Position paper accepted for CONVERSATIONS 2024 – the 8th International Workshop on Chatbots and Human-Centred AI, hosted by CERTH, Thessaloniki, Greece, December 4-5, 2024.

Explorations on Using Facial Expression Data for Studying Emotional Reactions during Chatbot Interaction

Lili Aunimo[1[0000-0002-3336-0948]], Ilona Pezenka[2[0000-0003-3963-1259]]

and David Dobrowsky[2[0000-0002-4646-0681]]

[1] Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland
[2] FHWien der WKW University of Applied Science, Währingergürtel 97, 1180 Vienna, Austria

Abstract. We present an exploratory study on emotional reactions during chatbot interaction. The focus is both on exploring points of maximal emotional intensity during chatbot interaction and on creating an annotation scheme for this type of studies. Facial expression data is collected and analyzed with the iMotions software. Initial results show that there is variation in emotional reactions of informants during different phases of interaction. In our study, we could identify one phase that was considerably more emotionally intensive than the others. Additionally, we identified three potential sources of emotion that were not related to the phase of the interaction. These were points when the chatbot answered wrong, took initiative or redirected the user to another media source. We also found that the scale of emotional intensity varies considerably between informants. Thus, we suggest that emotional expressions could be used as a part of chatbot interaction studies as they give valuable information regarding the user experience. Moreover, we suggest that emotion measures are normalized for each informant, making them comparable across individuals with varying emotional intensities.

Keywords: Chatbot, emotional reactions, facial expression analysis, interaction, user experience.

1 Introduction

Chatbots are becoming more popular and important in a variety of fields, including customer service (Haugeland, Følstad, Taylor & Bjørkli, 2022), health (Laranjo et al., 2018), education (Smutny & Schreiberova, 2020), foodservice (De Cicco et al., 2021), and tourism (Calvaresi et al., 2021), to mention a few. The online service ChatGPT had the fastest user growth in Internet history until July 2023. It reached 1 million users in just 5 days, and it was the fastest growing application until July 2023 when Threads took over (Duarte, 2024). Companies, on the other hand, saw and used the enormous variety of potential for this technology even before the creation of tools like ChatGPT, as studies have shown that interactive communication is more successful than static communication (Go & Sundar, 2019).

Chatbots are a text-based form of conversational agents (Nißen et al., 2022). They are defined as software-based systems designed to interact with humans via text-based natural language (Feine et al., 2020). Many chatbots are domain based and therefore have a limited knowledge base (Gnewuch et al., 2017). There is a relevant differentiation of chatbots by their interaction style. Task-oriented chatbots help users to execute a task or solve a problem, whereas social oriented chatbots try to create and maintain a good quality conversation or to establish relationships (Chattaraman et al., 2019; De Cicco et al., 2021; Nißen et al., 2022; 80; Rapp et al., 2021).

So far, we know very little about the emotions that arise during a chatbot interaction and that ultimately have an impact on factors such as trust, intention to further interact with the chatbot, and satisfaction. Thus, this study explores the emotional responses of informants during chatbot interaction using facial expression data. Traditional methods such as questionnaires regarding usability, user acceptance and perceived emotions complement the use of biometric data. Using biometric data may offer an objective and complementary view to the informants' emotions.

This research sets out to answer the following questions:
Are there some specific phases in the conversation with the chatbot where emotions arise? How to annotate and analyze biometric informant data to make the above study systematic and transparent?

To answer the above questions, the data annotation and analysis scheme must first be considered. Facial expression data often contains many variations because human expressiveness varies greatly from person to person, among others. Furthermore, an analysis of the entire recording may not be necessary when addressing the emotionally loaded points in the interaction. Thus, first an annotation scheme must first be developed. This paper addresses the development of the annotation scheme and gives an exploratory view into the interaction data.

In the remainder of the paper, we first describe related work, then address the data, the research setting and the creation of the annotation scheme. Results of the initial exploratory analysis are then presented. We conclude the paper with a short analysis and discussion of the significance of the results and draw the conclusions of our study.

