

BACHELOR THESIS  
ARTIFICIAL INTELLIGENCE

**Radboud University**



---

**The effect of a peer persona on the user  
experience for an educational chatbot**

---

*Author:*

Sanne Janssen  
s1018595

*First supervisor:*

M.A. De Sá Siqueira  
PhD student - Behavioral  
Science Institute  
marianna.desasiqueira@ru.nl

*Second supervisor:*

dr. M.H. Vastenburg  
Assistant professor -  
Design for Behavior Change  
martijn.vastenburg@ru.nl



January 28, 2022

## **Abstract**

In the last few years, the use of chatbots has become quite popular. One of the fields they can be used in is education, however, most implementations in this field have failed to compare to human counterparts for the same task. In this study, a chatbot with the persona of a student's peer was developed to educate users on the importance of safe internet practices. This peer persona, as well as a neutral persona, were compared to investigate which of the two conditions would lead to a higher user experience. The user experience was measured using a modified Unified Theory of Acceptance and Use of Technology. The results indicated that a peer persona compared to a neutral persona does not have a significant effect on the overall user experience. However, it does increase the Perceived Enjoyment of the user. The results also confirmed that the correlations within the Unified Theory of Acceptance and Use of Technology were accurate.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	Chatbots . . . . .	5
2.1.1	Educational chatbots . . . . .	5
2.2	Peer learning . . . . .	6
2.3	User experience . . . . .	6
2.4	Unified Theory of Acceptance and Use of Technology (UTAUT)	7
2.4.1	Effort Expectancy . . . . .	8
2.4.2	Performance Expectancy . . . . .	9
2.4.3	Perceived Enjoyment . . . . .	9
2.4.4	Trust . . . . .	9
2.4.5	Satisfaction . . . . .	9
2.4.6	Social Influence . . . . .	9
2.5	Research Model and Hypotheses . . . . .	9
<b>3</b>	<b>Method</b>	<b>11</b>
3.1	Chatbot Conditions . . . . .	11
3.2	Materials . . . . .	11
3.2.1	Chatbot Design . . . . .	11
3.2.2	Website Design . . . . .	13
3.3	Procedure . . . . .	15
3.3.1	Questionnaire . . . . .	16
3.4	Demographic . . . . .	17
3.4.1	Control Condition . . . . .	17
3.4.2	Experiment Condition . . . . .	18
3.4.3	Technology Experience . . . . .	18
<b>4</b>	<b>Results</b>	<b>19</b>
4.1	Statistical Analysis . . . . .	19
4.1.1	Outliers . . . . .	19
4.1.2	General Overview . . . . .	20
4.1.3	ANOVA Analysis . . . . .	21

4.1.4	Pearson Correlation . . . . .	23
4.2	Qualitative Analysis . . . . .	24
4.2.1	Control group . . . . .	24
4.2.2	Experiment group . . . . .	25
<b>5</b>	<b>Discussion</b>	<b>27</b>
5.1	Interpretation . . . . .	27
5.2	Limitations . . . . .	28
5.3	Recommendations . . . . .	29
<b>6</b>	<b>Conclusion</b>	<b>31</b>
<b>7</b>	<b>Acknowledgements</b>	<b>32</b>
<b>A</b>	<b>Appendix</b>	<b>38</b>
A.1	Informed Consent Form . . . . .	38
A.2	Task Instructions . . . . .	40
A.3	Questionnaire . . . . .	41
A.3.1	Check if the chatbot has been used . . . . .	41
A.3.2	Technology Experience . . . . .	41
A.3.3	UTAUT questions . . . . .	41
A.3.4	Open questions . . . . .	44
A.3.5	Demographic questions . . . . .	44
A.3.6	End of survey . . . . .	45

# Chapter 1

## Introduction

In the last few years, the use of chatbots has become quite popular: almost anyone who uses the internet regularly can name a website that makes use of one. Chatbots are mainly used by companies to engage directly with their customers through a natural language user interface, often powered by artificial intelligence and machine learning. They are seen as a great and promising alternative to traditional customer service, as they provide a way to reduce costs, increase revenue, and always have help available to customers [1][2].

Besides being used as a customer service tool, chatbots have recently emerged in other areas, such as the educational domain [3]. The goal of deploying chatbots in this field is to make the learning experience more motivating, interactive, and human-like. One study used an educational chatbot for language learning, where it imitates a language-learning partner [3]. This study, however, discovered that users found it difficult to work with the chatbot, and lost interest over time. Users preferred their human language-learning partner, which led the study to conclude that they did not perceive the chatbot as a human. Creating a human-like chatbot is a primary goal of chatbot development that can be measured by passing the Turing test [4]. Not perceiving the chatbot as a human is a sign of a lower user experience, since anthropomorphism in chatbots was found to lead to a higher behavioral intention to use the system [5]. This is thus an important quality to strive for.

In order to address the problem of a low user experience described above, this thesis aims to answer the following research question:

*What is the effect on the user experience of a peer persona for an educational chatbot?*

Research has shown that in general, people take more advice from a peer than they do from an expert or anyone else [6]. One study also presented

that people relate better with people who have similar personality characteristics as themselves [7]. Furthermore, it is assumed that humans prefer to attribute human traits to machines and interact with them the same way they would with other humans, for example giving names to inanimate objects [8].

In this study, an educational chatbot with peer characteristics is used to measure the user experience. The internet is becoming more and more integrated with our personal daily lives, and so is cybercrime [9]. This is why the subject of the educational chatbot is online user awareness and safe computing. The goal of this chatbot is to educate students about the dangers of the internet and encourage them to practice safe computer habits in a quiz-like conversation.

Specifically, this study is aimed at determining whether the user experience is greater when talking to the chatbot's peer persona, or when talking to the chatbot's neutral persona.

The user experience was analyzed using a technology acceptance model, as is common when dealing with user experience. The Unified Theory of Acceptance and Use of Technology (UTAUT) was used, first introduced by Venkatesh et al. [10], and modified by Chao & Cheng-Min [11]. The model has been extended with the classic parameter Social Influence [12][13].

## Chapter 2

# Background

### 2.1 Chatbots

The definition of a chatbot is “a computer program designed to simulate conversation with human users, especially over the internet” [14]. Chatbots are mainly used by companies to engage directly with their customers through a natural language user interface, often powered by artificial intelligence and machine learning. They are seen as a great and promising alternative to traditional customer service as they provide a way to reduce costs, increase revenue, and always have help available to customers [1][2].

For users, interacting with a chatbot might feel more natural and efficient than using a traditional web service, as they can converse through a text messaging interface [15].

The aim of chatbots is to have a naturally flowing conversation with the user, ideally in which the user cannot tell that they are speaking to an artificial agent [14]. This inability to distinguish between a real human and a machine was defined as the Turing test [4]. If one cannot distinguish between a robot and a human, it might be said that the machine can think like a human [4]. Thus far, chatbots have not passed the Turing test.

This study focuses on an educational chatbot deployed in the field of online user awareness and safe computing.

#### 2.1.1 Educational chatbots

Chatbots can be applied in many fields, one of such being in education. Recently, there has been an increase in the usage of chatbots that support students in their e-learning [16]. It is said that the use of chatbots in an educational context is the best way to bridge the gap between education and technology [17].

The use of chatbots in e-learning has been studied by Eric Hsiao-Kuang Wu et al. [18]. In this study, a chatbot was designed to not only help users with their e-learning materials but also to have an everyday conversation with

them. During the period where users followed e-learning courses, the chatbot would converse with them to combat the feeling of loneliness that arose as a consequence of the lack of interaction with teachers. Results showed that although the chatbot indeed diminished the feelings of loneliness and isolation, it still did not feel like a real human companion.

Another use case of educational chatbots is in the domain of language learning. Luke K. Fryer et al. implemented a chatbot that simulated a language-learning partner [3]. Participants were paired up with either a human language partner or a chatbot language partner, and engagement and motivation to keep learning the language were measured over time. This study discovered that users found it difficult to work with the chatbot and lost interest over time. Users preferred their human language partner and working with a human increased their interest in the language course.

The educational chatbot created in this study aims to teach users about online user awareness and safe computing, as cybercrime is, unfortunately, becoming more and more integrated with our daily lives [9]. The goal of this chatbot is to educate students about the possible dangers of the internet and encourage them to practice safe computer habits, by asking them questions in a quiz-like manner.

## 2.2 Peer learning

While learning new things in everyday life, people take advice from people around them. Most of the time, they don't ask a teacher or take a course; they consult friends, family, and colleagues, as this gets them the most information in a short period of time [19]. This way of gathering information is known as peer learning; it is often seen as an accessible and easy way to learn, as peers are or have been in a similar position to themselves. Laple & Barham showed that participants were more inclined to take advice from peers than from an expert on the topic, which supports the idea that people relate better with others who have similar personality characteristics as themselves [6][7].

Furthermore, it is assumed that humans prefer to attribute human traits to machines and interact with them the same way they would with other humans, which is important to note since this project is focused on creating a peer condition in a chatbot [8].

