Bachelor thesis

Artificial Intelligence

Radboud University

The effect of emoji's on user experience in educational chatbots

Author:
Serah Sommers
S1015986

First supervisor:
M. A. De Sá Siqueira:
PhD student - Behavioural Science Institute
marianna.desasiqueira@ru.nl

Second supervisor:
M. Vastenburg
Assistent Professor Design for Behavior change
martijn.vastenburg@ru.nl

January 26, 2022

Abstract

Chatbots have become more and more popular in the past few years, also in the field of education and e-learning. However, as this is such a recent development, not much research has been conducted about the improvement of the user experience in these educational chatbots. Optimizing the user experience of educational chatbots is important, as it might improve learning processes.
In this study, the influence of emoji use on the user experience was investigated. This was done by comparing the user experience of two chatbots: one using emojis versus one not using emojis. For this, an educational chatbot that teaches its user about cyber security was developed. The user experience was measured using the User Experience Questionnaire Model (UEQ Model).
Results suggest that emojis do not have a significant effect on the user experience. However, it was found that emojis did have a significant influence on the Novelty scale of the UEQ Model.
Future research could focus on using new features to improve user experience, as well as replicating this study with a different model, type of chatbot, or bigger (and demographically diverse) group of participants. It could also change the frequency of emoji use or the type of emoji. Besides that, future research could help to clear up the ambiguity about the effect of emojis on the Hedonic Quality.

Contents

Chapter 1

Introduction

Chatbots: you are an exception if you have never used one. A chatbot is a commonly known application, used by many companies to provide help to their customers, i.e. customer service. They have many advantages for both the company and the user. For companies, they reduce service, sales or support costs, which can lead to more profit [1]. Users also benefit from chatbots, as they help them to obtain timely and efficient assistance or information, social interaction, and amusement [2].
Aside from customer service, chatbots are also applied in many other fields, for example in healthcare, entertainment, or education [3][4][5]. The latter is a form of 'e-learning', which is becoming more popular and is changing the way people work and learn [6]. Chatbot technology is the most innovative solution in filling the gap between technology and education, and recently there has been an increase of chatbots for e-learning platforms to support student learning [7].
However, since chatbots are such a recent development, especially in the field of e-learning, much research should still be done on how they can be optimized to improve learning processes and learning outcomes [8].
One possibility for this learning optimization is to improve the user experience of the chatbot. The user experience is seen as a crucial aspect of a product, as it measures how well the product is experienced by the user. Developers should aim to design products with a high user experience and evaluate it well [9]. A higher user experience, caused by e.g. more emotional presence or social capability, leads to more user satisfaction and usability [10][11]. Previous research already found that educational chatbots that were socially capable, led to better learning results [12]. However, little in-depth research has been conducted on how to improve this user experience in chatbots [13], let alone in educational chatbots specifically.
So, if one wants to optimize the learning results in e-learning, one should desire to find new ways to improve the user experience. One way to do this is by using so-called emojis in the chatbot's dialog during communication.

This study will look at how the use of emojis might contribute to a higher user experience in an educational chatbot. By using emojis in the chatbot's dialog, the user will perceive more emotional expression [14] and the chatbot will be more socially attractive [15], which might improve the user experience. The following research question is proposed:

'How much do emojis influence the user experience of an educational chatbot?'

This study, which uses an educational chatbot that will teach the user about the domain of cyber security, will try to improve the user experience by using emojis in the chatbot's dialog. The user experience will be measured by the User Experience Questionnaire Model, also known as the UEQ Model [16][17][18]. It measures the user experience through 'Attractiveness'. This factor will indicate whether there is an effect of emoji usage in chatbots, which will consequently answer the research question of this study.

Chapter 2

Background

2.1 Chatbots

A chatbot, also known as a digital assistant or conversational agent, is a computer program that engages in a conversation with a human in natural language [19]. The conversation with the machine happens in a dialogic fashion and tries to achieve some particular result [20]. In this study, for example, the educational chatbot is ought to improve the users' knowledge about cyber security.

2.1.1 History of chatbots

The history of chatbots is only about 70 years old. It started with the Turing Test [21], created by Alan Turing in 1950, where the participant would have an interacting conversation with a computer. During the conversation, the participants did not know whether they were conversing with a computer or another human. The computer passed the Turing test if the participant afterward indicated they believed they had been talking to a human.
In 1960, one of the first chatbots called ELIZA was created [22], which only had limited knowledge and communication skills, but was an inspiration in the field of AI. The term 'Chatterbot' was first mentioned in 1991 [23], from which the term 'Chatbot' originated. The creation of ELIZA led to the creation of ALICE in 1995, which was the first online chatbot that could discuss any topic with knowledge of the Web [24].
In the years after that, up until now, chatbots are being improved and consequently can be used as a great tool for e.g. a company or some instance. Due to the invention of new tools and more open-source data, this is less challenging nowadays. However, as there is done little research about the affordances of chatbots [19], still a lot of research needs to be done in order to know in which directions these improvements should go, especially in specific domains, e.g. in education. Hence, the purpose of this study.

2.1.2 Educational chatbots

Along with the creation of chatbots, a specific type of chatbot emerged: the educational chatbot. An educational chatbot can be described as a helping aid to gain knowledge about some topics. With the growth of conversational technologies, the possibilities for integrating conversational agents in e-learning are receiving more attention [25].
Chatbots are now used in many different applications for e-learning: they can for example be used for tutoring [26], practicing conversations while learning a new language [27], as a learning companion [28], a reflection and meta-cognition skill tool [29] or a role-playing actor within a simulated experimental learning environment [30].
In this study, the domain in which the educational chatbot will teach the user something is cyber security. As cyber-attacks happen more frequently [31], cyber security becomes more and more important. Previous research showed that when users were informed and taught about cyber threats, they had more precautions behavior [32]. This shows the importance of educating users about cyber security, hence this study focuses on this specific domain. The chatbot will discuss the following topics of cyber security: passwords, public WiFi, phishing, and (illegal) downloading. These topics are chosen, as many people deal with these topics in daily life.
Applying chatbots in education seems to have a positive influence on the users learning ability; educational chatbots accelerate learning processes and enrich the learning experience by encouraging participation of the users [30]. This is partly caused by the engaging and human-like contact between the chatbot and the user. For this reason, it is important to do more research on this topic. For example, there should be investigated which features contribute to a better and more pleasant user experience. That is why in this study we will explore the effect of emojis.

2.2 Emojis

Before emojis were invented, emoticons were primarily used. The word emoticon is derived from the two English words 'emotion' and 'icon', as they are icons that try to convey emotion [33]. The typographic symbols are embedded in the text to add this extra emotion, using something other than just language.
A few emoticon examples are shown in Figure 2.1. Emoticons were first used in 1973 by the technicians that invented the typewriter [34]. However, those emoticons were limited, as the typewriters only contained a few characters.

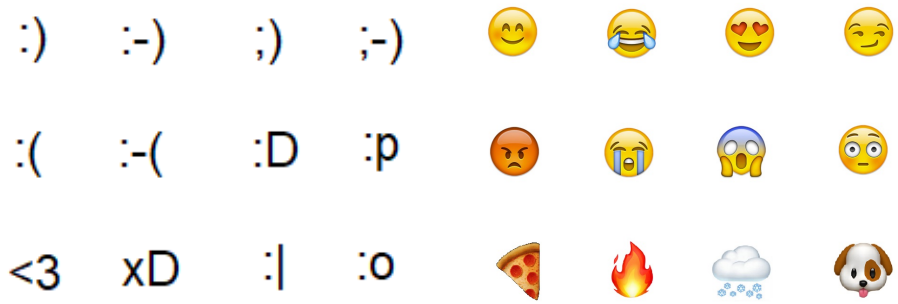

Figure 2.1: Examples of emoticons

Figure 2.2: Examples of emojis

To expand the idea of emoticons, emojis were invented. Emojis are small, colored pictographs that just like the emoticons convey emotions [35]. Next to that, they can also represent various other concepts, like objects, places, animals, or the weather. Some examples of emojis are illustrated in Figure 2.2.

The first development of emojis happened in the late 1990s by Shigetaka Kurita for Japanese mobile phone providers [36]. Nowadays, emojis are included in every iOS and Android mobile platform, because they are so appreciated. They are used more frequently than emoticons and are reported to have more positive motives and attitudes than emoticons [14]. Oxford Dictionary even made the "Tears of joy" emoji the 'Word of the Year' in 2015 [37], which recognizes the popularity of emojis nowadays.
Despite their popularity, however, emojis are still a poorly researched communication phenomenon as only a few studies have focused on it [38]. Given they are so well in demand, it might be convenient to add them to applications, e.g. chatbots, and observe the effect. Previous research found that readers interpreted text messages with (nonface) emojis as having greater levels of positive emotion, specifically as having more joy than messages without (nonface) emojis [39]. This is an example of the fact that the use of emojis in text-based communication has proven to add more emotional intensity and valence, and emojis can be seen as effective quasi-nonverbal cues [40]. Consequently, the user will perceive more emotional expression [14] and it makes the chatbot more social attractive [15]. Previous research already found that educational chatbots that were socially capable, led to a higher perceived user experience [41], as well as better learning results [12]. With this knowledge, this study will investigate if the use of emojis leads to an increased user experience in educational chatbots, intending to improve the learning results.

2.3 User experience

This study will measure the user experience of its chatbots. 'User experience', also abbreviated as 'UX', is a concept that is used a lot in the field of Human-Computer Interaction (HCI). The concept has multiple definitions [42], and the aim is still to develop a common understanding of the meaning and scope of user experience [43].
The definition ranges from traditional usability to beauty, hedonic, affective, or experiential aspects of technology use [44], so finding one definition is a complicated task. One research suggests that the concept of user experience is dynamic, subjective, and depends on context [45]. The measuring approach of the user experience should be evaluated carefully for an application such as a chatbot. In this study, the User Experience Questionnaire Model (UEQ Model) is applied.

2.4 User Experience Questionnaire Model

This study will measure the user experience of its chatbots with the User Experience Questionnaire Model, also known as the UEQ model [16][17]. The goal of the UEQ Model is to have a direct measure of the user experience. This model appears to be an easy to apply, reliable and valid measure for user experience [46].

