UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Alar Kirikal

Computational Simulation of How Emotions are Constructed in Our Brain According to the Theory of Constructed Emotion

Master's Thesis (30 ECTS)

Supervisor(s): Prof. Kuldar Taveter

Tartu 2021

Computational Simulation of How Emotions are Constructed in Our Brain According to the Theory of Constructed Emotion

Abstract:

Understanding emotion is one of the most difficult tasks for computers. Not only is it difficult for computers, but humankind itself even has different theories of how emotions are constructed in our mind and how exactly the brain constructs feelings. Most theories about emotions say that basic emotions are genetically endowed, whereas the theory of constructed emotion states that our brain constantly uses past experience to guide our actions and generate emotional feelings and, in each situation a new instance of emotion is generated. This takes away the generic approach to classify emotions and allows us to describe emotions in multidimensional values – valence, arousal, and dominance. By describing emotional sensations on these dimensions, we can start comparing emotional states based on input and reflect the emotional state on a computer and thus improve the current state of Human-Computer Interaction (HCI). This allows creating a simulation that is not trying to find the best output from learned data by finding common features of the input and expected output but is emotionally intelligent and generates output based on its current emotional state. This thesis analyses the possibilities of the given theory about emotions and its possibilities of implementing it in HCI applications. By improving the emotional intelligence of these applications, computer programs can become better at evaluating our emotional state and act accordingly.

Keywords:

Emotion recognition, neuroscience

CERCS: P175 Informatics, systems theory

Ehitatud emotsiooni teoorial põhinev emotsioonide ülesehituse simulatsiooni loomine

Lühikokkuvõte:

Emotsioonide mõistmine on arvutite jaoks üks raskemaid ülesandeid. Antud ülesanne ei ole raske ainult arvutitele vaid ka inimkonnale, kellel on erinevaid teooriaid sellest, kuidas meie aju töötleb ning mõistab emotsiooni. Enamik teooriaid emotsioonide kohta ütlevad, et põhiemotsioonid on meile sündides kaasa antud geneetiliselt. Ehitatud emotsiooni teooria (TCE) väidab, et meie aju pidevalt kasutab meie eelnevaid kogemusi selleks, et juhtida meie edasisi tegevusi ja genereerida peas emotsioone. Iga olukorra jaoks kus oleme tekitatakse uus emotsiooni juhtum. Selline lähenemine muudab üldist arusaama sellest, kuidas emotsioone klassifitseeritakse ja annab meile võimaluse kirjeldada emotsioone mitmedimensionaalsete väärtustega – valentsiga, erutusega ning domineerivus. Kirjeldades emotsionaalseid tundeid läbi nende väärtuste saame me hakata emotsioone võrdlema ning tänu sellele neid peegeldama läbi tarkvara, mis lubab meil luua rakendusi, mis parendavad inimeste suhtlust arvutitega. Sellist meetodit kasutades on võimalik luua inimese aju simulatsioon, mis ei ürita leida parimat vastet õpitud teksti seast nagu seda praegu teevad masinõppe mudelid. Antud simulatsioon oleks iseõppiv ning emotsionaalselt tark. Antud magistritöö uurib ehitatud emotsiooni teooriat ning üritab rakendada seda ehitades simulatsioonirakenduse. Parendades emotsionaalset intelligentsust rakendustes, mis suhtlevad inimestega, saavad arvutid hakata paremini hindama inimeste emotsionaalset seisundit ja käituda vastavalt.

Võtmesõnad:

Emotsiooni tuvastus, neuroteadus

CERCS: P175 Informaatika, süsteemiteooria

Table of Contents

1 Introduction 5
1.1 Research Objectives 6
1.2 Research Methods 6
2 Background 7
2.1 Understanding Emotion 7
2.2 Theory of Constructed Emotion 8
2.3 Words in the predictive brain 9
2.4 Emotion Classification 9
2.5 The Principles Underlying the Emotional Chatbot 10
3 Designing and Implementing an Emotional Chatbot 11
3.1 Architecture 11
3.2 Datasets 14
3.3 Data processing – Natural Language Processing 15
3.4 Expectation Generation 18
4 Validation 21
4.1 Setup 21
4.2 Results 23
4.3 Feedback 29
5 Future Work 32
5.1 Possible Improvements 32
5.2 Domains of use 33
6 Conclusions 35
7 References 36
Appendix 40
I. Glossary 40
II. License 41

1 Introduction

Human-Computer Interaction (HCI) is something that computer scientists have been working on since the beginning era of computers. As we are executing software on computers, we also need to interact with it and work on it. The paradigm of doing so has shifted from something that only computer engineers could do and understand to the daily life of everyone using their smartphones, laptops, and even talking to personal assistants. As life becomes more convenient with computers becoming more powerful, then more is also expected from computers and thus better level of communication is expected. By making computers more emotionally aware can improve the user experience of HCI [23].

As computers have become part of our lives and our interaction with them is more frequent than ever, it should also become more natural for us to communicate with computer programs. Currently, a lot of research and development has been done in the field of HCI and a lot of effort has been recently put into the field of artificial intelligence (AI). Using supervised learning algorithms, it is possible to write software that would generate contextual responses by one party of a chat so that the human user might not even understand that they are chatting with a computer application. This approach, however, is based on previous information that the AI-powered software is using and is not mimicking the way humans interact as humans have a lot more complex nervous system, including our brain, that analyses the situation to generate output – an action or in a conversational context a reply.

Emotion has always been an interesting topic to discover and understand and how our brain processes emotions because all of our communication is impacted by emotional states. As many around the world try to define emotion and build ideas around what it is, then there is still no scientific consensus on one single definition [49][50]. While it is definitive, that emotional changes in our mind trigger physiological changes in our brain, then it is still unclear what emotion is in its essence of how we perceive the world. If defining and understanding emotion in our brain has proven to be that difficult, then how are we supposed to make computers emotionally aware and develop emotional consciousness of their own to communicate with us? As we try to make our software seem more human by introducing automated chatbots and personal assistants like Siri [31] and Alexa [32], people expect them to talk back and be more humanlike. This however assumes that the software knows how to handle situations based on input and how to use historic interactional data to procure an action – for example, a reply. One theory for understanding emotion comes from Lisa Feldman Barrett, called the "theory of constructed emotion" [10][35]. This theory states that at all times our brain uses our past experience to help us choose actions to take and help us understand the situations we are in. These understandings can also be emotional concepts from which our brain constructs an emotional state based on the current situation using the input of different sensory information, historical concepts of similar situations, create an instance of emotion and our motivation for action and thus generates an expected outcome of how we choose to act on it [1].

This thesis reports on designing and implementing an experiment, named EmReflect, to build a chatbot application that would not generate responses based on historical data like current algorithms used for chatbots [36][37][38] but instead would try to understand the emotion expressed within the chat and reply with a response that would mimic the emotional state.

The benefits of applications that mimic emotions have already been demonstrated by other research papers [2][51] and have proven to be interesting for people to interact with. An emotionally aware chatbot could potentially help the users to better understand their emotional states by allowing them to write freely and get responses as if they were talking to a friend instead of a computer program. The World Health Organization (WHO) states that more than 264 million people worldwide are affected with depression and between 76% and 75% of the people suffering from depression in low- and middle-income countries do not receive treatment for their disorders. This is not always because of treatment not being available, but depression is often not easily diagnosed and misunderstood as just having a bad mood [3]. Emotionally intelligent chatbot computer application, which could be a part of a personal assistant, could be one of the solutions capable of diagnosing symptoms of depression based on the emotional state earlier than the subjects themselves.

If we can create a simulation that is accurate to reflect emotions based on communication with a real human by using similar input sources as humans do by using sensory input and predictive action determination, it is possible to start creating better artificial intelligence products that could be emotionally aware and intelligent. A simulation that can learn emotions and construct them based on the situations could learn over time and become emotionally mature by experiencing different situations – just like a human.

1.1 Research Objectives

The objective of this research and implementation of EmReflect is to build a simulation to verify that emotions can be constructed, represented, and used as per the theory of constructed emotion. Researching the topic and proposing future work are a few of the ways this thesis should prove to be useful. Validating how emotions are constructed and used in the implementation of EmReflect can be useful for others to help build emotional analysis solutions.

1.2 Research Methods

The first step of this research is to implement EmReflect – an agent, that is capable of understanding emotion and making predictions based on the situation, motivation, and emotional state of the input and the agent itself. Using created software an experiment is run with participants that will conduct a conversation with the chatbot every day to gather feedback about the agent's emotional intelligence and generated responses. Once the experiment is conducted, results from the created database are collected, analyzed, and a feedback form is sent out to the participants to better learn about the user experience of such a solution.

2 Background

This section will provide an overview of the most popular current theories of emotion. There are many ways how emotion can be defined and thus also used within computer science. The theory of constructed emotion will be described in more detail as this is the main theory used for this thesis. Also, the main characteristics of measuring emotions from the text will be described and dimensional values are explained on how emotions can be constructed from text input.

2.1 Understanding Emotion

Emotion can be described as a biological state within a human nervous system that is strongly affected by one's surroundings and thoughts and cause neurophysiological changes [4]. However, there is no single scientific definition of emotion on which everyone agrees. One thing is certain – emotions are extremely complex. Emotions have puzzled scientists for decades and still not having an agreed definition shows how little we understand where our emotions come from and how are they formed. Theories can be widely split into two bigger categories – discrete and dimensional models [5].

Theories of discrete emotional categories mostly state that humans have a basic set of emotions endowed genetically [6]. Paul Ekman stated that the six basic emotions according to his theory are anger, disgust, fear, happiness, sadness, and surprise [7]. The theory of basic emotions claims that these emotions exist in all humans and are universal regardless of culture or origin. The theory of basic emotions also states that these emotions should be universally detectable by facial expressions or other physiological measurements. This claim is illustrated by Figure 1 from [7].

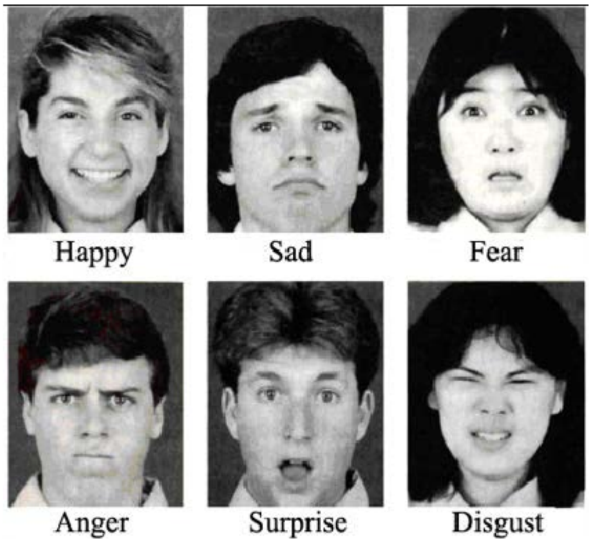

Figure 1. Six basic emotions according to Ekman. [7]

Dimensional models of emotions believe define emotions based on different dimensions. Different authors have stated different dimensions that are used to describe emotions. Examples of dimensions are “pleasure”, “arousal” and “strain” by Wilhelm Max Wundt and “pleasure”, “attention”, and “activation” by Harold Schlosberg [8]. These examples demonstrate that there is a common pattern of what dimensions can be used to describe emotional states. Therefore, different dimensional theories of emotion follow a similar path [9].

As emotions belong to the main characteristics of what makes us truly human in terms of interactions with each other, computer scientists have always tried to bring this closer to computers by making computers seem more humanlike as if they would understand and also show up emotions. Different supervised learning algorithms have been used to make computers respond to inputs in a human-like fashion. Such applications tend to be driven by neural networks or other machine learning algorithms that have extracted features from learned data and are using these features to generate an output for any yet unseen data [39]. This means that the user might think that the computer is smart and is responding by showing emotion whereas it is simply constructing a response based on learned datasets.

2.2 Theory of Constructed Emotion

This thesis is based on the theory of constructed emotion [10][35] proposes another take on what emotions are and how they exist within our life. The theory of constructed emotion considers emotions as concepts that are based on situations. According to the theory of constructed emotion, at any given moment we take several types of input into account when creating a concept for the situation at any given moment in our life [1].

