Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

EmoBot: Artificial emotion generation through an emotional chatbot during general-purpose conversations

Md Ehtesham-Ul-Haque [a], Jacob D'Rozario [a], Rudaiba Adnin [a], Farhan Tanvir Utshaw [a], Fabiha Tasneem [a], Israt Jahan Shefa [b], A.B.M. Alim Al Islam [a,*]

[a] Next-generation Computing (NeC) Research Group, Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

[b] Department of Computer Science and Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh

ARTICLE INFO

Action editor: Alexei V. Samsonovich

Keywords:
Emotion
Emotion-generation
Chatbot
Psychological theory
User study

ABSTRACT

Emotion modeling has always been intriguing to researchers, where detecting emotion is highly focused and generating emotion is much less focused to date. Therefore, in this paper, we aim to exploring emotion generation, particularly for general-purpose conversations. Based on the Cognitive Appraisal Theory and focusing on audio and textual inputs, we propose a novel method to calculate informative variables to evaluate a particular emotion-generating event and six primary emotions. Incorporating such a method of artificial emotion generation, we implement an emotional chatbot, namely *EmoBot*. Accordingly, *EmoBot* analyzes continuous audio and textual inputs, calculates the informative variables to evaluate the current situation, generates appropriate emotions, and responds accordingly. An objective evaluation indicates that *EmoBot* could generate more accurate emotional and semantic responses than a traditional chatbot that does not consider emotion. Additionally, a subjective evaluation of *EmoBot* demonstrates the appreciation of users for *EmoBot* over a traditional chatbot that does not consider emotion.

1. Introduction

There has been a long history of chatbots powered by various techniques (Lasecki et al., 2013), for example, deep learning-based techniques significantly outperform transitional rule-based models (Sutskever, Vinyals, & Le, 2014). These deep learning-based chatbots generate responses to user requests, showing their encouraging application perspectives (Xu, Liu, Guo, Sinha, & Akkiraju, 2017). However, most models in this regard are only simulated in some specific cases, such as for customer care (Lasecki et al., 2013; Xu et al., 2017), e-commerce (Cui et al., 2017), distance education (Heller, Proctor, Mah, Jewell, & Cheung, 2005), etc., being far from acting in a natural conversation context that subsumes the effect of emotion by default. In this light, emotion is inherent, yet an elegant property for human beings that varies from person to person and cannot be completely bounded by logic (Frijda, 2009). This eventually makes it difficult to compute emotion mathematically. Compared to emotion detection systems (Agrafioti, Hatzinakos, & Anderson, 2011; Garcia-Garcia, Penichet, & Lozano, 2017; Majumder et al., 2019; Sailunaz, Dhaliwal, Rokne, & Alhajj, 2018), an emotion generation model with substantial accuracy is still far beyond the reach of the researchers. There exist a rich number of emotion generation theories (Frijda, 1986; Marks, 1982; Plutchik, 1980; Scherer, Schorr, & Johnstone, 2001) proposed by researchers in Psychology, Physiology, Neurology, and other concerned fields. However, each of these theories includes determining a large set of unknown variables to evaluate an emotion-generating event, making the theories hard to implement in a computational system. At the same time, with the growing popularity of chatbots in industries such as entertainment, customer care service, etc. Hu et al. (2018) and Io and Lee (2017), it has become of utmost importance to make the chatbots more emotionally responsive and more human-like. As a result, artificial emotion generation has become necessary (Cano, González, Gil-Iranzo, & Albiol-Pérez, 2021; Hieida & Nagai, 2021), and utilizing a psychological theory for such emotion generation has the potential to generate human-like emotions (Kim & Kwon, 2010). Over the years, psychologists have proposed several theories (Frijda, 1986; Marks, 1982; Plutchik, 1980; Scherer et al., 2001) related to emotion. Among these theories, the Cognitive Appraisal Theory of Emotion Generation (Frijda, 1986; Lazarus & Lazarus, 1994; Ortony, Clore, & Collins, 1990; Scherer et al., 2001) is a popular one and

* Corresponding author.
E-mail address: razi_bd@yahoo.com (A.B.M.A. Al Islam).

https://doi.org/10.1016/j.cogsys.2023.101168
Received 2 March 2023; Received in revised form 5 August 2023; Accepted 29 August 2023

has the potential to generate human-like emotions in natural conversations (Kim & Kwon, 2010; Scherer et al., 2001). In this light, although the Cognitive Appraisal Theory (Ortony et al., 1990) has provided a theoretical foundation for the computational modeling of emotion generation, most of the existing models (Hudlicka, 2015a; Kim & Kwon, 2010) based on this theory are yet to be implemented in reality. Therefore, in our study, we explore implementing a method for artificial emotion generation that can generate basic emotions — namely joy, sadness, fear, anger, and surprise (Panksepp, 2004)–based on Cognitive Appraisal Theory (Ortony et al., 1990).

In addition, most of the existing computational emotional models (Cui et al., 2017; Lasecki et al., 2013; Xu et al., 2017) are designed only for limited specific scenarios, such as customer care, and do not provide an implementation strategy focusing on real-life general-purpose conversations perspective. Therefore, in our study, we focus on implementing a method for artificial emotion generation for general-purpose conversation. Here, general-purpose conversation means suitable to be used for natural or general conversation context, not being limited to any particular type of conversation. Furthermore, with the growing popularity of chatbots in industries such as entertainment, customer care service, etc., Hu et al. (2018) and Io and Lee (2017), it has become of utmost importance to make the chatbots more emotionally responsive and more human-like. Therefore, in our study, we implement a chatbot– *EmoBot*– for general-purpose conversations, utilizing our developed method for artificial emotion generation.

According to Cognitive Appraisal Theory (Ortony et al., 1990), the generation of emotion in an individual is determined by the way s/he evaluates the situation. This evaluation is achieved through measurement criteria called appraisal variables (informative variables to evaluate a particular emotion-generating event) (Hudlicka, 2015b). Consequently, to calculate such informative variables to evaluate a particular emotion-generating event to generate basic emotions in *EmoBot*, we focus on the audio and corresponding text and propose multiple approaches. In our calculations of emotion, we include the current mood and personality traits of *EmoBot*. For instance, one cannot expect to get a happy emotional response from a person immediately after making him angry. In such cases, the response from the person will depend on that specific person's current emotion. We consider this phenomenon emotion-awareness. We have implemented such emotional awareness in *EmoBot*. Accordingly, we explore the following research questions in this study.

- **RQ1:** Can we leverage a psychological theory for artificial emotion generation during general-purpose conversations? How can we implement a chatbot that maintains its own emotional state and is responsive to the emotions of humans using a psychological theory?
- **RQ2:** Do users appreciate such a chatbot that maintains its own emotional state and is responsive to the emotions of humans?

In the process of answering the research questions, we make the following set of contributions to this paper.

- From an interdisciplinary perspective — combining the knowledge of Computer Science and Cognitive Psychology, we propose a novel methodology for artificial emotion generation based on Cognitive Appraisal Theory (Hudlicka, 2015b; Ortony et al., 1990). Incorporating our proposed methodology of artificial emotion generation, we designed and developed a chatbot, namely *EmoBot*, which generates proper responses to user requests or comments – by identifying the emotions of the users and maintaining its own emotional state – during general-purpose conversations. Therefore, we leverage the Cognitive Appraisal Theory (Ortony et al., 1990) for artificial emotion generation and implementation of an emotional chatbot. This informs the researchers of a method related to how an emotion-aware chatbot can be implemented for general-purpose conversations based on a psychological theory.
- To investigate users' appreciation for our implemented chatbot, we perform both objective and subjective evaluations of *EmoBot*, and compare it with a non-emotional chatbot, BotLibre (2013). The evaluation reveals that *EmoBot* generates emotion-aware responses to user requests and can provide a satisfactory user experience. The results indicate that emotionally motivated responses provide the user with a feeling of satisfaction during interaction with an agent. Such user evaluation will assist researchers to understand users' appreciation for chatbots that are responsive to human emotions and facilitate the development of solutions in the realm of emotion generation in chatbots.

In this work, Section 2 represents the background of this study, Section 3 covers the related work. After that, Section 4 highlights method overview and setup, Section 5 describes the mapping and evaluation of this study, Section 6 details the Emobot evaluation phase. Finally, Section 7 highlights the discussion of this study and Section 8 summarizes the conclusion and future work.

2. Background

In this section, we look into the theory of emotion generation on which our work is based. Then we will define the relevant terminologies.

2.1. Emotion

A key characteristic of emotions is their multimodal nature. The cognitive modality is directly associated with the evaluation-based working definition of emotions emphasized in the cognitive appraisal theories of emotion generation. The emotions addressed by these theories are typically the *basic* emotions: joy, sadness, fear, anger, and disgust (Panksepp, 2004). These emotions are characterized by stable patterns of triggers, behavioral expression, and associated distinct subjective experiences (Hudlicka, 2015b). The dimensional perspective describes emotions in terms of two- or three dimensions. The most frequent dimensional characterization of emotions uses two dimensions: valence and arousal (Russell, 2003; Russell & Barrett, 1999). Valence reflects a positive or negative evaluation and the associated felt state of pleasure (versus displeasure). Arousal reflects a general degree of intensity or activation of the organism. The degree of arousal reflects a general readiness to act: low arousal is associated with less energy and high arousal with more energy. Since this 2-dimensional space cannot differentiate among emotions that share the same values of arousal and valence (e.g., anger and fear, both characterized by high arousal and negative valence), a third dimension is often added, termed dominance. The resulting 3-dimensional space is often referred to as the PAD space (Mehrabian, 1996) (pleasure (synonymous with valence), arousal, dominance). From this, the renowned PAD Model (Zhou, 2018) defines mood states by three variables — Pleasure, Arousal, and Dominance. Further, the generation of emotion is related to personality traits (Barańczuk, 2019; Cervone & Little, 2019) and the five-factor model of personality (Barańczuk, 2019) is defined by five major traits: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism.

2.2. An emotion-generating event (emotion-inducing situation)

Emotion-inducing situation refers to a particular event that influences the observer of that event to respond physiologically (heart rate, blood pressure, etc.,) and change the state of their mood (Hudlicka, 2015b).

2.3. The properties of input (stimuli)

A stimulus, whether real or imagined, is evaluated in terms of its meaning and consequences for the agent, to determine the appropriate reaction from an organism (Hudlicka, 2015b). Stimuli refer to *a set of properties* to be given as input to an emotion generation system as the conversational input (Abdul-Kader & Woods, 2015; Ciechanowski, Przegalinska, Magnuski, & Gloor, 2019). Audio is an important part of the conversation (Kim, Goh, & Jun, 2018; Roniotis & Tsiknakis, 2018; Zhu et al., 2022), therefore, we take audio as our primary source of conversational input here. Using speech-to-text, we get textual input properties, such as text sentiment and text emotion from conversations. Moreover, intensity, signal gap, the average gap between words, average pitch, volume, and speech rate from the audio and text sentiment are important properties of audio input for conversational agents (Abdul-Kader & Woods, 2015; Ciechanowski et al., 2019; Kim et al., 2018; Roniotis & Tsiknakis, 2018; Zhu et al., 2022). Therefore, we also include such properties as inputs in our study and use them to calculate appraisal variables.