## 2.3 User experience

According to the international standard on ergonomics of human-system interaction, ISO 9214-210, user experience is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" [20]. Analyzing and optimizing the user experience of the

two conditions of the chatbot is the primary goal of this study given that a good user experience ensures the users have an optimal experience while interacting with the product, incentivizing them to use it again [21]. User experience is usually measured using a technology acceptance model.

## 2.4 Unified Theory of Acceptance and Use of Technology (UTAUT)

A technology acceptance model that can be used to analyze the acceptance and intentions of users is the Unified Theory of Acceptance and Use of Technology, abbreviated to UTAUT. This model was developed by Venkatesh et al. and it is often used to analyze the user experience, as was also done in this study [10]. The theory suggests that there are four key constructs: Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions. The first three constructs would influence the behavioral intention and user behavior, whereas the fourth would influence user behavior directly. The UTAUT model was made to combine eight models that try to explain information technology usage and was thus deemed a good fit for this project as it takes the best features from each model and combines them. The original UTAUT model can be seen below in figure 2.1.

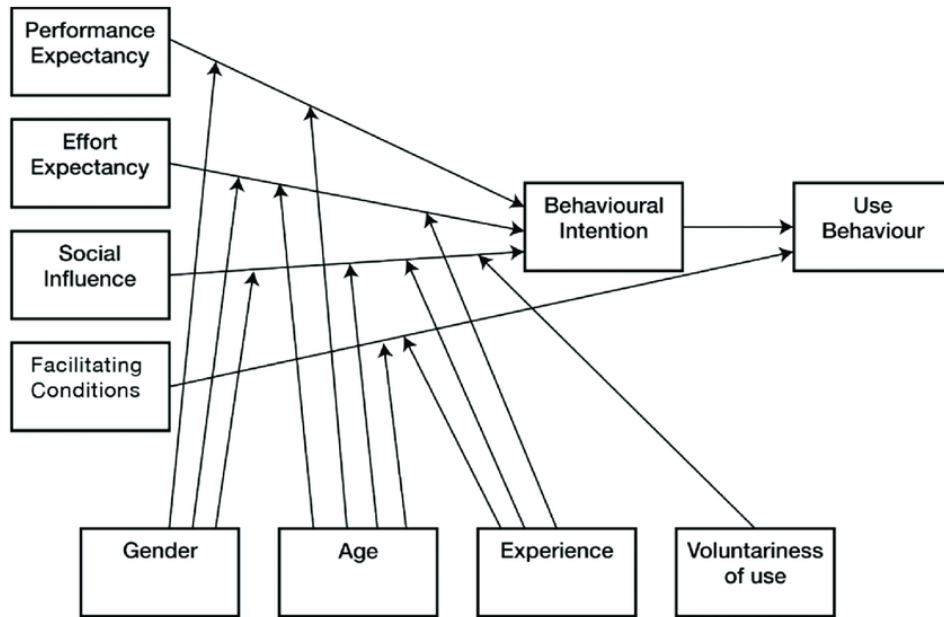


Figure 2.1: The classic UTAUT model

The baseline model used in this study was the UTAUT model as modified

by Chao & Cheng-Min, where a few extra predictors were added to the model, such as Perceived Enjoyment, Satisfaction, Trust, as well as Perceived Risk [11]. It was also proposed that Effort Expectancy, Performance Expectancy, and Perceived Enjoyment were directly influenced by external variables, such as the research condition. This model is displayed below in figure 2.2.

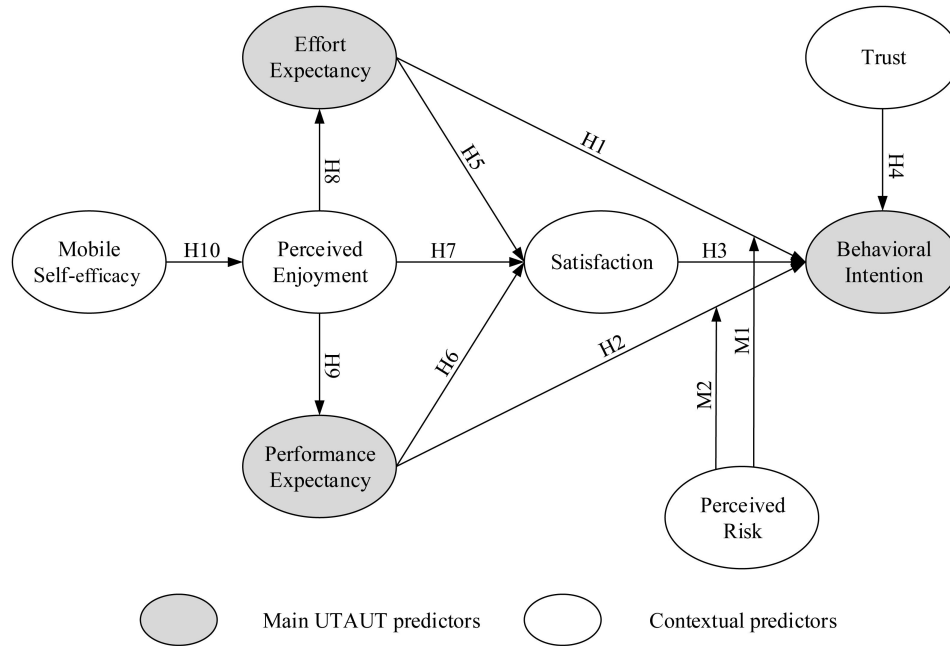


Figure 2.2: The UTAUT model as modified by Chao & Cheng-Min

The model in figure 2.2 was adjusted for the current study by removing Perceived Risk and Behavioral Intention, and adding Social Influence. Perceived Risk was removed due to the fact that using the chatbot does not carry a risk of any kind. Behavioral Intention was removed, considering the goal of the study was not necessarily to find out whether people would use the chatbot again but to find out whether they were satisfied with it, which is more in line with analyzing the user experience. Social Influence was added because the study compares a peer persona with a non-peer persona. Since this is a social aspect that is changing, Social Influence is an interesting aspect to measure. The different factors present in the final model will now be elaborated upon.

### 2.4.1 Effort Expectancy

Effort Expectancy is defined as “the degree of ease associated with the use of the system” [10][22]. According to Venkatesh et al., this factor was inspired

by the Perceived Ease of Use factor in the Technology Acceptance Model (TAM), as proposed by Davis [10][23].

#### **2.4.2 Performance Expectancy**

Performance Expectancy can be defined as “the degree to which an individual believes that using the system will help him or her to attain gains in a job” [10][22][24][25]. According to Shin, the three factors that affect Performance Expectancy are Perceived Usefulness, extrinsic motivation, and job fit [25][22].

#### **2.4.3 Perceived Enjoyment**

According to Park et al., Perceived Enjoyment can be described as “the extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use” [26][11].

#### **2.4.4 Trust**

Trust was defined as “students’ perceptions about the reliability and trustworthiness of the system” by Arpaci [27]. According to previous studies, three components make up trust: integrity, benevolence, and ability [27][28].

#### **2.4.5 Satisfaction**

The definition of Satisfaction is “users’ level of satisfaction with reports, websites, and support services” [29]. That study also showed that Satisfaction has a significant influence on whether people will use an information system.

#### **2.4.6 Social Influence**

Social Influence, according to Diaz & Loraas, can be described as “the degree to which an individual feels that it is important for others to believe he or she should use the new system” [30][22]. Given that this study is about peer conditions, Social Influence was hypothesized to be an important factor to add back into the model.

### **2.5 Research Model and Hypotheses**

As mentioned, the Unified Theory of Acceptance and Use of Technology as modified by Chao & Cheng-Min was chosen for this study, extended with Social Presence [11][12][13]. The corresponding research model is as follows:

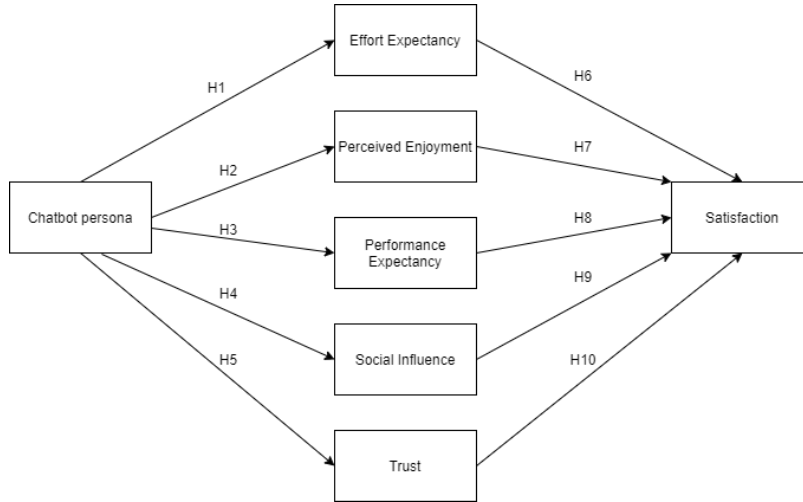


Figure 2.3: The research model

In this study, the external variable that influences the main predictors is the chatbot persona, of which there were two. The first is the control condition, where the chatbot is very formal, uses long and complicated words, and does not try to connect with the user. The second condition is the experimental chatbot, which is the peer persona. In this persona, the chatbot aims to relate and connect with the user through the use of informal, colloquial language, emojis, and slang commonly used by the target audience. The hypotheses are formulated as follows.

