies created for a specific study. Since they are not methodologically validated, these technologies are subject to bias and manipulation. In addition, the chatbot context impacts the user experience, raising a relevant question: is it ideal for the evaluation technology to consider the system's domain when evaluating UX, or should it serve any conversational system? A study should be carried out to understand the impact of considering the chatbot domain in the UX evaluation and whether specific or generic evaluations are the best.

4.4 SQ4. Method of data collection

According to our results in Part 1 of this SMS, the most applied method for UX data collection is the 7-point Likert scale (N = 12), followed by the 5-point Likert scale (N = 10) and open questions (N = 5). The Likert scale is a set of statements (items) in which participants are asked to demonstrate their level of agreement with the statement [Joshi *et al.*, 2015]. Park *et al.* [2021] applied two data collection methods (7- and 5-point Likert scale) to evaluate hedonic aspects such as Psychological Well-being, Pleasure, General Experience, and Psychological Impact. One example of a 7-point Likert scale item is "I feel hopeful about my future" (psychological well-being).

In the SMS extension (Part 2), the most applied method for UX data collection is the 5-point Likert scale (N = 17), followed by the 7-point Likert scale (N = 15) and open questions (N = 7). In El Hefny *et al.* [2021], the 5-point Likert Scale was used to evaluate hedonic aspects such as usefulness and satisfaction rate, and "the range of the user satisfaction scores is from -2 (not useful/not satisfying) to 2 (useful/satisfying)". Furthermore, the 5-point Likert data collection method is gaining more prominence, we can consider that this increase is mainly due to the increased use of CUQ (Chatbot Usability Questionnaire) technology that uses this data collection method.

4.5 SQ5. Evaluation technology composition

To answer this sub-question, we collected the characteristics of each assessment technology, such as interviews, questionnaires and metrics. In one of the studies which used interviews with questions designed by the authors Benke *et al.* [2020], they asked the participants "How was the perception of the chatbot?" and also "How was your experience with the appearance of the chatbot?" to evaluate the text-based chatbot to assist teams in previously identified challenges.

In the study conducted by Fiore *et al.* [2019], the User Experience Questionnaire (UEQ) was adopted, an assessment instrument that uses a 7-point Likert scale to measure various aspects of the user experience. This questionnaire addresses key elements including Attractiveness, perspicuity, efficiency, dependability, stimulation and novelty. The complete UEQ form used in the research can be viewed in Figure 4.

In the SMS extension (Part 2), a study is presented in Alazraki *et al.* [2021], where a user study was designed for the research. The study questionnaire contained multiple-choice questions that aimed to evaluate the chatbot's ability to exhibit empathy, the level of involvement of each user, the usefulness of the platform, and the chatbot's ability to identify emotions [Alazraki *et al.*, 2021].

Gambetta *et al.* [2021] used the Chatbot Usability Questionnaire (CUQ), a questionnaire based on a 5-point Likert scale, which aims to evaluate the user experience in general. Although the CUQ is primarily a usability questionnaire intended to measure chatbot effectiveness and ease of use, it can also be considered an instrument to evaluate user experience broadly (general UX). For example, ”the chatbot seemed very unfriendly". Figure 5 presents all the sentences of CUQ.

Due to space limitations, to better analyze the results and for an in-depth list of technology's characteristics, the technical report of Part 1 can be consulted in Souza *et al.* [2023] and of the Part 2 in Mariano *et al.* [2024].

4.6 SQ6. Type of analysis

Most technologies (Part 1) (79.31%, N = 23) extract quantitative data and 17.24% (N = 5) of them extract qualitative data (Table 9). Mixed data is used in only one technology [Elsholz *et al.*, 2019], in which participants answered five questions on a 7-point Likert scale and one open question. This result aligns with the outcomes of the SQ4 question, which states that quantitative scales are the most frequent response collection method.

In the SMS extension analysis (Part 2), we observed that the majority of extracted data continues to be quantitative, representing 75.51% (N=37) of the total (Table 9). However, the number of studies with mixed methods increased to four. An example is the work of Schmitt *et al.* [2022], in which 15 items were included in the user test to evaluate participants' perception of the Hermine system, a chatbot to support students in retrieving relevant course information and presenting information related to course questions. Quantitative data were assessed on a 7-point Likert scale, adapted from a previous source. In addition, three qualitative questions were asked.

When analyzing the SMS (Part 1 and 2), we have 76.92% (N=60) publications extracting quantitative data, 16.67% (N=13) qualitative data, and 6.41% extracting mixed data (Table 9).

4.7 SQ7. Chatbot function

To answer this sub-question (Part 1 of SMS), we qualitatively searched the studies to identify the characteristics of the chatbots evaluated in the papers. For example, Park *et al.* [2021] implemented two chatbots to instruct users to write about some of their most difficult experiences in life. They are based on the combination of three therapeutic techniques that can help the user to reflect on past feelings, social relationships, or situational circumstances as well as themselves. Portela and Granell-Canut [2017] presented two chatbots to understand the nature of emotional engagement between the individual psychological mindset and a chatbot during a conversation.

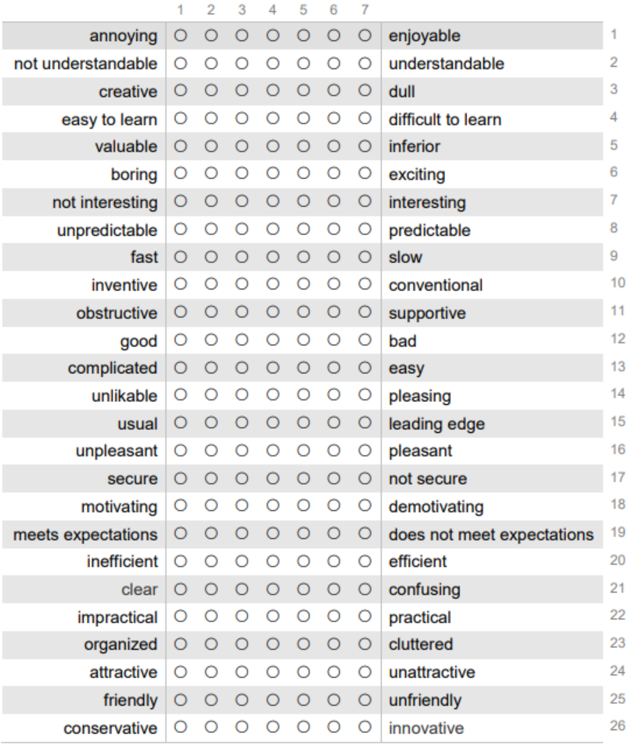
Some of the chatbot domains we found include conversational learning tools; motivational coaching for the elderly; collecting data on mental and physical well-being; managing banking transactions; simulating a visit to a doctor; IT support of a company; and a movie recommendation system. Most domains relate to user activities that, by their nature, arouse emotions in the users, such as psychological-related or education and learning interactions. However, current chatbot users are highly exposed to, for example, customer service chatbots which are more practical, productivity-oriented interactions. Because such domains do not evoke hedonic aspects directly, according to our results, they may have been neglected in the literature.