2.4. Cognitive appraisal theory

Cognitive Appraisal Theory (Frijda, 1986; Lazarus & Lazarus, 1994; Ortony et al., 1990; Scherer et al., 2001) is one of the most prominent psychological theories and has been widely used for designing models of emotion generation. It states that emotions are extracted based on our evaluations (information or estimates) of events that cause specific reactions in different people. It provides a set of domain-independent evaluative criteria that capture the current interpretation of the agent's internal and external environments as they relate to the agent's current goals. Some approaches have been proposed over the years to determine these evaluation criteria and model this evaluation process. Among them, two approaches are the most popular — list and hierarchical. In the list-approach (Scherer et al., 2001) the evaluation criteria is defined by a list of appraisal variables there, a reaction is evaluated in terms of its meaning and consequences for the agent. This evaluation assigns specific values to the individual appraisal variables. Once the values of these appraisal variables are determined by the agent's evaluation process, the resulting vector is then mapped onto a particular emotion. A related and overlapping set of evaluation criteria has been proposed by Ortony et al. (1990). They proposed a model, called the OCC model, which provides rich taxonomy of triggers and resulting emotions. This approach evaluates emotion primarily based on three types of triggers — consequences of events, actions of agents, and aspects of objects. The hierarchical (OCC) approach (Hudlicka, 2015b) can produce a rich number of emotions of different types. Thus, the hierarchical approach (appraisal variables) is used most often in designing emotion generation models. In our work, we also use the appraisal variable (hierarchical) approach (Hudlicka, 2015b) to generate emotion.

2.5. Informative variables to evaluate an emotion-generating event (appraisal variables)

Appraisal variables are a way of mapping a particular event that influences the observer of that event to respond physiologically to some numerical values. Using this, we can mathematically calculate the emotion of a person with the help of knowledge about the emotion-inducing situations at that time. Thus, emotion-inducing situation to appraisal variable mapping is the process of translating emotion-inducing situations to some numerical values which can be used in modeling emotion generation. Conforming to the Cognitive Appraisal Theory (Hudlicka, 2015b; Ortony et al., 1990), the information of a specific functional reaction is the principal task of emotion generation – which can be classified into three different factors – Relevance, Implication, and Coping Potential (Hudlicka, 2015b). These factors involve ten informative variables to evaluate an emotion-generating event (appraisal variables) (Hudlicka, 2015b). As a result, we are considering these ten informative variables to evaluate an emotion-generating event (appraisal variables) in this study for evaluating a particular situation.

2.5.1. Relevance

Given the constant barrage of stimulation, an organism must decide which stimuli are sufficiently relevant to warrant deployment of attention and possibly adaptive reaction or whether the status quo can be maintained and ongoing activity pursued.

- **Suddenness** Any sudden stimulus (characterized by abrupt onset and relatively high intensity) is likely to be registered as novel and deserving of attention.
- **Familiarity** The evaluation to determine the degree of familiarity with the object or event.
- **Predictability** Another important mechanism for relevance check is based on complex estimates (based on an observation of regularities in the world) of the probability and predictability of the occurrence of a stimulus.
- **Valence** The evaluation of whether a stimulus event is likely to result in pleasure or pain.

2.5.2. Implication

This is a central appraisal objective since it determines to what extent a stimulus or the situation is appraised as furthering or endangering an organism's survival and adaptation to a given environment, as well as satisfying its needs and attaining its goals.

- **Discrepancy from Expectation** The situation created by the event can be consistent or discrepant with the individual's expectation concerning that point in time or position in the action sequence leading up to a goal. For example, there would be some discrepancy in expectation if the father of the failed student gave him a present after learning of the exam results.
- **Conducive to goal** Most important, the organism needs to check the conduciveness of a stimulus event to help attain one or several of the current goals/needs. The consequences of acts or events can constitute the attainment of goals/needs, progress towards such attainment, or facilitation of further goal-directed action.
- **Urgency** Adaptive action in response to an event is particularly urgent when high-priority goals or needs are endangered, and the organism has to resort to specific responses when it is likely that delaying a response will make matters worse.

2.5.3. Coping potential

Successful coping with a stimulus event implies that the individual's concern with the eliciting event disappears. The coping potential check determines which types of responses to an event are available to the organism and which consequences will affect the organism under each option.

- **Control** One important dimension is to what extent an event or its outcomes can be influenced or controlled by natural agents (i.e., people or animals). For example, while the behavior of a friend or the direction of an automobile is generally controllable up to a degree, the weather or the incidence of a terminal illness is usually not.
- **Power** If control is possible, coping potential depends on the power of the organism to exert control or to recruit others to help. With the help of the power check, the organism evaluates the resources at its disposal to change contingencies and outcomes according to its interests.

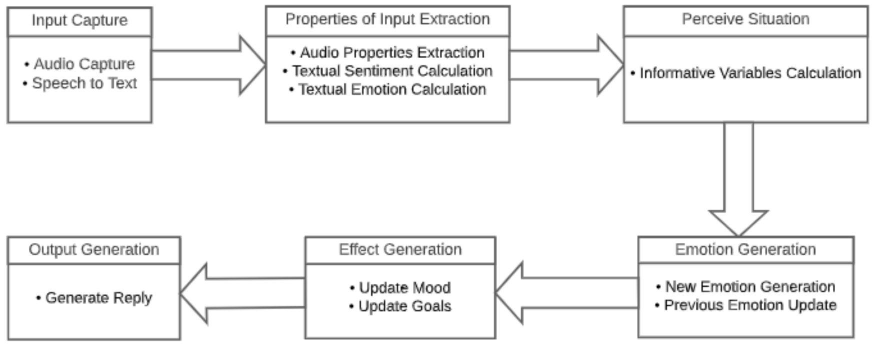


Fig. 1. Implementation steps of the emotion generation system.

- **Adjustment** It is particularly important to check how well one can adjust to the consequences of an event if the control and power checks yield the conclusion that it is not within one's power to change the outcomes. For example, a failed student might be able to live perfectly well with a terminal failing grade if he was convinced that his future should in any case be sought in the world of finance.

3. Related work

In this section, we discuss studies on emotion generation, the prevalence of chatbot systems, and emotion generation using chatbots.

3.1. Emotion generation

According to Damasio (2003), emotions are a series of physical reactions and changes in internal states. An emotion generation model was inspired by the theory of Cano et al. (2021) which has one module called the controller, which internally has internal values and action selection. The internal values of the controller are related to driving values, such as fatigue, hunger, homesickness, and curiosity, which are defined as primary states whereas, emotions such as fear, anger, boredom, and happiness are considered to be secondary states. Another study (fei Shi, liang Wang, Ping, & kun Zhang, 2011) proposed an artificial emotional model based on neuroscience in which emotion can be associated with past experiences. Velásquez (1997) introduced a computational model for emotion generation and it was initially designed for agent developers for providing enough functionality to design emotional agents that can be used in a variety of applications. Another study (Hieida & Nagai, 2021) acknowledged the implementation of social emotions in robots while considering their relationship with basic emotions as a major issue in the field of robotics. Cano et al. (2021) presented a study on designing social robots for children with autism spectrum disorder (ASD). They identified that social robots must be prepared to express and recognize emotions when interacting with children with ASD because they do not tolerate surprises and changes in the environment. Researchers have also worked on generating humorous text or jokes (Dybala et al., 2010; Yang, Lavie, Dyer, & Hovy, 2015). The full statistical approach in response generation has been introduced very recently (Candia-Rivera, Catrambone, Thayer, Gentili, & Valenza, 2022; Du, Jin, Yan, & Yan, 2023; Jain & Rath, 2023; Ritter, Cherry, & Dolan, 2011; Tu et al., 2022; Wang et al., 2022). However, there is hardly any statistical response generation that depicts the emotion of the user as per our knowledge. H-Cogg-Off (Tavakoli & Palhang, 2016) architecture introduced emotion generation on multiple levels. Using such architecture, Pudane, Lavendelis, and Radin (2017) presented a detailed block diagram of multi-level emotion generation. They adopted the BDI (Belief, Desire, and Intention) (Sloman & Chrisley, 2005) architecture in the routine level of the model. They showed the distinction and interaction between emotions on multiple levels. In addition, based on the Cognitive Appraisal Theory (Ortony et al., 1990), Kim and Kwon (2010) proposed a computational model of emotion generation. Designed for interactive tasks, it evaluates the current task-based situation to generate emotion correspondingly. Their model is for playing games of twenty questions. However, these studies did not present any implementation methods or calculation procedures for generating emotions during general-purpose conversations, whereas, we implemented a method for artificial emotion generation based on the Cognitive Appraisal Theory (Ortony et al., 1990). Moreover, our research focuses on a broader scope and generates emotions for a continuous audio input stream while adapting the state of the different emotions during general-purpose conversations.

3.2. Chatbot systems

Chatbots are powered by various techniques. Rule-based (Ritter et al., 2011) or retrieval based (Yu, Papangelis, & Rudnicky, 2015) chatbots were used often in the early days. These techniques are usually limited to small-scale data or closed application domains. An alternative to using in free-form conversations is the Crowdsourcing chatbot system (Huang, Chang, Swaminathan, & Bigham, 2017; Huang, Lasecki, Azaria, & Bigham, 2016) which is human manual operations dependent. Moreover, human evaluations are too time-consuming and involve human labor that cannot be reused. Hence, evaluating chatbots and natural language generation is an enormous challenge (Liu et al., 2017). The study in Papineni, Roukos, Ward, and Zhu (2002) showed a language-independent method of automatic machine translation evaluation that matches the response from a human agent (Xu et al., 2017). In addition, using deep learning, researchers have been able to work on open-domain large-scale conversations (Li, Monroe et al., 2016). The study in Xu, Szlam, and Weston (2021) released a long-term open-domain communication entitled Multi-Session Chat (MSC) that gives the opportunity for the conversation to develop and improve with time as the model has more context and more understanding of that specific user's interests. Using the sequence-to-sequence (seq2seq) model (Ha, Dai, & Le, 2016; Sutskever et al., 2014), researchers (Xu et al., 2017) created a chatbot system for customer care and were able to achieve good results. Researchers also reported a modified seq2seq model (Li, Galley et al., 2016) for generating sentences matching certain personas. The study in Adiwardana et al. (2020) introduced a generative chatbot using a seq2seq model using data from public-domain social media conversations. Huber, McDuff, Brockett, Galley, and Dolan (2018), an image-grounded conversational agent investigated the relationships between such image information and the generated language. However, chatbots often do not meet users' expectations (Jenneboer, Herrando, & Constantinides, 2022; Koivunen, Ala-Luopa, Olsson, & Haapakorpi, 2022; Liu, Hu, Yan, & Lin, 2023; Luger & Sellen, 2016; Misischia, Poecze, & Strauss, 2022; de Sá Siqueira, Müller, & Bosse, 2023). Luger and Sellen (2016) showed that there is a large gulf between peoples' expectations about the capabilities of chatbots, and what such systems can actually deliver. The work reveals multiple design challenges arising from this gulf between user expectation and experience, such as how a chatbot may reveal its current state. Similarly, researchers have studied the perceived intelligence of agents in-depth (Cassell, Bickmore, Campbell, Vilhjalmsson, & Yan, 2000). The work by Cassell et al. (2000) studied how multi-modal interactions affect the experience of using a chatbot. Whilst the focus of Cassell's work had been multi-modal representations of intelligence (physical gestures in addition to voice), the central concept of the users' need to locate intelligence, and thereby the need to represent intelligence to the user is an important concept in Conversation Agent research. As a result, it is necessary to correctly use the emotional responses and behavior of an agent to express, recognize, and understand emotion when interacting with a user (Cano et al., 2021). Gratch, Wang, Gerten, Fast, and Duffy (2007) show that in order to create rapport, agents should provide more positive, emotional feedback from time to time. These create opportunities to answer open questions: how can we add emotional context into natural conversation generation in a chatbot to make the responses more engaging and emotional? Therefore, in our study, we focus on implementing an emotional chatbot to make the responses more engaging and emotional for the users by generating artificial emotions.