H1: The peer persona of the chatbot is perceived as easier to use than the neutral persona of the chatbot.

H2: The peer persona of the chatbot is perceived as more enjoyable than the neutral persona of the chatbot.

H3: The peer persona of the chatbot is perceived as more useful than the neutral persona of the chatbot.

H4: The peer persona of the chatbot is more socially influential than the neutral persona of the chatbot.

H5: The peer persona of the chatbot is trusted more than the neutral persona of the chatbot.

H6: The higher the Effort Expectancy is, the higher the Satisfaction is.

H7: The higher the Perceived Enjoyment is, the higher the Satisfaction is.

H8: The higher the Performance Expectancy is, the higher the Satisfaction is.

H9: The higher the Social Influence is, the higher the Satisfaction is.

H10: The higher the Trust is, the higher the Satisfaction is.

# Chapter 3

## Method

### 3.1 Chatbot Conditions

As previously mentioned, this study focused on the user experience of an educational chatbot with a peer persona in comparison to a neutral persona. This was achieved by creating two versions of the chatbot where the experimental condition represents the peer persona and the control condition represents the neutral persona. The control condition was developed in a group, which consisted of my colleagues Spence van Asperdt, Andy Huang, Annemiek van der Leest and Serah Sommers. A chatbot was developed that has a conversation with the user about safe computing and cyber security. After the initial creation with the aforementioned colleagues, slight modifications were applied to the control condition to create a more neutral and formal persona. This results in a bigger contrast with the peer condition. The experimental condition was developed after the creation of the control condition.

Both the control condition and the experimental condition were tested by participants followed by a survey where the data was collected from.

### 3.2 Materials

#### 3.2.1 Chatbot Design

##### IBM Watson

IBM Watson is an online platform that allows the user to create their own virtual assistant or chatbot. It utilizes machine learning and natural language processing to classify and understand user inputs. IBM Watson was chosen for developing the chatbot due to its architecture, sophisticated code, and competent capabilities for our application [31]. IBM Watson's interface has no learning curve and is overall familiar to users who have used a chat

function before. IBM Watson also provides an API key that can be pasted into the HTML code of any website, making integration almost effortless.

### **Dialog Design**

The dialog was designed within IBM Watson. The interface allows the developer to make a dialog decision tree in which the questions and answers are stored. The answers of users were classified using intents, which are essentially the core meaning of what the user is trying to convey to the chatbot. On average, around 100 examples were declared for each intent. The natural language processing and machine learning from Watson can classify user inputs to one of these intents. The dialog was thoroughly tested for errors in the flow of the dialog. The dialog was designed in a way where the chatbot guides the conversation, this makes it easier to lead the conversation and predict possible intents the user might convey.

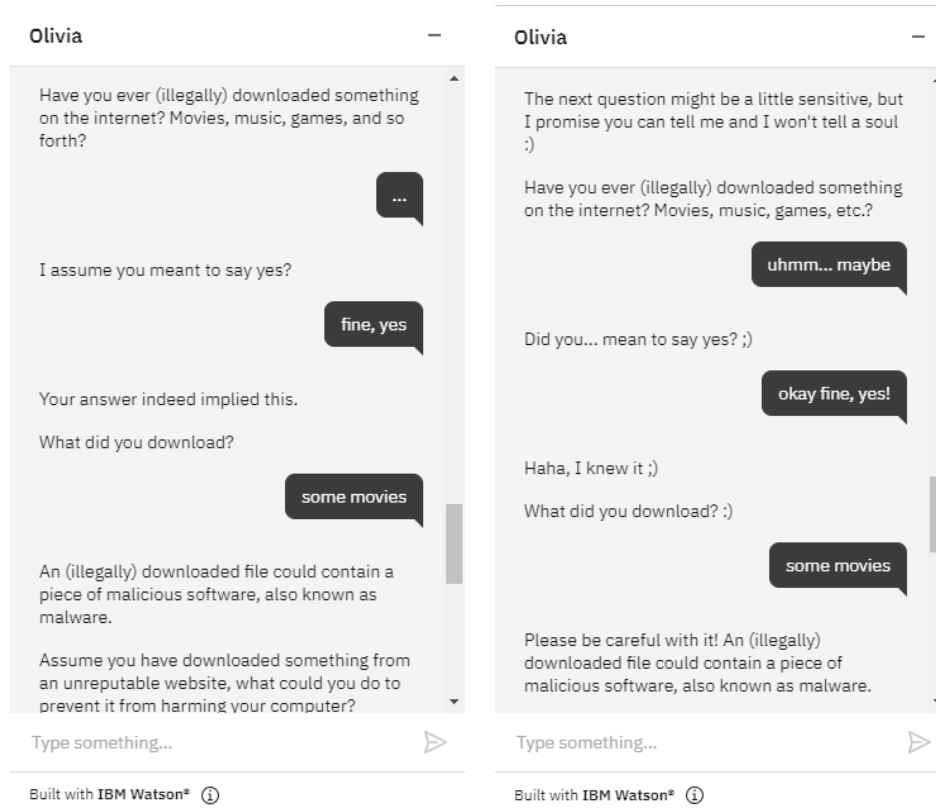
### **Peer Condition**

The research- or experimental condition of this study was the peer persona of the chatbot. Peers are “people belonging to the same societal group based on age, grade, or status” [32]. This condition’s personality is a stark contrast with the control condition; the control condition is a chatbot that is formal, distant, polite, and strictly informative. Because it does not have a specific target audience, it is harder for users to connect with it or feel like they can relate to the bot. It uses complicated and formal words, which makes its language convoluted to understand.

The experimental condition was designed to be a peer of the target audience. While designing the peer persona, it was kept in mind that the target audience for this experiment would be students aged 18-29 who are in college or university (from now on referred to as ‘students’), and the peer persona was developed as such. Students were chosen as the target audience because the chatbot would mostly be tested by confidants and colleagues, who currently satisfy all experimental criteria. Its personality is friendly, casual, colloquial, sassy, and funny, and it tries to connect with the user and make them feel comfortable. These characteristics were chosen because studies have shown that peer learning relations based on friendship can positively influence the learning process and outcome, as well as the motivation of the students [33]. This version of the bot also uses emojis and more punctuation marks, as this is more similar to how the target audience would interact with peers through text and adds a certain level of non-verbal communication to the chat [34]. For example, the addition of emojis and punctuation marks can make the bot sound more sincere, enthusiastic, or sad, which helps communicate its emotions and intent to the user. The target audience also typically uses a lot of emojis in their texts, so this helps the conversation feel more natural

and human-like [35].

An example chat at the same point in the conversation with both chatbot conditions can be seen below in figure 3.1.



(a) Example chat with the non-peer condition. Notice the formality and lack of connection with the user. (b) Example chat with the peer-condition. Notice how the chatbot tries to relate to the user and make them feel at ease.

Figure 3.1: Example chats with both research conditions

### 3.2.2 Website Design

The website that was used to host the chatbot was created with Blogger, a free online blogging system that allows you to set up multiple customizable blogs [36]. Two different web pages were created, one for each research condition.

An API key provided by IBM Watson was pasted on the website to provide a weblink. This way, the chatbot did not need to be hosted on a server and was hosted by IBM itself. To access the chatbot, users need to click on the chat icon in the lower right corner, similar to chatbots commonly used in webshops [37]. Instructions to return to the questionnaire after finishing the conversation were given to the participants through the text in the top

portion of the screen. A visualization of the website along with the chat button can be seen below in figure 3.2.