The SMS extension (Part 2) revealed a wide range of applications for chatbots, demonstrating their versatility and potential in different contexts. Some of the applications include using chatbots to improve the experience of disclosing disabilities and accessing support in higher education [Iniesto *et al.*, 2023], guiding users to be self-aware and express their feelings when listening to music [Cai *et al.*, 2023], motivational coaching through a fully data-driven conversational agent [Zorrilla and Torres, 2022; Alazraki *et al.*, 2021], support students in retrieving course-relevant information and presenting course-related questions [Schmitt *et al.*, 2022], customer service [Day and Shaw, 2021], combating misinformation during the Covid-19 pandemic [El Hefny *et al.*, 2021], introducing context awareness and emotion management to improve students' emotional confidence [Zhang *et al.*, 2022]. These applications highlight chatbots' ability to provide support, guidance, and interaction across various scenarios.

Although we found papers that address ChatGPT and other generative AI chatbots, it is interesting to note that none passed the first and second filters. This suggests that although ChatGPT is receiving considerable attention, studies that focus on evaluating the hedonic aspects of user experience in text chatbots are not yet being conducted. This research gap is notable as hedonic experience is essential in user adoption and satisfaction with conversational technologies.

4.8 SQ8. Chatbots created for a specific group

Considering the identified chatbots (Part 1 of SMS), only five (19.2%) were created for specific groups, while the majority (80.77%, N = 21) did not aim at a particular group of people. The specific groups we find are the elderly (N = 1), teenagers



	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Figure 4. User Experience Questionnaire [Laugwitz *et al.*, 2008]

in Korea (N = 1), and students (N = 3) (Table 9). For example, El Kamali *et al.* [2020] evaluates a chatbot developed as a motivational coach for the elderly. The chatbot presented in De Nieva *et al.* [2020] helps students relieve the stress of the academic workload. Finally, Kim *et al.* [2019] evaluated a chatbot to collect demographic information and questions about internet use by teenagers in Korea. Our findings reveal that most of these technologies are not developed for a target audience, and the lack of this specificity can affect the UX, since without knowing the user profile, the developers cannot anticipate the interaction. Users with more digital maturity may have fewer challenges navigating the conversation than the less experienced users.

In the extension of the SMS (Part 2), it is observed that the majority of chatbots were designed to serve a specific group of people, representing 52% (N=13) of the total (Table 9). This targeted approach allows chatbots to be adapted to the unique needs and characteristics of each user group. Examples of specific groups include the chatbot presented by Schmitt *et al.* [2022], developed for students, the chatbot aimed at employees in the hotel sector discussed by Flandrin *et al.* [2022], and the chatbot aimed at chemistry students mentioned in Sharma *et al.* [2021]. This specialization allows for more effective and personalized interaction, contributing to a more satisfactory and relevant user experience in different contexts and areas of activity.

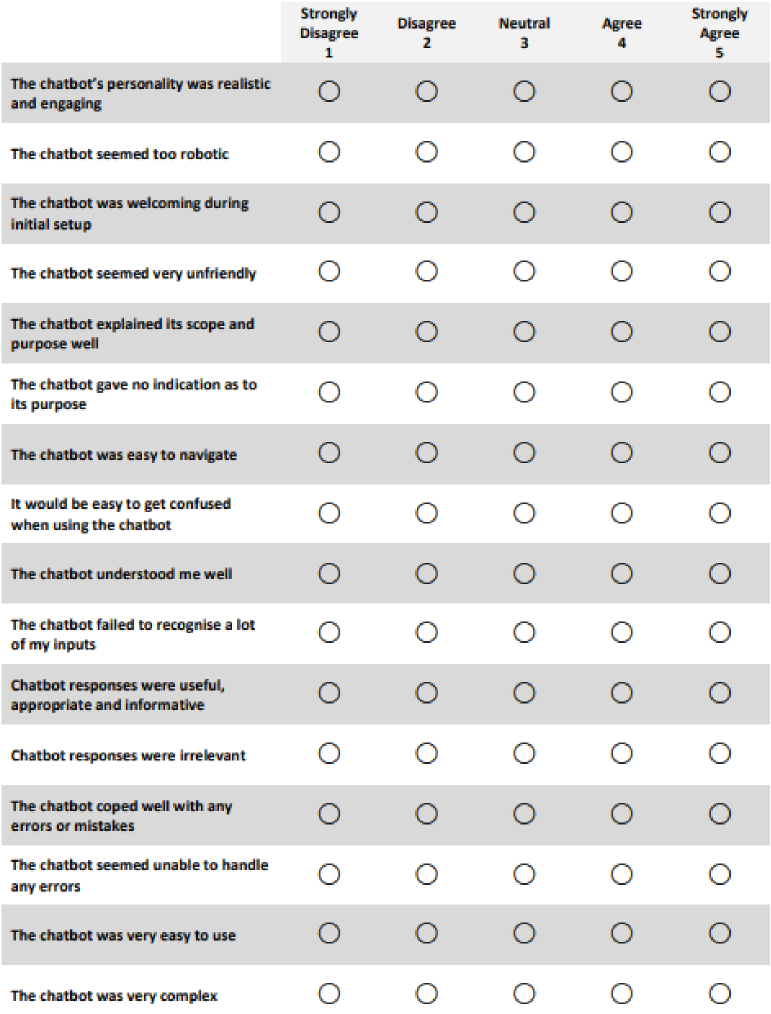
It is essential to recognize the importance of developing chatbots targeted at specific groups of users, as the user experience can vary considerably between different types of users. Proper contextualization is essential to ensure chatbots meet the specific needs, preferences, and skills of each demographic or user group. By taking into account factors such as age, gender, technology skills, and usage goals, developers can create more personalized and relevant chatbot experiences. This not only improves user satisfaction but also increases the overall effectiveness and usefulness of the chatbot for the specific target audience. Therefore, highlighting this contextualization in the design and development of chatbots is crucial to ensure an optimized and satisfactory user experience for all users.

4.9 SQ9. Type of chatbots

More than half of the chatbots (Part 1 of SMS) we found (69.2%, N = 18) (Table 9) are both conversation- and task-oriented. The remainder chatbots are conversation-oriented only. We did not find chatbots that are task-oriented only. One example of conversation-oriented chatbot was presented by Denecke *et al.* [2020] to assist the user in regulating their emotions. Jain *et al.* [2018] developed chatbots for both conversation and shopping.