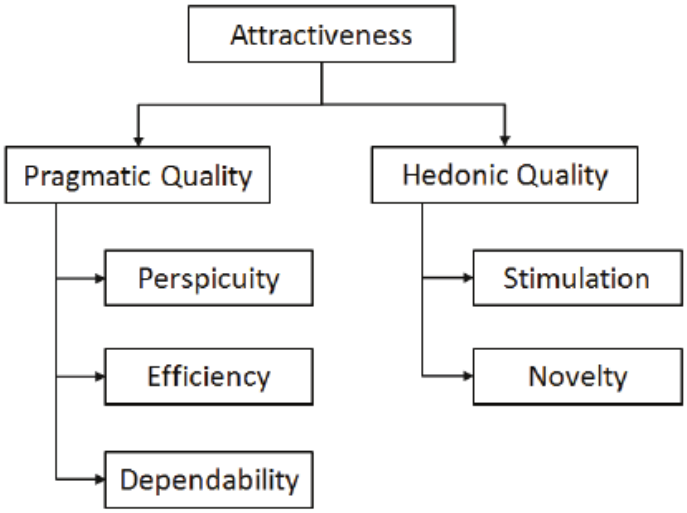

Figure 2.3: Assumed scale structure of the User Experience Questionnaire Model (UEQ Model)[16]

As is shown in Figure 2.3, the model contains multiple factors that work in a somewhat hierarchical way. The model tests the chatbot on multiple factors: Perspicuity, Efficiency, Dependability, Stimulation, Novelty, and Attractiveness. Perspicuity, Efficiency, and Dependability are factors to determine the Pragmatic Quality (goal-directed) of a chatbot. Stimulation and Novelty are factors to determine the Hedonic Quality (not goal-directed) of the chatbot. All factors of the Pragmatic Quality and the Hedonic Quality should influence the Attractiveness of the chatbot, which is a pure valence factor that indicates the user experience of the chatbot. It captures the general impression: the higher the Attractiveness, the better the user experience and vice versa.
The User Experience Questionnaire (Appendix A1) of the UEQ Model contains 26 item questions, which the participant should fill in. The questionnaire is currently available in over 30 different languages, including both Dutch and English. The list of 26 questions represents the six factors of the model: Perspicuity, Efficiency, Dependability, Stimulation, Novelty, and Attractiveness.
Appendix A.1 shows the list of questions that will be rated on a 7-point Likert scale [47]. The answers are scaled from -3 (fully agree with the negative term) to +3 (fully agree with the positive term), where the items are represented by two values with the opposite meaning. The items are given in randomized order, so half of the items start with the positive term, the other half with the negative term. Each factor of the model is measured by some of the item questions.

2.4.1 Attractiveness

Attractiveness is the overall impression of the product, i.e. the measurement of user experience.
The 6 items that correspond with Attractiveness are:

- annoying vs. enjoyable
- good vs. bad
- unlikable vs. pleasing
- unpleasant vs. pleasant
- attractive vs. unattractive
- friendly vs. unfriendly

2.4.2 Perspicuity

Perspicuity is how easy it is to get familiar with the product, i.e. whether it is clear, easy to learn, and understand. It is part of the Pragmatic Quality of the model.

The 4 items that correspond with Perspicuity are:

- not understandable vs. understandable
- easy to learn vs. difficult to learn
- complicated vs. easy
- clear vs. confusing

2.4.3 Efficiency

The Efficiency measures whether users can solve the task without unnecessary effort, whether the interaction is efficient and whether the chatbot reacts to the user input fast. It is part of the Pragmatic Quality of the model.
The 4 items that correspond with Efficiency are:

- fast vs. slow
- inefficient vs. efficient
- impractical vs. practical
- organized vs. cluttered

2.4.4 Dependability

Dependability is how much the user feels in control of the interaction with the chatbot. Is measures how much the user can predict the chatbots behavior and how safe the user feels when using it. It is part of the Pragmatic Quality of the model.
The 4 items that correspond with Dependability are:

- unpredictable vs. predictable
- obstructive vs. supportive
- secure vs. not secure
- meets expectations vs. does not meet expectations

2.4.5 Stimulation

Stimulation measures how exciting, fun, and motivating it is for the user to use the chatbot. It is part of the Hedonic Quality of the model.
The 4 corresponding items of the questionnaire are:

- valuable vs. inferior
- boring vs. exciting
- not interesting vs. interesting
- motivating vs. demotivating

2.4.6 Novelty

Novelty tries to indicate how innovative and creative the chatbot is and how much it captures the user's attention. It is part of the Hedonic Quality of the model.
It is measured by the following 4 items:

- creative vs. dull
- inventive vs. conventional
- usual vs. leading edge
- conservative vs. innovative

With the obtained results of the questionnaires, the convenient calculation tool that is included in the User Experience Questionnaire Model can be used for the data analysis. The tool calculates for each factor of the model a numeric value, that can also be visualized in a scale representation. When comparing e.g. two different products, the difference of these values can be calculated. This way, it is possible to conclude whether one product is significantly better than the other in one or more of the factors.

2.5 Research model and hypotheses

Applying the UEQ Model to the research question of this study, the following research model will arise:

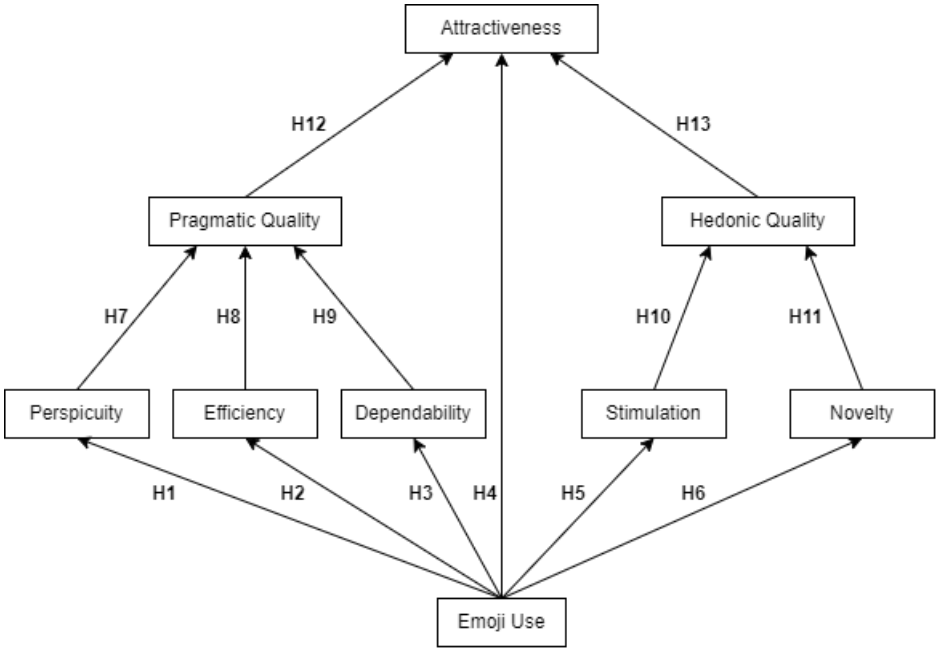

Figure 2.4: Research model

The corresponding hypotheses are:

H1: A chatbot that makes use of emojis has a higher perceived Perspicuity than a chatbot that does not make use of emojis.
H2: A chatbot that makes use of emojis has a higher perceived Efficiency than a chatbot that does not make use of emojis.
H3: A chatbot that makes use of emojis has a higher perceived Dependability than a chatbot that does not make use of emojis.
H4: A chatbot that makes use of emojis has a higher perceived Attractiveness than a chatbot that does not make use of emojis.
H5: A chatbot that makes use of emojis has a higher perceived Stimulation than a chatbot that does not make use of emojis.
H6: A chatbot that makes use of emojis has a higher perceived Novelty than a chatbot that does not make use of emojis.
H7: The higher the perceived Perspicuity in a chatbot, the higher the Pragmatic Quality in a chatbot.
H8: The higher the perceived Efficiency in a chatbot, the higher the Pragmatic Quality in a chatbot.
H9: The higher the perceived Dependability in a chatbot, the higher the Pragmatic Quality in a chatbot.
H10: The higher the perceived Stimulation in a chatbot, the higher the Hedonic Quality in a chatbot.
H11: The higher the perceived Novelty in a chatbot, the higher the Hedonic Quality in a chatbot.
H12: The higher the Pragmatic Quality in a chatbot, the higher the Attractiveness of the chatbot.
H13: The higher the Hedonic Quality in a chatbot, the higher the Attractiveness of the chatbot.

Chapter 3

Research Methodology

This study tries to give an answer to the following research question:
'How much do emojis influence the user experience of an educational chatbot?'.
To answer this question, an experiment should be conducted, where the user experience of two chatbots will be measured: one chatbot that uses emojis and a control chatbot that does not use emojis.
The control chatbot, which aims to teach its user more about cyber security, was developed by 5 co-students: Andy Huang, Annemiek van der Leest, Sanne Janssen, Serah Sommers, and Spence van Asperdt. Each student individually developed the experimental chatbot that would answer their personal research question. In this study, a chatbot that uses emojis was developed, and compared to the control chatbot on user experience. For this comparison, the User Experience Questionnaire Model [16] was used. This chapter will dive into how this study was carried out: which research model, hypotheses, materials, design, and procedure were used.

3.1 Materials

For this study, two chatbots were developed that could be accessed online and at home by participants. The participants filled in two corresponding questionnaires, which served as a way to obtain the required data for this study. In this section, the materials needed for obtaining this data will be discussed.

3.1.1 Chatbot

The first educational chatbot used in this study was developed by the research group together and teaches the user about the subject of cyber security. It has a somewhat fixed dialog, where the chatbot leads the conversation. It asks questions and gives feedback according to the user's response.

This way, the user will gain knowledge about the different topics of cyber security. The dialog contains 4 topics in the given order: Passwords, Phishing, Pubic Wi-Fi, and (Illegal) Downloading. After each topic, the chatbot also sends a website link with more information on the topic. This way, the user can dive into the subject even more if desired.
After the development of this chatbot, which was used in the control condition of the study, it was duplicated and adjusted to use emojis. This second chatbot was used in the research condition. The only difference between the two chatbots used in this study is that the control chatbot does not use emojis and the research chatbot does use emojis. The name and dialog of both chatbots are identical. Figure 3.1 and Figure 3.2 illustrate an example of the dialog in both chatbots.