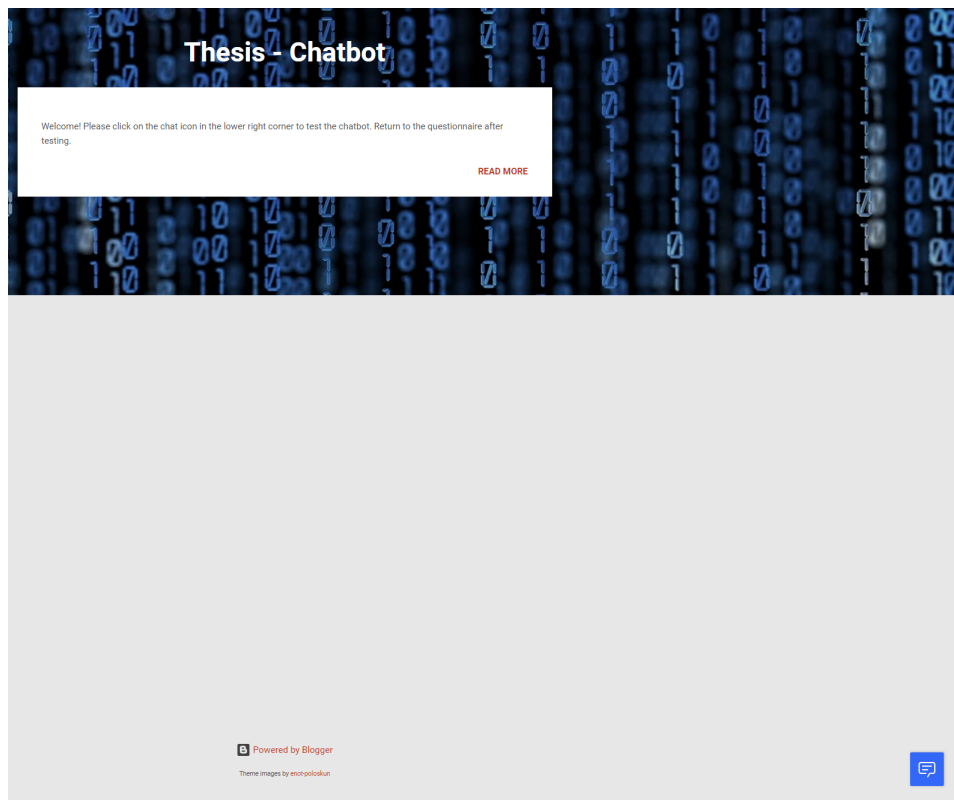
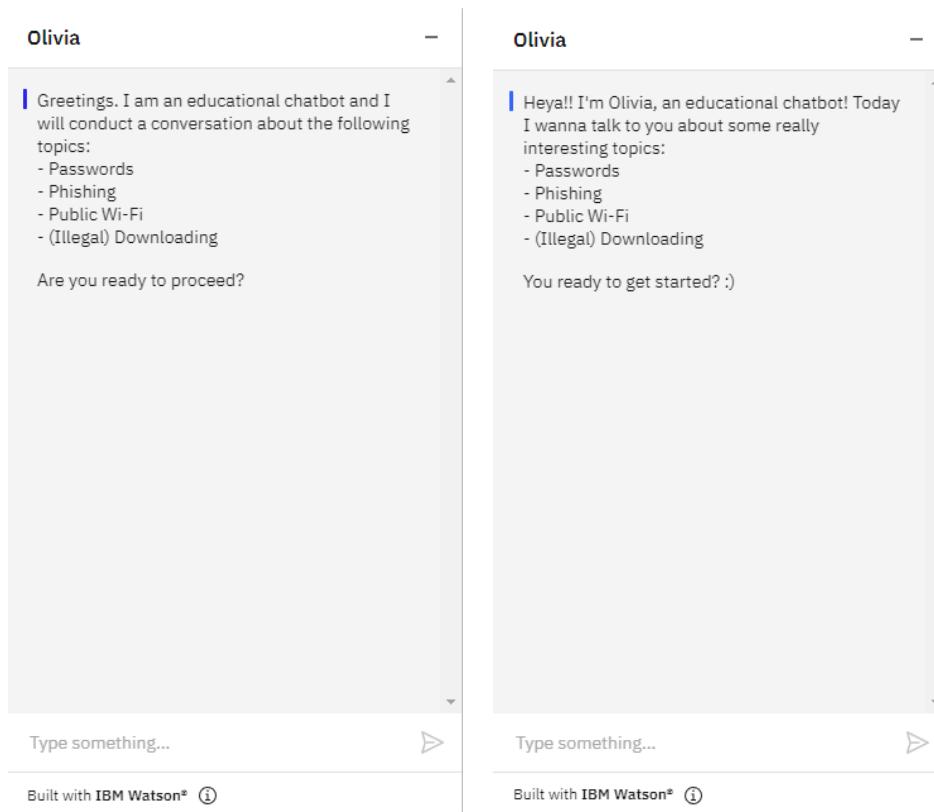


Figure 3.2: The website on which the chatbot is hosted, along with the chat button

The chat windows that are visible after clicking the chat button are shown below in figure 3.3. Note that both windows are identical in makeup. The only differing features are the personas.



(a) The chat window of the control chatbot (b) The chat window of the experimental chatbot

Figure 3.3: The chat windows of both research conditions

### 3.3 Procedure

For this study, a between-subject design was chosen. This method was determined to be fitting because it was preferable to have people give their full attention to one chatbot conversation and one survey, instead of making participants have the same conversation twice with the only difference being the persona, as well as doing the same survey twice. Research has shown that the prospect of multiple surveys can reduce the response rates [38]. The aim was to acquire at least 20 participants per research condition that use the chatbot and fill in the associated survey. The two research conditions were distributed evenly and randomly among the participants. They were instructed to have a conversation with the chatbot, which takes 5 to 10 minutes, and afterward fill in a questionnaire in which all factors presented in the research model were measured. The link to the chatbot and survey was distributed through social media such as LinkedIn and SurveyCircle, and with confidants such as family members and colleagues.

### 3.3.1 Questionnaire

The questionnaire was made using Qualtrics, which is a platform where intricate surveys can be developed for free, as provided by Radboud University [39]. The full questionnaire can be found in the Appendix. The setup of the questionnaire was as follows.

First, the participants were shown an informed consent form, which informed them about the study, and told them how their data would be processed, to keep in line with the GDPR law [40]. After giving consent, the participants were shown task instructions, which consisted of useful tips they should keep in mind while conversing with the chatbot. Next, users were directed to the website where they could converse with the chatbot. At the end of the conversation, a password was disclosed that allowed access to the survey, preventing participants from skipping the dialog with the chatbot. The participants were directed back to the survey, where the password was filled in. A question was asked about the users' experience with chatbots since this might affect the way participants interact with the system and their ease of use [41].

After this question, participants were asked to fill in four to five statements about each of the UTAUT predictors. The answers were given in a Likert-scale format, ranging from 1 to 5, with 1 being "Strongly disagree" and 5 being "Strongly agree". After the multiple-choice questions, users were asked some open questions about their experience, with the mandatory ones being what they did and did not like about the chatbot. Participants were also given the chance to give any other remarks they might have, though not mandatory.

Finally, some demographic questions were asked, such as age, gender, nationality, education, and work field. This was to get some insight into the participant sample for the study. In table 3.1, a quick overview of the questionnaire is given, along with the sources upon which the questions were inspired.

Aspect measured	Number of statements	Example statement	References
Technology Experience	1	“Have you ever used a chatbot before today?”	N/A
Effort Expectancy	5	“Learning how to use the chatbot was easy for me”	[11][42][43][44]
Perceived Enjoyment	5	“I find using the chatbot enjoyable”	[11][42][44]
Performance Expectancy	4	“The chatbot is a useful way to learn about cyber security”	[11][42][43][44]
Social Influence	4	“In general, people would support the use of the chatbot”	[42][43][44]
Trust	4	“I believe the chatbot is trustworthy”	[11]
Satisfaction	4	“Overall, I was satisfied with the chatbot”	[11]
Open Questions	3	“What did you like about the chatbot?”	N/A
Demographic Questions	5	“What is your age?”	N/A

Table 3.1: Overview of the buildup of the questionnaire

## 3.4 Demographic

In total, this study had 45 participants test the chatbot and fill out the questionnaire. As mentioned before, the participants were randomly assigned a condition, either the “control condition”, which is the neutral persona, or the “experiment condition”, which refers to the peer persona.

### 3.4.1 Control Condition

There were 23 participants in the control condition. Their ages ranged from 16 to 74, with the average age being 31, and the median age being 24. 47.8% of participants were male, 47.8% were female, and 4.4% were non-binary. One participant was English, everyone else was Dutch. As for their education, 26.1% had finished secondary school as their highest level of education, 17.4% had finished an HBO Bachelor, 13% an HBO Master, 17.4% a WO Bachelor, and 26.1% had finished a WO Master.

### 3.4.2 Experiment Condition

In the experiment condition, there were 22 participants, with the youngest participant being 18 and the oldest 51. The average age was 26 and the median age was 23. In this group, 31.8% was female and 68.2% was male. All of these participants were Dutch, and their highest level of finished education was as follows. 40.9% had finished secondary school as their highest level of education, 22.7% had finished an HBO Bachelor, 9.1% an HBO Master, 18.2% a WO Bachelor, and 9.1% had finished a WO Master.

### 3.4.3 Technology Experience

Since the question about technology experience was answered pretty much identically by both groups, the combined graph will be shown below with the statistics of how often participants used chatbots before, as can be seen in figure 3.4. The majority of people had used chatbots several times before, or at least once before. A small group of people had used it lots, and three participants had never used a chatbot before.