In the extension of the SMS (Part 2), the analysis revealed that the majority of chatbots found are of the conversation-oriented type, totaling 56% (N=14). Furthermore, 40% (N=10) of the chatbots identified are both conversational and task-oriented, while only one chatbot found is exclusively task-oriented, representing 4% of the total (Table 9). These results highlight the predominance of the chatbot approach focused on interaction through conversations, which reflects the growing emphasis on natural communication and the system's ability to understand and respond. However, the presence of both conversational and task-oriented chatbots indicates a trend towards more versatile systems, capable of offering both support in terms of information and in carrying out specific actions for users.

The findings show that chatbots usually perform at least



	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5
The chatbot's personality was realistic and engaging	○	○	○	○	○
The chatbot seemed too robotic	○	○	○	○	○
The chatbot was welcoming during initial setup	○	○	○	○	○
The chatbot seemed very unfriendly	○	○	○	○	○
The chatbot explained its scope and purpose well	○	○	○	○	○
The chatbot gave no indication as to its purpose	○	○	○	○	○
The chatbot was easy to navigate	○	○	○	○	○
It would be easy to get confused when using the chatbot	○	○	○	○	○
The chatbot understood me well	○	○	○	○	○
The chatbot failed to recognise a lot of my inputs	○	○	○	○	○
Chatbot responses were useful, appropriate and informative	○	○	○	○	○
Chatbot responses were irrelevant	○	○	○	○	○
The chatbot coped well with any errors or mistakes	○	○	○	○	○
The chatbot seemed unable to handle any errors	○	○	○	○	○
The chatbot was very easy to use	○	○	○	○	○
The chatbot was very complex	○	○	○	○	○

Figure 5. Chatbot Usability Questionnaire [Holmes *et al.*, 2019]

one defined function, which can be a general conversation or performing tasks. This finding raises the question: are there chatbots that perform neither conversation nor tasks? It should be investigated whether there are chatbots that do not perform any of these functions, and if so, what their use is.

4.10 SQ10. How the chatbot was evaluated

To answer this sub-question (Part 1 of SMS), we qualitatively analyzed the studies to identify how the researchers carried out the evaluations in terms of the tasks to users, the instructions to answer the questionnaires, and the order of completion (pre or post-interaction evaluation). For example, Denecke *et al.* [2020] performed the assessment by creating six tasks, and the users were asked to complete the tasks to evaluate specific functionality. Participants provided feedback on whether they were able to complete the task and possible issues that may have occurred. Additionally, the participants assessed concrete aspects of the user experience using the UEQ. Analyzing this result, it is noted that there is no standard in the evaluation, that is, each author defines his/her way of conducting the experiment.

In the extension of the study (Part 2), two main ways of evaluating the chatbot were identified: controlled experiment, representing 60% (N=15), and case study, totaling 40% (N=10). An example of evaluation through a controlled experiment is found in the paper of Law *et al.* [2022], in which a 2x3 factorial experiment was conducted with 251 participants. They were asked to perform three tasks with a chatbot for an online bank under one of six conditions, varying in humanity and conversational performance. As an example of a case study, we have the paper of Alazraki *et al.* [2021], in which a evaluate the application through a human trial with N=16 subjects from the non-clinical population, as well as two medical professionals specialised in mental health when interacting with a computational framework that augments a rule-based agent for the delivery of selfattachment technique (SAT). These evaluation approaches provide valuable insights into the performance and effectiveness of chatbots in different contexts and usage scenarios.

4.11 SQ11. Empirical evaluation

The results revealed that none of the UX assessment technologies in chatbots found were empirically evaluated (SMS Part 1). This was because none of the selected papers focused on the UX assessment technology, but using the technology to assess one or more chatbots.

In the extension of the SMS (Part 2), a technology was identified being empirically evaluated, representing 3.85% of the total (Table 9). An example of this type of evaluation can be found in the paper of Jin *et al.* [2021], in which the authors report that they used an empirical approach, applying psychometric methods to evaluate the reliability and validity of the proposed model. This methodology contributes to a more precise and well-founded understanding of the effectiveness and applicability of the technology in question.

It is relevant to conduct evaluations of the technologies as a way of validating them and ensuring that they they are consistent and reliable. Performing empirical evaluation requires seeking aspects such as verification of feasibility and validation, which are important steps to refine the technology and identify problems that can interfere with the quality of the evaluation [Shull *et al.*, 2001].

4.12 SQ12. Aspects of emotional health in UX evaluation

In Part 1 of SMS, we found that only a few studies (23.1%, N = 6) considered some aspect of the user's emotional health (Table 9). Those who considered it, most applied questionnaires aimed at assessing mental health, physical and psychological well-being, or considered the user's emotional health during the UX assessment. For example, Lee *et al.* [2021] evaluated self-reflection and self-awareness, and the chatbot in question had the function of leading the user to improve his writing.

In the extension SMS (Part 2), a small portion, representing only 11.54%, considered some aspect of the user's emotional health (Table 9). An example of this is found in the paper of Alazraki *et al.* [2021], which evaluates empathy during the study and the level of emotion of users. This approach demonstrates the importance of considering not only functional but also emotional aspects when interacting with chatbots, aiming to provide more humanized experiences adapted to users' emotional needs.

Although there are few studies that consider the user's emotional health, it is noted that there is already a concern on the part of the authors to include such aspects in the evaluation of UX. However, further investigations, that take this factor into account, should still be produced, since the investment in hedonic aspects has advantages, as in a commercial context, to retain customers [Chitturi *et al.*, 2008].

4.13 SQ13. How the user's mental state was assessed

We inspected the studies to identify how the user's mental state was assessed. In the UX evaluation carried out in the study by Yun *et al.* [2020], the following aspects are evaluated: revelation of emotional expression and usefulness of emotional expression. In Benke *et al.* [2020], the aspects perceived emotion and expression of emotion are considered. Both reflect the user's emotional state and were evaluated through the Affective Benefits and Costs of Communication Technology (ABCCT). Knowing that there are assessment technologies aimed at mental state, such as ABCCT, it is feasible to suggest that, in other studies of chatbot evaluation, a stage of concern for the well-being of the user is added, and that there is not even the need to create a new way of doing this, since there are methods that already do this.

In the SMS extension (Part 2), Cai *et al.* [2023] evaluates empathy during the study and the level of emotion, while Moilanen *et al.* [2022] addresses users' preference in relation to chatbots to find self-care solutions for mental health, through a classification question. Additionally, Cai *et al.* [2023] also investigates users' mental well-being, measured using a 7-item short version of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS), which assesses mental well-being based on in the experiences reported by interviewees over the past week. Figure 6 presents the WEMWBS questionnaire, providing a standardized instrument to assess users' mental well-being in relation to the use of chatbots for self-care and mental health.