Throughout life, we experience everything by our senses – sight, hearing, touch, smell, and taste. When we first feel something unique through our senses, we generate a concept of what that is and how we feel about it. For example, when we constantly feel cold little drops on our skin while walking outside, we associate it with rain. This means that next time we start sensing something similar, we do not need to process everything that is taking place, and our brain does not have to figure out on the spot that it is raining but instead can rely on some probability on prior experience. Once we have learned a concept, our brain memorizes the situation along with the emotions associated with it [10]. Next time we are in a similar situation our brain also associates the emotions with the given situation. This can also explain why we still feel strong emotions associated with simple memories from childhood.

Once we have learned a specific situation, similar sensory data can already give us information about the situation we are currently experiencing. This also means that our brain is working in a predictive manner. Similarly, to predicting situations from sensory input, we can also predict emotions in the same way [1]. Because of that, talking about sad moments in our life can trigger the same emotion that we felt during the experience we are talking about. This can also explain why some people can have difficulties with specific childhood memories shaping their lives and not being able to let go of some concepts they formed as children. Predicting situations based on sensory input can also explain how our brain works as fast as it does – it does not have to process all the sensory input all the time [11]. This means our brain is constantly processing different possible interpretations of our current situation and chooses the most probable outcome based on the current goal, motivation, and emotion. For example, this explains how “muscle memory” works. Muscles do not have memory, but our nervous system does. Once we have learned a specific action or muscle movement based on certain input, we can predict it happening faster compared to processing the situation every time. In the same way that our nervous system helps us to predict an outcome for muscle action, it can also predict and generate an instance of emotion. In the same way of generating emotions based on the input received from our senses, our brain is actively thinking ahead of situations about to happen and creates emotions about situations that have not yet occurred or for example, while waiting for something good to happen, we often start feeling good already before it takes place because our brain is already living in the expectation of this situation which is accompanied by the emotion we are already feeling. This also means that the more we think about something happening ahead we are also more likely to experience it and the other way round. If we feel very much down and not expecting to feel better soon, it becomes more difficult to feel better as we are not expecting positive emotions to occur in our current emotional state.

Constructed emotions are also not classified the same way as in simpler discrete emotional models. Since emotions are always constructed from different inputs, the feeling of a similar emotional state, for example, happiness, is never actually the same as it always differs in some respects related to the concrete situation we are experiencing. For the theory of constructed emotion, the Valence-Arousal-Dominance (VAD) model [20] helps us to understand how different emotions can be constructed and how different emotions can be perceived.

2.3 Words in the predictive brain

Words we use in our daily language strongly shape the way our brain thinks and how we process things around us [12]. Since birth humans are surrounded by language and words that start shaping our minds. In different linguistic environments and cultures, we are bound to learn to express ourselves in different ways as cultures have their own sets of words for describing the life around them. Once we start going through different situations in life, words help to shape how we start associating these situations and feelings with previous experiences. Once we are told that what we see is “sad”, it will start connecting with other memories in our mind that are associated with that emotion. This means that emotions and interpretations of these emotions are strongly connected with the language and words used to describe different situations throughout life experiences [1]. Moreover, psychological theories also claim that emotional words are used in our brain as cues for situations [40].

Linking words with emotions allows for flexibility and individual understanding of similar situations. People can perceive similar situations with different emotions based on their previous experiences. This explains how people, especially from different cultural backgrounds, can easily react to situations differently. Thus, these learned emotional descriptions are always linked to the concepts that our minds already use for understanding the given word.

2.4 Emotion Classification

The Valence-Arousal-Dominance model presents just one way for describing emotions in a multidimensional format proposed by Russell and Mehrabian [33]. Most models tend to use valence and arousal as the main dimensions of describing constructed emotions [13] and there is no set limit of dimensions or agreed standard to use. However, the valence-arousal-dominance constructionist model has been widely used in different studies [53][54][55] and it has been found support in research [52]. Valence is used to describe positivity, arousal - engagement, and dominance - control over the affective state [13]. For example, happiness is an emotion that is very positive, engaging, and usually a strong dominating feeling. On the other end of the spectrum - sadness is not only negative in terms of valence, but also low in engagement and without much a lot of control over its affective state. This can be compared to anger, which is also very negative, but also extremely engaging and dominant over our emotional state. Comparison of VAD values of the six basic emotions, as hypothesized by Ekman [7] can be seen in Figure 2.

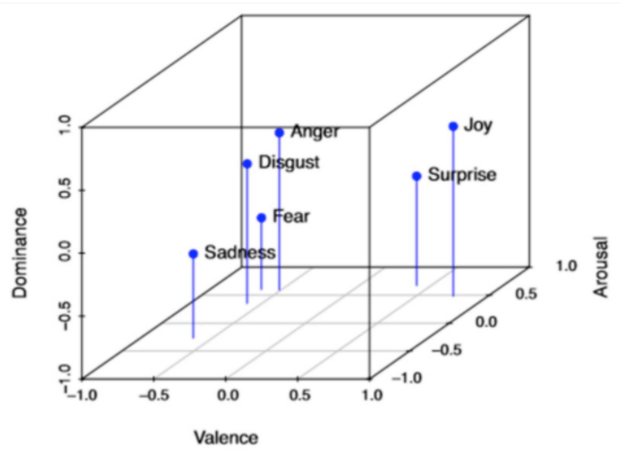

Figure 2. Six basic emotion Valence-Arousal-Dominance values. [34]

This thesis does not proceed from categorizing emotions into a fixed set of emotion types. The reason is that the theory of constructed emotion states that all emotions are instances created of concepts representing various situations. The theory of constructed emotion states that the brain is constantly interpreting the situations we perceive and thus calculating the expected actions to be performed. This means that every emotion we feel can differ based on our current experience [1]. For example, when we feel happy in two different situations, it will feel similar but might have some difference in how we experience it [1]. A typical machine learning algorithm would classify both situations as "happiness", whereas the theory of constructed emotion enables to show how it was constructed and how it differs from other situations by using multidimensional models that are based on the dimensions of valence, arousal, and dominance.

2.5 The Principles Underlying the Emotional Chatbot

In a dialog, both participants are engaging in predictions about each other's emotional states, which is also called emotion co-construction [1]. For that purpose, humans use verbal as well as non-verbal communication, such as facial expressions and body movements. When a human is interacting with a chatbot, the only form of information that a party can use for predicting emotions by the conversation partner is textual information. This means that a chatbot should be able to predict the emotional state by its human conversation partner based on the emotion-related words that the partner has used in the conversation. It can be shown that verbalizing as precisely as possible the emotion the partner has expressed may change the emotional state of the interaction partner by modifying its predictions about the partner. In psychotherapy, this kind of technique is known as the mirroring technique that is a conscious use of active listening by the therapist, accompanied by the reflection of the client's affect language to stimulate a sense of empathy [46]. In our case, the role of the therapist is played by the chatbot. For example, if the chatbot mirrors the textual input by the human by stating "you look sad", this impacts the next emotional state by the human conversation partner. The design and implementation of the emotional chatbot described in Chapter 3 is based on the mirroring technique, which is rooted in emotion co-construction by the conversation partners [1], which are, in this case, the human and the emotional chatbot with whom she or he is interacting.

3 Designing and Implementing an Emotional Chatbot

This chapter will describe the design and implementation of the chatbot application based on the theory of constructed emotion. Section 3.1 will provide an interview of the application's architecture. Section 3.2 describes the datasets used by the application. Section 3.3 addresses natural language processing by the chatbot and Section 3.4 deals with the generation of expectations by the chatbot.

3.1 Architecture

The architecture of the emotion-aware chatbot consists of four kinds of modules: controller, memory, and representations of situations and expectations. These components and relationships between them are represented in Figure 3. We will next describe how the architecture works. An instance of the situation component consists of all the input perceived by the agent about the situation the instance represents. In the architecture of the emotion-aware chatbot EmReflect, the situation consists of the question asked by the chatbot from its user and the reply given to the chatbot as a freeform text. This means that the situation itself contains all the data that is about to be processed by the chatbot and each situation requires a prediction of what action the chatbot should perform next. In EmReflect, the action would be generating and uttering a reply based on the situation.

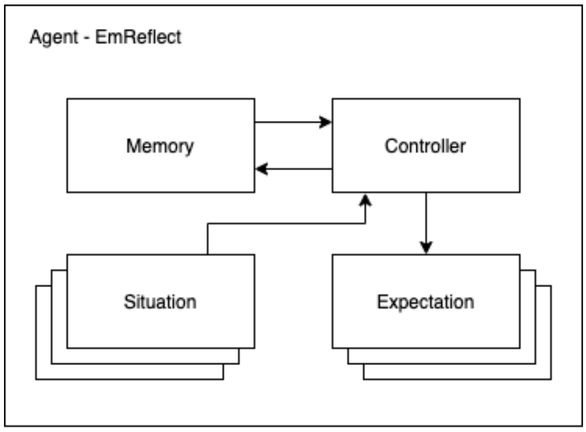

Figure 3. Generic architecture of EmReflect. [41]

A situation instance is the input that triggers a sequence of actions by the agent that calculate the motivation and emotion, while also updating the agent's memory. Each situation will result in a prediction that the agent will eventually perform. In our case, prediction is an emotionally reflective reply by which EmReflect chatbot responds to its user. Motivation and emotion calculations are done for every instance of a situation with the help by the memory. The memory module is also used as a knowledge base for motivation and emotion calculations. Since memory stores situations and expectations that were predicted in the past, these knowledge base entries are helpful in calculating future expectations by finding entries that have already performed predictions in similar situations. Motivation, emotion, and their usage within the controller module are described in greater detail in the following paragraphs that are accompanied by Figure 4.

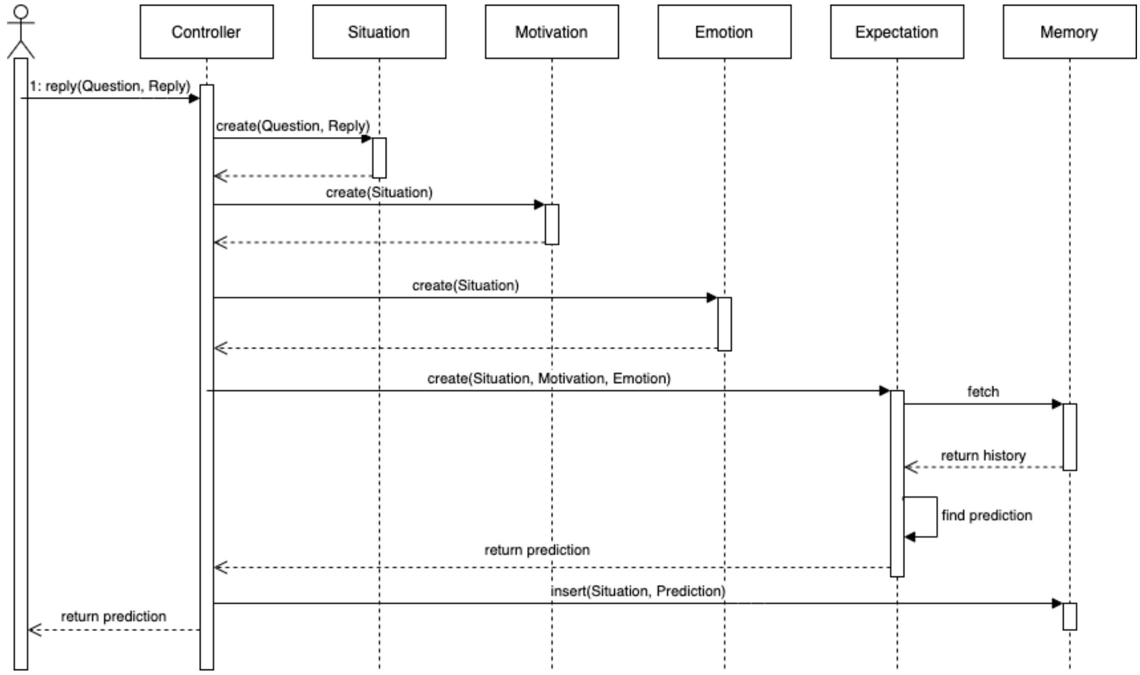

Figure 4. Execution flow of EmReflect agent

The first step of the execution of the EmReflect agent is creating the situation. Within the agent architecture, the situation consists of a question that the chatbot asked from the user and the reply that was given by the user. When a reply is sent to the controller, it will first create an instance of the situation, which will converge the information about the current situation – the question and the reply. Creating a situation instance is important for keeping in a coherent way data about the questions and replies. It also allows for easy storing of situation instances in the memory and passing on the same information when creating instances of motivation, emotion, and expectation.