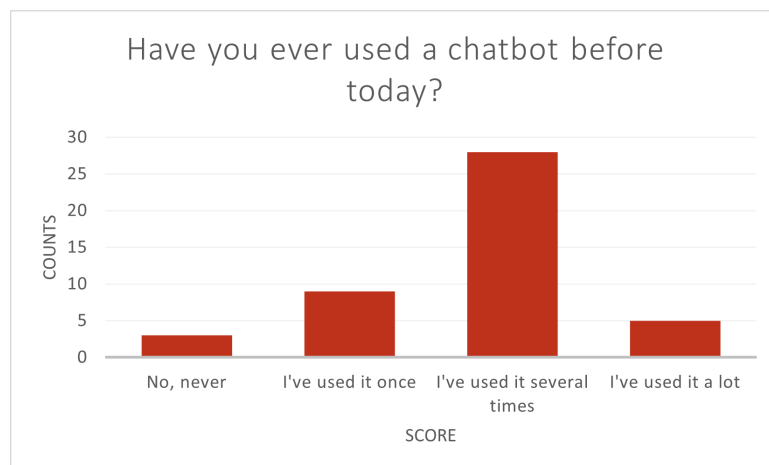


Figure 3.4: Technology Experience graph

# Chapter 4

## Results

The data from the questionnaire was used for the analysis of this study. Since each UTAUT-aspect in the questionnaire had four to five statements, all measuring approximately the same aspect, these statements were grouped and averaged to calculate the mean, standard deviation, and median [45]. This also allowed for ANOVA analysis and correlation analysis using Pearson's correlation coefficient, and allowed for the comparison of the scores for each predictor.

In order to answer hypotheses H1 through H5, one-way ANOVA was used, since the means of both conditions must be compared to draw conclusions [46]. ANOVA has three assumptions that must hold for it to be valid: independence, normality, and equal variance. However, these assumptions may be violated since ANOVA is assumed to be robust, which is described as "the extent to which the test will give the right answer even when assumptions are violated" [47]. It has been shown that accurate ANOVA analysis can be done on data that is not normally distributed, for example.

For hypotheses H6 through H10, Pearson's correlation coefficient was used. The significance level of this study was set at  $\alpha = 0.05$  [48].

### 4.1 Statistical Analysis

#### 4.1.1 Outliers

Since outliers can influence the overall analysis and accuracy of the experiment, all survey data was first checked for outliers. This was achieved by evaluating which data points fell outside of 1.5 times the interquartile range. This method was chosen because mean and standard deviations are relatively more influenced by outliers, while the interquartile range is more resilient [49]. In the control condition, two participants showed outliers in Effort Expectancy, three participants had outliers in Performance Expectancy, one had outliers in Trust, and two had outliers in Satisfaction. These outliers were all distributed among the same participants, so in total three partici-

pants with outlier values were found in the control condition.

In the experimental condition, one participant showed outliers in Effort Expectancy, two participants had outliers in Perceived Enjoyment, two had outliers in Performance Expectancy, two had outliers in Trust, and also two participants had outliers in Satisfaction. Again, these were the same participants across all factors, so in total two participants with outliers were found in the experimental condition.

In conclusion, five participants out of 45 participants were considered outliers. That is  $\frac{1}{9}$ , or roughly 11% of the total sample size. This was deemed too great a number to remove the outliers since it would shrink the sample size and thus alter the results significantly [50]. By this reasoning, the outliers were not removed.

#### 4.1.2 General Overview

In table 4.1, a general overview of the predictor’s means and their respective standard deviation can be seen.

<b>Predictor</b>	<b>Control</b>	<b>Experiment</b>
Effort Expectancy	4.226 ± 0.795	4.373 ± 0.776
Perceived Enjoyment	3.261 ± 1.163	3.700 ± 1.080
Performance Expectancy	3.435 ± 1.122	3.580 ± 1.111
Social Influence	2.946 ± 0.906	2.966 ± 1.011
Trust	3.587 ± 1.029	3.591 ± 0.978
Satisfaction	3.587 ± 1.007	3.693 ± 1.097

Table 4.1: Overview of the means and standard deviations

For all predictors, the mean of the experimental condition was higher than the control condition. The table above can be seen in graphical form in figure 4.1.

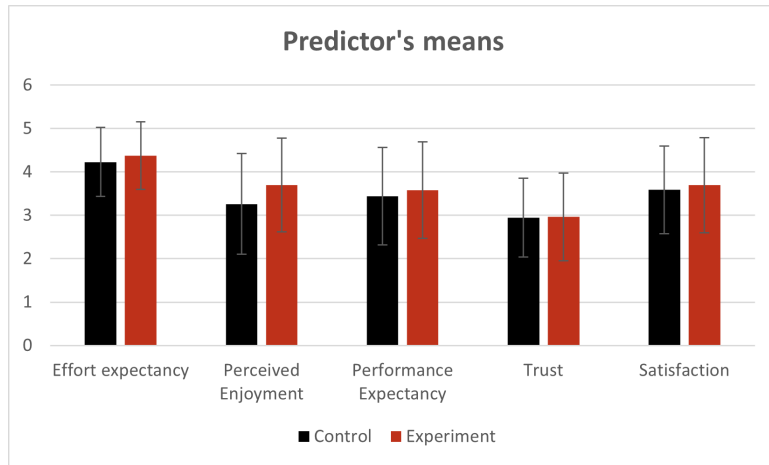


Figure 4.1: Predictor's means and standard deviations

The counts of the scores of each aspect can also be seen in graph form in figure 4.2.

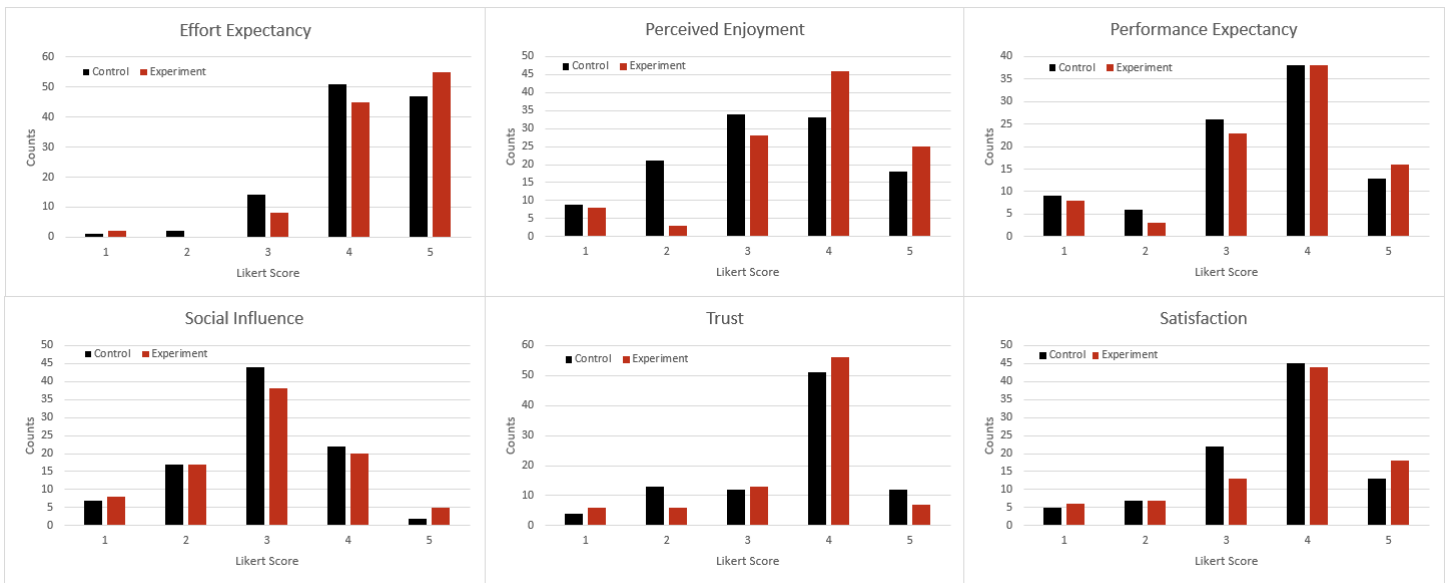


Figure 4.2: Likert Counts of the predictors

### 4.1.3 ANOVA Analysis

#### Assumptions

As mentioned before, hypotheses H1 through H5 had to be analyzed using one-way ANOVA. This meant analyzing Effort Expectancy, Perceived En-

joyment, Performance Expectancy, Social Influence, and Trust. The three ANOVA assumptions first had to be checked, otherwise, the analysis would not hold. The first assumption is independence. Since this was a between-subject design, and both the research conditions were randomly assigned to the participants, it can be assumed that the data is independent. Participants were also unknowing that there were two conditions, making this a blind study.

The second assumption is normality. The data from each predictor was evaluated for normality using the Shapiro-Wilk test. All predictors were found to have a p-value  $< 0.05$ , which meant the null hypothesis of Shapiro-Wilk was not accepted, and therefore the data was not normally distributed. However, as mentioned previously, ANOVA is robust, so an assumption may be violated and the analysis would still work [47].

The third assumption is that of equal variance. The Brown-Forsythe test was used to check if the population variances were equal. It used the median instead of the mean, which made it better suited for this study than for example Levene’s test due to the data not being normally distributed [51]. All factors were found to have a p-value  $> 0.05$ , which indicated the null hypothesis of Brown-Forsythe was accepted, and thus the variances were equal and the assumption was met.

## ANOVA Results

A one-way ANOVA analysis was performed on the first five concepts. The results of the test can be seen in table 4.2. The degrees of freedom are denoted with between-groups first, and within-groups second.

