A Review of Emotions in Human-Conversational Agent Interaction

Krutheeka Baskaran, Wei Cui, Atreyi Kankanhalli
National University of Singapore
kruth.b@nus.edu.sg, discw@nus.edu.sg, atreyi@comp.nus.edu.sg

Abstract
Conversational agents (CA), powered by natural language processing, have become increasingly popular across multiple domains. However, these agents often fail to communicate effectively with users, leading to poor adoption and task outcomes. Emotions are a fundamental aspect of such interactions, influencing use and adoption of digital artifacts. Despite their salience, understanding of the role of emotions in human-CA collaboration remains fragmented. Motivated thus, we review empirical studies on emotions in human-CA interactions. We synthesize the findings from the reviewed studies in terms of antecedents, emotion-related outcomes, and their relationships, in the form of a descriptive model. Based on the synthesis we identify knowledge gaps and propose directions for future research. Our analysis provides insights into the role of emotions in human-CA interactions and contributes to research in this area.

Introduction
In recent years, conversational agents (CAs) are increasingly being implemented in multiple domains like education, customer service, and healthcare due to their growing capabilities and potential for cost savings (Diederich et al. 2022). These agents draw on natural language processing and other techniques to interact and collaborate with users to perform varied tasks, such as personal assistance e.g., Siri (Rhee and Choi 2020) and customer service (Ashktorab et al. 2019). In spite of advances in CA capabilities, these agents often fail to communicate effectively with users (Weiler et al. 2022). Designing CAs to meet user expectations entails understanding the nature of emotions evoked in human-CA interactions (Rapp et al. 2021).

Emotions are a fundamental aspect of human communication and human-machine interaction (Brave and Nass 2002). They influence the way we think and interact with each other and with technology. Prior work suggests that CAs which evoke positive emotions are adopted by users (Hudlicka 2003). On the other hand, irrelevant responses from a chatbot resulted in negative user emotions and abandonment of the CA (Brendel et al. 2020). Thus, there is a need to understand the nature of emotions in human-CA interactions, with the aim to enhance their communication with users and improve CA performance.

However, we lack a holistic understanding of the role of emotions in human-CA interactions cumulated from studies in this area. In this regard, several literature reviews have examined CA-related studies broadly (Diederich et al. 2022; Zierau et al. 2020), while some reviews have addressed specific CA application contexts, like healthcare (e.g., ter Stal et al. 2020) and education (Weber et al. 2021). Others have reviewed research on specific CA characteristics such as their social cues (Feine et al. 2019) or text-based communication (e.g., Rapp et al. 2021). Among the CA research reviews, we lack a review of the role of emotions in human-CA interactions. Considering the importance of emotions in human-CA interactions, it would be valuable to provide a holistic view of research findings in this area.

Motivated thus, we perform a review of quantitative empirical studies in this area and synthesize the findings in terms of antecedents, emotion-related outcomes, and their relationships. We contribute by identifying and organizing the empirical literature and synthesizing the findings into a descriptive model. This allows us to derive knowledge gaps and propose future research directions on this topic.

Conceptual Background

Conversational Agents
Conversational agents are digital artifacts that interact with humans through natural language to collaborate on a variety of tasks (Diederich et al. 2022), e.g., the popular ChatGPT. These agents are sometimes powered by artificial intelligence (AI) and are referred to synonymously as AI agents or chatbots. With advances in CA design, they aim to mimic human communication with a high degree of interaction and intelligence (Feine et al. 2019). CA design elements can be classified into 5 types i.e., verbal, visual, auditory, invisible,
and interaction (Zierau et al. 2020). The verbal category includes elements that are expressed through words in either written or spoken form, such as CA conversation style and content. The visual and auditory categories include elements that interact via humans’ vision (e.g., CA appearance) and hearing (e.g., CA voice quality), respectively. CA design elements that cannot be sensed directly by hearing, seeing, or words, such as CA personality, are considered as invisible. The interaction category refers to CA elements representing the interaction structure e.g., communication mode being text or speech (Zierau et al. 2020). We use these five categories to classify CA characteristics in our review.

Emotion

Emotion is a fundamental aspect of human communication. Emotion refers to subjective feelings induced by or attributed to a specific stimulus (Zhang 2013), which is usually episode-based and short-lived (Russell 2003). In the past, emotions were categorized into a set of discrete basic emotions such as anger, disgust, fear, happiness, sadness and surprise (Ekman 1992). However, this view was found to be limited in its ability to explain other aspects of emotions. To address this limitation, researchers proposed the circumplex model of affect which suggests two additional dimensions i.e., valence and activation (Russell 2003). According to this model, valence denotes positive or negative reactions towards a stimulus, while activation reflects the intensity or arousal level.