Once a situation instance has been created, the agent is ready to start composing the corresponding motivation instance. Motivation represents the course of actions to be taken by the chatbot. For example, if a person is within a conversation and the conversation partner changes the topic, the person may lose the motivation to continue. The motivation value of the agent helps to determine the prediction. By understanding the motivation of the user based on the freeform text, the agent can decide how it should predict the next course of action based on the goal of the agent. If the goal is to reflect the user's behavior, as is the case with EmReflect, low motivation replies by the user would also lower the motivation of the agent. However, the predictions returned by the agent can also change the motivation of the user and this way, the agent can have a goal of maximizing the motivation of the user to talk to the chatbot [42]. However, motivation instance is implemented in EmReflect but currently not used. Due to the architecture of this kind of chatbot having multiple components that need tuning and validation, it was important to allow for step-by-step development and verification.

The next prerequisite necessary for generating an expectation is calculating the emotion parameter based on the existing situation instance. Understanding the emotion parameter from the user state represented by the chatbot helps the agent to generate a prediction with a high probability. As can be seen in Figure 4, expectation is initialized with the situation, motivation, and emotion, the latter being the most dominant factor for creating a prediction. The emotional state of the user informs the agent on how it should reply, considering the goal of the agent. If the user is feeling sad, an emotional helper agent could make a prediction that it should be as emotionally supportive as possible and try to increase the motivation for the user to write responses and talk about their problems. In case of EmReflect, the emotional parameter is used to reflect the emotional state of the user, as it has been perceived by the chatbot, meaning that if the user is using words that indicate happiness, the agent will also reply happily. The emotion instance should describe the emotional state of the user based on the current situation that has been used for the initialization of the instance. This means that historic data of similar situations is not yet used, and the emotional value is calculated solely based on the current text entered by the user. Emotion calculation is the most important part of the implementation of EmReflect as the goal of the application is to reflect the emotional state of the user based on their replies while talking to the chatbot. The emotional state of the user is not categorized as belonging to a set of predefined emotions. Instead of that, the chatbot categorized emotional states by the user based on the dimensional approach described in Section 2.4. According to this approach, emotions are categorized according to their three dimensions – valence, arousal, and dominance. Utilizing this model for categorization of emotions facilitates computational comparison of emotional states because the values for different dimensions can be normalized and combined with the historic data of the same kind stored in the agent's memory. These values are calculated based on the text input by the chatbot's user by means of a natural language processing algorithm, which is described in detail in section 3.3. Decomposing the user input and applying to it the natural language processing algorithm results in a set of emotion-related words for which mean Valence-Arousal-Dominance (VAD) values are calculated. The VAD values calculated about the situation represent the emotional state of the user necessary for the expectation calculation and are stored within the corresponding emotion instance. Once the calculation is complete and the values have been found, the controller continues with the execution flow and initializes the expectation instance, which will result in the final prediction.

The final step of processing a single input from the user is creating an instance of expectation. Expectation instance as shown in Figures 3 and 4 combines previously calculated parameters situation, motivation, and emotion and produces a prediction that will be sent back to the controller and ultimately to the user. Expectation instance combines the knowledge obtained from the situation, motivation and emotion parameters and also the knowledge stored in the memory for determining what kind of action the agent should perform according to the theory of constructed emotion. For EmReflect, the prediction is a text reply given by the agent to the user. Prediction generation heavily depends on the context of the agent and on the goal it is trying to achieve, together with other parameters used by the prediction algorithm. Due to the complexity of text generation as such, possible values for predicting the reply given to the user were predetermined and the best option of the set was selected as the chosen prediction. Possible replies to be predicted were chosen to be generic enough to be context-free so that they could be used as emotional replies in various contexts. VAD values for these replies were calculated with the same algorithm as EmReflect uses for user input. For generating a reply to the user, the EmReflect chatbot calculates the difference between the emotional VAD values of the user reply and possible predictions. Difference calculation is done by summing up the differences of valence, arousal, and dominance mean values, where finally the reply with the lowest difference will be used as the reply that is returned to the user by the agent. Reply generation could also be implemented through machine-learning algorithms that would use corpora of conversations as the learning data. Possible replies that share the context of user input can be generated from the ML model and the most fitting one can be selected either based on the differences between the VAD values like in the current architecture or by having VAD calculations as part of the input to the ML model. Once the prediction is selected, it is returned to the controller which sends it back to the user as a reply in the chat by the EmReflect chatbot.

As this section shows, the theory of constructed emotion can seem difficult to implement at first, but thanks to the VAD models for dimensional representation of emotional states, vast lexicons, and strong language processing capabilities in today's libraries it is possible to implement a chatbot based on this theory. Initial implementation of the EmReflect chatbot does not use the full capabilities of the architecture shown in Figure 4 but allows for verification and tuning that is necessary to ensure that the individual components of the implementation work as designed before increasing the complexity. Such an architectural approach is also modular, which allows working on each component separately and incremental improvement of the end-product.

3.2 Datasets

The execution of the architecture of the EmReflect chatbot described in section 3.1 requires data processing. Since a chatbot implementation of an emotionally aware agent only has a textual input by the user to work with, it requires rich corpora and previously documented data consisting of different words and phrases to interpret the textual input by the user. For dealing with motivation, the agent is required to know about words and phrases that would indicate higher and lower motivational levels expressed by the user. Depending on the usage context, there might also be the need to express the motivation by the agent if the agent is capable of being motivational. A similar need exists for emotion calculation – a rich dataset is required for the agent to understand the emotional meaning of the input by the user. EmReflect does not yet use motivation as part of expectation calculation but utilizes the emotional state of the user by finding the valence, arousal, and dominance values for the user input and replying with a prediction that has similar VAD values. Therefore, it was necessary to find a dataset that could be utilized by the EmReflect chatbot to interpret the emotional meaning of the freeform written text inserted by the human user.

Initially, the implementation of the EmReflect chatbot received inspiration from the BayesACT project [14][15]. This project used a similar approach to building emotional states but from the perspective of predicting the identities of the human participants. The BayesACT project has implemented algorithms that use VAD values for identities, behaviors, and actions, which are referred to as EPA values – evaluation, potency, and activeness. Datasets used by the BayesACT project [16] were initially chosen also for the EmReflect chatbot. However, soon it became clear that these datasets did not contain enough data for covering the VAD values for many popular emotion-related words and phrases – the dataset contained only about 1500 entries and originated in the Interact software dataset Indiana2002-04 [16][17]. Upon investigating different datasets provided by the Interact software [17], it became visible that most of the datasets were around the same size as Indiana2002-04.

Since the datasets from the Interact software [16][17] were not sufficient for the EmReflect chatbot, where freeform text analysis is the main use case for the VAD analysis, the author of this thesis started looking for and discovered other datasets of similar value. The most popular lexicon used so far in the research was based on Bradley and Lang ANEW norms lexicon from 1999 that contained about 1000 words [18]. This dataset, while being innovative, was still too small to be utilized for our chatbot agent.

Newer lexicons with a significantly higher numbers of included words and phrases have come out within the last 15 years with one of the biggest ones provided by Warriner, Kuperman, and Brysbaert, who extended the lexicon to nearly 14 000 English lemmas [19]. This dataset is also rich as for each included lemma it also contains information about gender, age, and educational differences. This allows for better accuracy of VAD values when analyzing a text in the given context because the agent is better aware of the user in terms of their gender, age, native language, etc., as VAD values can differ based on these parameters [19]. The dataset, which was introduced in 2013, was the largest dataset until 2018 when Mohammad introduced the NRC-VAD lexicon [12].

The NRC-VAD lexicon is currently the largest database providing VAD values for 20 007 English words. Moreover, it also includes translated values for over 100 languages [12][20], where the translations are provided by Google Translate [43]. While the NRC-VAD lexicon also provides translations to other languages, then EmReflect makes use of only English values. However, adding other languages to the EmReflect chatbot would require minimal changes, while preserving the current functionality. With the introduction of the NRC-VAD lexicon, the author also advanced research by comparing the new lexicon with existing ones that shows the superior reliability and accuracy of the lexicon compared to the previously largest Warriner et al. lexicon [12]. The size comparison of popular lexicons is presented in Table 1.

Table 1. Size comparison of different valence-arousal-dominance Lexicons

Lexicon	**Word Count**
Indiana2002-04	1500
ANEW	2477
Warriner et al.	13 915
NRC-VAD	20 007

3.3 Data processing – Natural Language Processing

Processing the user input is one of the most difficult tasks in the implementation of the EmReflect chatbot. Freeform text can introduce a lot of issues that need fine-tuning and problem-solving as they occur. People often write using figurative speech and slang words, and the language used by them can consist of phrases or terms that are cultural or demographical, which is hard to understand for any language processing software. Words can also have several meanings, where the most popular one is not always meant by the author of the text, which can be tricky to solve with language processing.

The first step of processing the input data represented in a natural language is to tokenize all input. Tokenization means simply dividing a text into smaller bits. Since our dataset contains VAD values per word, we tokenize all input into words. This is done in two stages - by first tokenizing the input into sentences and after that tokenizing every sentence into words. This approach is common in natural language processing. Although the current dataset does not include values for phrases, in the future we also plan to introduce VAD values for phrases to achieve even better accuracy. Currently, sentences are not split only by full stops but also by conjunctions.

The tokenization process also includes two important steps that help to better understand the actual meanings. The first step was to include the part-of-speech tags (POS tags) in each word of a sentence. This functionality has been built into the Python NLTK library [21] and it helps to correctly understand the meaning of a word. For example, the word "see" as a verb means "to perceive with the eyes", but as a noun, it can mean "a bishop's chair." Therefore, it is important to tag words correctly as it can lead to very different VAD values. The second step to better understand the emotional value of a sentence is considering negating words. This functionality, as well, has been added into the NLTK library by marking in a tokenized sentence every word with a special suffix that indicates a part of the sentence after considering a negation like "not", "no", "neither", etc. This step is necessary because otherwise a positive valence would be assigned to sentences that include a negating word in the middle. For example, consider the sentence "The movie was not good at all." This sentence would have received a positive valence score based on the words "movie" and "good", whereas the actual contextual interpretation of the sentence states something negative. Negating the words in the sentence resulted in an array of words where every word after the word "not" was marked by the negating suffix. The negating suffix is utilized later on when preparing words for finding VAD values in the dataset. The tokenization, POS tagging, and negation marking include everything necessary to prepare the data for actual processing and contextual interpretation. An example of the sentence tokenization with negating suffixes can be seen in Figure 5.

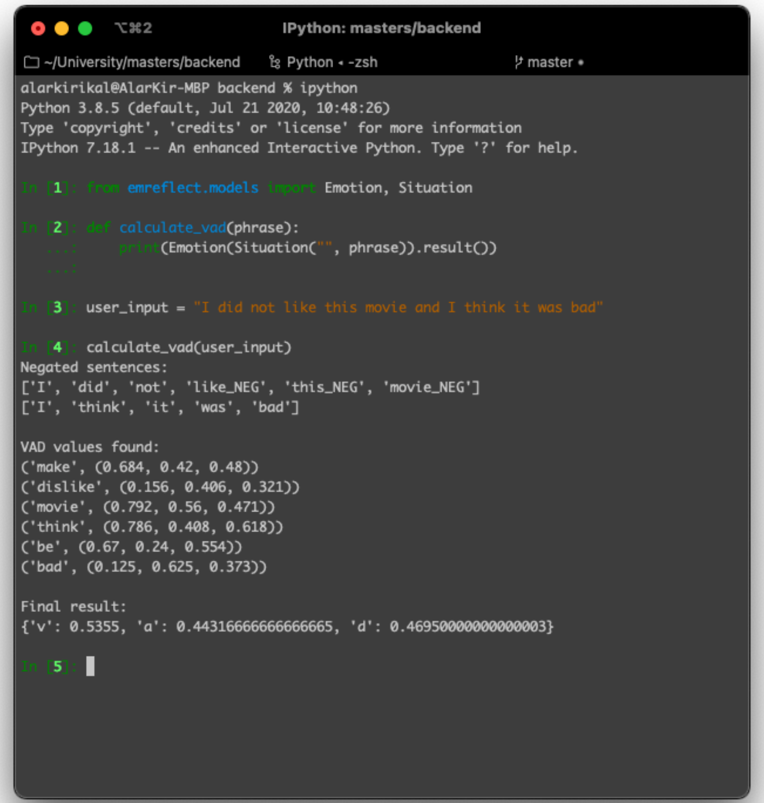

Figure 5. EmReflect sentence tokenization with negation.