2 Related Work

Only a few studies in the field of chatbot-human interaction have used facial expression analysis to date. Most studies in the field use quantitative methods and are based on surveys (Rapp et al., 2021; [8]. Nicolescu and Tudorache, (2022) report that most studies on human-computer interaction on AI chatbots in customer service are methodologically based on an experimental setting with chatbot-customer interaction, followed by questions related to the interaction experience. The work using biometric data to study chatbot-human interaction is scarce. Picoli at al. (2024) report initial results on how biometric data can be used to recognize the emotions of the users during interaction with a chatbot. Ciechanowski et al. (2019) [6] have used biometric data to study the uncanny valley effect in chabot-human interaction, but they do not use facial expression data. However, to the best of our knowledge this is the first study exploring how facial expression data could be used to study chatbot-human interaction.

Even though the work on human-chatbot interaction using facial expression data is scant, there is some previous research in the field of human-computer interaction using facial expression data. For instance, one of the early works on facial expression analysis in the context of human-computer interaction was done by Ward (2004). Ward's (2004) study investigates the relationship between facial and physiological reactions to computer-based events, as well as the feasibility of using facial expression analysis to detect and discern qualitative changes in users' facial movements while completing a web-based quiz. The findings that users' facial expressions do respond to common computer-based events suggest that facial expression analysis is becoming a feasible approach for such investigations. Sometime after Ward's (2004) study, Shrammel et al. (2009) analyzed eye movements, pupil size, and facial electromyographic (EMG) activity while participants interacted with virtual avatars who were either staring at them or at someone else. Furthermore, Bellur and Sundar (2010) studied users' psychophysiological responses to structural characteristics of user interfaces. They employ EEG to detect attention, GSR to detect arousal, and EMG of facial muscles to detect valence. More recently, Ciechanowski et al. (2018) employed electromyography and a series of questionnaires to analyze interactions with avatars that are graphically very similar to humans and a text chatbot without any avatar. Johanssen et al. (2019) were the first questioning the interpretation of data. They examined correlations between perceived emotions and usability problems in interactive mobile apps by analyzing facial expression data. Since users have diverse reactions to the same usability issues, the authors propose that user emotions be gathered and analyzed alongside other knowledge sources, such as user interaction events. Furthermore, the authors recommend monitoring only individual emotional reaction graphs and keeping a look out for dramatic changes for a short period of time, since such shifts typically indicate unexpected software behavior.

Samara et al. (2019) express similar concerns. They collected facial expression data while respondents engaged with software and completed various activities. The study's major findings show that users' facial expressions cannot accurately communicate their true emotions. Furthermore, participants' facial expressions vary more during active engagement tasks than during passive interaction. Further they found weak correlations, and in some cases inconsistency, between the individual and combined reported levels of valence and arousal derived from facial expressions. Thus, the results of this study also suggest that analysis on an individual level is necessary.

Overall, earlier research in the field of human-computer interaction that used facial expression analysis demonstrates that the data can provide quite meaningful results, but this also requires a thorough analysis. Thus, in this study, we use data from a human-chatbot conversation to demonstrate how to prepare and analyze this type of data in a useful way.

3 Data and Methods

3.1 The Data

We collected and analyzed data from ten informants. The informants were students in business administration and thus belonged to the target group of the chatbot. To be able to do the research, permission was requested from the data protection and ethics committee of the university in question. Before being asked to interact with the chatbot, the informants were provided an informed consent form to approve. After this, the informants were asked to turn on their webcams and to interact with the chatbot online to obtain information about a study program. The task is given in Table 1.

Table 1. The task that the informants should complete when interacting with the chatbot.

Imagine you are working full-time in a company as a specialist, and you would like to strengthen your expertise in digital marketing. You have no prior knowledge about the study program of Digital Communication & Marketing. You interact with the chatbot to figure out if the study program would suit you. Please try to find information on the following aspects through the bot:
1. Timing of the courses
2. Contents of the study program
3. Ways of applying to the study program
4. Criteria of acceptance to the study program
Afterwards, try to clarify any other question you have regarding the program with the bot.