Emotions underlie user experiences and decisions to adopt and use new information systems (Beaudry and Pinsonneault 2010; de Guinea and Markus 2009). For instance, positive emotions elicited by digital artifacts could motivate individuals to continue using them e.g., joy when interacting with a CA (Yang et al. 2019). In contrast, negative emotions evoked by digital artifacts could result in their abandonment e.g., frustration with using a CA (Riquel et. al 2021). Due to their salience, we focus on understanding the role of emotions in human-CA interactions.

Research Approach

We reviewed quantitative empirical literature on emotions in human-CA interactions as per procedures given in Okoli (2015), which include paper search, selection, coding, and analysis of the findings. First, a search query was formulated combining CA-related and emotion-related terms, followed by identifying relevant search outlets. Specifically, we searched journals and conference proceedings from related fields like information systems (IS), human-computer interaction (HCI) and computer science. Accordingly, we searched the AIS, ACM, IEEE, and Web of Science databases to identify relevant studies on the focal topic. Second, papers were screened based on our inclusion and exclusion criteria (see Table 1) – by going through the abstracts and then the full-text of the papers. Through this process, we finally selected 36 papers for our review. Most of these papers were from HCI and IS outlets. This is because the computer science papers mostly did not statistically evaluate the relationships between antecedents i.e., independent variables (IVs) and outcomes, i.e., dependent variables (DVs).

For the final set of 36 papers, we coded the IVs and DVs. For the studies with emotion aspects as DVs, we analyzed their relationships with the IVs. We used p < 0.05 as the requirement for significant relationships. We applied Zierau et al.’s (2020) classification of CA design elements to code our CA-related IVs, supplemented by our own categories where needed. In total, 57 relationships between the IVs and emotion as the DV were identified, of which 30 were positive, 9 were negative and 18 were not significant.

Results

We present our findings in four sections. The first section reports descriptives of the 36 papers reviewed. The second section summarizes papers where emotion appeared as antecedent (IVs). The third section sums up studies where emotion was an outcome (DV). The last section presents a descriptive model that synthesizes the findings between antecedents and emotion as outcome.

Descriptives

While we found papers quantitatively studying the role of emotions in human-CA interactions from 2010, our analysis showed a significant rise in studies from 2018 onwards,

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Paper examines human-CA interaction. AI agent or chatbot is often used as synonym for CA.</td>
<td>(1) Paper is not empirical (e.g., literature review)</td>
</tr>
<tr>
<td>(2) Paper assesses human users’ and/or CAs’ emotions quantitatively.</td>
<td>(2) Paper studies one-way communication between human and CA rather than two-way conversation (e.g., students watching a video of a pedagogical agent).</td>
</tr>
<tr>
<td></td>
<td>(3) Paper adopts Wizard-of-Oz setting where CAs are not truly automated (i.e., subjects perceive the responses as coming from CAs, but are actually replied by human researchers).</td>
</tr>
<tr>
<td></td>
<td>(4) Paper asks users to evaluate the CA by only imagining their interaction with it (e.g., through looking at CAs’ images or scenarios, not through actual dialogues).</td>
</tr>
<tr>
<td></td>
<td>(5) Paper performs a qualitative study.</td>
</tr>
<tr>
<td></td>
<td>(6) Paper does not statistically assess the relationships between antecedents and outcomes.</td>
</tr>
</tbody>
</table>

Table 1: Paper inclusion and exclusion criteria

The reviewed studies were situated in various application contexts. The most studied application was customer service with 13 papers, followed by social CAs with 5 papers and personal assistance CAs with 4 papers. Business services, healthcare, and team collaboration CAs were examined in 3 papers each. Further, education and gaming CAs were studied in 2 papers each. One paper explored CAs in the e-commerce context. In terms of research methods, experiments were most commonly used (26 papers), with 24 studies conducting lab experiments and 2 studies in field conditions. The second most common method was a combination of experiment and survey (5 papers) followed by survey (4 papers). One paper combined experiment, survey, and interviews.

**Emotion as Antecedent**

Of the 36 papers reviewed, 9 papers examined emotions between humans and CAs solely as antecedents. Among these, majority of the papers (7 papers) studied how CA detecting or expressing emotions influenced human perception towards the CA e.g., trust, intimacy, or attachment towards the CA (e.g., Huber et al. 2018). Further, one paper explored the moderation effect between pre-interaction emotion and CA’s social presence on the intention to use it (Mozafari et al. 2021). The remaining paper examined the relationship between human control over emotion-aware chatbots and its effect on trust in the chatbot and human autonomy (Benke et al. 2022).