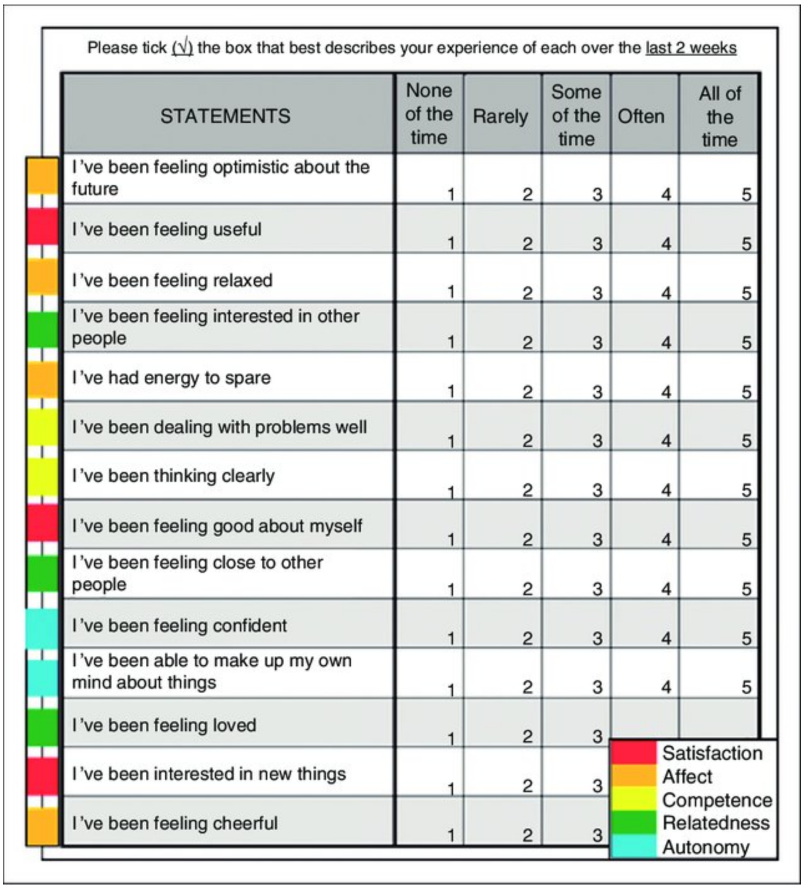
5 Discussion

The quantitative results obtained in this SMS (Part 1 and 2) were generated from the responses to each research subquestion, and some of them are presented in Table 9. Overall, 69 different evaluation technologies were found in the examined studies.

Regarding the hedonic aspects identified in the SMS, I would like to point out the increase in the number of hedonic aspects identified in Part 2. While Part 1, covering up to October 2021, found 66 hedonic aspects, the extension conducted (between November 2021 and June 2023) revealed 132 hedonic aspects, more than double in the initial total. Additionally, Part 1 highlighted ”trust” as the most recurrent hedonic aspect, whereas in the extension phase, ”general UX” emerged as the predominant aspect. The significant increase in the number of hedonic aspects and the shift in priorities indicate that a broader range of hedonic aspects are being considered in evaluating the UX of text-based chatbots. It is also noteworthy that, between Part 1 and Part 2, only 10 aspects were the same.

Our analysis shows there is little variety in the format of the technology, as most of the studies use questionnaires and interviews, almost the same result that Tubin *et al.* [2022] found, once he states that are extensive use of questionnaires created by the authors in the methods discovered in their study. In this SMS, only 30% of the technologies are specific to text-based chatbots, which shows little specificity in the evaluations of this type of system. In comparison with the result of Guerino and Valentim [2020] there is a notable discrepancy, as the authors found a balance in the presence of specific and non-specific technologies for conversational systems. It is worth mentioning that conversational systems can be chatbots, conversational agents, virtual assistants, applications with voice functions, among others. This can have an impact on the outcome of the evaluation, since in several studies particular characteristics of this type of conversational systems are not examined.

Regarding the way the answers are collected, most technologies use quantitative methods, which impacts the speci-



Please tick (√) the box that best describes your experience of each over the last 2 weeks

	STATEMENTS	None of the time	Rarely	Some of the time	Often	All of the time
🟧	I've been feeling optimistic about the future	1	2	3	4	5
🟥	I've been feeling useful	1	2	3	4	5
🟧	I've been feeling relaxed	1	2	3	4	5
🟩	I've been feeling interested in other people	1	2	3	4	5
🟧	I've had energy to spare	1	2	3	4	5
🟨	I've been dealing with problems well	1	2	3	4	5
🟨	I've been thinking clearly	1	2	3	4	5
🟥	I've been feeling good about myself	1	2	3	4	5
🟩	I've been feeling close to other people	1	2	3	4	5
🟦	I've been feeling confident	1	2	3	4	5
🟦	I've been able to make up my own mind about things	1	2	3	4	5
🟩	I've been feeling loved	1	2	3		
🟥	I've been interested in new things	1	2	3		
🟧	I've been feeling cheerful	1	2	3		

🟥	Satisfaction
🟧	Affect
🟨	Competence
🟩	Relatedness
🟦	Autonomy

Figure 6. Warwick-Edinburgh Mental Well-Being Scale [Watson, 2018]

ficity of the evaluations, since it does not allow the user to expose, in a detailed way, how their experience was. Considering the study by Ren *et al*. [2019], this is not positive, since in his research he states that for a chatbot usability assessment it is necessary to consider the context of the use of the system and in which situation it will be applied. The same need is noted for a UX evaluation study. Only through quantitative methods, it is not possible to have this depth in the analysis.

Considering the chatbots evaluated in the identified articles, in Part 1 of the SMS, it was recognized that the minority is specific to a group of users. This finding shows that almost all chatbots found in this SMS are designed and developed for any type of user. However, in Part 2 of the SMS, the results show a significant difference: the majority of the identified chatbots were targeted at specific user groups. This shift reflects an evolution in the development approach, with an increasing emphasis on tailoring chatbots to meet the unique needs and characteristics of distinct target audiences. We also identified that only one of the chatbots is directed only at tasks, contrary to what Rapp *et al*. [2021] work found, since most of the chatbots found in their search are task-oriented.

One of the papers examined presented an empirical evaluation of the technologies and it is understandable, as all the other papers deal with the evaluation of UX in chatbots and not with the evaluation of UX technologies in chatbots. This is a result similar to Guerino and Valentim [2020], since in his research less than 7% of technologies discovered were empirically evaluated. Studies in which technologies were created, they could have been minimally evaluated, to assure the quality of the proposed evaluation technology [Shull *et al*., 2001].