To communicate with the chatbot, participants needed to have a Facebook Messenger account. During the conversation with the bot, their faces were recorded online via a webcam. To analyze the data, we used iMotions, a software platform that includes several biometric technologies (eye tracking, facial expression recognition, galvanic skin reaction, and electroencephalography). The program employs Affectiva Inc.'s AFFDEX algorithm (El Kaliouby & Robinson, 2005) to post-process the videos to determine emotions. The technique is based on Ekman and colleagues' Emotional Facial Action Coding System (EMFACS) mappings (Ekman & Friesen, 2003; Ekman & Rosenberg, 1997) and use categorized images as a training database. The iMotions program identifies changes in important facial characteristics (such as brows, eyes, and lips) and predicts the likelihood of the seven main emotions, namely joy, surprise, anger, sadness, fear, disgust and contempt. In addition to the seven basic emotions, two other metrics are calculated, namely engagement and valence. These are depicted in Figure 1. Engagement is a general measure of total expressiveness. It is derived by taking the average of the greatest evidence scores from the higher (brow raise, brow furrow, nose wrinkle) and lower (lip corner depressor, chin raise, lip pucker, lip press, mouth open, lip suck, grin) facial regions. Valence is a measure of the participant's pleasant or bad experience. Smile and cheek lift are two factors that improve the chance of positive valence. Inner brow raise, brow furrow, nose wrinkle, upper lip raise, lip corner depressor, chin raise, lip press, lip suck are all factors that enhance the possibility of negative valence (iMotions Software, Facial, 2024).

For emotion recognition, the iMotions software has as the default setting at a threshold of 50 %. This is the likelihood that the recognized facial expressions match the specific emotions discovered. A 50% chance represents a reasonably strong manifestation of facial reaction. Only emotions above this threshold likelihood are considered.

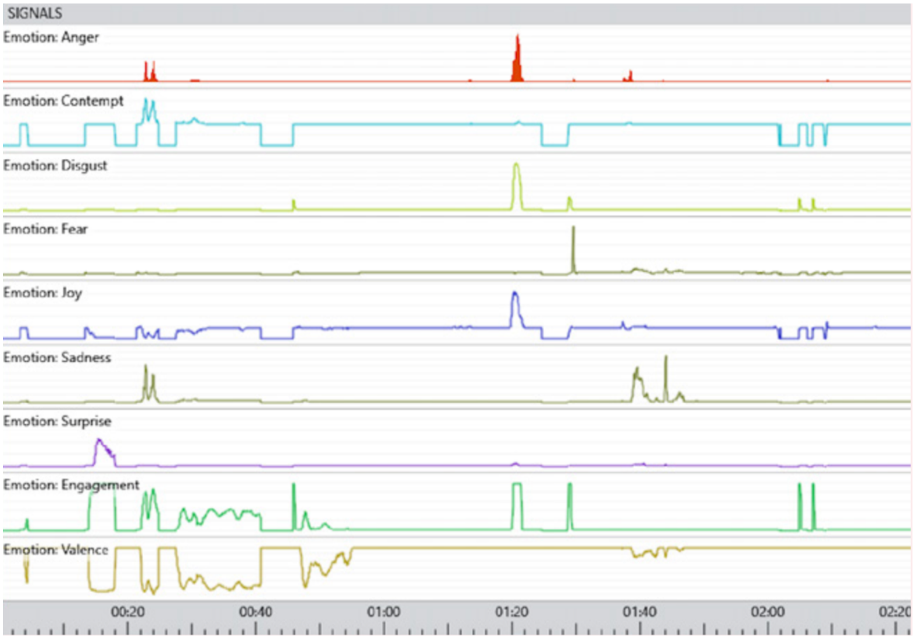

Figure 1. An imaginary example of the emotion analysis of an informant by iMotions during interaction with the chatbot. In addition to emotions, also engagement and valence are measured.

3.2 The Chatbot

The chatbot was constructed with Chatfuel (Chatfuel, 2024), Google Dialogflow (Dialogflow, 2024), and Janis (Janis, 2024) and implemented in Facebook Messenger. The Facebook Messenger chatbot design (graphics, buttons) was created using the Chatfuel program. Janis' software solution (as a Slack app) was used to transfer free text input from test subjects to Google Dialogflow through Facebook Messenger and Chatfuel. Janis then transmits the responses back to Chatfuel, which outputs them via the Messenger chatbot, after Google Dialogflow assigns the text input to the proper intents. Each Chatfuel button was mapped to an intent in Google Dialogflow. Texts saved in Google Dialogflow were used to answer further questions. The information on the master's degree program is organized into around 100 intents. The three main subject areas are the curriculum content, the organization of the studies, and application processes and costs.

The chatbot is available on the university's English website. Although it resembles an educational chatbot, it is more correctly defined as a service chatbot giving information on an educational product. The chatbot does not conduct any pedagogical conversations.