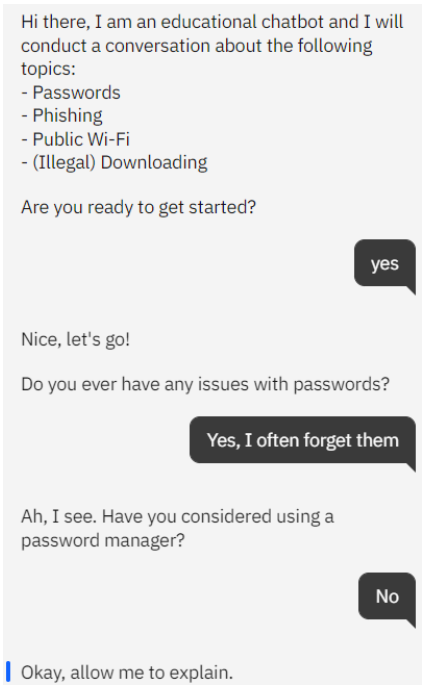

Figure 3.1: Example dialog control condition chatbot

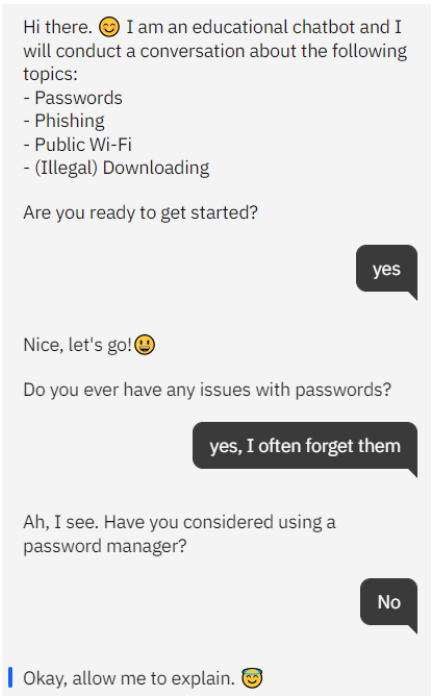

Figure 3.2: Example dialog research condition chatbot

3.1.1.1 IBM Watson

For the development of the chatbots, the IBM Watson [48] tools and applications were used. IBM Watson can be used to create multiple chatbot assistants. IBM Watson is a recognized, well-developed, easy-to-learn tool, and makes it easy to integrate chatbots on a website.
The research group started the process of creating a chatbot by working out a decision tree of the somewhat fixed dialog between the chatbot and user. The IBM Watson program tries to match the user's possible input to one of its so-called 'intents'. Each intent contains many, sometimes over 100, examples that match the intent's meaning. This way, the chatbot can recognize the user's intention and give an appropriate predetermined reaction to it.
Once the whole dialog structure and intents were constructed, the testing and improving phase was carried out. This was done with help of external testers, who gave more insight into different possible answers users can give to the chatbot. This is crucial since different formulations also had to be added as values to the intents. After many rounds of testing and improving the chatbot, the final control condition chatbot was finalized.
Each student of the research group copied and adjusted this chatbot to their individual research question to create the second chatbot for their research condition. In this study, the chatbot was adjusted to make use of emojis. The emojis were added on locations where one would normally naturally place them. As this study only tries to investigate one condition, namely the effect of emojis, nothing else was changed. After building and finalizing both chatbots, they had to be put on a certain platform. In this study, two websites were used.

3.1.2 Website

The two chatbots each were integrated in their own website. These websites were very simple, with only a small text field with instructions and contact information of the researcher. The websites were identical, to keep consistent conditions.
Both websites were developed with Blogger [49]. Blogger allows for an easy way to provide the participants with a platform where the chatbot could be integrated and accessed. On this platform, an account can be created, which allows the owner to design the website. On this website, blogs can be created that can include a chatbot function. After creating both websites and blogs, shown in Figure 3.3, they were uploaded so they were online and accessible to the participants.

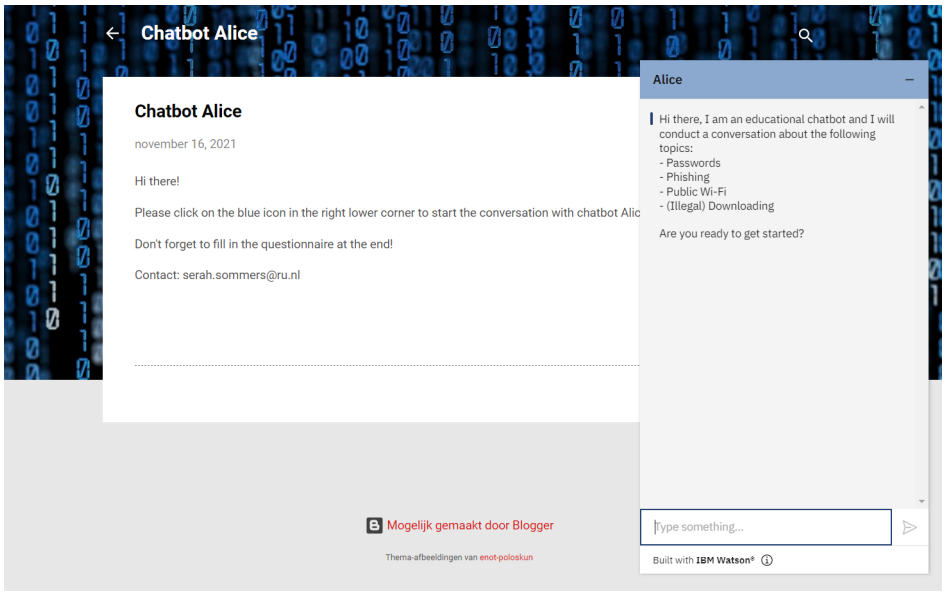

Figure 3.3: Website with integrated chatbot

3.1.3 Questionnaire

After the participants conversed with the chatbot and the end of the dialog was reached, they received a link to a questionnaire which they were asked to fill in. The questionnaires used for this study were developed with Qualtrics [50]. Qualtrics is a survey software tool, provided for free by Radboud University. The participants had to fill in two questionnaires: one after each chatbot conversation. The participant could choose between a Dutch or English version of the questionnaire. Normally it is preferred to keep all independent variables the same in both conditions and only change the dependent variable (i.e. emojis) [51]. However, it was important participants had the opportunity to fill in the questionnaire in their native or foremost language, which differed among the participants. This way, participants understood and interpreted the questionnaire better, which decreased the misinterpretation and misunderstanding of the language [52].
The questionnaire (Appendix A.2) started with asking for the Keyword which they obtained at the end of the conversation with the chatbot. The next question asked them to fill in an Identification Code, which they should keep the same in both chatbot questionnaires. The third question page asked the participant to (re)read the Informed Consent Form and give consent to participating in the study. After that, the questions that tested the user experience were posed. The 26 Likert Scale questions from the User Experience Questionnaire Model were divided between multiple pages, to make it more clear. Then, the last part of the questionnaire followed: the general questions about the participant itself. Only one of the two questionnaires contained these questions since it otherwise would be unnecessary work.

The general questions asked the user to give their age, gender, nationality, highest finished level of education, work/study field, and whether they ever had used a chatbot before or not.
During the whole questionnaire, participants were allowed to go back into the previous pages to adjust their answers to the questions. They also could take as long as they needed to finish the questionnaire.

3.2 Study Design

This study implemented the the within-subject design. In practice, this means that every participant in this study tested both chatbots and filled in a questionnaire for each chatbot. According to Charness et al. [53], the within-subject design has the advantage that it is less prone to noise than the between-subject design, as the environment in which the participant takes part in the experiment is consistent. They state that another advantage of the within-subject design is that you get double the amount of information compared to the between-subject design, with the same sample size.

One of the downsides of the within-subject design is that there might occur an order effect, according to Charness et al. This is an effect that originates when the same action is carried out sequentially, which influences the way a participant experiences the application. To limit the impact of the order effect, this research design made use of the across-subject counterbalancing method [54]. One group started with the chatbot that did not use emojis and secondly conversed with the chatbot that did use emojis. The second group did this vice versa. Participants were assigned a group in a randomized order.

With the use of the User Experience Questionnaire Model data analysis tool, the collected data was analyzed on differences between the two chatbots. This way, conclusions could be drawn about whether there is a significant difference in user experience between the chatbots. Any significant difference could be explained by the use of emojis since this is the only difference between both conditions.
To reduce biases, the study was constructed to be as consistent as possible in both conditions. First of all, both chatbots were given the same name, namely 'Alice'. Previous research found evidence that using female names for chatbots is preferred over using male names [55]. Keeping the same name in both conditions helped prevent any gender bias. Secondly, the websites in which both chatbots were integrated had identical appearances, to make sure the website would not influence the measurement of user experience when the user i.e. preferred one website of the other. Third, the responses of the chatbots were the same, with exception of the emojis. Again, this prevented bias, in this case, linguistic bias. Lastly, both chatbot conversa-

tions and questionnaires were supposed to be carried out on the same day, so right after each other and in the same environment. This reduced the effect of mood, environmental or circumstantial biases.
The sample size of participants required for this study was N$\geq$ 40 participants between the age of 18 to 35, as it was focused on young adults. The reason behind this is that this age group contains a lot of students that would benefit from better learning tools. Besides that, students also make use of emojis often.

3.3 Participants

Data was collected from in total 43 participants, who were found mainly through social circles. Because this study used a within research design, the participants were divided between two groups. The first group consisted of 22 participants, who first tested the control condition chatbot and afterward the research condition chatbot. The second group consisted of 21 participants, who first tested the research condition chatbot and then the control condition chatbot.
In this study, all participants had Dutch nationality. The participants were between 20 and 32 years old, with an average age of 23,3. Of the total 43 participants, 20 identified as 'male' (46,5%) and 23 identified as 'female' (53,5%). The distribution for their highest finished education level was as follows: 39,5% Secondary School, 23,3% HBO Bachelor, 25,6% WO Bachelor, and 11,6% WO Master. Almost all participants (97,7%), except for one, had used a chatbot before. The participants were from various work/study fields.