<b>Predictor</b>	<b>df</b>	<b>F</b>	<b>p</b>
Effort Expectancy	1, 223	1.956	0.163
Perceived Enjoyment	1, 223	8.595	0.004
Performance Expectancy	1, 178	0.756	0.386
Social Influence	1, 178	0.020	0.887
Trust	1, 178	0.001	0.979

Table 4.2: Overview of the predictor’s one-way ANOVA test

Table 4.2 displays that only Perceived Enjoyment has a  $p < 0.05$ . This meant ANOVA’s null hypothesis was rejected, and the alternative hypothesis, H2, was accepted; The peer persona of the chatbot was perceived as more enjoyable than the neutral persona of the chatbot. Hypotheses H1, H3, H4, and H5 were rejected.

#### 4.1.4 Pearson Correlation

In order to answer hypotheses H6 through H10, the correlation between all the aspects and Satisfaction had to be computed. The results are visible in table 4.3.

<b>Predictor</b>	<b>Pearson's r</b>	<b>p</b>
Effort Expectancy	0.282	0.0001
Perceived Enjoyment	0.695	< 0.00001
Performance Expectancy	0.695	< 0.00001
Social Influence	0.429	< 0.00001
Trust	0.543	< 0.00001

Table 4.3: Predictor's correlation with Satisfaction

The remaining hypotheses, H6 through H10, had to be accepted or rejected based on these values. H6, for example, was "The higher the Effort Expectancy is, the higher the Satisfaction is." The p-value must be  $< 0.05$  to be statistically significant and to be able to accept a hypothesis. Since all coefficients showed a  $p < 0.05$ , this meant the aspects were significantly correlated with the user's Satisfaction. It also meant that H6 through H10 were accepted.

Since all hypotheses had now been either accepted or rejected, the research model can now be shown with the hypotheses having either a red box or a green box around them. The red boxes represent rejected hypotheses while the green ones represent accepted hypotheses.

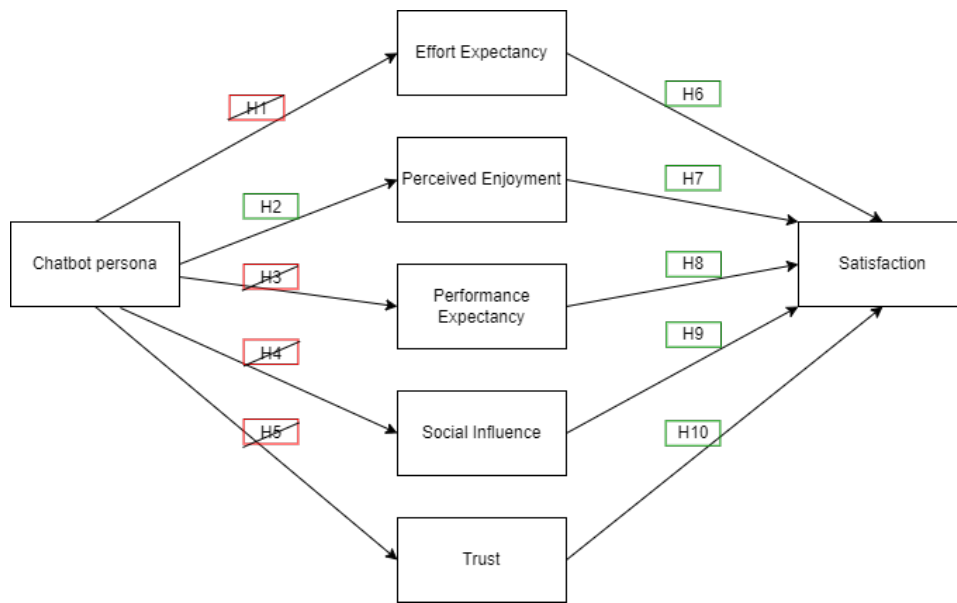


Figure 4.3: Research Model along with rejected and accepted hypotheses

## 4.2 Qualitative Analysis

The open questions shed some more light on how participants experienced the chatbot. Their experiences are summarized below, along with examples of what users commented.

### 4.2.1 Control group

- Participants often noted that the chatbot gave quick responses. They for example said “It was quick, gave elaborate answers,” “Quick responses,” “Quick answers.”
- The chatbot was also found to be clear, straight-to-the-point, efficient, concise, and informative by many participants. They illustrated this by saying “It provides clear and straight-to-the-point information,” “Clear answers”, “It is very informative”, “Clear use of language,” “Clear and concise information.”
- It was also commented that the chatbot gave relevant information about the topics that were discussed. One participant noted that “it looked like the chatbot knew a lot.” Others said “It gave me useful information, phrased in a way that’s easy to understand,” “It gave relevant and useful information about the topics.”
- The links given by the chatbot were found to be “very useful links to enhanced information about the topic” by one participant, while other

participants noted that they didn't trust the links the chatbot gave, as it had warned them about phishing beforehand. They said "warned me about not clicking links, but then suggested me links??" and "It is very controversial to give information about phishing, while at the same time sharing links with more information."

- Many participants noted that the chatbot felt "very impersonal," "Conversation didn't really feel personal, I had no connection to it," "It's not personal in its communication," "The fun was missing, the chatbot was very professional."
- It was found to be annoying that the chatbot sometimes didn't recognize an answer as a correct one. This also happened the other way around; some questions were marked as correct while they weren't. Users said "Some answers I gave were marked as correct even though they were not," "It did not reply correctly to the question: have you ever used public wifi before," "It did not understand me well," "First time using it, it went into a closed-loop, it kept on stating: I didn't understand please rephrase," "It didn't always understand the way I answered."

#### 4.2.2 Experiment group

- Similar to the control group, participants noted the chatbot gave quick, clear, useful information. They for example said "It was quick and gave clear answers," "It gave clear answers and simple explanations," "The chatbot was very clear with the answers it gave."
- The humor and emojis were found to be enjoyable by many participants. They said they liked "the humor :)" and "the smileys and human tone."
- It was also expressed frequently that the conversation felt natural, supple, casual, and pleasant to speak to. Users also expressed that the bot was easy to talk to and came across as friendly and helpful. This was illustrated by users commenting they liked that the bot was "very informal, it felt like an actual conversation. It understood 'ye, yuh', etc," "It was pleasant to speak to," "It came across as friendly and helpful," "I like the way the chatbot talks to me," "It was easy to talk to," "It responded quite positive about everything," "The conversation felt natural, not like a bot," "I liked that the chatbot was very casual in vocab and that I got accurate responses to my answers."
- Participants commented that the chatbot was a useful, easy and engaging way to learn about cyber security. They said "it gave useful

information about cyber security,” “It made learning about cyber security easy and fast,” “It was a strong method for learning because you are engaging more in the topic, than just reading the Wikipedia page.”

- Some participants found that the conversation still felt like a bot conversation. They said “it’s a bot and not a human,” “You can still feel that it’s a program. The chatbot is not always reacting to the answer of the person talking to it.”
- A few users said that the answer options were a bit limited. They commented that they didn’t like “the lack of answering possibilities aside from yes or no,” “The same answers over and over. Not much variation,” “The answers felt a bit generic.”
- Just like with the control condition, participants noted that sometimes the chatbot didn’t understand their answer. They said things like “It rated my answers as incorrect, while it wasn’t,” “It didn’t seem to handle complex sentences very well,” “When asked about the phishing email it did not understand my reply completely I think.”

This qualitative data could be used to support the results of the quantitative data and provided some insight into why some values were the way they were. This will be further elaborated upon in the discussion.

## Chapter 5

# Discussion

### 5.1 Interpretation

This study aimed to find out whether the user experience is higher when conversing with a chatbot that has a peer persona, than with a chatbot that has a neutral persona. The experiment condition was developed with this peer persona, while the control condition remained neutral. It was found that the persona of the chatbot did not have a significant effect on the Effort Expectancy of the chatbot. This indicates that the ease of use was relatively equal for both conditions, and a peer persona did not influence the way users handled the chatbot environment. This meant H1 had to be rejected.

Performance Expectancy was also found not to have a significant difference between the two chatbot conditions, indicating that participants did not show a preference for either one of the conditions concerning how useful they thought the chatbot was. This could have to do with the fact that both chatbots gave the same information about cyber security, just with a different tone of voice. This claim is supported by the comments made by participants since in both groups users noted the chatbot gave them useful and clear information. This led to H3 being rejected too.

Social Influence also did not have a significant difference between the experiment group and the control group, meaning H4 was rejected. Compared to the means of the other predictors, Social Influence had the lowest mean, with both the control-mean and experiment-mean falling below a value of 3. This might indicate that the usage of the chatbot is something relatively personal, and not influenced by other people around you. It is quite surprising that the mean of the experiment group was not significantly higher than the control group, since the experiment condition targeted students with the chatbot as its peer. It was expected that relating to the chatbot's personality would lead to the feeling that other students would advise a student to use the chatbot, but this was not the case. This might be indicative that talking to the chatbot did not come across as an actual human conversation. This is

supported by multiple participants in the experiment group who commented that the chatbot did not understand some of their simpler answers, with a few comments also emphasizing the conversation still felt like talking to a bot. These factors might have contributed to the lower scores given to the Social Influence factor. Improving the quality of the conversation as well as the anthropomorphism might lead to higher scores in this regard.