3.3. Emotion generation on chatbots

There has been a lot of work on the impact of emotional response on the quality of customer service and some of these have been implemented in customer service chatbots. Here, the user's attitude (Martin, 1985) towards a company and satisfaction (McCollough, 2000) depend on the correct emotional response. Over 40% of the user requests for customer services on social media are mainly emotional as shown by Xu et al. (2017). They have created a conversational system to automatically generate responses for user requests. Being much like human agents, their system shows empathy to help users cope with emotional situations. Much like their work, Tianran Hu et al. (2018) have worked on creating a tone-aware chatbot that is surprisingly more empathetic than human agents. Being focused on mainly eight types of tones — anxious, frustrated, impolite, passionate, polite, sad, satisfied, and empathetic, their research has shown the impact of tones on users' emotions, such as empathetic tone reduces users' negative emotions such as frustration and sadness. As their system is for customer care on social media, they have worked on identifying the ones that are most beneficial for that purpose. They have also shown that their tone-aware chatbot generates an appropriate response to user requests as human agents which shows how close a chatbot agent can get to humans. The emotional text has an impact on customers' perceptions of service agents and generally, in a positive direction (Zhang, Erickson, & Webb, 2011). The user's positive emotion is increased with the agent's cheerful emotion (Herzig et al., 2016) and customer stress is reduced with an empathetic tone (Prendinger & Ishizuka, 2005). Most of these works on emotion generation are focused on customer services and are not based on any psychological theory for generating emotions, which creates an open question: how we can implement an emotional chatbot using a psychological theory for general-purpose communication? Therefore, we have implemented a chatbot for everyday general-purpose communication. Whilst the goal of prior studies (Herzig et al., 2016; Hu et al., 2018; Prendinger & Ishizuka, 2005; Zhang et al., 2011) is mainly on establishing empathy with the user focusing fully on the users' side, our research focuses not only on the users' speech but also adapting the emotional state of the chatbot resulting in more natural conversation.

4. Method overview and setup

We have developed a chatbot named *EmoBot* which generates appropriate emotional responses to a user's speech. We deployed the chatbot system online and made it open for use.

Table 1
Extracted properties of input.

Properties of audio input	Properties of textual input
Intensity	Text emotion
Signal Energy	Text sentiment
Average Gap between Words	
Average Pitch	
Volume	
Speech Rate	

Table 2
Informative variables to evaluate an emotion-generating event.

Relevance	Implication	Coping potential
Suddenness	Discrepancy from Expectation	Control
familiarity	Conducive to Goal	Power
Predictability	Urgency	Adjustment
Valence		

4.1. Overview of steps

Here, we split the implementation process into six steps. The steps are shown in Fig. 1.

- **Input capture:** Taking audio as the main input, we convert it into its corresponding text with the help of a speech-to-text system.
- **Properties of input extraction:** From audio and text input, the corresponding auditory and textual properties are extracted. A list of the properties is presented in Table 1.
- **Informative variables to evaluate an event (appraisal variables):** Conforming to the Cognitive Appraisal Theory (Ortony et al., 1990), the information of a specific functional reaction is the principal task of emotion generation. Based on this theory, 10 informative variables to evaluate an event (appraisal variables) are considered (Hudlicka, 2015b; Scherer et al., 2001). These variables can be found in Table 2 (Hudlicka, 2015b). In this work, one of our main contributions is to find an unconventional mapping between the captured input and informative variables to evaluate an event. We evaluate multiple mappings to find the most prominent one. The acquisition of the mappings and corresponding evaluation is elaborated in the next section.
- **Emotion Generation:** Emotions are generated according to the informative variables to evaluate an event in the system. Here, we generate primary **emotions** (Panksepp, 2004)- *Joy*, *Sadness*, *Fear*, *Anger*, *Surprise*. If the value of emotion is less than a certain threshold, 0.2 then the emotion is labeled as unclassified in our system. Each emotion has arousal and decay rate as properties. These emotions are assessed using different mapping procedures between informative variables to evaluate an event and emotions. Mapping and evaluation are presented in the following section. We focus on finding a novel mapping that generates the most accurate emotions. A particular emotion is updated using the following equation proposed by the authors. The updated emotion will be a proportion of the previous and current emotions.

$$emotion_i(previous) = emotion_i(previous) - emotion_i(previous) \times time \times decayRate_i$$

$$updatedEmotion_i = emotion_i(previous) + (1 - emotion_i(previous)) \times emotion_i(current)$$

$$emotion_i(current) = updatedEmotion_i$$

- **Effect Generation:** The effect of a generated emotion includes the updated mood states, personality traits, and the previous emotion of the system. We use the renowned PAD Model where **mood states** are defined by three variables — Pleasure, Arousal, and Dominance (Zhou, 2018). We also use personality traits in our

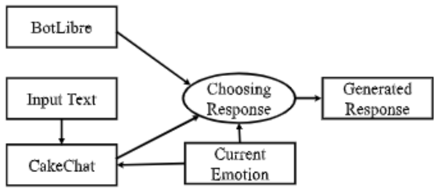


Fig. 2. Extended output generation process.

Table 3
PAD model of mood states (Zhou, 2018).

Trait combination	Mood type
+P +A +D	Exuberant
−P −A −D	Bored
+P +A −D	Dependent
−P −A +D	Disdainful
+P −A +D	Relaxed
−P +A −D	Anxious
+P −A −D	Docile
P +A +D	Hostile

Table 4
Change in mood for different emotions (Zhou, 2018).

Emotion	Pleasure	Arousal	Dominance
Joy	0.4	0.2	0.1
Surprise	0.4	0.5	−0.2
Fear	−0.64	0.6	−0.43
Anger	−0.51	0.59	0.25
Sadness	−0.3	0.1	−0.4
Unclassified	0	0	0

system. For modeling personality, we use the five-factor model of **personality** (Barańczuk, 2019). In this model, personality is defined by five major traits: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Different values of PAD variables determine different mood states. The combination is shown in Table 3. Mood values are updated according to the present emotion. If the value/arousal of the current emotion of the system is x and the corresponding weights for pleasure, arousal, dominance, openness, conscientiousness, extroversion, agreeableness, and neuroticism are $p_x, a_x, d_x, o_x, c_x, e_x, g_x, n_x$, then new values of the mood and personality factors can be calculated with the following equations-

$$pleasure(current) = pleasure(previous) + p_x \times x$$
$$arousal(current) = arousal(previous) + a_x \times x$$
$$dominance(current) = dominance(previous) + d_x \times x$$
$$openness(current) = openness(previous) + o_x \times x$$
$$conscientiousness(current) = conscientiousness(previous) + c_x \times x$$
$$extroversion(current) = extroversion(previous) + e_x \times x$$
$$agreeableness(current) = agreeableness(previous) + g_x \times x$$
$$neuroticism(current) = neuroticism(previous) + n_x \times x$$

- **Output Generation:** The response is generated as the output of the system and presented to the user in both audio and text formats. (see Table 4).

4.2. Setup of EmoBot from BotLibre

BotLibre (2013) is a free open-source chatbot. It presents an artificial intelligence-enabled platform for the web, mobile, social media, gaming, and the Metaverse. However, it does not consider the emotional states of users. Even then, it is popular with its users having a large community involved with it. BotLibre community has over 500 000 registered users, over 100 000 bots, over 100 million conversations, and over 1 million downloads (BotLibre, 2013). Besides, BotLibre has been used in several recent research studies (Easton et al., 2019; Simomukay, 2018; Somasiri et al., 2016). Examples of such existing research studies include developing a virtual agent to support individuals living with physical and mental comorbidities (Easton et al., 2019), a drug addiction recovery through mobile-based application (Somasiri et al., 2016), and data collection (Fossi, Dzwonkowski, & Othman, 2021). Thus, it is regarded as a well-known open-source chatbot platform in both research and use-case domains. Accordingly, we have adopted this platform as a non-emotional chatbot (chatbot that does not consider emotion) in our study and compared it with our developed emotional chatbot, Emobot.

We modified BotLibre (2013) and used CakeChat (2018) to implement EmoBot. EmoBot does not require any command from the user to perform. The user's speech is captured automatically when a sentence is finished based on a silence threshold. CakeChat (2018) produces an output sentence by taking the correct emotional state as its input. Thus, the response becomes more accurate that matches the correct emotional context. We generate the emotion and give it as input to Cake Chat. The generated response either comes from the modified (BotLibre, 2013) or CakeChat (2018). VaderSentiment (Hutto & Gilbert, 2014) is used to determine which response is to be accepted. The procedure of the system is exhibited in Fig. 2.

5. Mapping and evaluation

We provide a calculation method for informative variables to evaluate a particular emotion-generating event (appraisal variables) from a set of properties of audio input (stimuli). We present three qualitative approaches to acquire the mapping and the evaluation to uncover the best approach.

5.1. Mapping by the authors of this study

To begin with the task of emotion generation from the informative variables to evaluate a particular emotion-generating event (appraisal variables), we discreetly examine the dependent variables and address the most suitable way of calculating the variables from the independent factors-based on the understanding of the concept and personal experiences of the authors of this study.