Once the dataset has been tokenized, it is ready to be processed to find the most accurate VAD value per each emotion-related word in its correct meaning. For this, the agent will start looping through all the words that the preparation phase has prepared. Each word first needs to be lemmatized. A lemma is a dictionary form of a word. While it is possible to try and find a match for every word in the form in which it is included in the database, words usually have different spellings in different parts of the sentence. A lemma of a word helps to get the correct base form of the word. Due to not all emotion-related words having matches in the VAD dataset used, we also need to get synonyms for every word that is not found in the dataset. This way, even if the exact word used is not a part of the dataset, synonyms help to find matches in the dataset that have the same meaning with similar valence, arousal, and dominance values. Processing of synonyms is possible thanks to the WordNet lexical database, which is used in the EmReflect implementation. The WordNet database takes a lemma as the input and returns all contextual synonyms for the lemma. It is also important to use the POS tag in this search as the synonyms found should have the same part-of-speech tag associated with them. Otherwise, we could find synonyms that have a different tag that is possibly used differently in the sentence compared to what the user meant, resulting in different VAD values. The EmReflect agent goes through every synonym found to find VAD values in the lexicon until it finds a hit, which then breaks the loop, saves the value, and continues the search for the next word. The EmReflect chatbot only looks for synonyms to words that are tagged as nouns, verbs, adjectives, or adverbs as they provide the most meaningful contextual values for emotional states.

Words of a negated value, which have a special suffix as described previously, are treated differently for synonyms. Normally we go through the list of synonyms found by the

NLTK WordNet lexicon and try to find in the dataset a match for the synonym. Differently, for a negated word, we find the antonym for each synonym found for the negated word. Thus, if the sentence is "The movie was not good", the word "good" would get a negated suffix and antonyms would be used for finding the VAD values for the word "good", resulting in "bad" and "evil".

Once all the text that has been input by the user is processed and the VAD values found have been stored in the agent's memory, the data processing module for emotions returns the mean values of each parameter to the controller. This will allow the controller to continue with the process of finding the final expectation using the results.

3.4 Expectation Generation

The step of expectation generation of the EmReflect agent execution determines the prediction that will be returned, which represents the action taken by the chatbot to continue the interaction with the user. As Figure 4 reflects, expectation calculation by the EmReflect agent requires the agent to generate an instance of expectation based on the corresponding instances of situation, motivation, and emotion. For making a better prediction, the agent also uses the memory, which holds the conversation history.

The situation instance can help to improve the prediction from two different perspectives. Firstly, the situation can be used to find similar situations from the memory. Since the memory component holds information about previous situations and expectations, this can help to narrow down or choose the best new prediction to choose, as the instance includes knowledge on what predictions have been chosen in similar situations. This also represents one of the core ideas behind the theory of constructed emotion, which states that the human brain constantly makes predictions based on past experiences in life [1][10][35], which in this architecture are defined as situations. A second usage of the situation instance is generating context-aware replies, which can then be analyzed and modified to best match the goal of the agent. In case of reflecting the emotion of the user, it would try to find the reply with the most similar emotional value that would be returned as the prediction. Although this kind of approach does not improve the emotional accuracy of the final prediction, it helps to generate contextual replies, which improves the user experience. EmReflect does not yet generate text as this would increase the complexity of the development and validation of other features like emotion calculation. EmReflect uses previously defined answers that are analyzed for the emotional states represented by them and then compared to the current situation. The answer with the most similar emotional state is then predicted and used as the response by the chatbot to the user.

Motivation instance is used in the expectation based on what is the goal of the agent. For example, in case of the emotional helper applications that should comfort the user in emotionally difficult situations, it is obvious that the agent should try to increase the motivation of the user. People, who are experiencing stress or depression, often show up little motivation towards everything around them [44] and therefore a supportive emotionally intelligent agent should also have the goal to increase the motivation of the user during the chat. Secondly, if the goal is to mimic human behavior, a human can lose motivation to talk to the other party when the other party is going off-topic or introducing a topic of no interest. An emotionally aware chatbot, which has an emotional history with different topics and contexts, should be able to express different levels of motivation when talking about different topics based on its previous experience. If the agent is curious by design and wants to learn about new topics that it has not yet encountered, it should show up high levels of motivation to continue the conversation and learn more by talking to the user. EmReflect does not yet take the motivation into account as the main idea was to confirm the emotion reflection aspect of the agent, which is not directly affected by the motivation parameter.

Currently, emotion calculation is heavily used for finding the expectation within the proposed architecture. The present version of the EmReflect chatbot solely relies on the emotion parameter, in cooperation with the memory, to return a prediction. For the EmReflect chatbot, 90% of the VAD value is based on the last response by the user and 10% is based on the previous situations stored in the memory. We have made this design decision to make the chatbot's emotional transition from one reply to another smoother and to make the bot fluctuate less emotionally.

By introducing the analysis of historic situations and motivation, the calculation for finding the most probable expectation would be more accurate but could also introduce more points of failure to finding out the correct values for each parameter used. Considering this, for EmReflect, only the emotion module is currently used to generate predictions. Once this has been fine-tuned to an acceptable level from an emotional perspective, the other parameters can be implemented with ease as the application still fully follows the proposed architecture depicted in Figure 3 and goes through the whole execution cycle as is shown in Figure 4.

EmReflect uses a predefined set of responses as shown in Table 2, which are possible predictions, that will create an output for the expectation instance. These phrases were chosen to have emotional diversity. Once the values for the expectation are found, they are compared with the list of possible outputs as shown in Table 2. The item with the least total difference in all dimensions is used as the action taken and the corresponding reply is sent to the user.

Table 2. Predefined expectation EmReflect replies.

Reply	**Valence/Arousal/Dominance Scores**
Oh, that sounds bad!	**V**: 0.445, **A**: 0.624, **D**: 0.3855
That’s terrible..	**V**: 0.061, **A**: 0.624, **D**: 0.45
That’s awful..	**V**: 0.073, **A**: 0.786, **D**: 0.445
That sucks..	**V**: 0.344, **A**: 0.811, **D**: 0.377
Amazing!	**V**: 0.896, **A**: 0.706, **D**: 0.783
Got it.	**V**: 0.74, **A**: 0.843, **D**: 0.667
Cool!	**V**: 0.885, **A**: 0.594, **D**: 0.781
Nice!	**V**: 0.93, **A**: 0.54, **D**: 0.65
That sounds good.	**V**: 0.8515, **A**: 0.442, **D**: 0.466
That sounds nice.	**V**: 0.8475, **A**: 0.4955, **D**: 0.524
Good to hear.	**V**: 0.849, **A**: 0.5325, **D**: 0.489
That is awesome.	**V**: 0.788, **A**: 0.3985, **D**: 0.7015
That’s incredible.	**V**: 0.857, **A**: 0.5385, **D**: 0.902
Happy to hear that.	**V**: 0.88, **A**: 0.796, **D**: 0.608
Sorry to hear that.	**V**: 0.583, **A**: 0.3955, **D**: 0.328
Great!	**V**: 0.958, **A**: 0.362, **D**: 0.81
Good!	**V**: 0.938, **A**: 0.368, **D**: 0.534
Oh, okay	**V**: 0.823, **A**: 0.24, **D**: 0.647
That’s tough	**V**: 0.344, **A**: 0.588, **D**: 0.742
OK.	**V**: 5, **A**: 5, **D**: 5

4 Validation

Building a simulation that respects the theory of constructed emotion is an extremely difficult and time-consuming task. The theory of constructed emotion describes how the brain uses our memory to build an emotional state and takes input from various senses that feed our brain live information about our current situation. Building all this in a computer simulation in one go is most likely going to fail, as there are many possible routes for decision-making and several possible points of failure. Therefore, it was decided to narrow the validation experiment down to a chatbot application, which limits the source of inputs for the simulation to only textual input and has the expectation generated only based on the emotion and memory. While analyzing and understanding a written text is still a difficult task for any software application to accomplish, it can help us to understand if constructing the emotional value from a freeform text has any basis of being accurate for generating a prediction in the form of a reply.

While there are many possible solutions to deliver such an experiment, it was important from the beginning of the implementation to introduce a modular solution. This is why some architectural modules are not used as a part of this thesis but have still been implemented to demonstrate the possibility of further developing the application and improving it with other aspects and improving the user experience.

The following sections describe the validation experiment in more detail. Section 4.1 addresses the setup of the experiment with the technical details about the implementation. Section 4.2 describes some results found in the data generated from the input by users. Section 4.3 discusses the feedback from the experiment and provides explanations for the feedback items.

4.1 Setup

EmReflect was created as a chatbot application. Its core architecture was built as described in section 3.1 by implementing classes for the situation, motivation, emotion, expectation, and of course the controller and memory. These classes allow us to create instances of the given types and make specific calculations involving these instances, as is shown in Figure 4. The controller is a class that can be replaced based on how the application is used. The EmReflect controller works as an interface to the backend implementation but can be easily replaced with another controller using the same data classes that were described previously. In the next paragraphs we will give an overview of the main technical components and technologies used for the implementation.

The backend component of the setup was written in Python [45] using Flask [24], which is a Python WSGI web application framework. It is a lightweight solution that allowed setting up a backend service with the necessary API endpoints with approximately 160 lines of code for the EmReflect backend component. The Flask application exposes API endpoints for communication with the EmReflect through the EmReflect controller class instance, as is shown in Figure 6. The backend code is not aware of the EmReflect architecture and does not know about the data classes like situation and emotion. Its only interface with the EmReflect controller class and also accesses the database. Database connection to the memory was necessary for the backend component to allow for basic operations like registering new users and fetching the next question for the frontend. The database selected for the implementation was MongoDB [25], which is a document-oriented database that allows for easy storage of JSON-like objects [56]. With objects stored in the JSON format in the database, insertions and querying is simple and do not require mapping data between different structures. The frontend was written in React [26] as it is one of the most popular

```python
app = Flask(__name__)
app.config["MONGODB_SETTINGS"] = {
    "host": os.environ["MONGODB_HOST"],
    "username": os.environ["MONGODB_USERNAME"],
    "password": os.environ["MONGODB_PASSWORD"],
    "db": "apiserver_app",
}

CORS(app)
db = MongoEngine(app)

emreflect = ApiController()


@app.route("/api/user/add", methods=["POST"])
def user_add():⋯

@app.route("/api/entry/can_add", methods=["GET"])
def entry_can_add():⋯

@app.route("/api/emreflect/next_question", methods=["GET"])
def question_get():⋯

@app.route("/api/emreflect/add_result", methods=["POST"])
def result_add():⋯

if __name__ == "__main__":
    app.run(debug=True, port=5000)
```

Figure 6. EmReflect backend flask application API map.

solutions available for writing a reactive user interface (UI). The react frontend application communicates to the Flask backend component via the exposed API endpoints. All of these components were built as Docker containers [27] to separate them and make it possible to replace one component out with another one if necessary, as with this approach the components of the setup are agnostic to the technologies applied. Final list of docker services consists of "frontend", "backend" and "mongodb" images. The frontend image also hosts a configuration file, which sets up nginx – a web server that handles incoming HTTP requests [28] and forwards them to services listed in the configuration. This setup was uploaded to an instance of a DigitalOcean [29] virtual machine, where these docker services are executed and serve the application at the domain address http://www.emreflect.com/.

The home page of the application http://www.emreflect.com serves as a source of information to describe what the EmReflect application is about and how it works. A link to the registration page is provided after the informative introduction. For registration, the user must add the name and an e-mail address, which can later be used to contact the participant for feedback. It is also possible to stay anonymous by providing a nickname and an e-mail address that would not reveal the identity of the user. For the EmReflect chatbot, it is not necessary to know any identifiable information about the user. After completing the registration, the user is presented with their unique URL that they must continue using daily to talk to the EmReflect chatbot – their emotional companion.