3.4 Procedure

Participants taking part in this study received an email with the information needed: an Informed Consent Form (Appendix A.2) and the Instruction Page (Appendix A.3). The Informed Consent Form informed the participants about their rights and the data collection, which they could read before the experiment started. The Instruction Page extensively explained the steps of how the experiment had to be done. Since this experiment was done entirely at home and independently, these instructions had to be very explicit. After reading both documents, the participants could start executing the steps explained in the Instruction Page. If anything went wrong during the experiment or if they had any questions, they could contact the researcher via the email address given.
Figure 3.1 and Figure 3.2 show an example dialog of both chatbots. Note that the participants were divided into two groups: one group tested the chatbot without emojis first and then the chatbot with emojis, and the

other group did it vice versa. For this, two different Instruction Pages were used. The first step in the study was to open the website where the first chatbot could be found. On this website, there was a message saying they should click on a button to start the chatbot dialog. Once the dialog came to an end and the conversation was over, the participant received a link to the questionnaire and a Keyword. Participants were given the choice between a Dutch and an English questionnaire. See Appendix A.4 for the English version.
This questionnaire started by asking the participant to fill in the Keyword. This Keyword serves as a way to check whether the conversation had reached a successful end. As both chatbots give a different Keyword, it also serves as a way to distinguish the two chatbots in the collected data. The chatbot without the emojis had the Keyword 'Animals', the chatbot with the emojis had the Keyword 'Plants'.
After filling in the Keyword, the participant had to think of an Identification Code, in the format of 'ABC123'. This Identification Code should be filled in in the second questionnaire as well, so the participant was told the memorize it. The purpose of the Identification Code was to know which two questionnaires in the collected data did correspond to the same participant. This is needed for the within research data analysis.
The next page of the questionnaire asked the participant to consent for participating in the experiment. It also included a link to the Informed Consent Form, which they already received in the email, just in case the participant had not read it yet.
Next came the most essential questions of the questionnaire: the 26 questions of the User Experience Questionnaire Model (Appendix A.1). What followed were some general demographic questions about the participant: their age, gender, nationality, highest finished level of education, work/study field, and whether they ever had used a chatbot before or not.
After finishing the questionnaire, they were instructed to read the next steps of the Instruction Page. There they were told to go to the second website, chat with the chatbot and again fill in the questionnaire that was given to them. Note that only one of the two questionnaires did contain the general questions at the end, since filling those questions in twice would only mean unnecessary extra work for the participant. Once the second questionnaire was finished, the participant was done with the study and could close all windows and files. If they would like to know more about the purpose of the study or the result, they were told they could contact the researcher at the given email address.

Chapter 4

Results

This chapter will discuss the quantitative results of this study. Once the data was collected for the questionnaires, data analysis could be carried out. During the data analysis, the User Experience Questionnaire Model Handbook [56] was used. This handbook gave a very convenient and clear description of how to remove outliers, analyze data and interpret the results. The two Excel tools [57][58] included in the UEQ Model were also used to carry out the data analysis.

4.1 Data analysis

4.1.1 Preprocessing the data

During the data analysis in this study, the Excel tool of the UEQ Model was used for convenience. The Excel tool consisted of two parts: the 'Data Analysis Tool' [57] and the 'Compare Scales Tool' [58]. The 'Data Analysis Tool' does data analysis on one dataset where it transforms the data, checks it for consistency and outliers, and calculates all necessary statistics needed to interpret the results. The 'Compare Scales Tool' compares the data from both chatbots. This is done with the T-test.

Before using the Excel tool, the data of both the Dutch and English questionnaires had to be combined into one file. This resulted in two tables: one table with data of the control condition chatbot and one table with data of the research condition chatbot.

As mentioned before in Chapter 2.4, the questions from the UEQ Model used a 7-point Likert scale [39], where the item questions were represented by two values with the opposite meaning. The participant could choose a value from 1-6, which represented their feeling about the chatbot the best. For example: if the positive term 'Good' was on the left side of the Likert scale and the negative term 'Bad' on the right side, a participant who thought it was equally 'Good' as 'Bad' would fill in the fourth value. This would

then result in a value of 4. However, the item questions were put on a randomized side, so half of the items started with the positive term, the other half with the negative term. Because half of the items were switched, they had inconsistent positive and negative terms. To make the data usable for data analysis, these inconsistent values had to be converted. This was done by giving all positive terms a positive value (e.g. maximum +3) and all negative terms a negative value (e.g. minimum -3). In the case of the example above, the value would now be 0. The Excel tool had predefined tabs that could automatically convert these values, such that there were consistent for every item question. Applying this step to the obtained data, resulted in two new data tables that served as the initial datasets.

4.1.2 Outlier removal

Since the participants did the experiment at home, it was harder to check whether they filled in the questionnaire seriously. This, among other things, is reason enough to check the collected data on outliers.
To check for outliers the UEQ-Model Handbook [56] recommends using a heuristic that looks at the maximum distance between the Likert-scale points, to find possible outliers in the collected data. Each of the scales of the model (e.g. Perspicuity) has multiple items questions that are somewhat alike, so the participant's answer should not differ too much between these items. For every question item on a specific scale, the maximum difference between the lowest and highest value is measured. If this difference is 4 or more (e.g. -2 and 3), the answers given in this scale can be seen as inconsistent. If three or more scales (e.g. Attractiveness, Efficiency, and Novelty) had an inconsistency, the data from this participant could thus be seen as an outlier. In these cases, the data from this participant in both the control and research condition data should be removed.
The Excel tool of the UEQ Model has a predefined tab, which checks for this heuristic and calculates the number of inconsistencies. After using this tool with the data obtained in this study, 4 participants were considered as outliers. The control condition contained two outliers: one had inconsistencies in the scales Attractiveness, Efficiency, and Dependability. Another had inconsistencies in the scales Efficiency, Dependability, and Novelty. The research condition also contained two outliers: one had inconsistencies in the scales Attractiveness, Efficiency, and Dependability. Another had inconsistencies in the scales Efficiency, Stimulation and Novelty. Normally the data of the participants of these outliers would be removed from both the control condition as research condition dataset. However, as there are relatively many outliers, namely 4, compared to the sample size (N=43), this study kept the outliers in the dataset.

4.1.3 Quantitative results

After preprocessing the data and investigating if any outliers should be removed, the quantitative results were obtained. With the obtained datasets, which now contained values in the range [-3, 3] to represent the positive and negative term values, the Excel tool of the UEQ-Model automatically calculated the values of the average, standard deviation (STD), confidence, and the upper and lower confidence interval where the number of participants was N = 43. This was done on all 6 scales (i.e. Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty) in both chatbots. This resulted in the tables shown in Table 4.1 and Table 4.2 below.

	Control Condition				
	Mean	STD	Confidence	Confidence Interval	
Attractiveness	1,11	1,02	0,30	0,80	1,41
Perspicuity	1,78	0,89	0,27	1,51	2,05
Efficiency	1,48	0,83	0,25	1,23	1,73
Dependability	1,35	0,65	0,20	1,15	1,55
Stimulation	0,66	0,98	0,29	0,37	0,95
Novelty	0,17	1,14	-0,17	0,34	0,00

Table 4.1: Required T-test values Control Condition

	Research Condition				
	Mean	STD	Confidence	Confidence Interval	
Attractiveness	1,46	0,80	0,24	1,22	1,70
Perspicuity	1,77	0,74	0,22	1,55	1,99
Efficiency	1,62	0,84	0,25	1,37	1,87
Dependability	1,35	0,68	0,20	1,15	1,56
Stimulation	0,91	0,95	0,28	0,63	1,20
Novelty	0,69	1,12	0,34	0,35	1,02

Table 4.2: Required T-test values Research Condition

Plotting the average values of each scale next to each other results in the graph shown in figure 4.1.

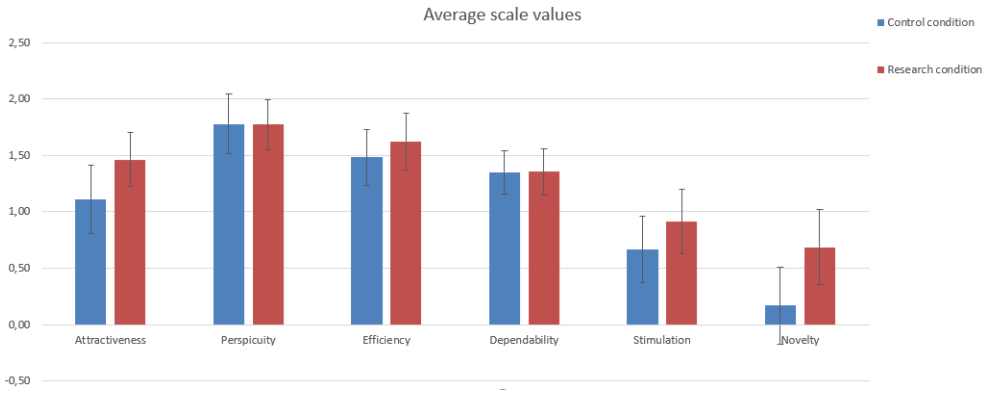

Figure 4.1: Histogram of average values per scale

This plot visualizes the average values, so there could be done an estimation on significance. It already stands out that there is quite a difference in the Novelty scale between the two conditions. To further investigate if there is indeed a significant difference in this scale or the other scales, this study carried out the T-test.

Before the T-test could be carried out, all assumptions had to be met. The T-test is possible when the data is randomly selected from the population. Since this research randomly selected its participants, this assumption is met. The data also had to have a continuous or ordinal value. As all values fell in the range of [-3, 3], this assumption was also met. Another assumption is that the data is normally distributed. A good way to check this assumption is to make a QQ-plot for every variable each condition [59]. In Figure 4.2, the QQ-plot of the Attractiveness in the control condition is shown as an example. A variable is considered normally distributed if the QQ-plot shows that the data points somewhat follow the linear line. The QQ-plots that were plotted, gave proof that all variables are normally distributed, although more data samples could have given more certainty about this. Due to the lack of many data samples, the last T-test assumption, namely that the data should consist of enough data samples, could not fully be met.

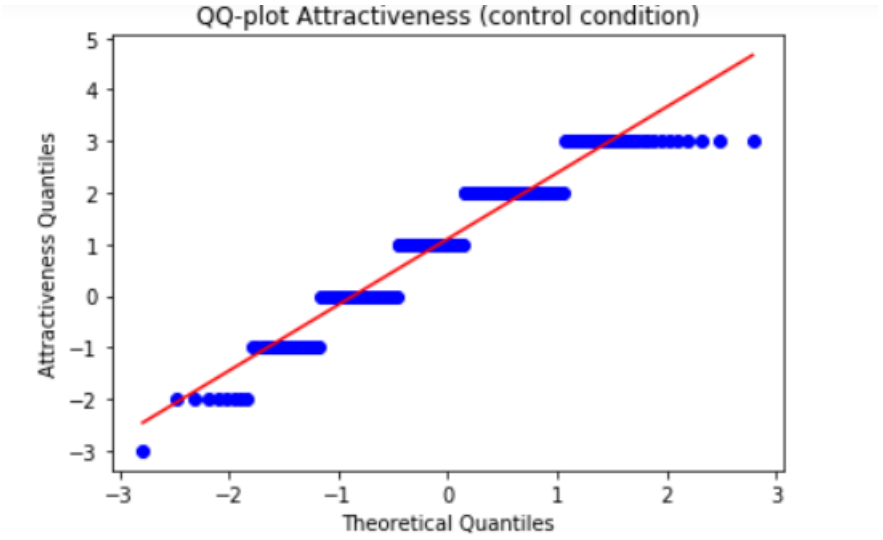

Figure 4.2: QQ-plot Attractiveness control condition

After checking all assumptions, and finding proof that most of them are met, the T-test was carried out. The T-test was done with a degree of freedom of 84 (N1 + N2 – 2) and a significance level of $\alpha = .05$. This significance level is a commonly used value in social sciences [60].