6 Threats to Validity

According to Ampatzoglou *et al*. [2019], SMSs and systematic literature reviews present threats to validity due to the volume of data, and whether by reading or data analysis. Hence, it is necessary to apply strategies to reduce the consequences of these threats. In this paper, we used an established protocol to conduct the SMS, provided by Kitchenham and Charters [2007], with the purpose of avoid threats related to the research process.

Another possible threat is the choice of search string terms. We identified many synonymous to the main terms and performed several tests in the digital libraries to find the ideal research string. To control for bias in the paper's selection, extraction and analysis, these steps were performed by three researchers. We conducted discussion rounds to find consensus as a strategy to reduce biases.

The absence of a wide-ranging database, such as Scopus, Web of Science, or Google Scholar, it is also a limitation of our SMS. In the future, we intend to increase the scope of the SMS done by adopting more wide-ranging databases.

7 Conclusion and Future Works

This SMS focused on investigating which technologies are used to evaluate hedonic aspects of UX in text-based chatbots. Our results include 52 papers that include 69 different technologies for UX assessment. Observing the years of publication of the studies evaluated, it appears that the topic, UX evaluation in chatbots, is recent, since the oldest work investigated is from 2017. This data leads us to conclude that there are still many possibilities for future work and that can explore the topics of user experience assessment technologies for chatbots and that the analyses around the subject are just beginning.

Our results revealed that there are gaps in the field of UX assessment technologies for chatbots. First, the literature lacks empirical studies to assess the reliability and consistency of technologies for evaluating hedonic aspects of UX for text-based chatbots. This has important implications for the validity of the results obtained by these technologies. Second, there is a lack of technologies that address the specific characteristics of human-chatbot interaction, which indicates that particular aspects of chatbots, such as identity and social interaction, are not properly considered when determining the user experience. Therefore, there is a need to create and validate text-based chatbot-specific UX technologies so that it becomes possible to extract target results that can contribute to the design of enriched UX for chatbots.

One of the most surprising results is the evaluation of emotional state in nine selected papers, which does not even represent a quarter of the total. This result is worrying, as it demonstrates the lack of research in examining the psychological well-being of users. The user's emotional state interferes UX with the chatbot and therefore should be considered in the evaluation of UX.

The extension of the SMS served mainly to reinforce the data found in the initial SMS, highlighting consistencies in relation to previous results. For example, the predominance of conversation-oriented chatbots, the lack of empirical evaluation of UX evaluation technologies, and the emphasis on quantitative evaluation methods remained consistent across both studies. Furthermore, the lack of specificity in chatbot evaluations and the lack of specific chatbots for user groups were issues that remained the same as what was observed in the initial SMS. These consistencies reinforce the importance of these aspects in chatbot research and highlight key areas that may require further attention and development in the future.

When comparing the quantitative data obtained in SMS Part 1 and SMS Part 2, notable differences are observed in the approaches adopted. While in SMS Part 1 most of the papers analyzed used a 7-point Likert scale to collect data (SQ4), in SMS Part 2, this scale was reduced to a 5-point Likert scale. Furthermore, when analyzing the presence of specific chatbots for groups of people (SQ8), a significant transition between the parties was noted. While in SMS Part 1 the smallest number of chatbots had this characteristic, in SMS Part 2, the majority of chatbots demonstrated that they were targeted at specific groups. In short, these discrepancies between SMS parts 1 and 2 of the SMS highlight the importance of comparative analysis to understand trends and developments in the field of human-computer interaction.

These results open opportunities for future research, including the definition of text-based chatbot-specific technologies to evaluate hedonic aspects of UX, as well as an empirical evaluation of the new and/or already in-use technologies. Moreover, our SMS will serve as a basis to continue the work involving UX in the context of text-based chatbots. Also, the development of an evaluation technology that fills the gaps found can be conducted. We hope to contribute to the scientific community, industry, and society in this context. Besides, we expect that our SMS can serve as a basis for future SMSs.

Acknowledgements

This research was funded Coordination for the Improvement of Higher Education Personnel (CAPES) - Program of Academic Excellence (PROEX).

References

Al-Emran, M., AlQudah, A. A., Abbasi, G. A., Al-Sharafi, M. A., and Iranmanesh, M. (2024). Determinants of using ai-based chatbots for knowledge sharing: Evidence from pls-sem and fuzzy sets (fsqca). *IEEE Transactions on Engineering Management*, 71:4985–4999. DOI: https://doi.org/10.1109/TEM.2023.3237789.

Alazraki, L., Ghachem, A., Polydorou, N., Khosmood, F., and Edalat, A. (2021). An empathetic ai coach for self-attachment therapy. In *2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI)*, pages 78–87. DOI: https://doi.org/10.1109/CogMI52975.2021.00019.

Altman, D. G. (1990). *Practical Statistics for Medical Research*. Chapman and Hall/CRC. DOI: http://dx.doi.org/10.1201/9780429258589.

Ampatzoglou, A., Bibi, S., Avgeriou, P., Verbeek, M., and Chatzigeorgiou, A. (2019). Identifying, categorizing and mitigating threats to validity in software engineering secondary studies. *Information and Software Technology*, 106:201–230. DOI: http://dx.doi.org/10.1016/j.infsof.2018.10.006.

Ashktorab, Z., Jain, M., Liao, Q. V., and Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19. ACM. DOI: http://dx.doi.org/10.1145/3290605.3300484.

Bae Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K., and Følstad, A. (2021). When the social becomes non-human: Young people's perception of social support in chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. ACM. DOI: http://dx.doi.org/10.1145/3411764.3445318.

Basili, V. R. and Rombach, H. D. (1988). Towards a comprehensive framework for reuse: A reuse-enabling software evolution environment. In *NASA, Goddard Space Flight Center, Proceedings of the Thirteenth Annual Software Engineering Workshop*, number UMIACS-TR-88-92.

Bawa, A., Khadpe, P., Joshi, P., Bali, K., and Choudhury, M. (2020). Do multilingual users prefer chat-bots that code-mix? let's nudge and find out! *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–23. DOI: http://dx.doi.org/10.1145/3392846.

Benke, I., Knierim, M. T., and Maedche, A. (2020). Chatbot-based emotion management for distributed teams: A participatory design study. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–30. DOI: https://doi.org/10.1145/3415189.

Brooke, J. *et al.* (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7. DOI: http://dx.doi.org/10.1201/9781498710411-35.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. DOI: https://doi.org/10.48550/arXiv.2005.14165.