All three of the mappings – a set of properties of input to relevance-implication, mood-personality traits to coping potential, and informative variables to evaluate a particular emotion-generating event (appraisal variables) to emotions – are populated by the discussion and verification among the authors of this study. The reasons behind this approach were - (i) we are only considering primary emotions rather than secondary emotions, (ii) we want to keep the calculations as simple as possible and get started with a system to test the accuracy of the method, (iii) the number of unknowns is relatively high and a learning model may make things complex at the initial stage. The mapping proposed by the authors of this study can be found in Tables 5, 6, and 7. For the mapping between informative variables to evaluate a particular emotion-generating event (appraisal variables) and emotions, we were inspired by Hudlicka (2015b). For example, in the case of suddenness calculation from the volume, the most probable emotion-generating event is that a higher value of volume increases the probability of something happening suddenly. Therefore, in the mapping between suddenness and volume, we put a value of 1 in 5.

If a variable x depends on independent variables $y_1, y_2, \dots, y_n$ and the corresponding mapping values for x in the mapping are $w_1, w_2, \dots, w_n$, then -

$$x = \sum_{i=1}^{n} w_i \times y_i \tag{1}$$

Table 5
Properties of input to informative variables (appraisal variables) of a particular emotion-generating event mapping by authors of this study.

Rel-Imp	Stimuli						
	Sentiment	Intensity	Signal Energy	Average Gap between Words	Average Pitch	Volume	Speech Rate
Suddenness	0	1	0	−1	1	1	0
Familiarity	1	−1	−1	0	−1	−1	−1
Predictability	1	−1	−1	0	0	−1	−1
Valence	1	−1	−1	0	−1	−1	−1
Discrepancy from Expectation	−1	0	1	−1	1	1	1
Conducive to Goal	1	0	1	0	0	−1	0
Urgency	0	1	1	−1	1	1	1

Table 6
Mood-personality traits to coping potential mapping by authors of this study.

Coping potential	Mood-personality traits							
	Pleasure	Arousal	Dominance	Openness	Conscientiousness	Extroversion	Agreeableness	Neuroticism
Control	0.1	0.1	0.3	0	0.5	0.30	0	−0.15
Power	0.5	0.25	0.25	0	0	0.1	−0.2	0.15
Adjustment	0.2	0	0.2	0.25	0.1	0.1	0	−0.25

Table 7
Informative variables to evaluate a particular emotion-generating event (appraisal variable) to emotions mapping by authors of this study.

Emotion	Appraisal Var									
	Suddenness	Familiarity	Predictability	Valence	Discrepancy from Expectation	Conducive to Goal	Urgency	Control	Power	Adjustment
Fear	1	−1	−1	−1	1	−1	1	−1	−1	−1
Joy	1	−1	−1	0	0	1	−1	0	0	−1
Sadness	−1	−1	0	0	0	1	−1	−1	−1	1
Surprise	1	0	−1	0	1	1	1	−1	0	0
Anger	1	−1	−1	1	1	−1	1	1	1	1

Table 8
Average correlation between the properties of audio input and relevance-implication for all participants.

Rel-Imp	Stimuli						
	Sentiment	Intensity	Signal Energy	Average Gap between Words	Average Pitch	Volume	Speech Rate
Suddenness	−0.2	0.95	0.6	−0.15	0.55	0.85	0.65
Familiarity	0.7	−0.3	0.2	0	0	−0.2	0.1
Predictability	0.1	0	0.15	−0.05	−0.05	−0.2	−0.05
Valence	0.5	−0.15	0.5	−0.1	0.35	0	0.1
Discrepancy from Expectation	−0.65	0.25	0.1	0.15	0.1	0.35	0.45
Conducive to Goal	0.9	−0.25	0.15	−0.05	−0.15	−0.25	−0.15
Urgency	−0.3	0.7	0.7	−0.2	0.25	0.65	0.75

Table 9
Average correlation between mood-personality traits and coping potential for all participants.

Coping potential	Mood-personality traits							
	Pleasure	Arousal	Dominance	Openness	Conscientiousness	Extroversion	Agreeableness	Neuroticism
Control	0.85	1	0.85	0.45	0.25	0.75	0.35	−0.7
Power	0.4	0.7	1	0.3	0.2	0.75	−0.35	−0.55
Adjustment	0.55	0.55	−0.15	0.65	0.05	0.2	0.65	−0.4

Using Eq. (1), firstly, we calculate all the informative variables to evaluate a particular emotion-generating event (appraisal variables) from the properties of input (stimuli), and secondly, we calculate all the emotions except the emotions labeled as unclassified. The emotion is considered to be unclassified only when the values of all the other emotions are lower than a certain threshold (0.2 in our implementation). For each input, the emotion with the highest value (arousal) is considered to be the current emotion of the system.

5.2. Correlation-based approach

Human beings can predict a situation by listening to a voice, such as hearing the tone of someone over the phone (Latinus & Belin, 2011; Park & Cameron, 2014). Therefore, here, we involved human participants to fill up the correlations between the properties of input (stimuli) and informative variables to evaluate a particular emotion-generating event (appraisal variables).

For the evaluation, we provided three Google sheets to the participants about the properties of audio input versus relevance-implication, mood-personality traits versus coping potential, and informative variables to evaluate a particular emotion-generating event (appraisal variables) versus emotions (each one is represented as independent factors versus dependent factors) respectively. The first two sheets are to calculate the correlation between the properties of audio input and informative variables to evaluate a particular emotion-generating event, while the third one is to find the correlation between informative variables to evaluate a particular emotion-generating event (appraisal variables) and emotions. During the evaluation, each user was asked to answer from their personal experience whether they think an escalation

Table 10
Average correlation between informative variables to evaluate a particular emotion-generating event (appraisal variable) and emotions for all participants.

Emotion	Appraisal Var									
	Suddenness	Familiarity	Predictability	Valence	Discrepancy from Expectation	Conducive to Goal	Urgency	Control	Power	Adjustment
Fear	0.95	−0.85	−0.8	−0.85	0.75	−0.65	0.7	−0.85	−0.75	−0.6
Joy	0.15	0.65	0.45	0.95	−0.65	0.95	0	0.85	0.75	0.55
Sadness	−0.15	−0.65	−0.6	−0.95	0.7	−0.8	0.05	−0.65	−0.65	−0.2
Surprise	0.85	−0.85	−0.7	0.1	0.75	−0.11	0.3	0	−0.32	−0.16
Anger	0.35	−0.6	−0.45	−0.8	0.65	−0.65	0.15	−0.6	−0.45	−0.75

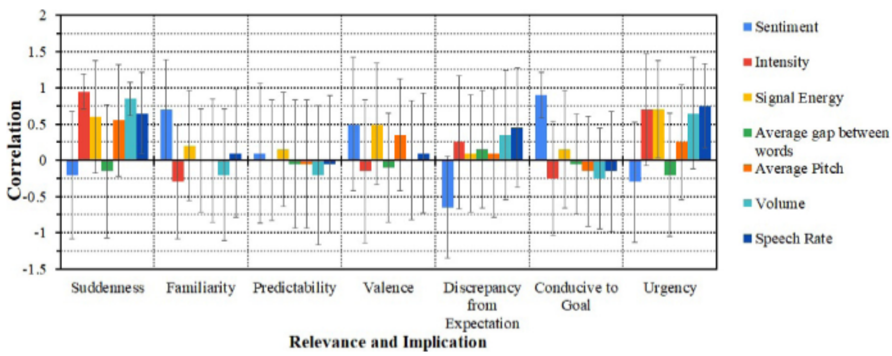


Fig. 3. Average correlation between the properties of input and relevance-implication with standard deviation for all participants.

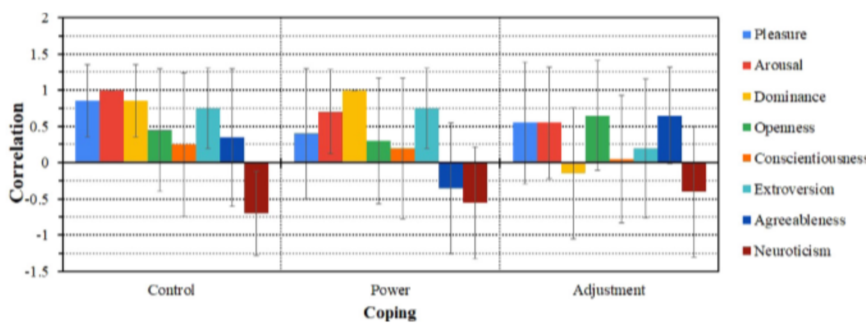


Fig. 4. Average correlation between mood-personality traits and coping potential with standard deviation for all participants.

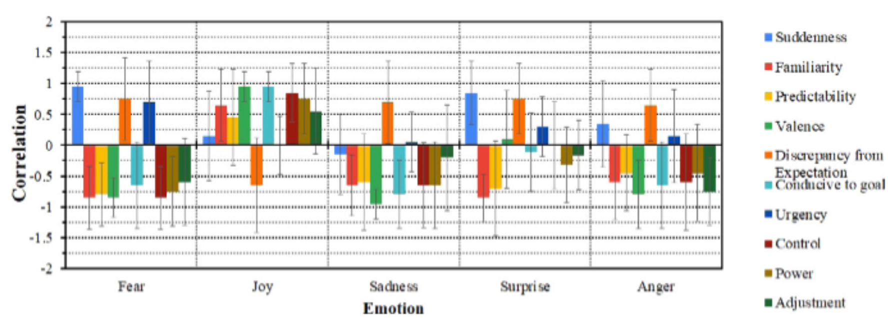


Fig. 5. Average correlation between informative variables to evaluate a particular emotion-generating event (appraisal variable) and emotions with standard deviation for all participants.

in the independent variable will increase (+1), decrease (−1), or no change (0) (the value of the dependent variable in a probabilistic sense). Therefore, the possible values are +1, −1, and 0, which signifies a positive, a negative, or no correlation between the independent and dependent variables.

The participants were instructed to consider a couple of things - (i) as the system can only generate primary emotions, they should think about the most probable examples for determining correlation, and (ii) they should also keep in mind that the goal of the system is to sustain an unclassified/happy state. A total of 20 participants provided data in this correlation-based evaluation: 16 via online (zoom) interviews and four in-person interviews. The demography of the participants is as follows: male — 12, female — 8, age below 25 years — 13, age equal to or above 25 years — 7. Each participant was given a brief overview for the evaluation. Then we presented the Google sheets to them sequentially. The participants were provided with a thorough explanation before filling up the correlations in the sheets. We provided both verbal and written descriptions of each of the variables to the participants for a better understanding of the factors. Tables 8, 9, and 10 present the average of the 20 correlations reported by the participants. We used these average values as weights to calculate the dependent variables using Eq. (1). Like before, unclassified emotion is shown when the

Table 11
Sample audio clip with properties.

Clip No.	Content	Sentiment	Intensity	Signal Energy	Average Gap between Words	Average Pitch	Volume	Speech Rate
1	I will not give your money back	−0.7	0.6	0.7	−0.5	0.3	0.6	0 .4

Table 12
Audio-based mapping of informative variables to evaluate a particular emotion-generating event (appraisal variable) and relevance-implication using linear regression.