For each of the six scales, the T-test was carried out to find if the use of emojis has any significant influence on the scale values. The results for the scale Attractiveness from the control condition ($M = 1.11, SD = 1.02$) and research condition ($M = 1.46, SD = 0.80$) indicate that emojis do not significantly increase the Attractiveness of an educational chatbot, $t(84) = -1.431, p = .078$.
Similarly, results for the scale Perspicuity from the control condition ($M = 1.78, SD = 0.89$) and research condition ($M = 1.77, SD = 0.74$) indicate that emojis do not significantly increase the Perspicuity of an educational chatbot, $t(84) = 1.968, p = .9738$.
The results for the scale Efficiency from the control condition ($M = 1.48, SD = 0.83$) and research condition ($M = 1.62, SD = 0.84$) indicate that emojis do not significantly increase the Efficiency of an educational chatbot, $t(84) = -1.726, p = .04402$.
The results for the scale Dependability from the control condition ($M = 1.35, SD = 0.65$) and research condition ($M = 1.35, SD = 0.68$) indicate that emojis do not significantly increase the Dependability of an educational chatbot, $t(84) = 1.875, p = .9679$.
The results for the scale Stimulation from the control condition ($M = 0.66, SD = 0.98$) and research condition ($M = 0.91, SD = 0.95$) indicate that emojis do not significantly increase the Stimulation of an educational chatbot, $t(84) = -0.732, p = .233$.
Finally, the results for the scale Novelty from the control condition ($M =$

$0.17, SD = 1.14$) and research condition ($M = 0.69, SD = 1.12$) indicate that emojis do significantly increase the Novelty of an educational chatbot, $t(84) = -1.808, p = .0371$.

So, it can be concluded that the scales Attractiveness, Perspicuity, Efficiency, Dependability, and Stimulation don't have any significant outcome, since those p-values are bigger than the significance level of .05. This means that there is no proof that the use of emojis does influence these scales significantly. However, since $p = .0371 \leq .05$, there can be concluded that there is found evidence that the use of emojis does influence the scale Novelty significantly. With this information, the hypotheses H1, H2, H3, H4, and H5 are rejected and H6 can be accepted.

To calculate the Pragmatic Quality and Hedonic Quality, the Excel tool was used again. It contains a predefined tab that calculates the average value of each scale. Each question item corresponds to a specific scale. The average is taken by summing the value of the questions that corresponded to a specific scale and then dividing them by their count. For example, the scale Perspicuity was represented by questions 2, 4, 13, and 21 in every participant. If the average values of all participants of these individual questions were e.g. 1, -2, 3, and 0, the average scale value of Perspicuity would be 0.5.
With the previously calculated average values of each scale, the average value of the Pragmatic Quality and the Hedonic Quality was determined. Again these average values fall in the range [-3, 3]. For the Pragmatic Quality, the average of the average values of Perspicuity, Efficiency, and Dependability was calculated. Summing the average values of these three scales and dividing them by 3, did result in a Pragmatic Quality value of 1,54 in the control condition and a value of 1,58 in the research condition. For the Hedonic Quality, the average of the average scale values of the Stimulation and Novelty scales was calculated. This was done in a similar way and resulted in a Hedonic Quality value of 0.42 in the control condition and a value of 0.80 in the research condition. Note that these values are all greater than zero, which implies that the qualities of the chatbots in both conditions are more positive than negative.
To check whether or not there is a significant difference in Pragmatic Quality or Hedonic Quality between the two conditions, the T-test was carried out again. For this, the standard deviation (STD) of the Pragmatic Quality and Hedonic Quality in the two conditions was calculated. These values, together with the average values, were used for the T-test. The obtained values are shown in Table 4.3.

	Control Condition				
	Mean	STD	Confidence	Confidence Interval	
Pragmatic Quality	1,54	1,01	0,30	0,81	2,11
Hedonic Quality	0,42	0,97	0,29	0,12	1,39
	Research Condition				
	Mean	STD	Confidence	Confidence Interval	
Pragmatic Quality	1,58	0,80	0,24	1,34	2,38
Hedonic Quality	0,80	0,81	0,24	0,56	1,61

Table 4.3: Required T-test values Pragmatic and Hedonic Quality

To make conclusions about the significance, the p-values are compared to a significance level of $\alpha = .05$. The results of the Pragmatic Quality from the control condition ($M = 1.54, SD = 1.01$) and research condition ($M = 1.58, SD = 0.98$) indicate that emojis do not significantly increase the Pragmatic Quality of an educational chatbot, $t(84) = -1.393, p = .08361$. The results of the Hedonic Quality from the control condition ($M = 0.42, SD = 0.97$) and research condition ($M = 0.80, SD = 0.81$) also indicate that emojis do not significantly increase the Hedonic Quality of an educational chatbot, $t(84) = -1.645, p = .0519$. However, since $p = .0519$ is quite close to a significant effect (see Chapter 5: Discussion).

To test whether hypotheses H7, H8, H9, H10, H11, H12, and H13 should be rejected or can be accepted, the correlation between the scales and the Pragmatic Quality or Hedonic Quality should be mapped. This was done with the Pearson Correlation Coefficient [61].
Before the Pearson Correlation Coefficient could be used, the assumptions for this method should be checked. It assumes that the two variables contain continuous values, that have a linear relationship. There also should not be any significant outliers, as well as being approximately normally distributed. As all variables that are considered contain continuous values between the range of [-3, 3], the first assumption is met. The proof of linearity was found with help of Figure 4.3 and Figure 4.4, which indeed shows that there is a linear relationship between all considered pairs of variables. Figure 4.3 and Figure 4.4 could also be used to look for outliers: a few points in some of the scatter plots might be considered as outliers (see e.g. the scatterplot of the Pragmatic Quality and Dependability in Figure 4.4). However, as this study contains only a few data points, it was decided to ignore those few outliers. As for the last assumption, all variables should be checked for normality. Previously in this study, normality was already checked with the use of QQ-plots. This showed that all variables were somewhat normally distributed, so the last assumption is also met.

After checking the assumptions, the Pearson Correlation Coefficient method was carried out. For each data point, the correlation coefficient values of the first variable (e.g. Perspicuity) and the second variable (e.g. Pragmatic Quality) were calculated. The correlation results of both the research and control condition are shown in Figure 4.3 and Figure 4.4.