**Emotion as Outcome**

Of the 36 papers, the remaining 27 papers explored emotions as an outcome. In almost all the papers - except one (Huber et al. 2018) - human emotion was captured as the outcome. A number of studies (10 papers) examined specific emotions as outcomes, wherein papers largely studied negative emotions (8 papers). For example, 3 papers examined human anxiety as an outcome due to interactions with the CA (e.g., Riquel et al. 2021). Another 2 papers assessed stress felt by humans during interactions with a CA. Further, 2 papers examined frustration due to dissatisfaction with the CA interaction (e.g., Brendel et al. 2020). Finally, 3 papers studied human emotions like guilt, anger, and shame when interacting abusively with a CA. On the other hand, only 2 papers examined positive emotions, one of which studied negative emotions as well. The first paper captured joy and interest as emotional outcomes when interacting with a CA (Yang et al. 2019) while the second paper tested if the phenomena of emotional contagion (e.g., transfer of happiness from CA to user) existed in human-agent interactions (Joby and Umemuro 2022).

Other dimensions of emotions, including valence and activation were also studied. One study assessed the emotional connection users felt towards the business based on the interaction with their customer service CA (Araujo 2018). In another study, emotional warmth was assessed following interaction with an embodied CA (Stein et al. 2020). Human emotional engagement with the CA was evaluated as an outcome in one paper (Shi et al. 2018). Interestingly, one study examined the CA’s emotion as the interaction outcome (Huber et al. 2018).

**Descriptive Model for Antecedents of Emotion**

In this section, we provide an overview of the antecedents for emotion as outcome. We present the relationships in the form of a descriptive model for a holistic understanding of the state of findings. The model also shows the consistency of findings across reviewed papers (see Figure 1). We used “inc” to indicate an antecedent with inconclusive findings, i.e., equal number of significant and non-significant relationships (50% each). If the findings about the antecedent were not significant (i.e., more than 50% findings were not significant), it was labeled as “ns”. We labeled the antecedent as “mix” if it was significant but had mixed findings (equal number of positive and negative findings). We marked the antecedent as “+” if the findings were significant and positive (i.e., more than 50% of the findings were positive) and “-” if the findings were significant and negative (i.e., more than 50% of the findings were negative). When there was only one finding for an antecedent, we marked it as “1” to show limited evidence. When the findings were significant and comparative e.g., comparing different conversation styles, we marked the antecedent as “*”.

**Categories of Emotion Antecedents**

To facilitate an efficient discussion of the antecedents, we classified the antecedents of emotion into four categories i.e., human, CA and task characteristics, as well as other factors. The first category is human characteristics, which captures user characteristics that impact emotions from human-CA interactions. The second category is CA characteristics,
which covers its basic design elements and other CA features. The basic design elements are classified as verbal, visual, auditory, invisible, and interaction (Zierau et al. 2020), as discussed before. Additionally, we identified four other CA features studied as antecedents. These antecedents, labelled as other features, are CA anthropomorphism, CA personalization, CA detecting human emotion, and CA expressing emotion. The third category of antecedents is task characteristics, which indicates the different task types influencing emotion. The fourth category is other factors, which includes antecedents not captured by the other categories, such as time and comparison between humans and CAs.

Overview of Findings

Figure 1 provides an overview of the relationships between the four IV categories and emotion outcomes. We synthesized 57 findings from the reviewed papers to derive this figure. Past research offers evidence that human characteristics influence emotions in human-CA interactions, with 6 findings. Among these, findings about the emotions of users after verbally abusing a CA were inconclusive. Also, findings about the effect of perceived affective quality of the CA (in this case sociability) on emotions were inconclusive. User culture and user satisfaction with the CA both appeared to impact emotions in single studies.

Further, our analysis indicates that task characteristics are determinants of emotion. One study found that users requesting basic information from a CA felt less positive emotion than when using it to access external services (Yang et al. 2019). In another study, a cooperative human-CA task had a positive impact on emotions of the users (Lim and Reeves 2010).

In terms of other factors, researchers compared the impact of interacting with CA versus human assistant on user emotions, where users preferred the human assistant (Drouin et al. 2022). They also examined the role of time, where users’ emotional arousal was found to increase over time when interacting with a CA displaying social cues (Huang and Lee 2022).

CA characteristics were the most common determinants of emotions, with 40 findings. Here, four basic CA design elements were studied i.e., verbal, invisible, visual, and interaction elements. Findings on the verbal elements were significant. For example, a CA that verbally counterattacked users increased negative emotions compared to an empathetic CA (Chin et al. 2020). Further, findings for invisible CA elements were also significant. For example, one study found that different CA qualities (hedonic and pragmatic) had a significant influence on emotions of users in different contexts (Yang et al. 2019). CA visual element i.e., CA embodiment did not have a significant impact on user emotion (Stein et al. 2020). Concerning interaction elements, CA interaction using facial expressions enhanced users’ emotional engagement as compared to speech interaction (Shi et al. 2018).