Rel-Imp	Stimuli						
	Sentiment	Intensity	Signal Energy	Average Gap between Words	Average Pitch	Volume	Speech Rate
Suddenness	−0.56	0.78	0.96	−0.96	1.06	1.35	1.17
Familiarity	0.50	−0.54	−0.69	0.76	−0.62	−0.70	−0.64
Predictability	0.45	−0.69	−0.80	0.58	−0.68	−0.73	−0.69
Valence	0.50	−0.28	−0.45	0.02	−0.38	−0.46	−0.25
Discrepancy from Expectation	−0.50	0.70	0.88	−0.60	0.89	0.95	0.57
Conducive to Goal	0.66	−0.53	−0.61	0.34	−0.44	−0.53	−0.54
Urgency	−0.59	0.58	0.75	−0.94	0.93	1.15	0.89

Table 13
Audio-based mapping of mood-personality traits and coping potential using linear regression.

Coping potential	Mood-personality traits							
	Pleasure	Arousal	Dominance	Openness	Conscientiousness	Extroversion	Agreeableness	Neuroticism
Control	0.20	0.15	0.19	0.01	0.01	0.01	0.00	−0.01
Power	−0.21	−0.11	0.08	0.00	−0.01	0.01	−0.01	0.01
Adjustment	0.14	0.07	−0.15	0.01	0.00	0.00	0.01	−0.01

Table 14
Audio-based mapping of informative variables (appraisal variables) of a particular emotion-generating event and emotions using linear regression.

Emotion	Appraisal Var									
	Suddenness	Familiarity	Predictability	Valence	Discrepancy from Expectation	Conducive to Goal	Urgency	Control	Power	Adjustment
Fear	0.0166	−0.0168	−0.0160	−0.0162	0.0167	−0.0191	0.0146	−0.1305	0.0814	−0.0009
Joy	−0.0054	0.0055	0.0068	0.0111	−0.0075	0.0100	−0.0084	0.0768	−0.0445	0.0217
Sadness	0.0022	−0.0044	−0.0033	−0.0068	0.0035	−0.0041	0.0025	−0.0303	0.0060	0.0109
Surprise	0.0156	−0.0160	−0.0163	−0.0126	0.0148	−0.0142	0.0152	−0.1152	0.0528	0.0100
Anger	0.0009	−0.0066	−0.0054	−0.0084	0.0068	−0.0080	0.0026	−0.0530	0.0198	0.0193
Unclassified	0.0081	0.0123	0.0127	0.0074	−0.0116	0.0106	−0.0082	0.0806	−0.0348	−0.0473

values of the other emotions are less than a specific threshold value (0.2). The graphical representation of the averaged correlation-based mappings, along with the standard deviations as error bars, are presented in Figs. 3, 4, and 5. It is important to understand that, as the perception of emotion is a highly subjective matter, the correlations reported by the participants will vary from person to person. This is evident from the figures as the error bars signify large variations in the data. To determine this variation, we suggest several mappings based on the participants.

5.3. Audio-based approach

A total of eight people participated in the audio-based approach, 4 female and 4 male. Having participated in the correlation evaluation, these participants were expected to provide accurate data because of their prior understanding of the concept. The average time for this evaluation is 56.3 min. The participants are aged between 20–25 years and provided data via online (zoom) interviews. Short audio clips of 3–5 s containing a single sentence were provided to the users and qualitative analysis of the collected data of the particular audio was performed. Before the analysis began, the participants were presented with three sample audio recordings and provided with qualitative sound properties (agreed upon by the authors) for each sample. A sample clip, with qualitative properties, is presented in Table 11. We followed this format in order to make them familiar with the calculated properties which helps a participant to understand the properties of inputs (stimuli) better and enables them to evaluate each sentence accordingly. Five separate audio clips were used as test audio clips after explaining the sample clips. For each audio clip, a participant was required to complete five tasks.

- **Task 1**: Report the qualitative audio properties (Content, Sentiment, Intensity, Signal Energy, Average Gap between Words, Average Pitch, Volume, and Speech Rate) on a scale of −1 to 1.
- **Task 2:** Report the relevance and implication factors (Suddenness, Familiarity, Predictability, Valence, Discrepancy from Expectation, Conducive to Goal, Urgency) on a scale of 0 to 100.
- **Task 3**: Report the coping potential considering two different moods (Exuberant, Hostile) on a scale of 0 to 100
- **Task 4:** Report the coping potentials for five different personality traits (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) with varying percentages on a scale of 0 to 100.
- **Task 5**: Report the probability of the generated emotions (Fear, Joy, Sadness, Surprise, and Anger) on a scale of 0 to 100.

For mapping generation, we used Linear Regression and calculated the slopes of the regression lines that can be used as weights. Mapping of the properties of input to relevance-implication is determined by the data from Task 1 and Task 2. For example, to find the weight of sentiment (an input) for suddenness (a relevance factor), we calculate the slope of the regression line that fits the (sentiment, suddenness) data points for all audio clips. The mapping of mood and coping potential is calculated from the data generated from Task - 1 and Task - 3, and the mapping of personality traits and coping potential from Task - 1 and Task - 4. The mapping of informative variables to evaluate a particular emotion-generating event and emotions requires the aggregation of the data from Tasks – 2, 3, 4, and 5. However, the calculation procedure is the same – finding the regression line that fits the data points. These mappings are presented in Tables 12, 13, and 14. The calculation of the variables from the mapping is similar to the previous approaches. Each dependent variable can be calculated from the mapping using Eq. (1). Note the presence of the unclassified emotion in Table 14. Unlike the previous methods, the unclassified state is directly calculated using the mapping.

5.4. Choosing the best approach

We implemented each mapping in *EmoBot* and experimented with a fixed set of sentences (Chen, Hsu, Kuo, Ku, et al., 2018) to find the most efficient mapping. Our data set (Chen et al., 2018) consists of conversations from the popular series "Friends", where each sentence is labeled with associated emotion. We chose 15 of these sentences from the dataset and corresponding responses. We tried to choose sentences such that each sentence can be used in a natural conversation without prior context and the responses of the sentences can cover the primary emotions considered in the paper. Some sentences required prior context in the dataset. Therefore, we chose 7 extra sentences from the real world to balance out the distribution of the response emotions. The resulting 22 response emotions contain at least 3 of each type of emotion except the unclassified emotion, which is the response emotion for 7 sentences. We emphasized that the sentences chosen from the real world are frequent sentences with an unambiguous response emotion in most of the cases. These sentences are verified by three people so that confusing sentences can be discarded. In summary, we get 22 sentences (at least 3 of each primary emotion type) with two properties - (1) the emotion of the sentence and (2) the emotion of the response.

For each approach, we experimented with *EmoBot* using these sentences and recorded the output responses of *EmoBot*. To mitigate the effect of one sentence on another, we avoided a single conversation for all the sentences. Rather, we narrated each sentence separately and tried to deliver the sentences in a way that supports the emotion of the input sentence, so that the output emotions from *EmoBot* can match the response emotions of the sentences. Table 15 shows each sentence, and the generated output of *EmoBot* using different mappings.

The evaluation shows that the authors of this study's Approach provides the highest exact emotion match and lowest mismatch among all three models. Besides, the adjacent emotion match is also substantial for the authors of this study's approach compared to the other approaches. The correlation-based approach provides better performance in terms of adjacent emotion matching. The audio-based mapping produces the highest incorrect emotions, although the exact match is similar to that of the correlation-based approach. Possible explanations of the results are -

- A single emotion-generating event can be evaluated differently by different people. The authors of this study's discussion-based approach may provide a unified mapping due to the discussion as well as verification of all the components, and thus,d generated the highest exact matches. For the audio-based and correlation-based approaches, the perceptions of different users contribute to the final mappings. Different people may have different judgmental views of the same situation. Therefore, the generalization leads to a higher adjacent emotion match.
- For the audio-based mapping, the situation to be evaluated by the users was narrowed down, compared to the broad spectrum of examples the participants came up with from personal experiences. As a result, the audio-based mappings were less generalized than the correlation-based mapping, resulting in the highest mismatch. From a different perspective, the audio-based approach is more unified than the correlation-based mapping and produces at least equal or better performance in terms of exact emotion match.

Therefore, in this paper, we consider the authors of this study's discussion-based mapping approach and implement this mapping for the final evaluation of *EmoBot*.

6. EmoBot evaluation

We compare our implemented emotional chatbot *EmoBot* with a non-emotional chatbot. Therefore, we have conducted both objective and subjective evaluations on *EmoBot* (emotional chatbot) and BotLibre (2013) (non-emotional chatbot). Accordingly, We compared the evaluation results found from these two chatbots.

6.1. Objective evaluation

We used the Cornell Movie-Dialogs Corpus (Danescu-Niculescu-Mizil & Lee, 2011) as our dataset for this evaluation. Sentences were paired up as tuples (input sentence, response). We fed a large number of input sentences from the tuples to both *EmoBot* and BotLibre to find the responses from the chatbots. For a particular (input sentence, response) tuple, we have three sources to get three separate responses - (i) the actual response of the tuple, (ii) the response from *EmoBot* for the input sentence, and (iii) the response from BotLibre for the input sentence. We refer to them as sources (i), (ii), and (iii) for further discussion. We analyzed the responses from sources (ii) and (iii) against source (i). We ran three types of analysis – semantic similarity, sentiment difference, and emotional accuracy – in a laptop with an intel core i5 processor, 8 GB RAM, and a 64-bit Ubuntu operating system.

6.1.1. Evaluation results

For **semantic similarity**, we converted the responses from source (i), (ii), and (iii) to vectors using the conceptnet numberbatch (Speer, Chin, & Havasi, 2017) semantic vectors. Then we calculated the cosine similarity between the vectors from source (i) - (ii) and source (i) - (iii). Fig. 6(a) exhibits the average similarity for different sample sizes for both chatbots, where *EmoBot* scores higher than BotLibre.

For **sentiment analysis**, we calculated the sentiment for the responses from sources (i), (ii), and (iii). VaderSentiment (Hutto & Gilbert, 2014) was used to calculate the sentiments. We then calculated the numeric differences in sentiment between the source (i)-(ii), and (i)-(iii). Fig. 6(b) shows the average sentiment difference for different sample sizes for both the chatbots. It reveals that for a lower number of samples, the difference is higher in *EmoBot*, but as sample size increases, the difference goes lower for *EmoBot* than for BotLibre.

For **emotion accuracy**, we calculated the emotion of the responses from all the sources using Paralleldots API. Then we compared the response emotion of sources (ii) and (iii) against source (i). Fig. 6(c) presents the accuracy of exact emotion matching and Fig. 6(d) presents the accuracy of adjacent emotion matching for different sample sizes. Adjacent emotions are shown in Table 16. The adjacency of the emotions is defined by the position of the emotions in the valence-arousal space (Jin X., 2005). The figures confirm that for both exact and adjacent emotion matching, *EmoBot* has higher accuracy than BotLibre.