Cai, W., Jin, Y., and Chen, L. (2022). Impacts of personal characteristics on user trust in conversational recommender systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3491102.3517471.

Cai, W., Jin, Y., Zhao, X., and Chen, L. (2023). "listen to music, listen to yourself": Design of a conversational agent to support self-awareness while listening to music. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3544548.3581427.

Campos, T. P. d., Damasceno, E. F., and Valentim, N. M. C. (2022). Proposal and evaluation of a collaborative is to support systematic reviews and mapping studies. In *XVIII Brazilian Symposium on Information Systems*, pages 1–8. DOI: https://doi.org/10.1145/3535511.3535531.

Candello, H. and Pinhanez, C. (2016). Designing conversational interfaces. *Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*.

Ceha, J., Lee, K. J., Nilsen, E., Goh, J., and Law, E. (2021). Can a humorous conversational agent enhance learning experience and outcomes? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14. DOI: https://doi.org/10.1145/3411764.3445068.

Chen, E. (2022). The effect of multiple replies for natural language generation chatbots. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3491101.3516800.

Chen, J., Chen, C., B. Walther, J., and Sundar, S. S. (2021). Do you feel special when an ai doctor remembers you? individuation effects of ai vs. human doctors on user experience. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7. DOI: https://doi.org/10.1145/3411763.3451735.

Chitturi, R., Raghunathan, R., and Mahajan, V. (2008). Delight by design: The role of hedonic versus utilitarian benefits. *Journal of marketing*, 72(3):48–63. DOI: https://doi.org/10.1509/jmkg.72.3.48.

Dahiya, M. (2017). A tool of conversation: Chatbot. *International Journal of Computer Sciences and Engineering*, 5(5):158–161.

Daniel, R., Purwarianti, A., and Lestari, D. P. (2022). Interaction design of indonesian anti hoax chatbot using user centered design. In *2022 Seventh International Conference on Informatics and Computing (ICIC)*, pages 1–6. DOI: https://doi.org/10.1109/ICIC56845.2022.10007024.

Day, M.-Y. and Shaw, S.-R. (2021). Ai customer service system with pre-trained language and response ranking models for university admissions. In *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 395–401. DOI: https://doi.org/10.1109/IRI51335.2021.00062.

De Nieva, J. O., Joaquin, J. A., Tan, C. B., Marc Te, R. K., and Ong, E. (2020). Investigating students' use of a mental health chatbot to alleviate academic stress. In *6th International ACM In-Cooperation HCI and UX Conference*, pages 1–10. DOI: https://doi.org/10.1145/3431656.3431657.

De Souza, A. C. R., Mariano, P. A. D. L., Guerino, G. C., Chaves, A. P., and Valentim, N. M. C. (2024). Technologies for hedonic aspects evaluation in text-based chatbots: A systematic mapping study. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*, IHC '23, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3638067.3638089.

Denecke, K., Vaaheesan, S., and Arulnathan, A. (2020). A mental health chatbot for regulating emotions (sermo)-concept and usability test. *IEEE Transactions on Emerging Topics in Computing*, 9(3):1170–1182. DOI: https://doi.org/10.1109/tetc.2020.2974478.

Dopler, F. and Göschlberger, B. (2022). Assessing expectations and potential of domain independent corporate learning chatbots. In *2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, pages 135–140. DOI: https://doi.org/10.1109/ICETA57911.2022.9974903.

El Hefny, W., El Bolock, A., Herbert, C., and Abdennadher, S. (2021). Chase away the virus: A character-based chatbot for covid-19. In *2021 IEEE 9th International Conference on Serious Games and Applications for Health(SeGAH)*, pages 1–8. DOI: https://doi.org/10.1109/SEGAH52098.2021.9551895.

El Kamali, M., Angelini, L., Lalanne, D., Abou Khaled, O., and Mugellini, E. (2020). Multimodal conversational agent for older adults' behavioral change. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 270–274. DOI: https://doi.org/10.1145/3395035.3425315.

Elsholz, E., Chamberlain, J., and Kruschwitz, U. (2019).

Exploring language style in chatbots to increase perceived product value and user engagement. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 301–305. DOI: https://doi.org/10.1145/3295750.3298956.

Essop, L., Singh, A., and Wing, J. (2023). Developing a comprehensive evaluation questionnaire for university faq administration chatbots. In *2023 Conference on Information Communications Technology and Society (ICTAS)*, pages 1–7. DOI: https://doi.org/10.1109/ICTAS56421.2023.10082753.

Fadhil, A., Schiavo, G., Wang, Y., and Yilma, B. A. (2018). The effect of emojis when interacting with conversational interface assisted health coaching system. In *Proceedings of the 12th EAI international conference on pervasive computing technologies for healthcare*, pages 378–383. DOI: https://doi.org/10.1145/3240925.3240965.

Fahn, V. and Riener, A. (2021). Time to get conversational: Assessment of the potential of conversational user interfaces for mobile banking. In *Proceedings of Mensch Und Computer 2021*, MuC '21, page 34–43, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3473856.3473872.

Fiore, D., Baldauf, M., and Thiel, C. (2019). " forgot your password again?" acceptance and user experience of a chatbot for in-company it support. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–11. DOI: https://doi.org/10.1145/3365610.3365617.

Flandrin, P., Hellemans, C., van der Linden, J., and Van de Leemput, C. (2022). Smart technologies in hospitality: effects on activity, work design and employment. a case study about chatbot usage. In *Proceedings of the 17th "Ergonomie et Informatique Avancée" Conference*, ErgoIA '21, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3486812.3486838.

Flohr, L. A., Kalinke, S., Krüger, A., and Wallach, D. P. (2021). Chat or tap?–comparing chatbots with 'classic'graphical user interfaces for mobile interaction with autonomous mobility-on-demand systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, pages 1–13. DOI: https://doi.org/10.1145/3447526.3472036.

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., von Wolff, R. M., Hobert, S., and Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 103:2915–2942. REGULAR PAPER. DOI: https://doi.org/10.1007/s00607-021-01016-7.

Gambetta, Z. A., Dessi Puji, L., and Ginar Santika, N. (2021). Calla beauty assistant: Beauty advisory chatbot. In *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, pages 1–6. DOI: https://doi.org/10.1109/ICAICTA53211.2021.9640281.

Guerino, G. C. and Valentim, N. M. C. (2020). Usability and user experience evaluation of conversational systems: A systematic mapping study. In *Proceedings of the XXXIV Brazilian Symposium on Software Engineering*, pages 427–436. DOI: https://doi.org/10.1145/3422392.3422421.

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human–Computer Interaction*, 19(4):319–349. DOI: https://doi.org/10.1207/s15327051hci1904_2.

Hassenzahl, M., Platz, A., Burmester, M., and Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 201–208. DOI: https://doi.org/10.1145/332040.332432.