Once accessing the unique URL provided for the user, the chatbot view is opened. Introductory comments are first shown to the user, which request the user to write as thorough replies as possible. After that, the first question is asked from the user. Questions to the user are the same for every day and they are designed to ask the participant about their day – starting from the morning and finishing with how the user spent their evening. EachEv input by the user is processed by the EmReflect chatbot via the API calls and an emotionally reflective response is predicted by the chatbot for the user after every reply to a question. Each participant could answer all the questions only once every day. The participants did not see any calculation values during the experiment and while they were aware of the idea of the validation experiment, the exact calculation results for their emotional state behind were purposefully left unknown to the user to stop people from trying to break the bot by trying out different emotionally strong wordings to see how it changes the values of the chatbot's emotional state values. A conversation example is in Figure 7.

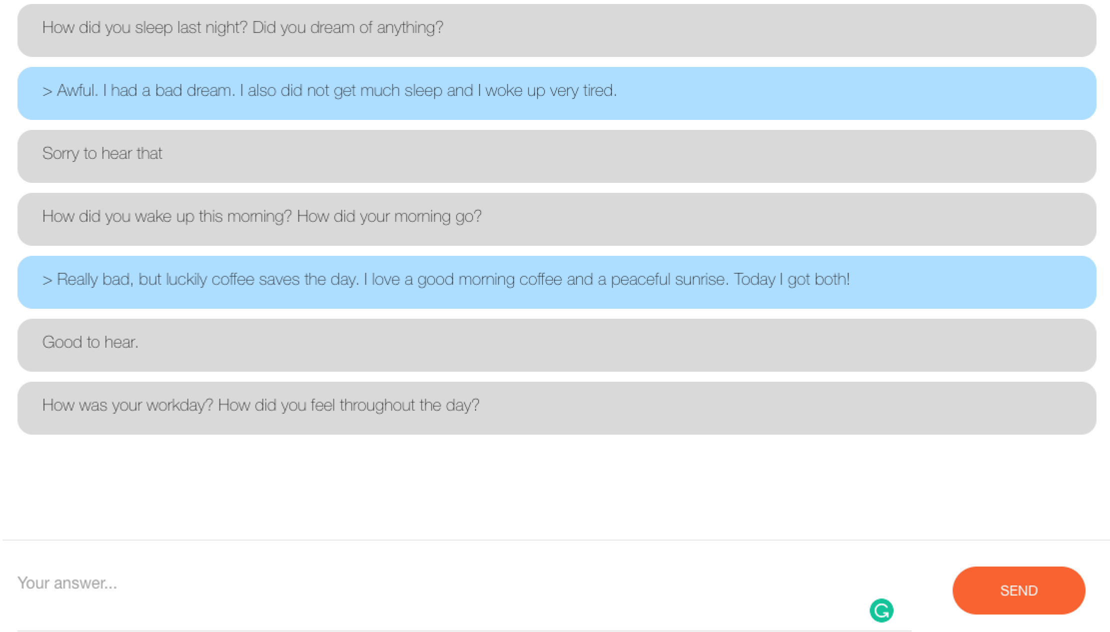

Figure 7. Example EmReflect conversation.

4.2 Results

The home page of the experiment with an introduction of the thesis was sent to a few hundred people in open chat rooms. Out of those, 17 visitors registered for the experiment, and they were instructed to access their unique URL every day before going to bed to go through a conversation with the EmReflect chatbot. All participants had at least a week to go through the questionnaire by the chatbot every day.

Out of 17 participants, 2 of them did not give any entries after registration and were excluded from the results. The remaining 15 users filled the questionnaire 3.5 times on average, whereas only 5 users had a conversation with the EmReflect chatbot for more than 5 times. Since the EmReflect chatbot did not include a reminder functionality as the e-mails collected from registration were only used for contact purposes, if necessary, several users often forgot about opening the conversation every day. Some of the participants also gave feedback during the experiment and expressed pity for missing out a day or two. A reminder functionality would most likely have increased the number of results collected during the experiment. Figure 8 reflects how many people entered some data but forgot to add any data later during the experiment.

One of the common feedback statements expressed by some of the participants during the experiment was that the same replies were generated by the chatbot repeatedly. This resulted in the idea that the emotion calculation from the input might be too generic,

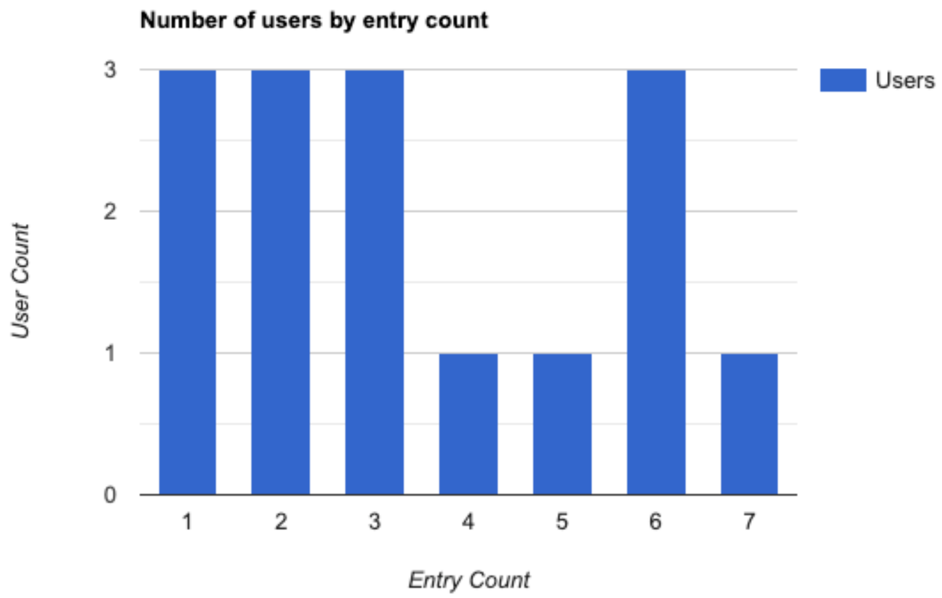

Figure 8. Number of users by data entry count, where entry was counted when all questions for the day were answered.

resulting in generic predicted replies by EmReflect. As was mentioned in Section 3, one of the main difficulties regarding the implementation of language processing is that emotional state is difficult to understand based on a written text and often requires fine-tuning to allow for better accuracy. If we look at all the words in a sentence, the mean VAD values of the sentence are often neutral as a sentence normally includes many keywords that have different emotional values associated with them. This can lead to a situation where specific keywords that should have, for example, given the sentence a strongly negative valence has as much of weight as any other words and therefore do not result in assigning a negative valence value to the input as a whole. Similarly, a sentence with one strong emotional keyword can easily become a neutral sentence and then result in a neutral response by EmReflect. By looking at the input data collected we can verify if that was the case.

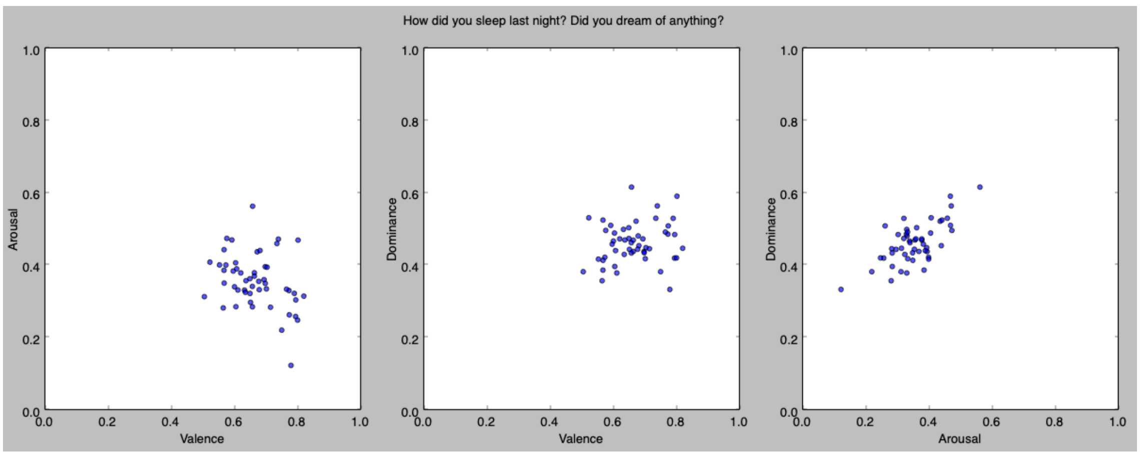

Figure 9. VAD values for first question.

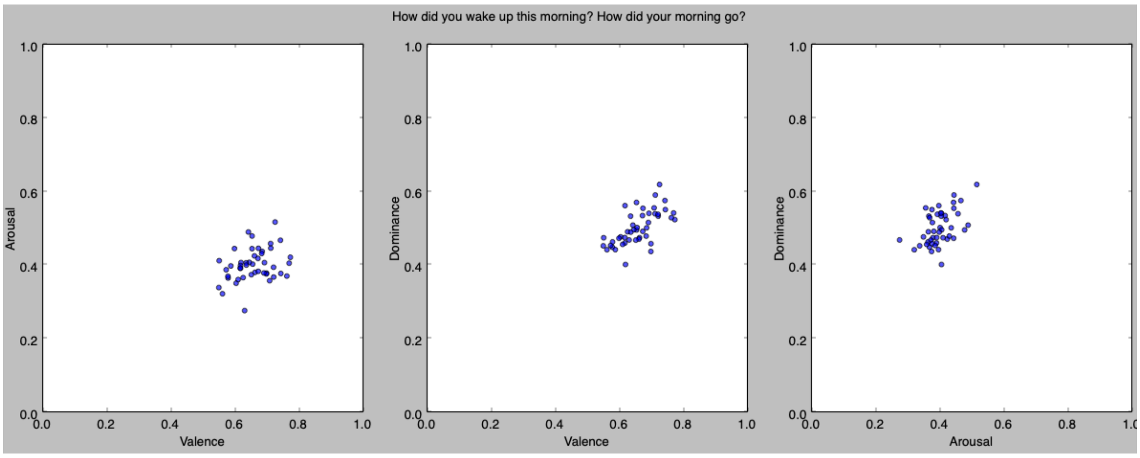

Figure 10. VAD values for second question.

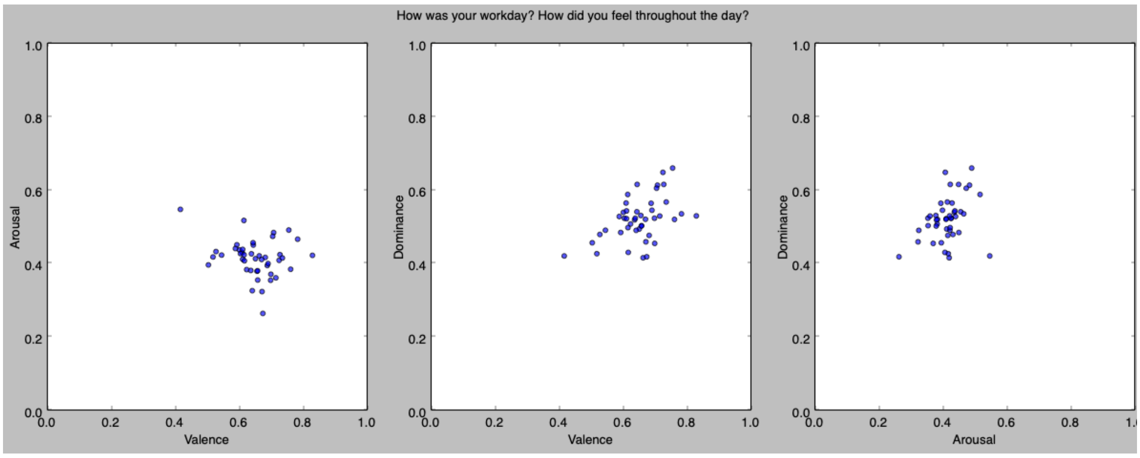

Figure 11. VAD values for third question.