6.2. Subjective evaluation

82 human participants were selected through a snowball sampling to use *EmoBot* and BotLibre. A total of 82 participants participated in the evaluation process, ages ranging from 16 to 63 years, with an average of 26.575 years. Among the 82 participants, 29 were female, and the rest were male Fig. 7.

In the first phase, the participants could see the emotion and mood state generated in the *EmoBot*. 70 participants participated in the first phase. In the second phase, 52 participants engaged in the evaluation where they could not see the emotion and mood state generated in the *EmoBot*. Therefore, only the response was different between the bots. 20 participants participated in the second phase. All data was taken in a quiet room to mitigate the effects of noise. Participants were encouraged to have a conversation with the bots in random order for 3–5 min.

Table 15
Evaluation result of different mappings.

Serial	Sentence	Source	Sentence emotion	Response emotion	Authors of this study's approach	Correlation approach	Audio approach
1	Hi.	dataset	unclassified	unclassified	surprise	joy[a]	anger
2	How do you feel?	dataset	unclassified	unclassified	unclassified[b]	joy[a]	unclassified[b]
3	What are you doing?	dataset	unclassified	unclassified	sadness[a]	joy[a]	unclassified[b]
4	Oh my God! Are you serious?	dataset	surprise	unclassified	surprise	joy[a]	fear[a]
5	Hi! Nice to meet you!	dataset	joy	unclassified	joy[a]	joy[a]	fear[a]
6	I got no sleep last night!	dataset	anger	unclassified	sadness[a]	anger	fear[a]
7	So, are you ready to go?	dataset	unclassified	unclassified	joy[a]	joy[a]	fear[a]
8	I'm so proud of you.	dataset	joy	joy	joy[b]	joy[b]	fear
9	Great, I'll see you then.	dataset	joy	joy	sadness	joy[b]	fear
10	Glad to have you back.	real-world	joy	joy	surprise[a]	joy[b]	fear
11	I'm sorry.	dataset	sadness	sadness	sadness[b]	joy	unclassified[a]
12	Oh my God, What happened?	dataset	surprise	sadness	sadness[b]	unclassi-fied[a]	fear[a]
13	Does it still hurt?	dataset	sadness	sadness	sadness[b]	anger	unclassified[a]
14	Man, I didn't think we were gonna make it!	dataset	surprise	fear	unclassified[a]	anger	fear[b]
15	Look out! There is a snake.	real-world	fear	fear	surprise	fear[b]	fear[b]
16	You are going to be expelled.	real-world	anger	fear	fear[b]	sadness[a]	fear[b]
17	Go!	dataset	anger	surprise	surprise[b]	anger	fear
18	Oh, look at the little cat!	dataset	surprise	surprise	unclassified	joy	fear
19	You have won the first place in the lottery!	real-world	surprise	surprise	surprise[b]	joy[a]	fear
20	You are so stupid	anger	real-world	anger	fear	fear	fear
21	I'll not give your money back.	anger	real-world	anger	surprise[a]	fear	fear
22	I will break the window of you car.	real-world	anger	anger	surprise[a]	sadness	fear
				Exact	8	4	5
				Adjacent	8	9	7
				Incorrect	6	9	10

[a] Adjacent match.
[b] Exact match.

Table 16
Adjacency list of emotions in valence-arousal space.

Emotion	Adjacent Emotion(s)
Anger	Disgust and Surprise
Disgust	Fear, Surprise, and Anger
Fear	Sadness, Disgust, and unclassified
Happiness	Surprise and unclassified
Sadness	Fear and unclassified
Surprise	Happiness and Disgust

Table 17
Questionnaire for rating of *EmoBot* and BotLibre.

Metrics	Questions
Efficiency	1. This chatbot can understand what I am trying to say just like a human 2. This chatbot can answer my questions efficiently 3. This chatbot can handle unexpected input
Effectiveness	4. This chatbot can understand my personality 5. This chatbot can understand my emotion 6. This chatbot can generate more realistic and appropriate response 7. This chatbot can generate semantically correct response
Satisfiction	8. I can converse with this chatbot easily 9. I prefer the responses of this chatbot 10. I find this chatbot easy to use 11. I will use this chatbot often 12. I will recommend a friend to use this chatbot

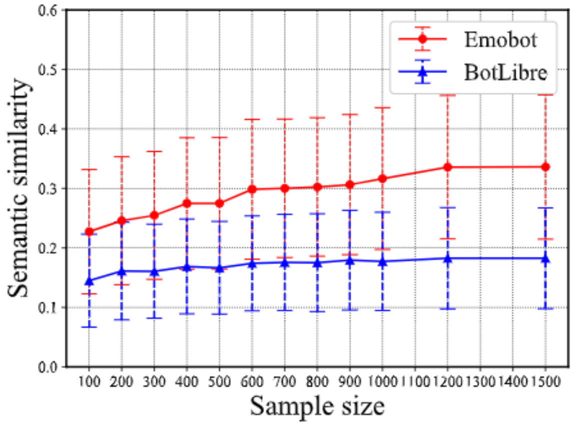
After the conversation is over, we provided each participant with a log of the conversation. The participants rated each response from both bots on a scale of 1 to 5, 1 being very poor and 5 being excellent. Also, participants evaluated the overall performance of the bots in three metrics (Table 17) - Efficiency (question answering, handling unexpected situations), Effectiveness (personality, emotion, realism), and Satisfaction (ease of use, functionality). In the first phase, we took scores for each response from the user after the whole conversation has ended. In the second phase, with the 20 users, we took scores for each response immediately after that response.

We define three metrics following the work of Pamungkas (2019) as follows:

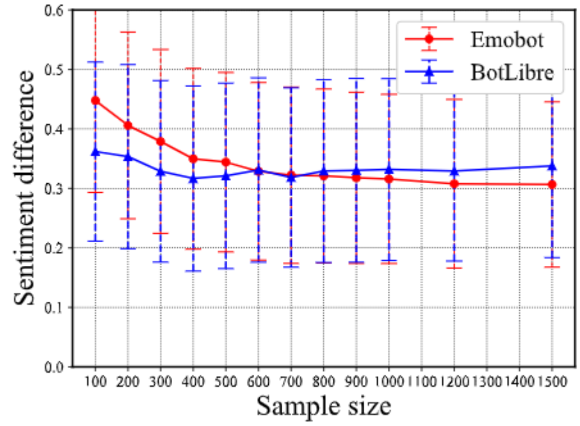
- **Efficiency** means that the chatbot is robust to manipulation and unexpected input (Pamungkas, 2019). Apart from those, it also means the chatbots' ability to control damage and inappropriate utterance.
- **Effectiveness** indicates functionality and humanity. From a functional point of view, a study by Eeuwen (2017) proposes to assess how a chatbot can interpret commands accurately. Chatbots' ability to execute the task as requested and the output linguistic accuracy (Cohen & Lane, 2016) are some other criteria that fall under this metric. Meanwhile, the human aspect means the machine should pass the Turing test (Weizenbaum, 1966).
- **Satisfaction** has three categories which are affect, ethics and behavior, and accessibility. Affect indicates chatbots' ability to convey personality, provide conversational cues, provide emotional information through tone, inflection, and expressivity, entertain and/or enable the participant to enjoy the interaction, and also read and respond to moods of human participant (Meira & Canuto, 2015). The ethics and behavior category indicates how a chatbot can protect and respect privacy (Eeuwen, 2017). Accessibility focuses on assessing the chatbot's ability to detect meaning and, also respond to social cues (see Table 18).

6.2.1. Evaluation results

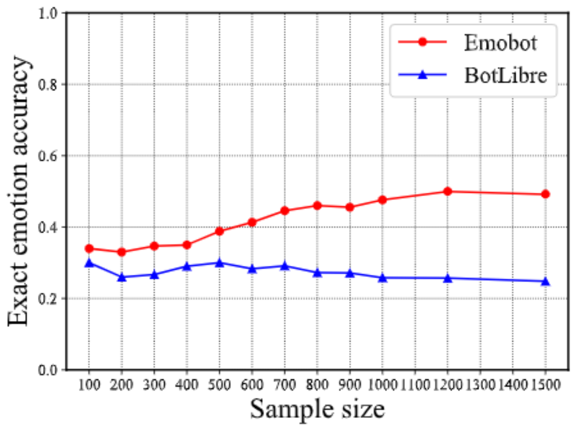
Score comparisons of different evaluation metrics for the two bots, BotLibre, and *EmoBot* are shown in figure Fig. 8. Fig. 8(a) presents the scores of the two bots. when users could see the emotions generated in the *EmoBot*, and Fig. 8(b) shows the scores where the user could not. When emotion is visible, the scores of *EmoBot* are higher than BotLibre.



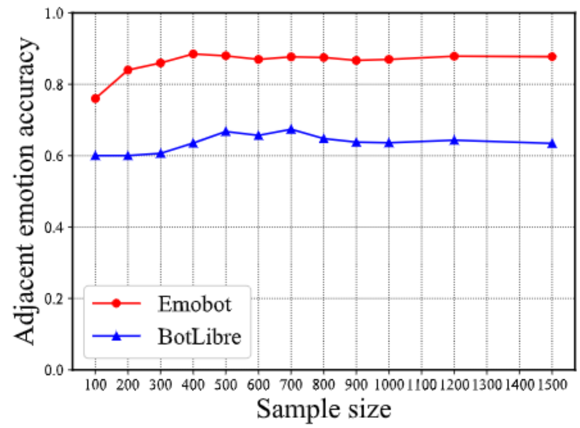
(a) Semantic Similarity between Responses



(b) Sentiment Difference between responses

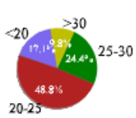


(c) Exact emotion accuracy between Responses

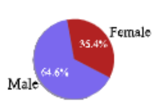


(d) Adjacent emotion accuracy between Responses

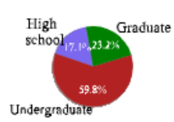
Fig. 6. Objective evaluations.



(a) Age



(b) Gender



(c) Education

Fig. 7. Demographic information.

Table 18
Overall improvement of *EmoBot* over BotLibre.

Objective evaluations		Subjective evaluations	
Metric	Improvement	Metric	Improvement
Sentiment difference	9%	Efficiency	8%
Semantic similarity	247%	Effectiveness	25%
Exact emotion accuracy	38%	Satisfaction	17%
Adjacent emotion accuracy	92%	Response scores	7%

This result indicates that the users appreciated seeing the emotions of *EmoBot*. They found the expression of emotion more effective and satisfying.