Höhn, S. and Bongard-Blanchy, K. (2021). Heuristic evaluation of covid-19 chatbots. In *Chatbot Research and Design: 4th International Workshop, CONVERSATIONS 2020, Virtual Event, November 23–24, 2020, Revised Selected Papers*, pages 131–144. Springer. DOI: https://doi.org/10.1007/978-3-030-68288-0_9.

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., and Mctear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? In *Proceedings of the 31st European Conference on Cognitive Ergonomics*, ECCE '19, page 207–214, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3335082.3335094.

Iniesto, F., Coughlan, T., Lister, K., Devine, P., Freear, N., Greenwood, R., Holmes, W., Kenny, I., McLeod, K., and Tudor, R. (2023). Creating 'a simple conversation': Designing a conversational user interface to improve the experience of accessing support for study. *ACM Trans. Access. Comput.*, 16(1). DOI: https://doi.org/10.1145/3568166.

ISO (2019). Ergonomics of human-system interaction — part 210: Human-centred design for interactive systems.

Jain, M., Kumar, P., Kota, R., and Patel, S. N. (2018). Evaluating and informing the design of chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, page 895–906, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3196709.3196735.

Jia, M. and Jyou, L. (2021). The study of the application of a keywords-based chatbot system on the teaching of foreign languages. *Journal of Intelligent & Fuzzy Systems*, pages 1–10. DOI: https://doi.org/10.48550/arXiv.cs/0310018.

Jin, Y., Chen, L., Cai, W., and Pu, P. (2021). Key qualities of conversational recommender systems: From users' perspective. In *Proceedings of the 9th International Conference on Human-Agent Interaction*, HAI '21, page 93–102, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3472307.3484164.

Jin, Y., Zhang, X., and Wang, W. (2019). Musicbot: Evaluating critiquing-based music recommenders with conversational interaction. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–22. DOI: https://doi.org/10.1145/3357384.3357923.

Joshi, A., Kale, S., Chandel, S., and Pal, D. (2015). Likert scale: Explored and explained. *British Journal of Applied Science amp; Technology*, 7(4):396–403. DOI:

http://dx.doi.org/10.9734/bjast/2015/14975.

Jung, J.-Y., Qiu, S., Bozzon, A., and Gadiraju, U. (2022). Great chain of agents: The role of metaphorical representation of agents in conversational crowdsourcing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3491102.3517653.

Kattenbeck, M., Kilian, M. A., Ferstl, M., Alt, F., and Ludwig, B. (2018). Airbot: using a work flow model for proactive assistance in public spaces. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 213–220. DOI: https://doi.org/10.1145/3236112.3236142.

Kernan Freire, S., Niforatos, E., Wang, C., Ruiz-Arenas, S., Foosherian, M., Wellsandt, S., and Bozzon, A. (2023). Lessons learned from designing and evaluating claica: A continuously learning ai cognitive assistant. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, page 553–568, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3581641.3584042.

Kim, S., Lee, J., and Gweon, G. (2019). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12. DOI: https://doi.org/10.1145/3290605.3300316.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report 2.3, EBSE, Ver. 2.3 EBSE Technical Report.

Lankton, N., McKnight, D. H., and Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, 16:880–918. DOI: https://doi.org/10.17705/1jais.00411.

Laugwitz, B., Held, T., and Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Holzinger, A., editor, *HCI and Usability for Education and Work*, pages 63–76, Berlin, Heidelberg. Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/978-3-540-89350-9_6.

Law, E. L.-C., FØLstad, A., and Van As, N. (2022). Effects of humanlikeness and conversational breakdown on trust in chatbots for customer service. In *Nordic Human-Computer Interaction Conference*, NordiCHI '22, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3546155.3546665.

Lee, Y.-C., Yamashita, N., and Huang, Y. (2021). Exploring the effects of incorporating human experts to deliver journaling guidance through a chatbot. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–27. DOI: https://doi.org/10.1145/3449196.

Liu, C., Zhou, S., Zhang, Y., Liu, D., Peng, Z., and Ma, X. (2022). Exploring the effects of self-mockery to improve task-oriented chatbot's social intelligence. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, page 1315–1329, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3532106.3533461.

Liu, Y., Kim, D.-j., Miao, T., and Chuang, Y. (2020). Slumberbot: An interactive agent for helping users investigate disturbance factors of sleep quality. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pages 1–4. DOI: https://doi.org/10.1145/3419249.3420091.

Mafra, M. G. S., Nunes, K., Rocha, S., Braz Junior, G., Silva, A., Viana, D., Silva, W., and Rivero, L. (2024). Proposing usability-ux technologies for the design and evaluation of text-based chatbots. *Journal on Interactive Systems*, 15(1):234–251. DOI: https://doi.org/10.5753/jis.2024.3856.

Mariano, P., Chaves, A. P., and Valentim, N. (2024). Technical report. Technical report, Federal University of Paraná. DOI: http://doi.org/10.6084/m9.figshare.25493947.v2.

mobiletime (2022). Mapa do ecossistema brasileiro de bots 2022. https://www.mobiletime.com.br/pesquisas/mapa-do-ecossistema-brasileiro-de-bots-2022/. Access: 14 August 2024.

Moilanen, J., Visuri, A., Suryanarayana, S. A., Alorwu, A., Yatani, K., and Hosio, S. (2022). Measuring the effect of mental health chatbot personality on user engagement. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*, MUM '22, page 138–150, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3568444.3568464.

Mudofi, L. N. H. and Yuspin, W. (2022). Evaluating quality of chatbots and intelligent conversational agents of bca (vira) line. *Interdisciplinary Social Studies*, 1(5):532–542. DOI: https://doi.org/10.55324/iss.v1i5.122.

Park, S., Thieme, A., Han, J., Lee, S., Rhee, W., and Suh, B. (2021). "i wrote as if i were telling a story to someone i knew.": Designing chatbot interactions for expressive writing in mental health. In *Designing Interactive Systems Conference 2021*, pages 926–941. DOI: https://doi.org/10.1145/3461778.3462143.

Portela, M. and Granell-Canut, C. (2017). A new friend in our smartphone? observing interactions with chatbots in the search of emotional engagement. In *Proceedings of the XVIII International Conference on Human Computer Interaction*, pages 1–7. DOI: https://doi.org/10.1145/3123818.3123826.

Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., and Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, 62(6):785–797. DOI: https://doi.org/10.1016/j.bushor.2019.08.005.

Rapp, A., Curti, L., and Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151:102630. DOI: https://doi.org/10.1016/j.ijhcs.2021.102630.

