Bachelor thesis

Artificial Intelligence

Radboud University 

The influence of a matching personality of a chatbot and its user on the user experience.

Author:
Tamara Verbeek
S1021005

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

Second supervisor:
dr. R.J.L. Boumans
Behavioural Science Institute
roel.boumans@ru.nl

June 25, 2021

Abstract

Since the last decade, chatbots are used more and more in society. They already have a fascinating past that dates back to the 1950s when Alan Turing began to wonder whether machines could think. He created a test called The Turing Test and many people had this test as a foundation to create a chatbot. In the modern day, chatbots are upcoming as a replacement of humans in conversational services, such as a question and answering system or recommendation system. About 57% of companies are willing to implement a chatbot or have already implemented one in order to improve the efficiency of the conversational services and decrease the costs. However, users still have mixed feelings about chatbots and they find them often merely unintelligent or pointless. In this research there will be looked whether having a matching personality of the chatbot and its user might improve the user experience. The condition in which a user has a matching personality with the chatbot will be compared to the condition in which a chatbot does not have any personality. The user experience will be measured with the Technology Acceptance Model, and one additional concept Enjoyment. The results suggest that having a matching personality with the chatbot increases the enjoyment of the conversation with the chatbot. However, there have not been found any other significant difference between the two conditions. There has been found that the ease of use, usefulness and enjoyment all have a positive correlation with respect to the attitude toward using the chatbot. There are quite some limitations to this research, since the sample size is very small and the sample might not be very representative. This of course influences the results and might make the results not very reliable.

Contents

Chapter 1

Introduction

Since the last decade, the usage of chatbots has increased in society. Chatbots have a fascinating past that dates back to the 1950s. Alan Turing, a pioneering British computer scientist who was well ahead of his time, began to ponder whether machines could think and later on whether machines could be intelligent (Salecha, 2020). Turing did not have a specific definition of thinking, he did not think it was important to agree on a definition for it (Copeland, 2000). However, according to Turing, a machine is intelligent if it can impersonate a human and persuade the other person in a real-time conversation that he is communicating with a human rather than a machine (Salecha, 2020). His proposal as a way of dealing with the question if machines could think is based on this assumption and it is called "The Turing Test". The suggestion was a particular kind of test that could be applied to a machine. Turing explained the test in the following manner:

" The idea of the test is that the machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing. A considerable proportion of a jury, who should not be expert about machines, must be taken in by the pretence. They aren't allowed to see the machine itself – that would make it too easy. So the machine is kept in a far away room and the jury are allowed to ask it questions, which are transmitted through to it: it sends back a typewritten answer. ... [The questions can concern] anything." (Copeland, 2000, p. 524)

This test was the foundation of multiple chatbots created in the future. In the 1960s, MIT professor Joseph Weizenbaum developed the first ever chatbot called ELIZA (Salecha, 2020). Weizenbaum was particularly interested in Alan Turing's work and used it to create ELIZA. ELIZA attempted to carry out conversations with humans. The conversations were meant to simulate a conversation someone might have with a psychiatrist (Weizenbaum, 1983). Later models of ELIZA were designed to pick up new words as the conversation

progressed and use them to assist the person in solving word problems.

After the origination of ELIZA, multiple more chatbots were created. In 1988, the first chatbot that uses Artificial Intelligence was arisen called Jabberwacky. It had less limited knowledge than ELIZA and it responds with contextual pattern matching based on previous conversations (Adamopoulou & Moussiades, 2020). Just like Jabberwacky and Eliza, most chatbots at that time were initially build and used for fun (Shawar & Atwell, 2007).

In the modern day, chatbots are upcoming as a replacement of humans in conversational services. These chatbots can help customers with buying products, they can answer questions that customers have, or they can recommend a product to a customer. The current success of chatbots is most likely due to major developments in computing technology and widespread acceptance of mobile messaging apps. With the widespread use of personal computers, the desire to connect and the desire of their creators to have natural language interfaces has grown the need for conversational agents urgently (Shawar & Atwell, 2007). Research has shown that 57% of companies are willing to implement a chatbot in the near future or have already implemented a chatbot. Chatbots are projected to save $8 billion by 2022 (Smestad & Volden, 2018). Two key reasons for the adoption of chatbots inside businesses are given in a report published by Garter (2017). The first is that chatbots have the ability to handle customer progress more efficiently than humans by using powerful decision trees. The second advantage is cost savings, as chatbots can provide customer loyalty at a fraction of the cost of human customer service agents (Smestad & Volden, 2018).