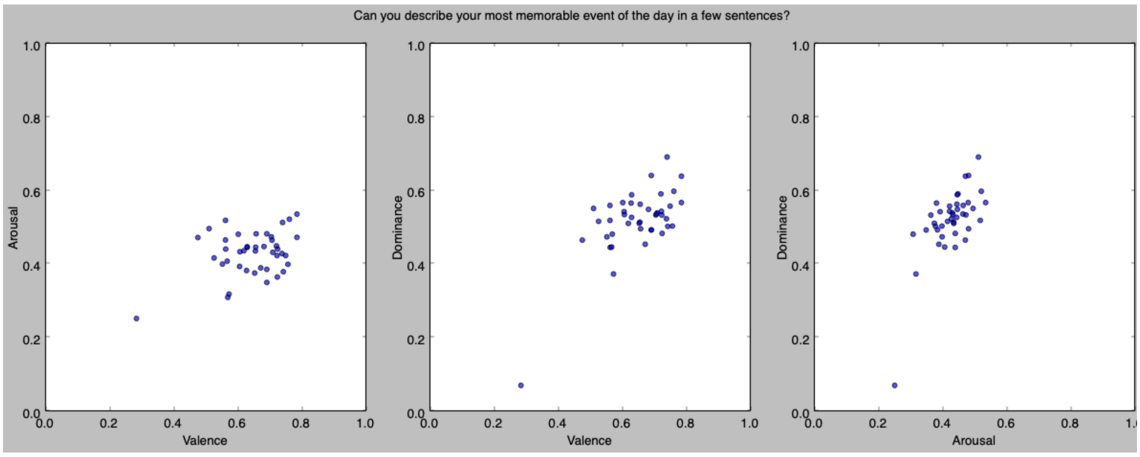

Figure 12. VAD values for fourth question.

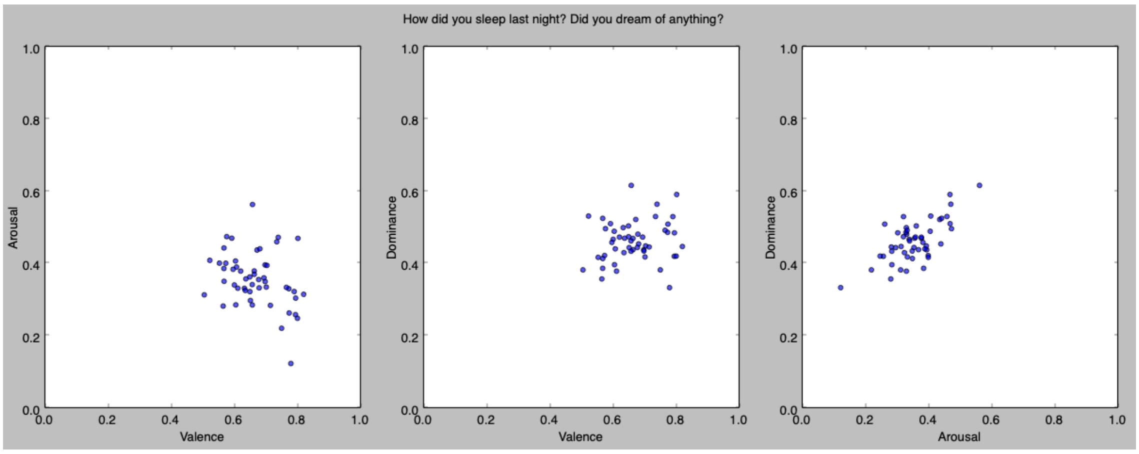

Figure 13. VAD values for fifth question.

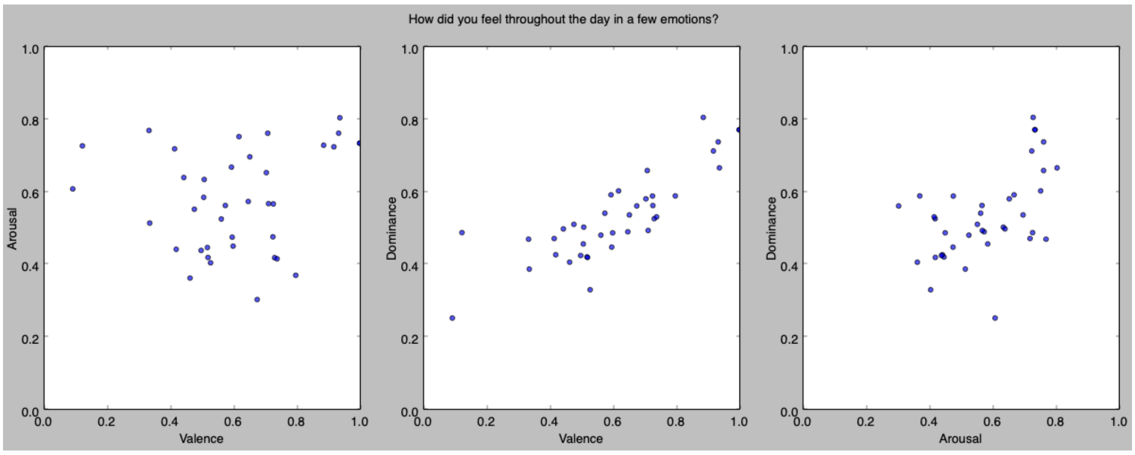

Figure 14. VAD values for sixth question.

The data represented in Figures 9, 10, 11, 12, 13, and 14 shows VAD values represented as two-dimensional data for valence, arousal, and dominance. This data shows if the chatbot identified different participants to have different emotional states based on asking from them the same questions. It is important to note here that Figures 9-14 represent the results for all collected entries by all participants. It can be seen from Figures 9-14 that most of the results are indeed in the same area. The valence is mostly around 0.6, arousal around 0.4, and dominance near 0.5. However, Figure 14 shows an interesting result related to the responses to the sixth question. The sixth question read as follows: "How did you feel throughout the day in a few emotions? (happy, sad, surprised, bad, fearful, angry, disgusted, …)" As we can see in Figure 14, the result for this question differs a lot between the participants, unlike the results to the other questions reflected by Figure 9-13. This most likely indicates that the emotion calculation for the first five questions resulted in a too generic emotional state and therefore the algorithm used for the emotional state calculation was not effective enough for accurate describing of the emotional states by the participants.

Another way to establish if the text processing resulted in too many neutral evaluations is looking at expectation calculations and specifically at what predictions were mostly returned to the participants? The data presented in Table 3 shows that the reply "Sorry to hear that" was by far the most popular reply given by the chatbot to the users, constituting 32.8% of all the replies provided by the chatbot during the experiment. It is pity to see that a few badly predefined responses made up the main reason why the other responses did not get predicted enough to create a positive diverse chatting experience for the users. Table 3 shows that the most popular replies by the EmReflect chatbot were not negative in their valence scores. However, the most popular reply by the chatbot was an emotionally negative statement "Sorry to hear that". This was the case because the precalculated VAD values for this statement were based on two trigger words – "sorry" and "hear". The mean VAD values for the given statement have the VAD scores shown in Tables 2 and 3. This, however, is not how we understand the given sentence in conversations but here it was considered by the algorithm as representing a negative emotional state. Therefore, emotional analysis from only calculating mean VAD values for all the words in a sentence is not enough to identify an accurate emotional state for the chatbot to generate an expectation and predict the next action to perform.

Table 3. Expectation frequency

Count	Expectation	Valence/Arousal
102	Sorry to hear that	V: 0.583, A: 0.3955
79	Good to hear.	V: 0.849, A: 0.3985
50	Got it.	V: 0.74, A: D: 0.667
24	That sounds nice.	V: 0.8475, A: 0.5325
12	Oh, okay	V: 0.823, A: 0.24
10	That sounds good.	V: 0.8515, A: 0.4955
10	Oh, that sounds bad!	V: 0.445, A: 0.624
5	That is awesome.	V: 0.788, A: 0.5385
5	OK.	V: 5, A: 5
3	That sucks..	V: 0.344, A: 0.706
3	Great!	V: 0.958, A: 0.665
3	Amazing!	V: 0.896, A: 0.843
2	That’s terrible..	V: 0.061, A: 0.786
1	Nice!	V: 0.93, A: 0.442
1	Cool!	V: 0.885, A: 0.54
1	That’s tough	V: 344, A: 0.588

4.3 Feedback

Feedback about the user experience of interacting with EmReflect was asked after the participants had had the opportunity to talk to the chatbot for a week. Some participants also gave feedback during the experiment, which resulted in possible different ideas for similar projects and constructive criticisms. After keeping the experiment open for a week, a feedback form was sent out to all participants asking about the software and how the interactions went. The next paragraphs will discuss the main feedback received from the users who participated in the validation experiment.

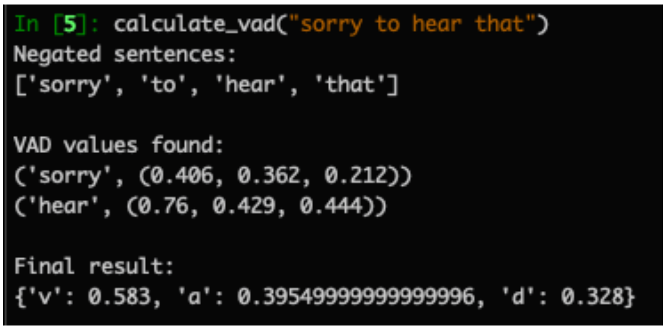

Figure 15. Emotion analysis for "Sorry to hear that"

First, the expectation "Sorry to hear that" was already mentioned in Section 4.2 for several times as being overly used by the EmReflect chatbot as a reply to something that was not actually of negative valence. Many participants sent examples of positive inputs, which resulted in the simulation to reply with this phrase. An example of this behavior can be seen in Figure 16. This was one of the earliest and most popular feedback items for EmReflect which was repeated on several occasions on the feedback form by the participants. It was also mentioned on the feedback forms that the opposite happened, where a negative input was given but something positive was returned as a response. This is most likely again the fault of some of the prepared replies having an inaccurate predefined VAD scores either in the user input or for the prediction response by the agent. An example can be made of the same sentence where "Sorry to hear that" results in a positive valence as is shown in Figure 15. In this example, the word "hear" has a positive valence score and the mean is still positive, whereas "sorry" should have had more weight in this phrase in order to be scored the way we perceive this in human-to-human communication.

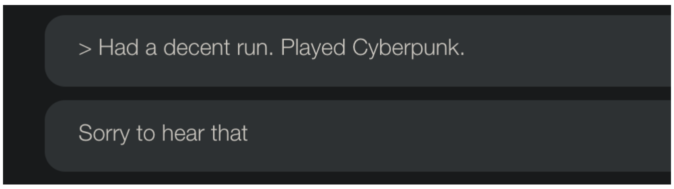

Figure 16. Example of a false-positive "Sorry to hear that" response.

During the implementation of the chatbot it became obvious that in order to achieve higher accuracy of predicting the emotional state by the chatbot, the input must contain adjectives and detailed enough descriptions, which would result in enough matches in the NRC-VAD lexicon [20]. At the start of every daily conversation, it was reminded to the participants that the replies should consist of multiple sentences, if possible, and consist of as many descriptions and adjectives as possible. However, it was quickly seen from the collected data that people are not used to writing long answers to a chatbot. This is particularly true in the case the chatbot agent itself responds with short answers. The average sentence count within the results was 3.9 sentences after splitting the input by conjunctions. Many people combined several statements in one sentence without providing any extra descriptions about these statements, making it very difficult for the emotional analyzer to extract from the given input any emotional value. This often resulted in neutral emotional scores of the input calculation and therefore answers by the EmReflect chatbot that were not in emotional correlation with the input by the participant.

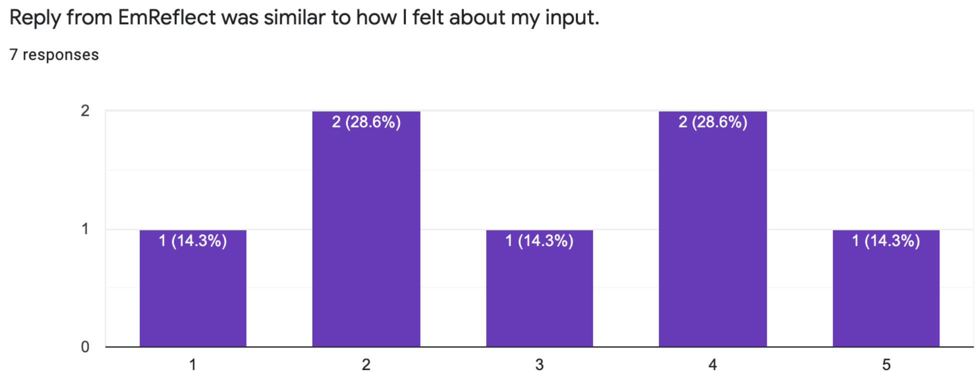

Figure 16. Feedback summary about reflection accuracy. 1 – Disagree; 5 – Agree.