4 The Analysis Framework

An analysis framework was developed with the goal of enabling the systematic analysis of emotions related to the different phases of the task and to certain potentially emotionally significant behaviors of the chatbot such as the bot giving a wrong or incomplete answer, taking the initiative or redirecting to external content such as videos or pdf documents. The task is depicted in Table 1 and the annotations related to it are those listed under *Task-related features* in Table 2.

To discover emotionally significant events, we inspected the maxima of each emotion in the interaction with the chatbot of each informant. We inspected these points with detail to find out if something unexpected or something that typically arises emotions happened at that point in the conversation. From studies that have analyzed the impact of chatbot service errors on user emotions (Huang & Dootson, 2022; Zhang et al, 2024), we know that wrong or incorrect answers by the chatbot can be sources of negative emotions. This analysis was conducted by first classifying each maximum into one of the five phases of the task: *Timing, Content, Application*, *Acceptance* and *Free*. Thereafter, the maximum point in the interaction was inspected for other possible sources of emotional reaction. This analysis resulted in creating three new annotations: *Wrong, Initiative* and *Redirection*. We call these annotations *other features*. They are depicted in Table 2. These annotations are not related to the task and co-occur with task-related annotations.

Table 2. The scheme for annotating the data of each informant. The *task-related features* refer to the phases on the task described in Table 1. The *other features* refer to actions of the studied chatbot that resulted in emotional reactions.

	Task-related features
Timing	Conversation related to the timing of the study program.
Contents	Conversation related to the course contents of the study program.
Application	Conversation related to the application procedure and schedule.
Acceptance	Discussion related to the acceptance criteria to the program.
Free	Any discussion on the topic freely chosen by the informant.
	Other features
Wrong	The utterance produced by the chatbot is content wise wrong or incomplete.
Initiative	The chatbot has taken the initiative regarding the topic of the conversation.
Redirection	The chatbot redirected the informant to some external site, such as a video in YouTube or a pdf document.

Figure 2 shows the results of the emotion analysis of an imaginary informant while performing the task given in Table 1. The data is annotated with the annotations described in Table 2. We can see in the figure for instance, that the maximum value for *anger* appears when the informant is searching for information about the *contents* of the study program. The results of our study in the next section are presented in a tabular format. For illustration purposes, we show in Table 3 how the imaginary information in Figure 2 would be represented in tabular format.

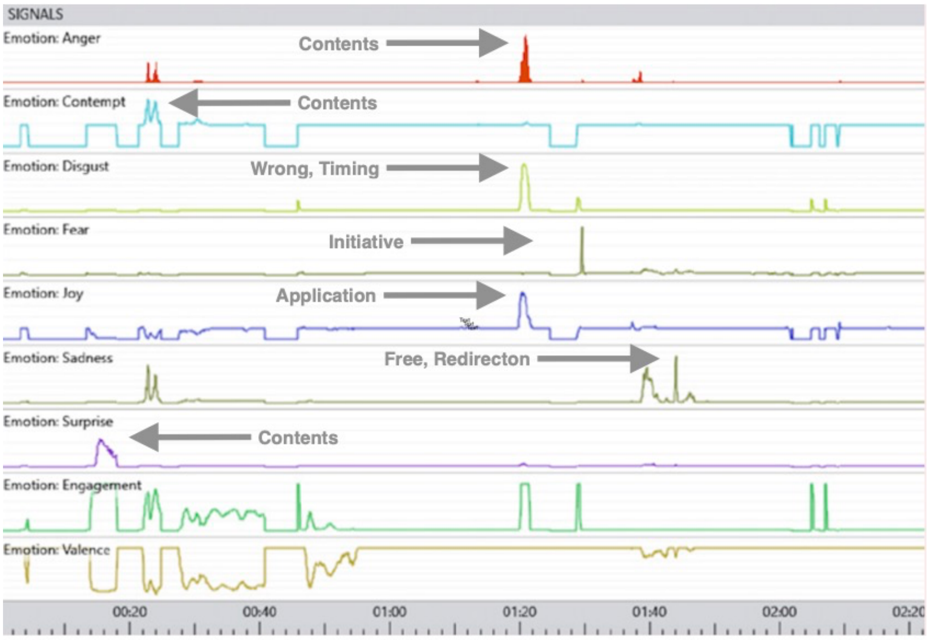

Figure 2. Example of applying the proposed annotation scheme to imaginary emotion recognition data.