The final hypothesis that had to be rejected was H5, which measures the Trust aspect. It was unexpected that the experimental group did not have higher scores for this aspect than the control group, since the experiment condition was developed with one of the key concepts being trustworthiness, and making the user feel at ease. This was also supported by the comments made by participants. A few participants in the control group noted that they did not trust the chatbot because it told the participants to be careful with phishing emails and links but still offered them links with additional information. It is noteworthy that these comments were not made in the experiment group. The trustworthiness of the chatbot might be improved by the aforementioned aspects such as a more robust dialog and a bot that is (even more) anthropomorphic.

One hypothesis that had to do with the difference between the chatbot conditions was accepted, namely H2. This hypothesis stated that the peer condition leads to a higher Perceived Enjoyment than the control condition. The significant difference found between the means of the control- and experiment-group might have to do with the phrasing that the chatbot used. Participants in the experiment condition also noted the chatbot was friendly and helpful, while participants in the control group found the bot impersonal.

All other hypotheses, H6 through H10, were accepted. This was expected since these are based on the UTAUT model and confirm that the correlations inside the model are indeed accurate. In the next section, a few limitations of the conducted study are discussed.

## 5.2 Limitations

Firstly, this study was conducted during the COVID-19 pandemic. This meant that the development experienced more challenges than initially anticipated. The pandemic also meant that the testing of the chatbot, as well as the questionnaire, had to be done without any contact between the researcher and the participants. This meant the participants could not be monitored while doing the survey and therefore might take it less seriously or rush through it. If this study was to be repeated, it might be a good idea to have the participants be in the same room as the researcher to simulate better experimental conditions.

Second, it was noted by multiple participants that the answering options in

the dialog were a bit limited and that the chatbot didn't always understand what the user was trying to say. The chatbot was developed in a little over a month, which of course is an extremely short time, and could be one of the reasons the chatbot wasn't as robust as the researchers would have liked. The bot was tested beforehand by multiple people, but not enough to simulate all user inputs that one might give. If the testing phase had been more extensive, it might have improved the accuracy of the chatbot.

Furthermore, five participants with outliers were found during the data analysis. It was decided not to omit these participants from further analysis, as the sample size would shrink too much. However, when looking at the individual responses from these participants, it gives a strong impression that they most likely did not take the study seriously. Most of them answered "Strongly Disagree" to all questions, which doesn't match up with any answers from other participants. This data was kept, and thus influenced the result of the study. In the future, this could be circumvented, for example by being in the same room as the participants which might make them take the study more seriously, or by significantly increasing the number of participants so that outlier removal doesn't have such a big impact on the sample size.

Continuing on the sample size, increasing the number of participants could increase the reliability and statistical power of the study. This research only had 22 or 23 participants per condition and it is recommended that the amount at least be doubled to produce more reliable results.

Finally, it is noteworthy that the participants did not represent the general population, as everyone except one participant was Dutch. Most of them also had experience with using chatbots, which likely doesn't represent the entire population. As a result, it is difficult to generalize the findings of this study. A more diverse group of participants might produce a more general outcome.

### 5.3 Recommendations

Educational chatbots, as well as peer personas in chatbots, are still relatively new and largely unexplored. The research conducted could be expanded upon, and some directions future research may take will now be proposed. As discussed previously, it might be advisable to conduct a similar study with a more diverse group of participants by including people of other nationalities, ages, and varying experiences with technology. This study was focused on students, so it may be interesting to see what a similar study would result in, in other target audiences such as adults aged 30-65 or children aged 10-14. Of course, the chatbot's personality would have to be adjusted to align with the targeted audience. Most participants had also used chatbots before, so a group of users who have no experience whatso-

ever might react differently in a chatbot study.

Moreover, as mentioned in the previous section, a more robust dialog with more testing and a longer development cycle might enhance the robustness of the chatbot dialog and provide users with a better experience.

As far as use cases go, it might be worthwhile to see how an educational chatbot that converses with the user could be employed in schools and other learning institutions. Participants noted that the chatbot felt like an easy and accessible medium to learn about cyber security, so it might prove useful to apply and integrate such a chatbot in a course curriculum.

## Chapter 6

# Conclusion

In this thesis, a study was conducted about the effect of a peer persona on the user experience for an educational chatbot. The results of this study suggest that conversing with a chatbot that has a peer persona does not improve the overall user experience, however, it does increase the Perceived Enjoyment of the participants. Apart from the Perceived Enjoyment factor, no significant difference was found between a chatbot that had a neutral persona and a chatbot that had a peer persona. The data does suggest that the UTAUT model's factors indeed predict the satisfaction of users. Previous research suggests that talking to a peer increases feelings of relatability, so the outcome that the overall user experience was not higher in a peer persona was not expected. Possibly, the surprising outcome of the study was caused by the limited dialog ability of the chatbot, the relatively small number of participants, or outliers that were kept in the analysis. In the future, it is recommended that researchers develop a chatbot with a more robust and well-constructed dialog that can handle all types of answers. This type of chatbot can then be used in education to aid students with their learning.

## Chapter 7

# Acknowledgements

I would like to thank Marianna de Sá Siqueira for her supervision over my bachelor thesis research. Her assistance and insights were invaluable. I would also like to thank Spence van Asperdt, Andy Huang, Annemiek van der Leest and Serah Sommers for their collaboration with developing the chatbot and being my colleagues. I want to thank Martijn Vastenburg for being the second reader of my thesis. Finally, a thank you to all participants of my study.

# Bibliography

- [1] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 3506–3510, 2017.
- [2] Darius Zumstein and Sophie Hundertmark. Chatbots—an interactive technology for personalized communication, transactions and services. *IADIS International Journal on WWW/Internet*, 15(1), 2017.
- [3] Luke K Fryer, Mary Ainley, Andrew Thompson, Aaron Gibson, and Zelinda Sherlock. Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75:461–468, 2017.
- [4] Alan Mathison Turing. Mind. *Mind*, 59(236):433–460, 1950.
- [5] Tim Rietz, Ivo Benke, and Alexander Maedche. The impact of anthropomorphic and functional chatbot design features in enterprise collaboration systems on user acceptance. 2019.
- [6] Doris Läßle and Bradford L Barham. How do learning ability, advice from experts and peers shape decision making? *Journal of Behavioral and Experimental Economics*, 80:92–107, 2019.
- [7] Akihiro Yorita, Simon Egerton, Jodi Oakman, Carina Chan, and Naoyuki Kubota. Self-adapting chatbot personalities for better peer support. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 4094–4100. IEEE, 2019.
- [8] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166, 2003.
- [9] Nabie Y Conteh and Malcolm D Royer. The unprecedented rise in cybercrime and the role of the human vulnerability factor. In *Ethical Hacking Techniques and Countermeasures for Cybercrime Prevention*, pages 32–43. IGI Global, 2021.

- [10] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003.
- [11] Cheng-Min Chao. Factors determining the behavioral intention to use mobile learning: An application and extension of the utaut model. *Frontiers in psychology*, 10:1652, 2019.
- [12] Tony Wibowo. Study of user acceptance and satisfaction of a mandatory government-regulated information system. *CommIT (Communication and Information Technology) Journal*, 11(1):41–44, 2017.
- [13] Mohammed-Issa Riad Mousa Jaradat and Mamoun S Al Rababaa. Assessing key factor that influence on the acceptance of mobile commerce based on modified utaut. *International Journal of Business and Management*, 8(23):102, 2013.
- [14] Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 373–383. Springer, 2020.
- [15] Petter Bae Brandtzaeg and Asbjørn Følstad. Why people use chatbots. In *International conference on internet science*, pages 377–392. Springer, 2017.
- [16] Giuseppe D’Aniello, Angelo Gaeta, Matteo Gaeta, and Stefania Tomasiello. Self-regulated learning with approximate reasoning and situation awareness. *Journal of Ambient Intelligence and Humanized Computing*, 9(1):151–164, 2018.
- [17] 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.
- [18] Eric Hsiao-Kuang Wu, Chun-Han Lin, Yu-Yen Ou, Chen-Zhong Liu, Wei-Kai Wang, and Chi-Yun Chao. Advantages and constraints of a hybrid model k-12 e-learning assistant chatbot. *IEEE Access*, 8:77788–77801, 2020.
- [19] David Boud. Making the move to peer learning. *Peer learning in higher education: Learning from and with each other*, pages 1–21, 2001.
- [20] International Organization for Standardization. *Ergonomics of Human-system Interaction: Part 210: Human-centred Design for Interactive Systems*. ISO, 2010.