Last, regarding other CA features, anthropomorphism had a positive effect on emotion, for example CAs with higher anthropomorphism increased user’s satisfaction and enhanced pleasure (Riquel et al. 2021). Personalization also enhanced user emotion towards the CA and increased emotional valence. Further, both emotion detection capability and emotion expression capability of CAs were found to have a positive effect on user emotions. The single study with CA’s emotion as outcome found that detecting user emotions led to higher emotional intensity/activation in CA’s responses (Huber et al. 2018).

Discussion and Future Directions

A holistic understanding of the role of emotions in human-CA interaction is vital to design effective CAs (Yang et al. 2019). For this purpose, we conducted a literature review to
identify and organize findings from extant research on emotions in human-CA interactions. Based on the literature analysis we synthesized the findings into a descriptive model that depicts the relationships between antecedents and emotional outcomes. We suggest future areas of research arising from inconclusive or understudied, yet promising, antecedents and relationships.

First, when emotion was studied as an outcome, our review reveals that only two studies examined positive emotions as outcomes. It is valuable to examine the different tasks and design elements that evoke positive emotions in humans while interacting with digital artifacts in order to promote usage and adoption (Beaudry and Pinsonneault 2010). Thus, future research could examine CA characteristics and task types that elicit positive user emotions, which can increase CA adoption in the long term.

Second, with respect to the antecedents of emotion, we found that studies on the effects of human characteristics like age, gender, personality and culture on emotion were relatively scant. These characteristics can significantly influence how humans interact with and perceive CAs (Diederich et al. 2022). Hence, future research could investigate how emotions vary with user characteristics, which could also aid in the personalization of CAs. While verbal abuse towards CAs was a common reaction to frustrations with different CA designs (e.g., Brendel et al. 2020), our analysis showed inconclusive findings for verbal abuse as an antecedent of user emotion. Future research can examine CA design elements that reduce the possibility of negative interaction cycles between the user and CA. While user satisfaction with the CA has a significant impact on their emotions, this was examined in a single study. Future research can study user emotions arising from satisfaction or dissatisfaction with different CA design elements. While the effect of CA’s affective quality was studied in one paper, the results were inconclusive. As different users may perceive the same stimulus to have different affective qualities (Russell 2003), it would be useful to examine how users perceive the same CA design element.

Third, we found limited research on how different task characteristics influence emotional responses (2 studies). Specific CA usage scenarios and their effects on user emotion were examined in only one study. The study showed more complex tasks like accessing external services through the CA (e.g., accessing the TV via Alexa) had more positive effects on emotions than simple search tasks (e.g., weather forecast). With CAs being implemented in multiple domains for varied tasks it would be useful to examine how different types of tasks influence emotional outcomes during human-CA collaborations. This can help organizations assess for which tasks they can obtain cost savings by deploying CAs. Fourth, regarding antecedents under other factors, time had a positive influence on user emotion, but this was examined in a single study. Obtaining stronger evidence requires more longitudinal studies to assess how users’ emotions towards a CA can change with time. Analyzing repeated measures of emotions over a period of time can give insights into how users’ attitude towards the CA develop.

Fifth, with respect to CA characteristics, a single study examined the effect of the content of the CA’s verbal response on user emotions. It would be useful for further research to explore systematically what kind of content in CA responses induces positive emotions in users and improves CA performance. Further, only one study examined the effect of CA’s social cues (intimacy) on user emotions. Future studies can examine what types of social cues (e.g., eagerness of the CA to continue the conversation or build a relationship) are positively perceived by users. Additionally, the effect of CA embodiment on user emotions was not significant. As CAs are being increasingly designed with varied embodiments, it would be valuable for future studies to understand how users respond to different embodiments. Further, we found a single study on personalization of the CA, which found that personalization is a significant determinant of user emotion (Shin et al. 2022). Future research on CA personalization could bolster the evidence on its impacts.

**Conclusion**

Designing AI systems, like CAs, that collaborate with humans in a natural way requires understanding human’s emotional responses to these artifacts. Examining emotions in human-AI interactions is challenging due to its temporal and subjective nature. A number of studies have focused on this topic, yet we lack a holistic view of empirical findings on the antecedents of emotions in human interactions with CAs. To address this gap, we review and analyze extant quantitative studies on the determinants of user emotions in human-CA interactions. We present a descriptive model that synthesizes the relationships between emotion and its antecedents. Based on the findings of our literature review, we identify knowledge gaps in the research on emotions in human-CA interactions and propose directions for future research in this area. This can promote the study and design CAs that collaborate with humans in ways that enhance their emotional responses and outcomes.

**Acknowledgments**

This research/project is supported by the Ministry of Education Singapore under the grant No: T1 251RES2201.

**References**