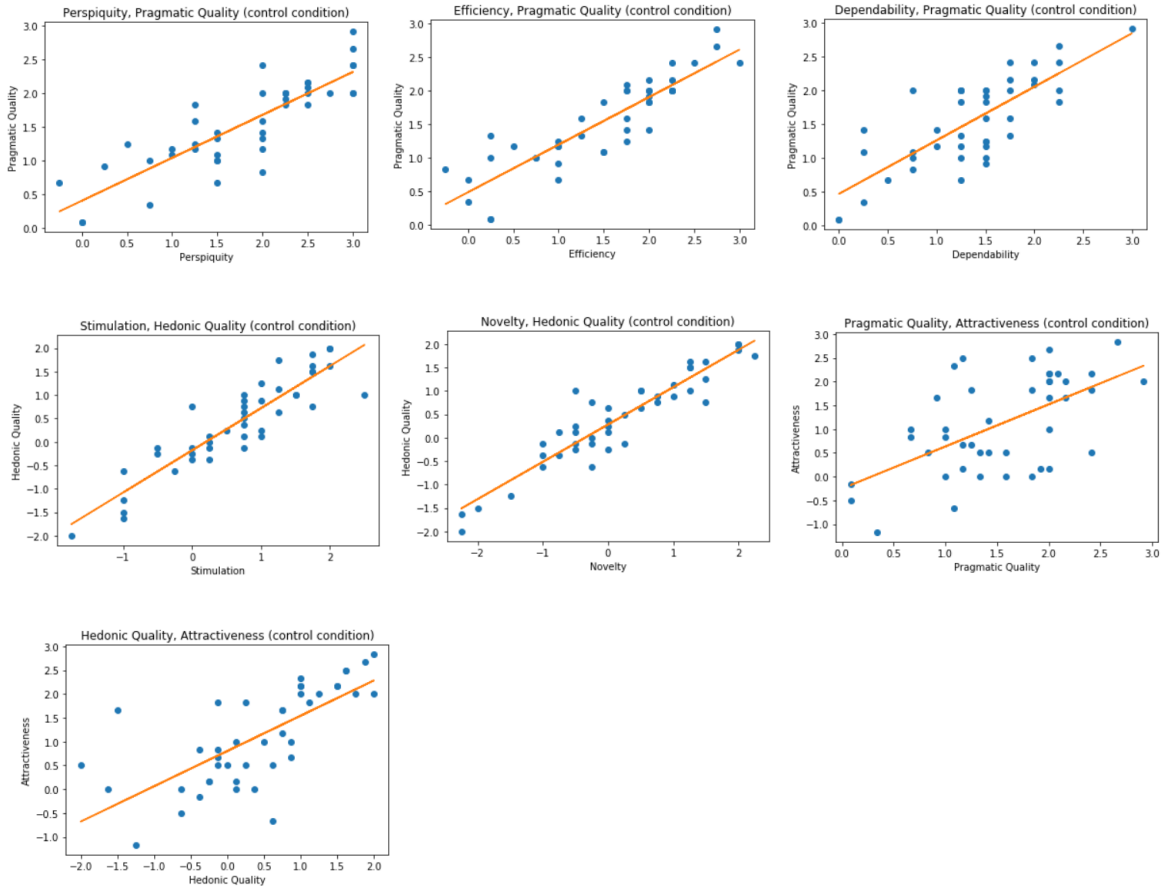

Figure 4.3: Pearson Correlation Coefficient scatterplot of control condition

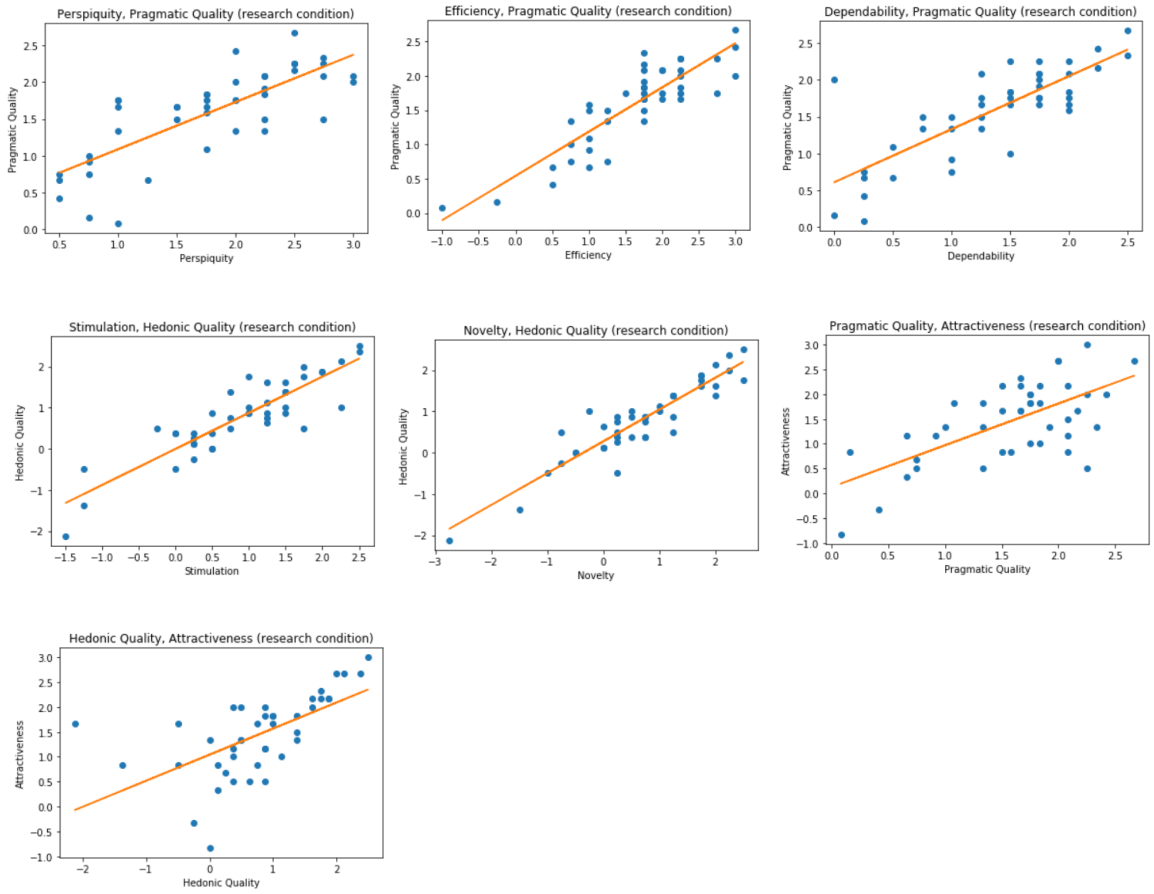

Figure 4.4: Pearson Correlation Coefficient scatterplot of research condition

A Pearson correlation coefficient was computed to asses the linear relationship between Perspicuity and the Pragmatic Quality. In the control condition, there was a positive correlation between the two variables, $r(41) = .85$, $p = .00001$. In the research condition, also a positive correlation was found, $r(41) = .77$, $p = .00001$.
Similarly, A Pearson correlation coefficient was computed to asses the linear relationship between Efficiency and the Pragmatic Quality. In the control condition, there was a positive correlation between the two variables, $r(41) = .89$, $p = .00001$. In the research condition, also a positive correlation was found, $r(41) = .87$, $p = .00001$.
Next, Pearson correlation coefficient was computed to asses the linear relationship between Dependability and the Pragmatic Quality. In the control condition, there was a positive correlation between the two variables,

$r(41) = .78$, $p = .00001$. In the research condition, also a positive correlation was found, $r(41) = .79$, $p = .00001$.
Again, a Pearson correlation coefficient was computed to asses the linear relationship between Stimulation and the Hedonic Quality. In the control condition, there was a positive correlation between the two variables, $r(41) = .91$, $p = .00001$. In the research condition, also a positive correlation was found, $r(41) = .89$, $p = .00001$.
A Pearson correlation coefficient was also computed to asses the linear relationship between Novelty and the Hedonic Quality. In the control condition, there was a positive correlation between the two variables, $r(41) = .93$, $p = .00001$. In the research condition, also a positive correlation was found, $r(41) = .92$, $p = .00001$.
A Pearson correlation coefficient was computed to asses the linear relationship between the Pragmatic Quality and the Attractiveness. In the control condition, there was a positive correlation between the two variables, $r(41) = .58$, $p = .000043$. In the research condition, also a positive correlation was found, $r(41) = .65$, $p = .00001$.
Lastly, A Pearson correlation coefficient was computed to asses the linear relationship between the Hedonic Quality and the Attractiveness. In the control condition, there was a positive correlation between the two variables, $r(41) = .71$, $p = .00001$. In the research condition, also a positive correlation was found, $r(41) = .61$, $p = .000013$.

The previous analysis shows that all Pearson correlation coefficients are between $r = 0.5$ and $r = 1.0$. In regards to the interpretation of these values [62], this means that all pairs of variables have a high positive degree of correlation. The corresponding calculated p-values all are $p \leq .05$, which is below the significance level of $\alpha = .05$. Based on this analysis, there can be concluded that hypotheses H7, H8, H9, H10, H11, H12, and H13 are accepted in both the control condition and the research condition. Figure 4.5 illustrates which hypotheses of the research model were accepted or rejected in this study.

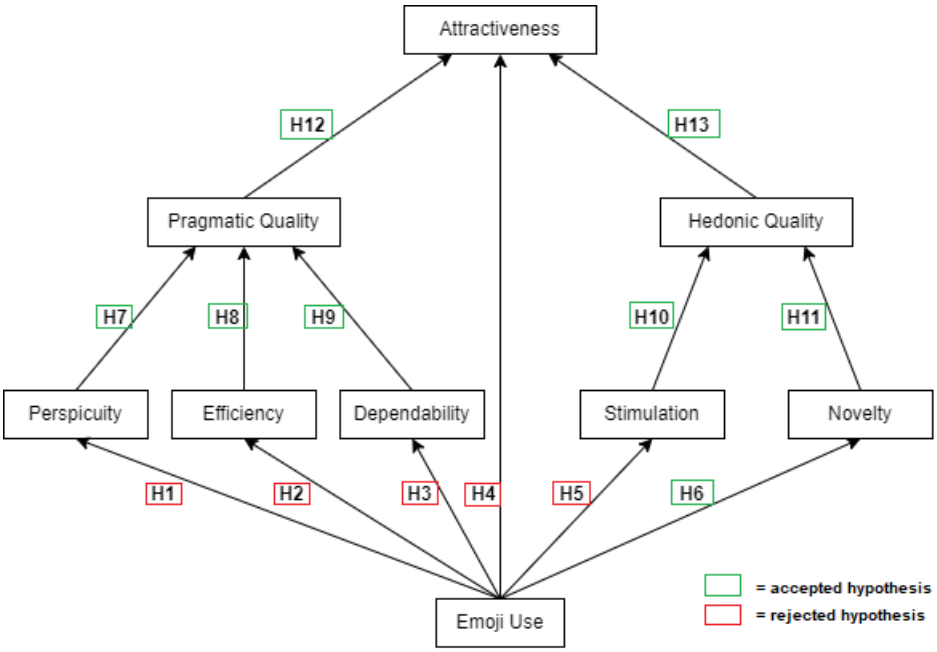

Figure 4.5: Accepted and rejected hypotheses in research model

Chapter 5

Discussion

Aiming at gaining a higher user experience in educational chatbots, this study made use of emojis. The influence of emojis was measured by applying the User Experience Questionnaire Model [16]. Obtaining the values of the six scales (e.g. Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty), the T-test was carried out.
Those results showed that, compared to the significance level, there is no significant difference between the control and research conditions in the scales Attractiveness, Perspicuity, Efficiency, Dependability, and Stimulation. However, there was found a significant difference in the Novelty scale. Because of this, hypotheses H1, H2, H3, H4, and H5 were rejected and hypothesis H6 was accepted. From this, it was concluded that there was found no evidence that the scales Attractiveness, Perspicuity, Efficiency, Dependability, and Stimulation are significantly influenced by the use of emojis. The use of emojis is in these scales not an appropriate way to increase (or decrease) their perceived value. However, there was found evidence that the Novelty scale is significantly influenced by the use of emojis. Thus, to increase the perceived Novelty (i.e. how innovative and creative the chatbot is), emojis could be used.
The Pearson Correlation Coefficient showed that there was found high correlation between all investigated pairs of variables. This led to accepting hypotheses H7, H8, H9, H10, H11, H12, and H13, as expected. Concluded from this was that there was found evidence that if the Perspicuity, Efficiency or Dependability increases, the Pragmatic Quality also increases. There was also found evidence that if the Stimulation or Novelty increases, the Hedonic Quality increases. Lastly, there was found evidence that if the Pragmatic Quality or Hedonic Quality increases, the Attractiveness of the chatbot increases. Note that in all these cases, a decrease in the first variable also led to a decrease in the second variable.
These findings show that if there is found a way to increases the Perspicuity, Efficiency, Dependability, Stimulation and/or Novelty, there it is likely that

the Attractiveness of the chatbot increases as well. As Attractiveness is a way to measure the user experience of the chatbot, increasing the Attractiveness will improve the user experience.
This study also carried out a T-test to see whether the use of emojis (in)directly lead to a significant difference in the Pragmatic Quality or Hedonic Quality in the two chatbots. There was found no evidence that there was a significant difference. However, for the Hedonic Quality $p = 0.519$ was close to a significant result. Future research could be conducted to give more certainty about this finding. As for this study, it can be concluded that emojis do not influence the Pragmatic Quality or Hedonic Quality.
The found results of this study can answer the main research question: 'How much do emojis influence the user experience of an educational chatbot?'. As the Attractiveness scale was not significantly higher with the use of emojis, it can be concluded that emojis do not influence the user experience of educational chatbots. However, there is found evidence that emojis influence the Novelty of the chatbot, which itself influences the Hedonic Quality, which again influences the Attractiveness. In this study, the Novelty does not influence the Hedonic Quality enough to say that the Novelty influences the Attractiveness and thus changes the user experience significantly.