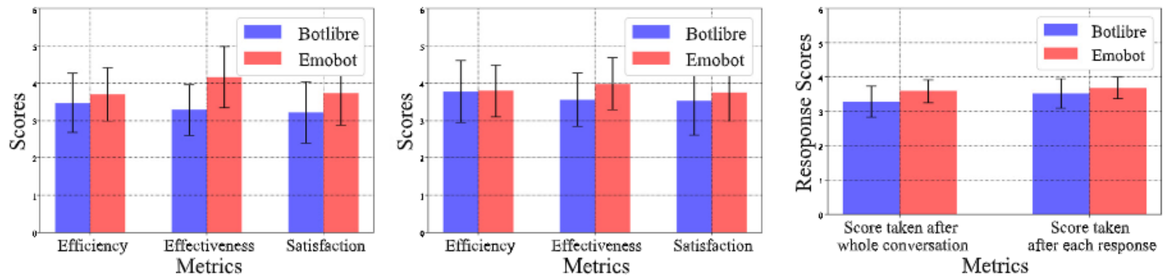
Without the emotions being visible, the scores are much closer. In this case, users could not see the emotions generated by the bots, but they received emotionally motivated responses from *EmoBot*, rather than generic, non-emotional responses from BotLibre. Though the margin is smaller, users still appreciate *EmoBot* over Botlibre according to the result. The efficiency scores for the bots are almost the same, but the effectiveness and satisfaction scores of *EmoBot* are better than that of BotLibre. Fig. 8(c) shows the average scores of each response generated by the bots. The first set of bars shows the scores when they were taken after the end of the whole conversation. The second set shows the scores when they were taken immediately after each response. The scores are a little higher when the scores were taken immediately after each response, rather than after the whole conversation ended. However, in both cases, *EmoBot* scored higher than BotLibre. As a result, users appreciated the responses from *EmoBot* more than the responses from BotLibre.

7. Discussion

In this section, we discuss how our study extends existing work done on emotion generation on chatbots and answer the research questions we set on to explore earlier in this paper.

7.1. Implementation of an emotional chatbot, EmoBot : Using a psychological theory (RQ1):

Our work contributes to creating a chatbot that maintains its emotional state so that the conversation feels as natural as possible. In this



(a) Scores with showing emotion to the users

(b) Scores without showing emotion to users

(c) Scores of the generated responses

Fig. 8. Score comparison between BotLibre and *EmoBot*.

light, based on Cognitive Appraisal Theory (Ortony et al., 1990), we implement an emotion-aware chatbot *EmoBot* by assessing a situation where we interpret various aspects of an event and arrive at an emotional response based on the interpretation. Our chatbot's emotional response not only reflects on a user's emotion but also the system's current emotional state which will let the participants converse in a more natural way. Here, our artificial emotion generation system first converts the interpretation of an event to a numerical value. We have used informative variables to evaluate an emotion-generating event (Hudlicka, 2015b) for artificial emotion generation. For calculating such informative variables to evaluate an emotion-generating event (appraisal variables), we propose three different approaches including the authors of this study's' discussion-based, correlation-based, and audio-based approaches. From all of these approaches, we generate three mappings — input to relevance-implication, mood-personality to coping potentials, and informative variables to evaluate an event to emotions (appraisal variables). As we have proposed three novel approaches for calculating the variables, we contribute to the prior work (Hudlicka, 2015a) of mapping informative variables to evaluate an emotion-generating event (appraisal variables). In addition, based on the previous emotion, personality traits, and emotional state of our chatbot, *EmoBot* generates a response inducing appropriate emotion. As a result, the users of *EmoBot* do not get the same emotional response every time for a specific speech, rather the generated responses will be dependent upon the state of the emotion and personality trait of the bot at the particular time of conversation, which is more natural. For example, if the bot is currently in an angry emotion and the user says 1I have had a fever for three days', our bot's emotional response will depend upon how angry it was at that moment and the speech's sentimental value. In this context, one cannot expect to get a happy emotional response from someone immediately after making him/her angry. Thus, instead of focusing on what type of response the user will like in general situations, we focused on creating a system as human-centric as possible by focusing on the particular situation. In this way, we introduce a novel way of implementing a chatbot that generates emotion during the general-purpose conversation, while valuing the specific temporal impact during the conversation. As a result, we not only focus on establishing empathy with the user but also our chatbot incorporates the emotional state of the system.

Our study builds upon Cognitive Appraisal Theory (Ortony et al., 1990) for artificial emotion generation and this extends prior research (Cano et al., 2021; Pudane et al., 2017; fei Shi et al., 2011; Tavakoli & Palhang, 2016; Velásquez, 1997) on artificial emotion generation model. We further implement our emotion generation model in a chatbot for general-purpose communication. Thus, our work extends previous research studies on emotion generation that are focused on a specific use case, such as customer service on social media (Hu et al., 2018; Lasecki et al., 2013; Xu et al., 2017), generating humorous text or jokes (Dybala et al., 2010; Yang et al., 2015), and system creation to predict and generate specific emotions on the addressee's mind (Hasegawa, Kaji, Yoshinaga, & Toyoda, 2013). As we created an emotion-aware chatbot for natural conversation, we also extend prior research (Cano et al., 2021) that focused on recognizing emotions when interacting with children with ASD. Our research further extends prior study (Huber et al., 2018) on an image-grounded conversational agent because we propose and implement an emotion-aware chatbot using audio input.

7.2. Appreciation for a chatbot with emotion generation capability (RQ2)

The listener must respond to the reactive emotion in a way that makes sense as per the speaker's emotional state instead of playing just a question-answer role performed by traditional chatbots (Li, Ishi, Inoue, Nakamura, & Kawahara, 2019). This is particularly important for specific types of people. For example, depressed people need a human touch, and in the case of reducing depression, emotion-aware chatbots can be of great use (Patel, Thakore, Nandwani, & Bharti, 2019). According to the human evaluation results, our chatbot can embed human-perceivable emotions in responses. As such, emotional tones used by chatbots in their responses could significantly affect user experience, our work is of much practical value. Meanwhile, the emotion-aware chatbot performs reasonably in terms of appropriateness and helpfulness levels during our evaluation there. We summarize and identify some major emotions that commonly occur in conversations, and these emotions are of great value to users. The benefits of using these emotions include reducing users' negative emotions, increasing their positive emotions, and eventually increasing user satisfaction. Statistically, our evaluation indicates that the passionate responses generated by *EmoBot* are appreciated by the users. Interestingly, according to previous work on chatbot (Lasecki et al., 2013; Xu et al., 2017), to a certain extent, human responses still outperform chatbot-generated responses on appropriateness which is probably due to the emotional tone embedded in the human responses. During natural conversation, people tend to bring up emotional thoughts and ideas. Thus, the emotional tone has a significant positive effect on the user experience (Li et al., 2019). Similarly, from our subjective evaluation, we can interpret that users appreciate an emotional chatbot over a non-emotional one. This extends the previous research studies on how emotion-aware chatbots are preferable among users (Hu et al., 2018; Li et al., 2019; Patel et al., 2019). Therefore, we predict that, in the upcoming days, the use of emotional chatbots will continue to grow and the usage of traditional question-answer chatbots will be very limited. However, we also feel that an emotional chatbot that generates inaccurate emotion might have potentially negative consequences.

8. Conclusion and future work

In this work, we propose a computational method of emotion generation following Cognitive Appraisal Theory and execute it in a chatbot for general-purpose conversations. We design our method of artificial emotion generation by combining the knowledge of Computer Science and Cognitive Psychology. Our goal in this work has been twofold — discovering a path to generate primary emotions, and making the generation process computationally feasible so that it can be implemented in systems such as chatbots. The main challenge with the Cognitive Appraisal Theory is to determine a way to calculate the informative variables to evaluate an emotion-generating event. In this regard, we propose three different approaches to deal with this challenge — the authors of this study's discussion-based, correlation-based, and audio-based approaches. To the best of our knowledge, we are the first to implement and provide an artificial emotion generation method based on a psychological theory in such a computationally feasible manner.

Furthermore, we demonstrate the computation by designing and developing an emotional chatbot, namely *EmoBot*, based on our proposed method. We carry out both subjective and objective evaluations of our developed *EmoBot* and compare it with a classical chatbot BotLibre, where, users appreciated our implemented emotional chatbot: *EmoBot*. *EmoBot* offers a new opportunity to provide individualized attention to users at scale and encourage interaction between humans and computers. This will help in achieving social, information, and economic benefits if we use the bot in corresponding places. In the future, we plan to obtain complex emotions. We are also considering emotion generation by exploring other psychological theories to find a better mapping of emotion generation that can express emotions. We intend to compare the *EmoBot* system with other currently available emotion-considering chatbots in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article

Acknowledgments

This work has been conducted at and supported by the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.

References

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, *6*(7).

Adiwardana, D., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., et al. (2020). Towards a human-like open-domain chatbot. arXiv:2001.09977.

Agrafioti, F., Hatzinakos, D., & Anderson, A. K. (2011). ECG pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, *3*(1), 102–115.

Barańczuk, U. (2019). The five factor model of personality and emotion regulation: A meta-analysis. *Personality and Individual Differences*, *139*, 217–227.

BotLibre (2013). BOT libre! - free chat bot hosting. URL http://www.botlibre.com.

CakeChat (2018). CakeChat: Emotional generative dialogue system. URL https://cakechat.replika.ai/.

Candia-Rivera, D., Catrambone, V., Thayer, J. F., Gentili, C., & Valenza, G. (2022). Cardiac sympathetic-vagal activity initiates a functional brain–body response to emotional arousal. *Proceedings of the National Academy of Sciences*, *119*(21), Article e2119599119.

Cano, S., González, C. S., Gil-Iranzo, R. M., & Albiol-Pérez, S. (2021). Affective communication for socially assistive robots (SARs) for children with autism spectrum disorder: A systematic review. *Sensors*, *21*(15), 5166.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., & Yan, H. (2000). Human conversation as a system framework: Designing embodied conversational agents. *Embodied conversational agents*, 29–63.

Cervone, D., & Little, B. R. (2019). Personality architecture and dynamics: The new agenda and what's new about it. *Personality and Individual Differences*, *136*, 12–23.

Chen, S.-Y., Hsu, C.-C., Kuo, C.-C., Ku, L.-W., et al. (2018). Emotionlines: An emotion corpus of multi-party conversations. arXiv preprint arXiv:1802.08379.

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.

Cohen, D., & Lane, I. (2016). An oral exam for measuring a dialog system's capabilities. *vol. 30*, In *Proceedings of the AAAI conference on artificial intelligence*.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). Superagent: A customer service chatbot for e-commerce websites. In *Proceedings of ACL 2017, system demonstrations* (pp. 97–102).

Damasio, A. R. (2003). *Looking for spinoza: Joy, sorrow, and the feeling brain*. Orlando: Harcourt.

Danescu-Niculescu-Mizil, C., & Lee, L. (2011). Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. arXiv preprint arXiv:1106.3077.

Du, Y., Jin, X., Yan, R., & Yan, J. (2023). Sentiment enhanced answer generation and information fusing for product-related question answering. *Information Sciences*.

Dybala, P., Ptaszynski, M., Maciejewski, J., Takahashi, M., Rzepka, R., & Araki, K. (2010). Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, *2*(1), 31–48.

Easton, K., Potter, S., Bec, R., Bennion, M., Christensen, H., Grindell, C., et al. (2019). A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. *Journal of Medical Internet Research*, *21*(5), Article e12996.