- [21] 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.
- [22] Andreas Chang. Utaut and utaut 2: A review and agenda for future research. *The Winners*, 13(2):10–114, 2012.
- [23] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.
- [24] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. Extrinsic and intrinsic motivation to use computers in the workplace 1. *Journal of applied social psychology*, 22(14):1111–1132, 1992.
- [25] Dong-Hee Shin. Towards an understanding of the consumer acceptance of mobile wallet. *Computers in Human Behavior*, 25(6):1343–1354, 2009.
- [26] Yoora Park, Hyojoo Son, and Changwan Kim. Investigating the determinants of construction professionals’ acceptance of web-based training: An extension of the technology acceptance model. *Automation in construction*, 22:377–386, 2012.
- [27] Ibrahim Arpaci. Understanding and predicting students’ intention to use mobile cloud storage services. *Computers in Human Behavior*, 58:150–157, 2016.
- [28] Ali Abdallah Alalwan, Yogesh K Dwivedi, and Nripendra P Rana. Factors influencing adoption of mobile banking by jordanian bank customers: Extending utaut2 with trust. *International Journal of Information Management*, 37(3):99–110, 2017.
- [29] William H DeLone, Ephraim R McLean, et al. Information systems success measurement. *Foundations and Trends® in Information Systems*, 2(1):1–116, 2016.
- [30] Michelle Chandler Diaz and Tina Loraas. Learning new uses of technology while on an audit engagement: Contextualizing general models to advance pragmatic understanding. *International Journal of Accounting Information Systems*, 11(1):61–77, 2010.
- [31] Jack Cahn. Chatbot: Architecture, design, & development. *University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science*, 2017.

- [32] Merriam Webster. Peer definition & meaning. <https://www.merriam-webster.com/dictionary/peer>, last accessed on 19/01/2022.
- [33] Hanne Riese, Akylina Samara, and Sølvi Lillejord. Peer relations in peer learning. *International Journal of Qualitative Studies in Education*, 25(5):601–624, 2012.
- [34] Tianran Hu, Han Guo, Hao Sun, Thuy-vy Nguyen, and Jiebo Luo. Spice up your chat: the intentions and sentiment effects of using emojis. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 2017.
- [35] María Guadalupe Veytia-Bucheli, José Gómez-Galán, and Diego Vergara. Presence of new forms of intercultural communication in higher education: Emojis and social interactions through whatsapp among graduate students. *Education Sciences*, 10(11):295, 2020.
- [36] Google. Blogger.com. <https://www.blogger.com>, last accessed on 26/01/2022.
- [37] Catherine A Wells. Location, location, location: The importance of placement of the chat request button. *Reference & User Services Quarterly*, 43(2):133–137, 2003.
- [38] Stephen R Porter, Michael E Whitcomb, and William H Weitzer. Multiple surveys of students and survey fatigue. *New directions for institutional research*, 2004(121):63–73, 2004.
- [39] Qualtrics. Qualtrics.com. <https://www.qualtrics.com>, last accessed on 26/01/2022.
- [40] European Commission. 2018 reform of eu data protection rules. [https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes\\_en.pdf](https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf), last accessed on 26/01/2022.
- [41] Subhasish Dasgupta, Mary Granger, and Nina McGarry. User acceptance of e-collaboration technology: An extension of the technology acceptance model. *Group Decision and Negotiation*, 11(2):87–100, 2002.
- [42] Ton AM Spil and Roel W Schuring. The utaut questionnaire items. In *E-Health Systems Diffusion and Use: The Innovation, the User and the Use IT Model*, pages 93–98. IGI Global, 2006.
- [43] Ricardo de Sena Abrahão, Stella Naomi Moriguchi, and Darly Fernando Andrade. Intention of adoption of mobile payment: An analysis in the light of the unified theory of acceptance and use of technology (utaut). *RAI Revista de Administração e Inovação*, 13(3):221–230, 2016.

- [44] Ruth Madigan, Tyron Louw, Marc Wilbrink, Anna Schieben, and Natasha Merat. What influences the decision to use automated public transport? using utaut to understand public acceptance of automated road transport systems. *Transportation research part F: traffic psychology and behaviour*, 50:55–64, 2017.
- [45] 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.
- [46] Hae-Young Kim. Analysis of variance (anova) comparing means of more than two groups. *Restorative dentistry & endodontics*, 39(1):74–77, 2014.
- [47] Geoff Norman. Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education*, 15(5):625–632, 2010.
- [48] Michael Cowles and Caroline Davis. On the origins of the .05 level of statistical significance. *American Psychologist*, 37(5):553, 1982.
- [49] Steven Walfish. A review of statistical outlier methods. *Pharmaceutical technology*, 30(11):82, 2006.
- [50] Herman Aguinis and Harry Joo. 10 debunking myths and urban legends about how to identify influential outliers. *More Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences*, 2014.
- [51] Brown-Forsythe Test. of variances (simulation).

# Appendix A

## Appendix

### A.1 Informed Consent Form

#### Information

##### Purpose of the Study

This study is conducted within the Bachelor Thesis of the Artificial Intelligence Programme of Radboud University.

The study involves an educational chatbot that will teach you about Cyber Security. Afterward, you will fill out a questionnaire.

##### 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 you want to quit. You can leave the study at any moment if you want.

##### Anonymity

All collected information and data are processed anonymously: the results cannot be traced back to you later on. Therefore, this also means that we 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: [sanne.janssen@ru.nl](mailto:sanne.janssen@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 voluntarily.
- I understand that I am allowed to stop at any time during this research study, should I wish to do so.
- 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.
- I consent to participate in this research study.

By clicking on 'I consent' below, you consent to participating in this study.

I consent

I do NOT consent

## A.2 Task Instructions

You are going to have a short conversation with a chatbot for about 5-10 minutes. The chatbot is an educational chatbot that aims to teach you about cyber security. Your task is to converse with the chatbot and fill in a questionnaire about the chatbot. **Please read the following instructions very carefully:**

- Please make sure you are running the chatbot in Google Chrome. Other browsers have not been tested.
- Use your computer or laptop for this session. Mobile phones and other devices are not supported.
- Please make sure you are in an environment that is distraction-free. Close any unnecessary program on your computer, interact with no other persons 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 unclear to you.
- Do NOT close the chatbot window during the conversation session.
- Open any links given during the conversation in a separate tab.
- As this is an English chatbot, it expects English responses 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 dialogue will otherwise shut down.

**Keep this page open and return to it once you are finished conversing with the chatbot.**

The link to the chatbot is:

<https://thesischatbot.blogspot.com> *for the experiment condition*

<https://chatbotthesis.blogspot.com> *for the control condition*

For any questions regarding the task, please contact the researcher at: [sanne.janssen@ru.nl](mailto:sanne.janssen@ru.nl)

## A.3 Questionnaire

### A.3.1 Check if the chatbot has been used

Please enter the password the chatbot gave you at the end of your conversation. \_\_\_\_\_

### A.3.2 Technology Experience

Have you ever used a chatbot before today?

- No, never
- I've used it once
- I've used it several times
- I've used it a lot

### A.3.3 UTAUT questions

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Neither agree nor disagree</b>	<b>Agree</b>	<b>Strongly agree</b>
Learning how to use the chatbot was easy for me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Interacting with the chatbot is clear and understandable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I find the chatbot easy to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It would be easy for me to become skilful at using the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the chatbot makes it easier to learn about cyber security	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.1: Effort Expectancy

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Neither agree nor disagree</b>	<b>Agree</b>	<b>Strongly agree</b>
I find using the chatbot enjoyable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is fun to talk to the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learning about cyber security from the chatbot was pleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chatbot is entertaining	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had fun using the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.2: Perceived Enjoyment

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Neither agree nor disagree</b>	<b>Agree</b>	<b>Strongly agree</b>
The chatbot is a useful way to learn about cyber security	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chatbot allows me to learn about cyber security more quickly than other methods	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the chatbot increases my productivity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the chatbot increases my chances of learning about cyber security	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.3: Performance Expectancy

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Neither agree nor disagree</b>	<b>Agree</b>	<b>Strongly agree</b>
People who influence my behaviour think I should use the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
People who are important to me would encourage me to use the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
People who influence my behaviour could assist me in using the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In general, people would support the use of the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.4: Social Influence

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Neither agree nor disagree</b>	<b>Agree</b>	<b>Strongly agree</b>
I believe that the chatbot is trustworthy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I trust the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not doubt the honesty of the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I trust the chatbot gives me reliable information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.5: Trust

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Neither agree nor disagree</b>	<b>Agree</b>	<b>Strongly agree</b>
I was very content while using the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was satisfied with the chatbot efficiency	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt delighted with the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, I was satisfied with the chatbot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.6: Satisfaction

### A.3.4 Open questions

What did you like about the chatbot? \_\_\_\_\_

What did you not like about the chatbot? \_\_\_\_\_

Are there any other things you would like to say about the chatbot? (optional) \_\_\_\_\_

### A.3.5 Demographic questions

What is your age? \_\_\_\_\_

What is your gender?

- Female
- Male
- Non-binary
- Prefer not to say

What is your nationality?

- Dutch
- German
- French
- English

Other: \_\_\_\_\_

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

- None
- Primary school
- Secondary school
- Higher education bachelor (HBO)
- Higher education master (HBO)
- University bachelor (WO)
- University master (WO)
- PhD

What is your work- or studyfield?

- Artificial Intelligence
- Computer Science
- Psychology
- Other: \_\_\_\_\_

### **A.3.6 End of survey**

Thank you for your time spent taking this survey, your response has been recorded.

For SurveyCircle users ([www.surveycircle.com](http://www.surveycircle.com)): The Survey Code is: 5YS7-V2U7-QE9P-MZX6