5.1 Limitations

During this study, there were a few limitations that could be the cause of the outcome of this study.
Firstly, this study was specifically about the small scale of educational chatbots. Thus, it is uncertain that the results and conclusions of this study can be extended to any other type of chatbot. Replicating the idea of this study on more types of chatbots or a bigger scale would be needed.
Secondly, the emojis used in this study were chosen to reflect the 'typical' and 'everyday' usage of emojis. The frequency of use and type of emojis in the dialog was rather subjective and represented the way the researcher would use them. By looking at what kind of emojis and the frequency of using them would be optimal, the user experience might be improved more.
Additionally, a big limitation to take into consideration was the time capacity. More time could be taken to carry out the multiple steps in the process. As testing and implementing the chatbot's dialog is crucial for its behavior and thus user experience, having more time for this would be more convenient. Also having more time to find participants to participate in the experiment would lead to higher sample size and thus higher statistical power.
This small sample size also required keeping outliers in the data set that otherwise would have been removed. It should be kept in mind that the

outliers may have influenced the outcome of this study.
Another limitation is that this study had a participation group consisting of all Dutch participants, in the age range of 18-35 years old. This is a somewhat specific group of participants, which may make it hard to generalize the results of this study to the total population.
During the data analysis, only a quantitative analysis was carried out. It would have been much better to also include qualitative analysis, with the use of open answer questions in the questionnaire and by evaluating the chatbot conversations.
Lastly, the study was set up and conducted during the Covid-19 pandemic. Consequently, the participants had to do the study online and at home, instead of in real life and with the researcher present to guide them. This may have led to more distractions or noise in the environment, which could have influenced the data.

5.2 Future work

As this study has not found evidence for a way to improve the user experience of educational chatbots, more future research work is needed.
One take-home message from this study that can be used for future work, is that emojis are a possible way to improve the Novelty aspect of educational chatbots. This means there is at least some influence by emoji use on one of the aspects of the user experience. There could be done future research finding similar or new ways to increase the other scale values to improve the user experience. This can be done by e.g. investigating what the optimal frequency of emoji use is, or investigating what specific type of emojis increase the scale values. As this study also concluded that there was a high positive correlation between the Attractiveness and the other 5 scales of the model, this might result in a higher user experience.
Another conclusion that was drawn in this study was that the p-value of the effect of emojis on the Hedonic Quality was very close to significant. As this was quite an ambiguous conclusion, future research should investigate and check if this insignificance was correct or not.
Besides that, future research can try to increase the scale values by focusing on a different demographic population group, as this study contained only Dutch participants with an age between 18 and 35 years old. Also replicating this study while using more participants might give a different result.
Furthermore, future research can also investigate if there are any significant differences in user experience caused by emoji use in a different chatbot domain. Instead of the domain of cyber security, future research can focus on chatbots used in e.g. the healthcare or entertainment domain.
There can also be investigated if a replication of this study with another

model gives a different result. Of course, the User Experience Questionnaire Model uses only one way to measure user experience, and many other ways are possible.

Chapter 6

Conclusions

This paper provides insight into the role of emojis on the user experience in educational chatbots. The findings in this study suggest that the use of emojis does not influence the user experience in educational chatbots.
There was found evidence that emojis do influence the Novelty scale of the User Experience Questionnaire Model. No such significant result was found for the Attractiveness, Perspicuity, Efficiency, Dependability, or Stimulation scale.
Additionally, there was also found evidence that there is a high positive correlation between the Perspicuity, Efficiency and Dependability scale and the Pragmatic Quality. This high positive correlation was also found between the Stimulation and Novelty scale and the Hedonic Quality.
Also, a high positive correlation was found between the Pragmatic Quality and Hedonic Quality and the Attractiveness.
Lastly, there was found no significant difference in Pragmatic Quality or Hedonic Quality when using emojis versus when not using emojis. However, the proof for this for the Hedonic Quality was quite ambiguous.
Future research could focus on using a new feature to improve the user experience in educational chatbots. It can also focus on replicating this study while using a different model to measure user experience, a bigger (demographically diverse) group of participants, changing the emoji use frequency or emoji type, or using a different type of chatbot. Future research could also assess whether the effect of emojis on the Hedonic Quality is significant or not, as this study found some ambiguous results.

Chapter 7

Acknowledgements

Firstly, I would like to thank Marianna De Sá Siqueira for her great guidance and involvement as a supervisor in this thesis. Secondly, I would like to thank my fellow thesis group members Andy Huang, Annemiek van der Leest, Sanne Janssen, and Spence van Asperdt for their great collaboration. They helped with developing the chatbot, website, and questionnaire of this thesis. I also would like to thank Martijn Vastenburg for becoming the second reader of this thesis. Next, I want to express my gratitude to my friends and family who were there to emotionally support me in times of stress and understood when I had no time to join them. Finally, I would like to thank all participants for their crucial contribution to this study.

Bibliography

[1] Darius Zumstein and Sophie Hundertmark. Chatbots – an interactive technology for personalized communication, transactions and services. *IADIS International Journal on WWW/Internet*, 15:96–109, 11 2017.

[2] Petter Bae Brandtzaeg and Asbjørn Følstad. Why people use chatbots. In *International conference on internet science*, pages 377–392. Springer, 2017.

[3] Silvia García-Méndez, Francisco De Arriba-Pérez, Francisco J González-Castaño, José A Regueiro-Janeiro, and Felipe Gil-Castiñeira. Entertainment chatbot for the digital inclusion of elderly people without abstraction capabilities. *IEEE Access*, 9:75878–75891, 2021.

[4] Silvia García-Méndez, Francisco De Arriba-Pérez, Francisco J. González-Castaño, José A. Regueiro-Janeiro, and Felipe Gil-Castiñeira. Entertainment chatbot for the digital inclusion of elderly people without abstraction capabilities. *IEEE Access*, 9:75878–75891, 2021.

[5] Fabio Clarizia, Francesco Colace, Marco Lombardi, Francesco Pascale, and Domenico Santaniello. Chatbot: An education support system for student. In *International Symposium on Cyberspace Safety and Security*, pages 291–302. Springer, 2018.

[6] Attila Nagy. *The Impact of E-Learning*, pages 79–96. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[7] Francesco Colace, Massimo De Santo, Marco Lombardi, Francesco Pascale, Antonio Pietrosanto, and Saverio Lemma. Chatbot for e-learning: A case of study. *International Journal of Mechanical Engineering and Robotics Research*, 7(5):528–533, 2018.

[8] Gwo-Jen Hwang and Ching-Yi Chang. A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, pages 1–14, 2021.

[9] Carine Lallemand, Vincent Koenig, and Guillaume Gronier. How relevant is an expert evaluation of user experience based on a psychological

needs-driven approach? In *Proceedings of the 8th Nordic conference on human-computer interaction: Fun, fast, foundational*, pages 11–20, 2014.

[10] Timo Jokela. The early phases of ux: Why they are important (more than evaluation), and what they are? In *International Workshop on the Interplay between User Experience (UX) Evaluation and System Development (I-UxSED 2012)*, page 49. Citeseer, 2012.

[11] Khaled Hassanein and Milena Head. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International journal of human-computer studies*, 65(8):689–708, 2007.

[12] Rohit Kumar, Hua Ai, Jack L Beuth, and Carolyn P Rosé. Socially capable conversational tutors can be effective in collaborative learning situations. In *International conference on intelligent tutoring systems*, pages 156–164. Springer, 2010.

[13] Asbjørn Følstad and Marita Skjuve. Chatbots for customer service: user experience and motivation. In *Proceedings of the 1st international conference on conversational user interfaces*, pages 1–9, 2019.

[14] Marília Prada, David L Rodrigues, Margarida V Garrido, Diniz Lopes, Bernardo Cavalheiro, and Rui Gaspar. Motives, frequency and attitudes toward emoji and emoticon use. *Telematics and Informatics*, 35(7):1925–1934, 2018.

[15] Austin Beattie, Autumn P Edwards, and Chad Edwards. A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, 71(3):409–427, 2020.

[16] Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. User experience questionnaire, 2018. `https://www.ueq-online.org/`.

[17] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Construction of a benchmark for the user experience questionnaire (ueq). *Int. J. Interact. Multim. Artif. Intell.*, 4(4):40–44, 2017.

[18] Andreas Hinderks, Martin Schrepp, Francisco José Domínguez Mayo, María José Escalona, and Jörg Thomaschewski. Developing a ux kpi based on the user experience questionnaire. *Computer Standards & Interfaces*, 65:38–44, 2019.

[19] Qingtang Liu, Jingxiu Huang, Linjing Wu, Ke Zhu, and Shen Ba. Cbet: design and evaluation of a domain-specific chatbot for mobile learning. *Universal Access in the Information Society*, 19(3):655–673, 2020.

[20] Robert Dale. The return of the chatbots. *Natural Language Engineering*, 22(5):811–817, 2016.

[21] Alan M Turing. Computing machinery and intelligence. In *Parsing the turing test*, pages 23–65. Springer, 2009.

[22] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[23] Michael L Mauldin. Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In *AAAI*, volume 94, pages 16–21, 1994.

[24] Richard S Wallace. The anatomy of alice. In *Parsing the turing test*, pages 181–210. Springer, 2009.

[25] Alice Kerry, Richard Ellis, and Susan Bull. Conversational agents in e-learning. In *International conference on innovative techniques and applications of artificial intelligence*, pages 169–182. Springer, 2008.

[26] Hien D Nguyen, Vuong T Pham, Dung A Tran, and Trung T Le. Intelligent tutoring chatbot for solving mathematical problems in high-school. In *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, pages 1–6. IEEE, 2019.

[27] Luke Fryer and Rollo Carpenter. Bots as language learning tools. *Language Learning & Technology*, 10(3):8–14, 2006.

[28] Marc Cavazza, Raúl Santos De La Cámara, Markku Turunen, Jose Relaño-Gil, Jaakko Hakulinen, Nigel Crook, and Debora Field. 'how was your day?' an affective companion eca prototype. In *Proceedings of the SIGDIAL 2010 Conference*, pages 277–280, 2010.

[29] Alice Kerly, Richard Ellis, and Susan Bull. Calmsystem: a conversational agent for learner modelling. In *International conference on innovative techniques and applications of artificial intelligence*, pages 89–102. Springer, 2007.