Ren, R., Castro, J. W., Acuña, S. T., and Lara, J. d. (2019). Usability of chatbots: A systematic mapping study. In *Proceedings of the 31st International Conference on Software Engineering and Knowledge Engineering*, SEKE2019. KSI Research Inc. and Knowledge Systems Institute Graduate School. DOI: http://dx.doi.org/10.18293/seke2019-029.

Ruane, E., Farrell, S., and Ventresque, A. (2021). User perception of text-based chatbot personality. In Følstad, A., Araujo, T., Papadopoulos, S., Law, E. L.-C., Luger, E., Goodwin, M., and Brandtzaeg, P. B., editors, *Chatbot Research and Design*, pages 32–47, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-68288-0_3.

Santos, G., Rocha, A. R., Conte, T., Barcellos, M. P., and Prikladnicki, R. (2012). Strategic alignment between academy and industry: A virtuous cycle to promote innovation in technology. In *2012 26th Brazilian Symposium on Software Engineering*, pages 196–200. DOI: https://doi.org/10.1109/SBES.2012.31.

Schmitt, A., Wambsganss, T., and Leimeister, J. M. (2022). Conversational agents for information retrieval in the education domain: A user-centered design investigation. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2). DOI: https://doi.org/10.1145/3555587.

Schrepp, M., Hinderks, A., and Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.*. DOI: https://doi.org/10.9781/ijimai.2017.09.001.

Sharma, M., Yadav, S., Kaushik, A., and Sharma, S. (2021). Examining usability on atreya bot: A chatbot designed for chemical scientists. In *2021 International Conference on Computational Performance Evaluation (ComPE)*, pages 729–733. DOI: https://doi.org/10.1109/ComPE53109.2021.9752288.

Shawar, B. A. and Atwell, E. (2007). Chatbots: are they really useful? *Journal for Language Technology and Computational Linguistics*, 22(1):29–49. DOI: https://doi.org/10.21248/jlcl.22.2007.88.

Shull, F., Carver, J., and Travassos, G. H. (2001). An empirical methodology for introducing software processes. *ACM SIGSOFT Software Engineering Notes*, 26(5):288–296. DOI: https://doi.org/10.1145/503271.503248.

Skjuve, M., Haugstveit, I. M., Følstad, A., and Brandtzaeg, P. B. (2019). Help! is my chatbot falling into the uncanny valley? : An empirical study of user experience in human-chatbot interaction. *Human Technology*, 15(1):30–54. DOI: https://doi.org/10.17011/ht/urn.201902201607.

Souza, A., Guerino, G., and Valentim, N. (2023). Technical report. Technical report, Federal University of Paraná. DOI: http://doi.org/10.6084/m9.figshare.23145257.v2.

Telner, J. (2021). Chatbot user experience: Speed and content are king. In *Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2021 Virtual Conferences on Human Factors in Software and Systems Engineering, Artificial Intelligence and Social Computing, and Energy, July 25-29, 2021, USA*, pages 47–54. Springer. DOI: https://doi.org/10.1007/978-3-030-80624-8_6.

Torkamaan, H. (2023). Mood measurement on smartphones: Which measure, which design? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(1). DOI: https://doi.org/10.1145/3580864.

Tubin, C., Mazuco Rodriguez, J. P., and de Marchi, A. C. B. (2022). User experience with conversational agent: A systematic review of assessment methods. *Behaviour & Information Technology*, 41(16):3519–3529. DOI: https://doi.org/10.1080/0144929x.2021.2001047.

Veglis, A., Maniou, T. A., *et al.* (2019). Chatbots on the rise: A new narrative in journalism. *Studies in Media and Communication*, 7(1):1–6. DOI: https://doi.org/10.11114/smc.v7i1.3986.

Völkel, S. T. and Kaya, L. (2021). Examining user preference for agreeableness in chatbots. In *Proceedings of the 3rd Conference on Conversational User Interfaces*, pages 1–6. DOI: https://doi.org/10.1145/3469595.3469633.

Wald, R., Heijselaar, E., and Bosse, T. (2021). Make your own: The potential of chatbot customization for the development of user trust. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 382–387. DOI: https://doi.org/10.1145/3450614.3463600.

Wambsganss, T., Kueng, T., Soellner, M., and Leimeister, J. M. (2021). Arguetutor: An adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–13. DOI: https://doi.org/10.1145/3411764.3445781.

Wambsganss, T., Zierau, N., Söllner, M., Käser, T., Koedinger, K. R., and Leimeister, J. M. (2022). Designing conversational evaluation tools: A comparison of text and voice modalities to improve response quality in course evaluations. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2). DOI: https://doi.org/10.1145/3555619.

Watson, K. (2018). Establishing psychological wellbeing metrics for the built environment. *Building Services Engineering Research and Technology*, 39:014362441875449. DOI: https://doi.org/10.1177/0143624418754497.

Xiao, Z., Zhou, M. X., and Fu, W.-T. (2019). Who should be my teammates: Using a conversational agent to understand individuals and help teaming. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 437–447. DOI: https://doi.org/10.1145/3301275.3302264.

Xu, H., Dinev, T., Smith, H., and Hart, P. (2008). Examining the formation of individual's privacy concerns: Toward an integrative view. *ICIS 2008 Proceedings - Twenty Ninth International Conference on Information Systems*.

Yu, D., Tian, J., Su, T., Tu, Z., Xu, X., and Wang, Z. (2021). Incorporating multimodal sentiments into conversational bots for service requirement elicitation. In *2021 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, pages 81–90. DOI: https://doi.org/10.1109/SOSE52839.2021.00014.

Yuen, M. (2022). Chatbot market in 2022: Stats, trends, and companies in the growing ai chatbot industry. https://www.insiderintelligence.com/insights/chatbot-market-stats-trends/. Access: 14 August 2024.

Yun, H., Ham, A., Kim, J., Kim, T., Kim, J., Lee, H., Park, J., and Jang, J. (2020). Chatbot with touch and graphics: An interaction of users for emotional expression and turn-taking. In *Proceedings of the 2nd Confer-*

ence on Conversational User Interfaces, pages 1–5. DOI: https://doi.org/10.1145/3405755.3406147.

Zhang, C., Li, G., Hashimoto, H., and Zhang, Z. (2022). Digital transformation (dx) for skill learners: The design methodology and implementation of educational chatbot using knowledge connection and emotional expression. In *2022 IEEE Global Engineering Education Conference (EDUCON)*, pages 998–1003. DOI: https://doi.org/10.1109/EDUCON52537.2022.9766384.

Zorrilla, A. L. and Torres, M. I. (2022). A multilingual neural coaching model with enhanced long-term dialogue structure. *ACM Trans. Interact. Intell. Syst.*, 12(2). DOI: https://doi.org/10.1145/3487066.