Eeuwen, M. v. (2017). *Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers*. Master's thesis, University of Twente.

Fossi, J., Dzwonkowski, A., & Othman, S. (2021). Analyzing music genre popularity. In *Intelligent computing* (pp. 284–294). Springer.

Frijda, N. H. (1986). *The emotions.*. Cambridge University Press.

Frijda, N. H. (2009). Emotion experience and its varieties. *Emotion Review*, *1*(3), 264–271.

Garcia-Garcia, J. M., Penichet, V. M., & Lozano, M. D. (2017). Emotion detection: a technology review. In *Proceedings of the XVIII international conference on human computer interaction* (pp. 1–8).

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *International workshop on intelligent virtual agents* (pp. 125–138). Springer.

Ha, D., Dai, A., & Le, Q. V. (2016). Hypernetworks. arXiv preprint arXiv:1609.09106.

Hasegawa, T., Kaji, N., Yoshinaga, N., & Toyoda, M. (2013). Predicting and eliciting addressee's emotion in online dialogue. In *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 964–972).

Heller, B., Proctor, M., Mah, D., Jewell, L., & Cheung, B. (2005). Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ innovate learning* (pp. 3913–3918). Association for the Advancement of Computing in Education (AACE).

Herzig, J., Feigenblat, G., Shmueli-Scheuer, M., Konopnicki, D., Rafaeli, A., Altman, D., et al. (2016). Classifying emotions in customer support dialogues in social media. In *Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue* (pp. 64–73).

Hieida, C., & Nagai, T. (2021). Survey and perspective on social emotions in robotics.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., et al. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–12).

Huang, T.-H. K., Chang, J. C., Swaminathan, S., & Bigham, J. P. (2017). Evorus: A crowd-powered conversational assistant that automates itself over time. In *Adjunct publication of the 30th annual ACM symposium on user interface software and technology* (pp. 155–157).

Huang, T.-H. K., Lasecki, W. S., Azaria, A., & Bigham, J. P. (2016). " Is there anything else I can help you with?" challenges in deploying an on-demand crowd-powered conversational agent. In *Fourth AAAI conference on human computation and crowdsourcing*.

Huber, B., McDuff, D., Brockett, C., Galley, M., & Dolan, B. (2018). Emotional dialogue generation using image-grounded language models. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–12).

Hudlicka, E. (2015a). Computational analytical framework for affective modeling: towards guidelines for designing computational models of emotions. In *Handbook of research on synthesizing human emotion in intelligent systems and robotics* (pp. 1–62). IGI Global.

Hudlicka, E. (2015b). Computational analytical framework for affective modeling: Towards guidelines for designing computational models of emotions.. In *Psychology and mental health: Concepts, methodologies, tools, and applications*. IGI.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *vol. 8*, In *Proceedings of the international AAAI conference on web and social media* (pp. 216–225).

Io, H., & Lee, C. (2017). Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 215–219). IEEE.

Jain, T., & Rath, S. (2023). *Impact of emotional intelligence on customer loyalty towards online fashion stores in India*: *Technical Report EasyChair*, EasyChair.

Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The impact of chatbots on customer loyalty: A systematic literature review. *Journal of theoretical and applied electronic commerce research*, *17*(1), 212–229.

Jin X., W. Z. (2005). An emotion space model for recognition of emotions in spoken Chinese.. *Affective Computing and Intelligent Interaction.*, *3784*, 397–402.

Kim, S., Goh, J., & Jun, S. (2018). The use of voice input to induce human communication with banking chatbots. In *Companion of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 151–152).

Kim, H.-R., & Kwon, D.-S. (2010). Computational model of emotion generation for human–robot interaction based on the cognitive appraisal theory. *Journal of Intelligent and Robotic Systems*, *60*(2), 263–283.

Koivunen, S., Ala-Luopa, S., Olsson, T., & Haapakorpi, A. (2022). The march of chatbots into recruitment: Recruiters' experiences, expectations, and design opportunities. *Computer Supported Cooperative Work (CSCW)*, *31*(3), 487–516.

Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., & Bigham, J. P. (2013). Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on user interface software and technology* (pp. 151–162).

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145.

Lazarus, R. S., & Lazarus, B. N. (1994). *Passion and Reason: Making sense of our emotions*. USA: Oxford University Press.

Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016). A persona-based neural conversation model. arXiv preprint arXiv:1603.06155.

Li, Y., Ishi, C. T., Inoue, K., Nakamura, S., & Kawahara, T. (2019). Expressing reactive emotion based on multimodal emotion recognition for natural conversation in human–robot interaction. *Advanced Robotics*, *33*(20), 1030–1041.

Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., & Jurafsky, D. (2016). Deep reinforcement learning for dialogue generation. arXiv preprint arXiv:1606.01541.

Liu, Y.-l., Hu, B., Yan, W., & Lin, Z. (2023). Can chatbots satisfy me? A mixed-method comparative study of satisfaction with task-oriented chatbots in mainland China and Hong Kong. *Computers in Human Behavior*, Article 107716.

Liu, C.-W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., & Pineau, J. (2017). How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. arXiv:1603.08023.

Luger, E., & Sellen, A. (2016). " Like having a really bad pa" the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5286–5297).

Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., & Cambria, E. (2019). Dialoguernn: An attentive rnn for emotion detection in conversations. *vol. 33*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 6818–6825).

Marks, J. (1982). A theory of emotion. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *42*(2), 227–242.

Martin, J. L. (1985). Consumer perception of and response to corporate complaint handling....

McCollough, M. A. (2000). The effect of perceived justice and attributions regarding service failure and recovery on post-recovery customer satisfaction and service quality attitudes. *Journal of Hospitality & Tourism Research*, *24*(4), 423–447.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament.. *Current Psychology.*, *14*(4), 261–292.

Meira, M. d. O., & Canuto, A. d. P. (2015). Evaluation of emotional agents' architectures: an approach based on quality metrics and the influence of emotions on users. *vol. 1*, In *Proceedings of the world congress on engineering* (pp. 1–8).

Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. *Procedia Computer Science*, *201*, 421–428.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge University Press.

Pamungkas, E. W. (2019). Emotionally-aware chatbots: A survey. arXiv preprint arXiv:1906.09774.

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 311–318). Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, http://dx.doi.org/10.3115/1073083.1073135, URL https://aclanthology.org/P02-1040.

Park, H., & Cameron, G. T. (2014). Keeping it real: Exploring the roles of conversational human voice and source credibility in crisis communication via blogs. *Journalism & Mass Communication Quarterly*, *91*(3), 487–507.

Patel, F., Thakore, R., Nandwani, I., & Bharti, S. K. (2019). Combating depression in students using an intelligent chatbot: a cognitive behavioral therapy. In *2019 IEEE 16th india council international conference (INDICON)* (pp. 1–4). IEEE.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3–33). Elsevier.

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users'affective states. *Applied artificial intelligence*, *19*(3–4), 267–285.

Pudane, M., Lavendelis, E., & Radin, M. A. (2017). Human emotional behavior simulation in intelligent agents: processes and architecture. *Procedia Computer Science*, *104*, 517–524.

Ritter, A., Cherry, C., & Dolan, W. B. (2011). Data-driven response generation in social media. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 583–593).

Roniotis, A., & Tsiknakis, M. (2018). Detecting depression using voice signal extracted by chatbots: A feasibility study. In *Interactivity, game creation, design, learning, and innovation: 6th international conference, ArtsIT 2017, and second international conference, DLI 2017, Heraklion, Crete, Greece, october 30–31, 2017, proceedings 6* (pp. 386–392). Springer.

Russell, J. A. (2003). Core affect and the psychological construction of emotion.. *Psychological review*, *110*(1), 145.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant.. *Journal of personality and social psychology*, *76*(5), 805.

de Sá Siqueira, M. A., Müller, B. C., & Bosse, T. (2023). When do we accept mistakes from chatbots? The impact of human-like communication on user experience in chatbots that make mistakes. *International Journal of Human–Computer Interaction*, 1–11.

Sailunaz, K., Dhaliwal, M., Rokne, J., & Alhajj, R. (2018). Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, *8*(1), 1–26.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

fei Shi, X., liang Wang, Z., Ping, A., & kun Zhang, L. (2011). Artificial emotion model based on reinforcement learning mechanism of neural network. *The Journal of China Universities of Posts and Telecommunications*, [ISSN: 1005-8885] *18*(3), 105–109. http://dx.doi.org/10.1016/S1005-8885(10)60071-4, URL https://www.sciencedirect.com/science/article/pii/S1005888510600714.

Simomukay, E. (2018). Experiência na construção pedagógica de bots para o ensino de química. *Revista de Educação do Vale do Arinos-RELVA*, *5*(2), 17–24.

Sloman, A. S. M., & Chrisley, R. (2005). The architectural basis of affective states and processes.. In *Who needs emotions: The brain meets the machine.* (pp. 203–244). Oxford University Press.

Somasiri, L., Galabada, S., Wijethunga, H., Dayananda, H., Nugaliyadde, A., Thelijjagoda, S., et al. (2016). D-REHABIA: A drug addiction recovery through mobile based application.

Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge.. In *Proceedings of conference on artificial intelligence.* (pp. 4444–4451). AAAI.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104–3112).

Tavakoli, M. K. M., & Palhang, M. (2016). Three levels of information processing: improvement using personality.. In *Proceedings of iranian conference on intelligent systems.* (pp. 1–6). ICIS.

Tu, Q., Li, Y., Cui, J., Wang, B., Wen, J.-R., & Yan, R. (2022). MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. arXiv preprint arXiv:2203.13560.

Velásquez, J. D. (1997). Modeling emotions and other motivations in synthetic agents. In *AAAI'97/IAAI'97, Proceedings of the fourteenth national conference on artificial intelligence and ninth conference on innovative applications of artificial intelligence* (pp. 10–15). AAAI Press, ISBN: 0262510952.

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., et al. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3506–3510).

Xu, J., Szlam, A., & Weston, J. (2021). Beyond goldfish memory: Long-term open-domain conversation. arXiv preprint arXiv:2107.07567.

Yang, D., Lavie, A., Dyer, C., & Hovy, E. (2015). Humor recognition and humor anchor extraction. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 2367–2376).

Yu, Z., Papangelis, A., & Rudnicky, A. (2015). TickTock: A non-goal-oriented multimodal dialog system with engagement awareness. In *2015 AAAI spring symposium series*.

Zhang, L., Erickson, L. B., & Webb, H. C. (2011). Effects of "emotional text" on online customer service chat.

Zhou, Q. (2018). Multi-layer affective computing model based on emotional psychology.. *Electronic Commerce Research.*, *19*(1), 109–124.

Zhu, Q., Chau, A., Cohn, M., Liang, K.-H., Wang, H.-C., Zellou, G., et al. (2022). Effects of emotional expressiveness on voice chatbot interactions. In *Proceedings of the 4th conference on conversational user interfaces* (pp. 1–11).