Table 3. The information in Figure 2 presented in a tabular format.

		Task-related features						*Other features*		
		Timing	Content	Application	Acceptance	Free	Sum	Wrong	Initiative	Red.
Emotions	Anger (1)	0	1	0	0	0	1	0	0	0
	Contempt (1)	0	1	0	0	0	1	0	0	0
	Disgust (1)	1	0	0	0	0	1	1	0	0
	Fear (1)	0	0	0	0	0	1	0	1	0
	Joy (1)	0	0	1	0	0	1	0	0	0
	Sadness (1)	0	0	0	0	1	1	0	0	1
	Surprise (1)	0	1	0	0	0	1	0	0	0
	Sum	1	3	1	0	1		1	1	1

5 Results

We found out that the informants discussed the topics related to the task several times during their interaction with the chatbot. Additionally, the order in which the task was accomplished was not necessarily the one given in the task description and the time spent discussing one topic varied. We validated the annotation scheme presented in Table 2 by using it to analyze the videos on chatbot interaction of the informants. A summary of the results is presented in Table 3.

Table 3. The number of times each emotion reaches its maximum value in different phases of the conversations. For example, the emotion *Anger* reaches its maximum value most often when the *content* of the study program is under discussion. All in all, the maximum is reached nine times as can be observed in the column *Sum*. In one case *anger* reaches its maximum value when the chatbot gives a wrong answer. This is denoted by the number 1 in the column *Wrong*. The figures in parenthesis after each emotion illustrate the number of informants whose facial expressions demonstrated the corresponding emotion during the conversation with the chatbot.

		Task-related features						*Other features*		
		Timing	Content	Application	Acceptance	Free	Sum	Wrong	Initiative	Red.
Emotions	Anger (9)	0	4	2	0	3	9	1	0	0
	Contempt (9)	1	3	3	0	0	7	0	0	0
	Disgust (9)	1	2	3	2	1	9	3	0	0
	Fear (10)	2	4	1	2	1	10	0	1	1
	Joy (10)	0	4	1	3	1	9	0	0	0
	Sadness (7)	1	1	0	1	1	**4**	1	0	0
	Surprise (9)	1	2	1	3	2	9	1	1	0
	Sum	6	**20**	11	11	9		**6**	2	1

Table 3 shows that *sadness*, appearing in the data of 7 informants, is the emotion that the software most seldom recognizes on the faces of the informants. The other emotions: *anger, contempt, disgust, fear, joy* and *surprise* with 10 or 9 occurrences are more common and appear in the interaction data of almost every informant. In this data, the maximum number of occurrences that one emotion may get is 10, which corresponds to the number of informants. This is explained by the annotation procedure: First, the maximum for each emotion for each informant is sought. If no maximum point is found, it means either that there are several points with the same maximum value or that the emotion in question has not been detected at all because it is under the confidence threshold of 50% that was set in the iMotions software. If the emotion has several maximum points, it is marked as appearing, but it cannot be classified under any tag. Secondly, each maximum point is classified according to the task-related tag and checked if it also represents one of the other annotations.

Table 3 also shows that the discussion related to contents of the study program is associated with more emotions than the other topics. This is denoted by the number 20 on the row *Sum* in the column *Content* of the table. However, this may also be just because this may be the predominant topic of discussion, and it may in terms of time spent take a large part of the entire interaction time with the chatbot. To determine more accurately the importance of the emotion of the informant when interacting with the chatbot on a specific task, the total time spent discussing each task should be calculated for each informant. This also applies for the other annotations: *wrong*, *initiative* and *redirection*. For example, if the discussion on acceptance criteria to the study program only accounts for 5% of the total interaction time of an informant, but for 20% of the emotional maxima, then we could state that the interaction on this topic is arising more emotions in the informant than would be expected, 5 % being the baseline if the emotions were evenly distributed among topics.

These results demonstrate how the annotation scheme created may be applied in a case where chatbot interaction videos are analyzed to discover emotionally loaded phases in the interaction between the user and the chatbot.