Because of the big volume of feedback about emotional mismatch of the replies by the agent, a deeper analysis was conducted about how it would be possible to improve the accuracy of scoring emotions by means of the VAD model. It became apparent that by considering all words in a sentence as input for the VAD analysis, the results were often close to a generic mean. This meant that the predicted replies by the expectation instance of the agent considered the emotional state as close to neutral. An improvement opportunity would be to start putting different weights on differently tagged words. In other words, not all nouns and verbs should get the same weight in the VAD analysis. For example, adjectives or adverbs should have higher weights when calculating the dimensional values for an emotion instance. This shows how tricky language processing can be and how easy it is to do mistakes that heavily influence the accuracy of the emotion analysis. This also shows that analyzing the emotional state only based on a textual input can easily result in multiple interpretations of the same sentence. This results from the fact that human-to-human interaction involves a lot more senses that take part in the decision-making process of the brain than pure text analysis.

Feedback about the user experience of using EmReflect can be seen in Figure 16. It included questions about whether EmReflect was accurate to reflect their emotional state with its replies and if the chatbot helped them to reflect on their day. Application as such can prove to be useful for many people because most of the participants agreed that the EmReflect chatbot helped them reflect on their daily activities and emotional state. While this result does not directly connect to simulating the theory of constructed emotion, which was another goal of the research work reported in this thesis, it indicates that similar software applications would possibly be used by people to help them to reflect on their emotional states on their own. Some ideas discussed in Section 5 about the future work brings are also related to some of the feedback provided by the participants. We can conclude this section by stating that improving the emotional accuracy of generating responses by the chatbot and changing the style of obtaining input from users can further enhance the user experience and pave a road for a better user experience.

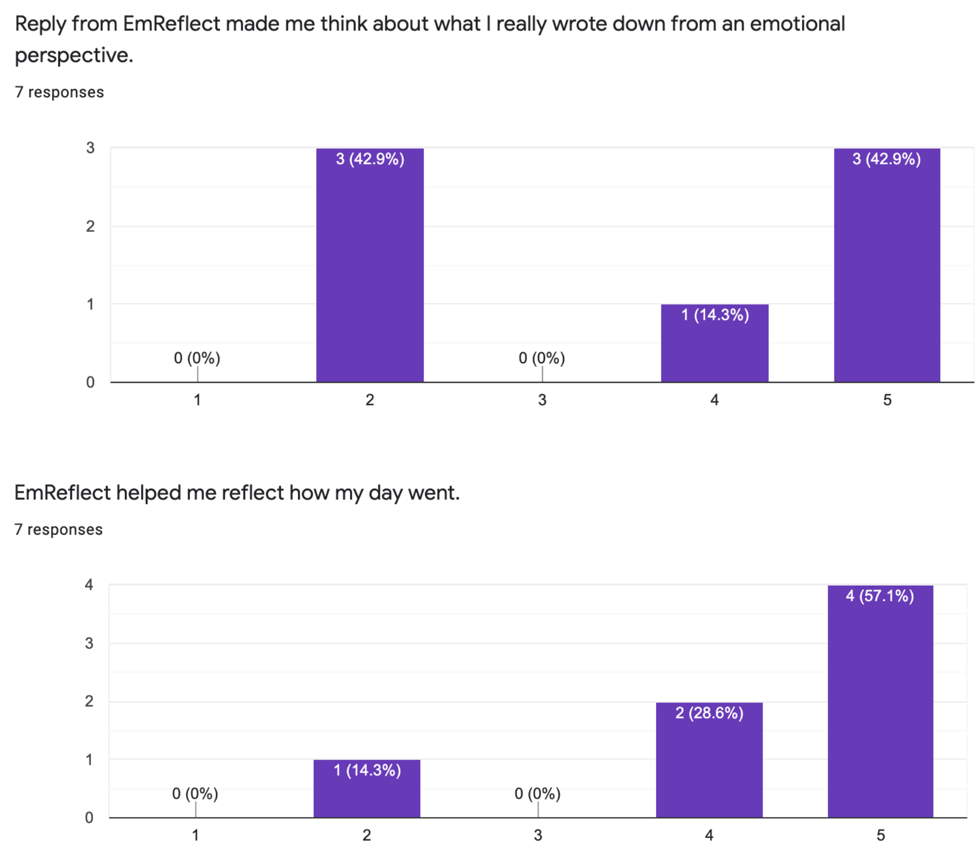

Figure 17. Feedback summary about reflection accuracy. 1 – Disagree; 5 – Agree.

5 Future Work

The research performed and the validation experiment conducted for this thesis provided valuable insight into how the theory of constructed emotion works and how it can be used to create better software for human-computer interaction. The feedback received in Section 4 with the implementation findings show how this topic can be further investigated and how the results can be used, which would not only help to discover new possibilities with continuing the implementation of EmReflect but also use parts of the work to create software that could improve our daily life.

5.1 Possible Improvements

A chatbot solution for analyzing the emotional state of the participants and responding with emotional awareness is a complex task to realize without making mistakes. Some possible improvement ideas appeared early in the validation experiment, and some also emerged during the experiment, when changing the course of the experiment would have ruined the results by changing the experience for the participants on the fly. The following paragraphs will discuss some improvement ideas for EmReflect that can refine its accuracy.

Firstly, as was mentioned in Section 3.2, motivation is currently not used as a part of the EmReflect implementation. The development and verification of using motivation for predictions are complex and need careful consideration for the design and implementation. Since EmReflect is only the first version of implementing an emotionally aware chatbot for reflecting emotions, the chatbot's accuracy of emotion prediction needs to be further tested and improved before other modules affecting the predictions can be implemented. Therefore, as soon as an acceptable level of emotional accuracy has been achieved, motivation should be considered when generating an emotion prediction as part of the expectation calculation. The motivation factor should be extracted from the user input first and based on that, the agent should decide which action to take. Since the execution logic of this agent is a part of the expectation calculations, the goal of the agent should be considered during this phase. By introducing a motivation factor to the chatbot agent, it can start understanding the motivation of the user and change its predictions based on how the agent would want to influence the motivational level of its user.

Using the data about previous situations for every expectation calculation would also be a strong next step forward to create a more accurate simulation. Moreover, that would much more precisely follow the theory of constructed emotion. The current solution calculates the emotional state based on the VAD values of words as included by the provided lexicon. This approach does not make the EmReflect chatbot to learn over time as it will not improve in generating predictions faster and based on previous situations, as the theory of constructed emotion states. If the EmReflect chatbot would use a long-term memory for generating similar responses to situations it has previously experienced and would combine it with the current emotional state calculation, it could provide a better simulation in terms of its self-learning capabilities. In this scenario, the agent would still process the input data every time, but it would improve its predictions by also relying on previous experiences to generate an expectation quicker.

Also, adding a supervised learning model with a learning dataset of random conversations could make the interaction with the EmReflect software a better experience as the replies would become context-aware if the learning dataset is big enough and the model is trained so that it would be capable of generating context-aware text from the input. Currently, the list of possible responses that EmReflect uses for predictions is fixed and while the list can be updated and improved with more possibilities, the answers will still stay generic because the expectation algorithm is not capable of pulling out the context of the user input and making a prediction based on that. Therefore, the current implementation can create a situation where the experience of a conversation would not feel natural. To create a simulation that would be close to a real conversation between two humans, it is necessary to introduce the context to the replies that the agent is returning. This does not mean, that the learning set should include conversations about all topics in the world, as the agent can be designed for a particular topic or area. In terms of EmReflect, this would mean that, for example, conversations between psychotherapists and clients could be used as example input data. Since users of EmReflect talk with it about their day, this can resemble a conversation with a therapist. Using this approach, the EmReflect chatbot can start generating better answers based on the goal of the agent.

5.2 Domains of use

Based on the experiment and ideas gathered during this thesis work, it is obvious that emotional facet of HCI applications can be used in many ways, be it in education, training, health, or entertainment domains. There is a possibility of developing applications that improve our daily lives individually or even for organizations [23]. The following paragraphs will discuss some of the ideas that were collected during the research and implementation of the EmReflect chatbot.

Stress and depression can be very difficult to notice [30] even for the people who are actively experiencing it. This means that analyzing the emotional state of these individuals can help to find signs of stress or depression before they could seriously harm the person. A similar emotional state analyzing module could be introduced to our personal assistant software, which could start spotting signs of stress before we know them. As time goes on, the popularity of the software like Siri [31] and Alexa [32] increases, and therefore it would be a seamless experience for the people to allow such analysis software to help them spot signs of stress. The resulting representation of the user's emotional state could then be visible through mobile or internet applications, which could also provide an overview and statistics about the fluctuations of one's emotional state. The same kind of software can also be used to elevating one's general mood and well-being, in addition to identifying signs of stress or depression.

Another possibility of finding signs of stress originates in the feedback to using the EmReflect chatbot. Namely, three people participating in the experiment stated that reflecting on their day every evening helped them to be more mindful about their feelings, and how they are doing in general. One participant expressed that this experiment provided them with the idea of starting to fill in a journal about their days to continue the self-reflection that this experiment started. Based on this feedback, an application could be built, using the same EmReflect module with a different user interface, that would act as a journal, where the users can write entries, which are then automatically analyzed in the background. This means that the application would constitute a private journal that would not be sent anyone else for ensuring privacy, but the emotional state based on the written text could still be analyzed by the software.

One of the problems that surfaced with the implementation of the chatbot agent was that although it was emphasized to the users in the validation experiment that adjectives should be used as much as possible with replies that extend to multiple sentences, the participants still replied rather shortly, as was expressed in Section 4.3. This is because communication by chat tends to be short and concise, which hurts the accuracy of reading the emotion of any given sentence. In the case of journal entries, this problem can be avoided organically, as the users are not chatting with anyone, but rather documenting their daily activities with adjectives and longer descriptions. Having this kind of natural way of using adjectives and descriptive expressions is a good method to get enough keywords necessary for extracting the emotional state of the author, without telling them how they should be writing replies to the agent. Therefore, a journal application could be designed and implemented that serves multiple purposes – a private journal for self-reflection and analysis that would also keep an eye on the emotional state of its user.

Apart from a personal gain from emotional state analysis, this kind of application can also be used to improve teamwork in professional teams and organizations. While many teams measure their success in terms of work accomplished or business profit gained, they receive little input of how the team is performing from the teamwork perspective. People are happier to work with coworkers with whom they feel comfortable. Unpleasant experiences at the workplace can decrease the quality of their work. Due to the complexity of measuring happiness on a large scale, it is also difficult to prove the necessity of improvement. Therefore, this generated an idea for an emotional state analysis within feedback forms about efforts of teambuilding. Teams working in sprints or on projects already conduct regular retrospective meetings and collect feedback using generic forms to improve future work. Often these forms are only read by the manager and if there are no people in the organization who could make improvement suggestions based on the results, these forms are rendered useless. A form with questions steered towards emotional and descriptive answers of work done with the help of the emotional analysis could take these forms to another level and help the employees to understand how the members of the team actually *feel* about the work they performed. Not only is it possible to achieve it by feedback forms, but it can actually be done with any input where employees use written means of communication. Applications like Slack [47] or Microsoft Teams [48] can be the main channels of communications for the whole organizations, which would also make it possible to analyze texts written in such applications to see what emotions are embedded in the texts written by them. It might naturally be difficult to hold people accountable for being neutral, but it could be helpful to see if people start communicating in an angry or hostile way.

6 Conclusions

The research on the theory of constructed emotion and implementation of the EmReflect chatbot has been continuously evolving. Emotion is a very complex concept to understand and coming to a single agreed understanding of how humans perceive, process, and construct emotions will require a lot of further research as it will currently continue to baffle psychologists, neuroscientists, and computer scientists. Until an agreed consensus is reached, researchers and software engineers will be engaged in devising brain-inspired computational architectures and will also try to make human-computer interaction as smooth and pleasant experience as possible, as computers have become an integral part of our daily lives. Smart speakers can have built-in microphones and can give you reminders to turn off the light before leaving home. Phones can control media and call your contacts without you touching the device. A car can sense your presence and unlock the vehicle before you reach to grab the handle, while also setting your seat to your preferred setting. These are all examples of using intelligent computer systems that have been created to offer their users with the best possible user experience. Improving the emotional and contextual intelligence of these computer systems that we interact with in our daily lives is extremely important as it will make the experience of using these products smooth and more pleasant.