[30] Claudia Roda, Albert Angehrn, and Thierry Nabeth. Conversational agents for advanced learning: Applications and research. *Proceedings BotShow*, pages 1–7, 2001.

[31] Andreea Bendovschi. Cyber-attacks–trends, patterns and security countermeasures. *Procedia Economics and Finance*, 28:24–31, 2015.

[32] Kevin F McCrohan, Kathryn Engel, and James W Harvey. Influence of awareness and training on cyber security. *Journal of internet Commerce*, 9(1):23–41, 2010.

[33] Albert H Huang, David C Yen, and Xiaoni Zhang. Exploring the potential effects of emoticons. *Information & Management*, 45(7):466–473, 2008.

[34] Maja Katarina Tomić, Marijana Marinez, and Tedo Vrbanec. Emoticons. *FIP-Financije i pravo*, 1(1):35–42, 2013.

[35] Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.

[36] SooJin Lee. Emoji at moma: Considering the'original emoji'as art. *First Monday*, 2018.

[37] Oxford word of the year 2015, 2015. `https://languages.oup.com/word-of-the-year/2015/`.

[38] Nikola Ljubešić and Darja Fišer. A global analysis of emoji usage. In *Proceedings of the 10th Web as Corpus Workshop*, pages 82–89, 2016.

[39] Monica A Riordan. Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology*, 36(5):549–567, 2017.

[40] Thorsten M Erle, Karoline Schmid, Simon H Goslar, and Jared D Martin. Emojis as social information in digital communication. *Emotion*, 2021.

[41] Juha Leino and Tomi Heimonen. Improving evaluation honesty and user experience in e-learning by increasing evaluation cost and social presence. In *IFIP Conference on Human-Computer Interaction*, pages 597–615. Springer, 2013.

[42] Jodi Forlizzi and Katja Battarbee. Understanding experience in interactive systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 261–268, 2004.

[43] Effie Law, Virpi Roto, Arnold POS Vermeeren, Joke Kort, and Marc Hassenzahl. Towards a shared definition of user experience. In *CHI'08 extended abstracts on Human factors in computing systems*, pages 2395–2398. 2008.

[44] Marc Hassenzahl and Noam Tractinsky. User experience - a research agenda. *Behaviour & Information Technology*, 25(2):91–97, 2006.

[45] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold POS Vermeeren, and Joke Kort. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 719–728, 2009.

[46] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer, 2008.

[47] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4):396, 2015.

[48] Imb watson, 2022. `https://www.ibm.com/watson`.

[49] Blogger, 2022. `https://www.blogger.com/about/?bpli=1`.

[50] Qualtrix xm, 2022. `https://www.qualtrics.com/uk/?rid=ip&prevsite=en&newsite=uk&geo=NL&geomatch=uk`.

[51] David E Gray. *Doing research in the real world*. sage, 2021.

[52] Abdul Momin Kazi and Wardah Khalid. Questionnaire designing and validation. *Journal of the Pakistan Medical Association*, 62(5):514, 2012.

[53] Gary Charness, Uri Gneezy, and Michael A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior Organization*, 81(1):1–8, 2012.

[54] Joseph L Brooks. Counterbalancing for serial order carryover effects in experimental condition orders. *Psychological methods*, 17(4):600, 2012.

[55] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. Gender bias in chatbot design. In *International Workshop on Chatbot Research and Design*, pages 79–93. Springer, 2019.

[56] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Ueq model handbook. 2019.

[57] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Ueq model data analysis tool. 2021.

[58] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Ueq model compare products tool. 2020.

[59] Keya Rani Das and AHMR Imon. A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*, 5(1):5–12, 2016.

[60] Neil J Salkind. *Encyclopedia of measurement and statistics*. SAGE publications, 2006.

[61] Karl Pearson. On first power methods of finding correlation. *Biometrika*, pages 459–469, 1925.

[62] Patrick Schober, Christa Boer, and Lothar A Schwarte. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, 2018.

Appendix A

Appendix

A.1 Question Items UEQ Model

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

Question items of the User Experience Questionnaire Model

A.2 Informed Consent Form

Informed Consent Form

Information

Purpose of the Study

This study is part of a Bachelor Thesis of the Artificial Intelligence Program of the Radboud University.
The study involves an educative chatbot that will teach the user (you) about Cyber Security. You will use this chatbot twice and fill in a questionnaire both times.

Voluntary Participation
Your participation should be voluntary. During the study you can indicate at any time that you want to quit participating. You can stop without having to explain why and at any moment you want.

Anonymity
All collected information and data is processed anonymously: the results cannot be traced back to you later. This also means that we therefore cannot inform you about your personal results when the study is completed. However, if you are interested, we can inform you about the results of the complete study. If you wish to be informed about these results, please let us know.

Data Usage
The (anonymous) data that is collected about you will be used as part of data sets, articles, and presentations. It will be accessible to other researchers for a period of at least 10 years. However, as this is anonymous data, it cannot be traced back to you. The data is collected from the chat history and the questionnaire forms.

Contact
If you have any further questions about this research or if you are interested in the results of the study, you can contact the following email address: serah.sommers@ru.nl

Consent
By giving consent as a participant of this study, you agree with the following statements:

- I have read the provided information about this research study and understand it.
- The aim of this research study has been made clear to me.
- I have been given the opportunity to ask questions regarding the research study.
- I participate in this research study on voluntary basis.
- I understand that I am allowed to stop at any time during this research study, without having to give a reason.
- I understand that the data gathered from this research study is anonymous.
- I understand how the data gathered from this research study will be stored and used.

A.3 Instruction Page

Task instructions

In this study you are going to have two short conversations with a chatbot for about 5-10 minutes. It is an educative chatbot that aims to teach you about Cyber Security. Your task is to conversate with the chatbot and afterwards fill in a questionnaire about the chatbot. You will do this two times sequentially and without taking a break. The total time for participating in this study is 20-30 minutes.

Please read the following instruction rules very carefully:

- Both mobile phones and computers/laptops support the participation of this research, but we advise you to use your computer/laptop.
- Please make sure you are in an environment that is distraction-free. Close any unnecessary program on your computer, interact with no other person and put your mobile phone on silent mode.
- Keep this instruction page open at all times, so you can re-read the instructions if things are unsure to you.
- Do NOT close the chatbot window during the conversation session.
- Open any link given during the conversation in a separate tab. You can do this by: Hold CTRL + click Left Mouse OR Click Right-Mouse and choose 'Open link in new tab'.
- As this is an English chatbot, it expects English responds from the user. Make sure you ONLY answer in English. You are allowed to quickly look up translations.
- Make sure NOT to respond to the chatbot before it is done typing itself (indicated by the '…').
- Only respond with ONE message at a time.
- Do not wait too long (more than a minute) to respond to the chatbot, as the dialog will otherwise shutdown.
- Should you not receive the link for the questionnaire within 15 minutes of testing a chatbot, please contact the researcher (serah.sommers@ru.nl)

Now you have read the instruction rules, you can start the actual research. Make sure to **read the following steps first, and then follow those steps in GIVEN ORDER**:

1. Open the chatbot for the first conversation with the following link: *https://chatbotalice.blogspot.com/2021/11/chatbot1.html*
2. Have a conversation with the chatbot until it reaches the end of the dialog and the chatbot sends you the link to the questionnaire
3. Fill in the whole questionnaire and go back to this instruction page
4. Now open the chatbot for the second conversation with the following link: *https://chatboted.blogspot.com/2021/11/chatbot2.html*
5. Have a conversation with the chatbot until it reaches the end of the dialog and the chatbot sends you the link to the questionnaire
6. Fill in the whole questionnaire
7. You are now done with this research!

Thank you for participating in this research study and I hope you had a little fun with it! If you still have any questions regarding this study or if you are interested in knowing it's purpose, please contact the researcher at: serah.sommers@ru.nl

A.4 User Experience Questionnaire

Block 1

What is the Keyword that the chatbot gave to you?

Fill in a identification code, in the format of '123ABC' (so 3 numbers, followed by 3 letters). Use the same identification code in both questionnaires! If this is the first questionnaire you fill in, think of a new original (not 123ABC) one without personal information in it and remember it. If this is the second questionnaire you fill in, please us the same one as in the last questionnaire.

Before you start this questionnaire, please read the following informed consent page thoroughly:
Informed consent form
I consent by participating in this research:

- I consent
- I do not consent

Chatbot Questions:

I find the chatbot... (1/5)

	1	2	3	4	5	6	7	
annoying								enjoyable
not understandable								understandable
creative								dull
easy to learn								difficult to learn
valuable								inferior
boring								exciting

I find the chatbot... (2/5)

	1	2	3	4	5	6	7	
not interesting								interesting
unpredictable								predictable
fast								slow
inventive								conventional
obstructive								supportive
good								bad

I find the chatbot... (3/5)

	1	2	3	4	5	6	7	
complicated	○	○	○	○	○	○	○	easy
unlikable	○	○	○	○	○	○	○	pleasing
usual	○	○	○	○	○	○	○	leading edge
unpleasant	○	○	○	○	○	○	○	pleasant
secure	○	○	○	○	○	○	○	not secure
motivating	○	○	○	○	○	○	○	demotivating

I find the chatbot... (4/5)

	1	2	3	4	5	6	7	
meets expectations	○	○	○	○	○	○	○	does not meet expectations
inefficient	○	○	○	○	○	○	○	efficient
clear	○	○	○	○	○	○	○	confusing
impractical	○	○	○	○	○	○	○	practical
organized	○	○	○	○	○	○	○	cluttered
attractive	○	○	○	○	○	○	○	unattractive

I find the chatbot... (5/5)

	1	2	3	4	5	6	7	
friendly	○	○	○	○	○	○	○	unfriendly
conservative	○	○	○	○	○	○	○	innovative

General Questions:

What is your age?

[ ]

What is your gender?

- ○ Female
- ○ Male
- ○ Prefer not to say
- ○ Other: [ ]

What is your nationality?

[ ]

What is the highest level of education that you have finished?

- ○ None
- ○ Primary School

- ○ Secondary School
- ○ Post-secondary Vocational Education (MBO)
- ○ Higher Education (HBO) Bachelor
- ○ Higher Education (HBO) Master
- ○ University (WO) Bachelor
- ○ University (WO Master)
- ○ PhD
- ○ Other [ ]

In which field do you work or study?

[ ]

Have you ever used a chatbot before today?

- ○ Yes
- ○ No