6 Conclusions

This paper presented an exploratory study on emotional reactions during chatbot interaction. The focus was two-fold: Firstly, we attempted to identify emotionally loaded phases in the interaction and secondly, we created an annotation scheme for the systematic analysis of affective data on chatbot interaction. Initial results show that there is variation in the emotional reactions of the informants during different phases of interaction. We also found out that the scale of emotional intensity varies between informants. Thus, we suggest that either the emotion measures are normalized for each informant or that just the maximum values of the emotions are used. These two approaches make the values recorded by the iMotions software comparable across individuals with varying emotional intensities.

The body of work using facial expression analysis to examine emotions experienced by the user during chatbot interaction is scant. This study shows how facial expressions may be used in such a study. We also demonstrate how the maximum values of each emotion may be used to determine the phase where the emotion was most salient during the interaction. The use of maxima instead of the absolute numeric values enables us to study in a uniform manner informants with large and small variations in facial expressions. This is useful as some informants showed strong facial expressions while other had very minimal facial expressions.

Implications of our findings for the design of chatbots can be derived to aim for an improved user experience: Emotional reactions can vary from individual to individual depending on many factors such as the relevance of the content with regard the task that the user is trying to accomplish or style of interaction such as the chatbot taking the initiative and directing the conversation. It would make sense for chatbots to understand the emotional reactions of users and adapt their interaction, for example by trying to increase the level of interactivity for the contents in question. This study presented a framework that could be used both for the long-term development of chatbot-human interaction or in real-time for modifying the interaction of the chatbot with the current user. For the real-time modification of interaction, the webcam and the facial recognition software would have to be accepted by the users as they are very privacy-invasive technologies.

Our work has several limitations. Firstly, a sample of ten informants is relatively small. However, it may serve as a validation of the annotation scheme developed and as a preliminary study on the data. It may be noted that the number of informants in studies based on biometric data are often small, and the focus is rather on a detailed study of the data. Studies on chatbot-human interaction based on facial expression data are scant and thus even a study based on a small number of informants may contribute to the field. Naturally, generalizations regarding the sources of emotional arousal in the conversation cannot be made with such a small sample size.

A second limitation of our study is that in cases where the variance within one informant in the intensity of the recognized emotion is small, the maxima may not be of importance. This may be true, and in a future study it could be beneficial to observe also the variance in the emotions of an informant, or even the distribution of the emotional values, rather than just observing the maxima. However, it should be noted that we did not consider emotion values whose likelihood was very low. Rather, we kept the threshold value of 50%, which is the default threshold for an emotion to be considered in the iMotions software. It means that we only considered those emotions that were recognized on the respondents face with a certainty of 50% or more.

In future work, we could study chatbot-human interaction with a larger number of informants. This would show if the presented analysis scheme covers the phases in the conversation that are emotionally important or if completely new sources of emotional arousal could be found. This would also enable us to gain more generalizable results on the sources of emotion and their types in chatbot-human interaction.

Acknowledgments. This study was partly funded by the Finnish Ministry for Education and Culture through the AIDriver! project.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Bellur, S., & Sundar, S. S. (2010). Psychophysiological responses to media interfaces. ACM SIGCHI, Atlanta, GA, USA.

Calvaresi, D., Ibrahim, A., Calbimonte, J.-P., Schegg, R., Fragniere, E., & Schumacher, M. (2021). The Evolution of Chatbots in Tourism: A Systematic Literature Review. In W. Wörndl, C. Koo, & J. L. Stienmetz (Eds.), Information and Communication Technologies in Tourism 2021 (pp. 3–16). Cham: Springer International Publishing.

Chatfuel homepage, https://chatfuel.com/, last accessed 25.9.2024

Chattaraman, V., Kwon, W. -S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. Computers in Human Behavior, 90, 315–330. https://doi.org/10.1016/j.chb.2018.08.048

Ciechanowski, L., Przegalinska, A., & Wegner, K. (2018). The necessity of new paradigms in measuring human-chatbot interaction. In Advances in Cross-Cultural Decision Making: Proceedings of the AHFE 2017 International Conference on Cross-Cultural Decision Making, July 17-21, 2017, Los Angeles, California, USA 8 (pp. 205-214). Springer International Publishing.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. Future Generation Computer Systems, 92, 539–548. https://doi.org/10.1016/j.future.2018.01.055

De Cicco, R. de, Da Silva, S. C. L. C. e., & Alparone, F. R. (2021). “It’s on its way”: Chatbots applied for online food delivery services, social or task-oriented interaction style? Journal of Foodservice Business Research, 24(2), 140–164.