EmReflect project has proven, that there is a need for research on the topic of emotion-aware computational architectures and simulations. While the brain has many sensory inputs to consider when analyzing the current situation and creates instances of our emotional state based on the current situation [1], the computer software that tries to mimic human behavior is still far from this. The theory of constructed emotion has a great potential for providing feasible theoretical foundations for designing and implementing simulation software that can construct emotional states from any input given, possibly including the emotional state of the software itself. Many aspects, like normalized motivation scaling and the VAD model included by the theory of constructed emotion are representable in computational form, which allows engineers to build computer simulations of this behavior and improve their accuracy over time to help researchers around the world to find out how our brain works. Moreover, while computation simulations based on the theory of constructed emotion can help to solve the question of how our brain processes emotions, they can also improve our daily life by solving important problems with which the humankind fights. By introducing accurate emotional mapping of the multimodal user input, it is possible to build applications that, for example, improve our mental health, even if the theory behind the implementation does not end up being correct. The findings of this thesis have shown that people around the world fight with mental problems that are either not diagnosed or not even noticed until it requires urgent attention and treatment that could have been prevented. An emotionally intellectual agent architecture can help to create solutions that help us in the form of a chatbot, a journal analysis software, or even a personal assistant application. If these solutions help us fight depression, stress and improve mental health, it can be counted as a success for this research.

7 References

[1] M. Gendron, L. F. Barrett, "Emotion Perception as Conceptual Synchrony," *Emotion Review*, vol. 10, Issue 2, 2018. doi:10.1177/1754073917705717.

[2] N. Rajcic, J. McCormack, "Mirror Ritual: An Affective Interface for Emotional Self-Reflection", *CHI 2020 Paper*, Paper 498, April 2020. doi:10.1145/3313831.3376625

[3] who.int, "Depression," https://www.who.int/news-room/fact-sheets/detail/depression.

[4] J. Posner, J. A. Russell, A. Gerber, D. Gorman, T. Colibazzi, S. Yu, Z. Wang, A. Kangarlu, H. Zhu, B. S. Peterson, "The neurophysiological bases of emotion: An fMRI study of the affective circumplex using emotion-denoting words," *Human Brain Mapping*, 2009, doi:10.1002/hbm.20553.

[5] S. Hamann, "Mapping discrete and dimensional emotions onto the brain: controversies and consensus," *Trends in Cognitive Sciences*, Volume 16, Issue 9, pp. 458-466, 2012. doi:10.1016/j.tics.2012.07.006.

[6] L. F. Barrett, B. Mesquita, K. N. Ochsner, J. J. Gross, "The Experience of Emotion," *Annual Review of Psychology*, Vol 58, pp. 373-403, 2007. doi:10.1146/annurev.psych.58.110405.085709.

[7] P. Ekman, "Are there basic emotions?", *Psychological Review*, 99(3), pp. 550-553. doi:10.1037/0033295X.99.3.550.

[8] M. Sarprasatham, "Emotion Recognition: A Survey," *International Journal of Advanced Computer Research*, 3(1), pp. 14-19, January 2015.

[9] K. Hoemann, Z. Khan, M. J. Feldman, C. Nielson, M. Devlin, J. Dy, L. F. Barrett, J. B. Wormwood, K. S. Quigley, "Context-aware experience sampling reveals the scale of variation in affective experience," *Scientific Reports*, Vol 10, Article nr. 12459, 2020. doi:10.1038/s41598-020-69180-y.

[10] L. F. Barrett, "How Emotions are Made: The Secret Life of the Brain," *Houghton Mifflin Harcourt*, 2017.

[11] J. Kwisthout, I. van Rooij, "Computational Resource Demands of a Predictive Bayesian Brain," *Computational Brain & Behaviour*, 3, pp. 174-188. doi:10.1007/s42113/019/00032/3

[12] S. M. Mohammad, "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, July 2018

[13] E. A. Kensinger, "Remembering Emotional Experiences: The Contribution of Valence and Arousal," *Reviews in the Neurosciences*, 15, pp. 241-251, 2002. doi:10.1515/REVNEURO.2004.15.4.241

[14] bayesact.ca, "Affect Control Processes (Bayesian Affect Control Theory)," http://bayesact.ca/

[15] J. Hoey, T. Schröder, A. Alhothali, "Bayesian Affect Control Theory," *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, doi:10.1109/ACII.2013.34

[16] affectcontroltheory.org, "Affect Control Theory, Research and Teaching Resources for ACT," http://affectcontroltheory.org/resources-for-researchers/data-sets-for-simulation/

[17] affectcontroltheory.org, "Affect Control Theory, INTERACT 2.1 (Beta)," https://affectcontroltheory.org/resources-for-researchers/tools-and-software/interact/

[18] M. M. Bradley, P. J. Lang, "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings," 1999

[19] A. B. Warriner, V. Kuperman, M. Brysbaert, "Norms of valence, arousal and dominance for 13,915 English lemmas," *Behaviour Research Methods*, 45, pp. 1191-1207, 2013. doi:10.3758/s13428-012-0314-x

[20] saifmohammad.com, "The NRC Valence, Arousal, and Dominance (NRC-VAD) Lexicon," https://saifmohammad.com/WebPages/nrc-vad.html

[21] nltk.org, "Natural Language Toolkit," https://www.nltk.org/

[22] nltk.org, "Categorizing and Tagging Words," http://www.nltk.org/book/ch05.html

[23] E. W. Pamungkas, "Emotionally-Aware Chatbots: A Survey," *Proceedings of ACM Conference (Conference'17)*, 2018, doi:10.1145/1122445.1122456

[24] flask.palletsprojects.com, "Flask web development," https://flask.palletsprojects.com/en/1.1.x/

[25] mongodb.com, "MongoDB, The application data platform," https://www.mongodb.com/1

[26] reactjs.org, "React, A Javascript library for building user interfaces," https://reactjs.org/

[27] docker.com, "Empowering App Development for Developers | Docker," https://www.docker.com/

[28] nginx.com, "NGINX | High Performance Load Balancer, Web Server, & Reverse Proxy," https://www.nginx.com/

[29] digitalocean.com, "DigitalOcean - The Developer Cloud," https://www.digitalocean.com/

[30] R. A. Bell, P. Franks, P. R. Duberstein, R. M. Epstein, M. D. Feldman, E. F y Garcia, R. L. Kravitz, "Suffering in Silence: Reasons for Not Disclosing Depression in Primary Care," *Annuals of Family Medicine*, Vol 9, No 5, 2011. doi:10.1370/afm.1277

[31] apple.com, "Siri - Apple," https://www.apple.com/siri/

[32] developer.amazon.com, "Amazon Alexa Voice AI | Alexa Developer Official Site,", https://developer.amazon.com/en-US/alexa

[33] J. Russell, A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, Volume 11, Issue 3, 1977. doi:10.1016/0092-6566(77)90037-X

[34] O. Balan, G. Moise, L. Petrescu, A. Moldoveanu, M. Leordeanu, F. Moldoveanu, "Emotion Classification Based on Biophysical Signals and Machine Learning Techniques," *Symmetry*, Volume 12, Issue 1, 2019. doi:10.3390/sym12010021

[35] L. F. Barrett, "The theory of constructed emotion: an active inference account of interoception and categorization," *Social Cognitive and Affective Neuroscience*, Volume 12, Issue 1, pp. 1-23, January 2017. doi:10.1093/scan/nsw154

[36] T. Nadarzynski, O. Miles, A. Cowie, D. Ridge, "Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study," *Digital Health*, Volume 5, 2019. doi:10.1177/2055207619871808

[37] M. Adam, M. Wessel, A. Benlian, "AI-based chatbots in customer service and their effects on user compliance," *Electronic Markets*, 2020. doi:10.1007/s12525-020-00414-7
[38] G. Battineni, N. Chintalapudi, F. Amenta, "AI Chatbot Design during an Epidemic like the Novel Coronavirus," *Healthcare*, 2020. doi:10.3390/healthcare8020154
[39] J. Hill, W. R. Ford, I. G. Farreras, "Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations," *Computers in Human Behaviour*, Volume 49, pp. 245-250, August 2015. doi:10.1016/j.chb.2015.02.026
[40] J. B. Best, "Cognitive Psychology," Fifth Edition, Wadsworth, 1999.
[41] L. Sterling, K. Taveter, "The Art of Agent-Oriented Modeling," *MIT Press*, 2009
[42] M. Shvo, J. Buhmann, M. Kapadia, "An Interdependent Model of Personality, Motivation, Emotion, and Mood for Intelligent Virtual Agents," *Iva'19*, 2019. doi:10.1145/3308532.3329474
[43] google.com, "Google Translate," https://translate.google.com/
[44] N. P. Jones, G. J. Siegle, D. Mandell, "Motivational and Emotional Influences on Cognitive Control in Depression: A Pupillometry Study," *Cogn Affect Behav Neurosci.*, pp. 263-275, 2016. doi:10.3758/s13415-014-0323-6
[45] python.org, "Welcome to python.org," https://www.python.org/
[46] Mirror technique. American Psychology Association, 2021. Retrieved from https://dictionary.apa.org/mirror-technique on 15.05.2021.
[47] slack.com, "Where work happens | Slack," https://slack.com/intl/en-ee/
[48] microsoft.com, "Video Conferencing, Meetings, Calling | Microsoft Teams," https://www.microsoft.com/en-ww/microsoft-teams/group-chat-software
[49] D. Keltner, S. J. Lerner, "Emotion.," *Handbook of social psychology*, pp. 317-352, John Wiley & Sons, Inc. doi:10.1002/9780470561119.socpsy001009
[50] K. R. Scherer, P. Ekman, "Approaches to Emotion," *Psychology Press*, Taylor & Francis Group, 1984
[51] F. Rehman, A. Munawar, A. Iftikhar, A. Qasim, J. Hassan, F. Samiullah, M. B. A. Gilani, N. Qasim, "Design and Development of AI-based Mirror Neurons Agent towards Emotion and Empathy," *International Journal of Advanced Computer Science and Applications*, Volume 11, Issue 3, 2020. doi:10.14569/IJACSA.2020.0110349
[52] M. Wyczesany, T. S. Ligeza, "Towards a constructionist approach to emotions: verification of the three-dimensional model of affect with EEG-independent component analysis," *Exp Brain Res*, 233(3), pp. 723-733, 2015. doi:10.1007/s00221-014-4149-9
[53] Y. Liu, O. Sourina, "EEG-based Dominance Level Recognition for Emotion-Enabled Interaction," *IEEE International Conference on Multimedia and Expo*, 2012, doi:10.1109/ICME.2012.20
[54] B. Reuderink, C. Mühl, M. Poel, "Valence, arousal and dominance in the EEG during game play," *International Journal of Autonomous and Adaptive Communications Systems,* 2012. doi:10.1504/IJAACS.2013.050691
[55] M. Mäntylä, B. Adams, G. Destefanis, D. Graziotin, M. Ortu, "Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?," *MSR '16: Proceedings of the 13th International Conference on Mining Software Repositories*, pp. 247-258, May 2016. doi:10.1145/2901739.2901752

[56] mongodb.com, "Why a Data Model Based on JSON-Like Documents," https://www.mongodb.com/collateral/why-a-data-model-based-on-jsonlike-documents

Appendix

I. Glossary

HCI – Human-Computer Interaction

ML – Machine Learning

VAD – valence-arousal-dominance

AI – Artificial Intelligence

TCE – Theory of Constructed Emotion

WSGI – Web Server Gateway Interface

API – Application Programming Interface

UI – User Interface

JSON – JavaScript Object Notation

URL - Uniform Resource Locator (web address)

HTTP – Hypertext Transfer Protocol

II. License

Non-exclusive license to reproduce thesis and make thesis public

I, Alar Kirikal____,

(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive license) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Computational simulation of how emotions are constructed in our brain according to the theory of constructed emotion _______________________________________,

(*title of thesis*)

supervised by Prof. Kuldar Taveter_______________________________________.

(*supervisor's name*)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons license CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive license does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Alar Kirikal

17/05/2021