Dobrowsky, D., Aunimo, L., Janous, G., Pezenka, I., & Weber, T. (2021). The influence of interactional style on affective acceptance in human-chatbot interaction–A literature review. Proceedings of the HHBIC 2021. Esignals Research. ISSN 2736-9323.

Gnewuch, Ulrich & Morana, Stefan & Maedche, Alexander. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service.

Google Dialogflow. https://dialogflow.cloud.google.com/. Last accessed 24.9.2024.

El Kaliouby, R., & Robinson, P. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In B. Kisačanin, V. Pavlović & T. S. Huang (eds.), Real-time Vision for Human-Computer Interaction (pp. 181–200). New York, Springer.

Ekman, P., & Friesen, W. V. (2003). Unmasking the face: A guide to recognizing emotions from facial cues. 3rd Edition. Cambridge, Malor Books.

Ekman, P., & Rosenberg, E. L. (1997). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford, Oxford University Press.

Fabio Duarte (2024). Number of ChatGPT Users (2024). Exploding topics. https://explodingtopics.com/blog/chatgpt-users (last visited 25.9.2024)

Facial expression analysis, iMotions Software. https://imotions.com/affectiva/ . Last accessed 25.9.2024.

Feine, J., Adam, M., Benke, I., Maedche, A., and Benlian, A. (2020). "Exploring Design Principles for Enterprise Chatbots: An Analytic Hierarchy Process Study". In: Proceedings of the 15th International Conference on Design Science Research in Information Systems and Technology (DESRIST), Kristiansand, Norway

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior, 97, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. International Journal of Human-Computer Studies, 161, 102788. https://doi.org/10.1016/j.ijhcs.2022.102788

Huang, Y.-S., & Dootson, P. (2022). Chatbots and service failure: When does it lead to customer aggression. Journal of Retailing and Consumer Services, 68, 103044. https://doi.org/10.1016/j.jretconser.2022.103044

Janis software, https://www.janis.ai/. Last accessed 24.9.2014.

Johanssen, J. O., Bernius, J. P., & Bruegge, B. (2019). Toward usability problem identification based on user emotions derived from facial expressions. In 2019 IEEE/ACM 4th International Workshop on Emotion Awareness in Software Engineering (SEmotion) (pp. 1-7). IEEE.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., . . . Lau, A. Y. S. (2018). Conversational agents in healthcare: a systematic review. Journal of the American Medical Informatics Association, 25(9), 1248–1258.

Nicolescu, L., & Tudorache, M. T. (2022). Human-computer interaction in customer service: the experience with AI chatbots—a systematic literature review. Electronics, 11(10), 1579.

Nißen, M., Selimi, D., Janssen, A., Cardona, D. R., Breitner, M. H., Kowatsch, T., & Wangenheim, F. V. (2022). See you soon again, chatbot? A design taxonomy to characterize 690 user-chatbot relationships with different time horizons. Computers in Human Behavior, 127,107043. https://doi.org/10.1016/j.chb.2021.107043

Picoli, H. S., Oliveira, C. L. T. D., de Araujo, L. P. M., Manzato, M. G., & Rodrigues, K. R. D. H. (2022, November). Human-ChatBot Interaction: measuring the psychophysiological reactions of chatbot users. In Proceedings of the Brazilian Symposium on Multimedia and the Web (pp. 377-380).

Rapp, A., Curti, L., & 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.

Samara, A., Galway, L., Bond, R., & Wang, H. (2019). Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10, 2175-2184.

Schrammel, F., Pannasch, S., Graupner, S.-T., Mojzisch, A., Velichkovsky, B.M. (2009). Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. Psychophysiology 46, 922–931

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. Computers & Education, 151, 103862.

Ward, R. (2004). An analysis of facial movement tracking in ordinary human–computer interaction. Interacting with Computers, 16(5), 879-896.

Zhang, J., Lu, X., Zheng, W., & Wang, X. (2024). It's better than nothing: The influence of service failures on user reusage intention in AI chatbot. Electronic Commerce Research and Applications, 67, 101421. https://doi.org/10.1016/j.elerap.2024.101421