A study has shown that the most frequently reported motivational factor driving chatbot use for users is productivity. In this study, a target group of chatbot users was conducted and the goal was to understand users' motivation to use a chatbot. A majority of the participants, 68%, emphasized how easy, fast, and convenient it is to use chatbots (Brandtzaeg & Følstad, 2017). However, there are still some mixed feelings about the chatbots. Some users find chatbots merely unintelligent or pointless and this has an effect on the user experience (Smestad & Volden, 2018). Misunderstood requests, irrelevant answers, and weak integration with human service agents cause users to be frustrated with chatbots (Shumanov & Johnson, 2021). In order to improve the user experience, a study explored whether human-computer experiences can be made more personalized by using the language of the conversation to suit user personality with congruent machine personality (Shumanov & Johnson, 2021). The study shows that user personality can be predicted during contextual interactions, and that the chatbots can be adapted to seem to have the same personality. The response language of the chatbot is then comparable to the language the user uses. The research concluded that for interactions involving social benefit, matching customer personality with congruent chatbot personality had a positive effect on consumer engagement with chatbots and buying outcomes(Shumanov & Johnson, 2021).

This study inspired me to explore whether matching user personality with the chatbot personality has a positive impact on the user experience. The research question is therefore: "What is the influence of a matching personality of a chatbot and its user on the user experience?". This research question will be answered by means of the Technology Acceptance model, created by Davis (Davis, 1985). This model explains how the behavior of a user of information technology can be predicted (Davis, 1985). In the model, the attitude of the user toward using the system is influenced by the ease of use and usefulness of the system. Besides these two concepts, I have decided to add one more concept since in this study the ease of use and usefulness are probably not the only two concepts influencing the attitude toward using. Therefore the concept enjoyment is added. Research by Koufaris et al. found that enjoyment has an impact on the attitude toward using the system (Koufaris, Kambil, & LaBarbera, 2001). The chatbot used in this research will be a song recommendation chatbot. The outcome of this research can benefit for example web shops that have a chatbot recommending a product. With the results they can decide to include a matching personality such that the user experience becomes better.

In the further chapters I will firstly give some background information about chatbots, personality, recommendation systems and the technology acceptance model. Thereafter, I will explain the research methodology, which explains what research model and materials I used, what the hypotheses are and what the design and procedure of the experiment are. After this, I will give you the results of the experiment and tell you something about the interpretation and description of the significance of my findings in the discussion. Lastly, I will end this research with a conclusion. After the conclusion there is an appendix with some parts explained in detail.

Chapter 2

Background

2.1 Chatbot

In essence, a chatbot is a program designed to mimic smart communication with a user on a text or spoken ground (McCrae & John, 1992). The chatbot recognizes the input of the user and will give an appropriate reaction. Most chatbots are implemented using pattern matching, this implies that the chatbot recognizes patterns in words or parts of a sentence and will give a response out of a database to them (McCrae & John, 1992). We can describe the design of a chatbot in figure 2.1.

Sansonnet et. al. provides a simple structure that describes the tasks that

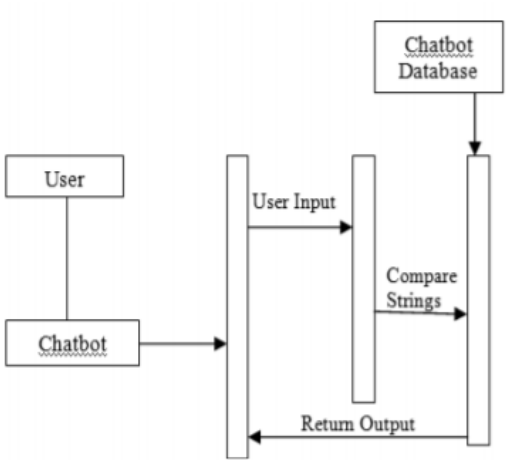

Figure 2.1: Sequence diagram representing the design of a chatbot (Dahiya, 2017)

modern chatbots are supposed to perform:

- *The Dialogical Agent*: this supplies the function of comprehension. The

agent should be able to lay hold of the problem of the user. The user should be able to enter textual or oral input that allows the user to give requests. These requests should then be interpreted by the chatbot using Natural Language Processing and converted into formal requests (Sansonnet, Leray, & Martin, 2006). Extended mixed-initiative interaction is possible with dialogue-based interfaces. After human collaborative problem solving, this approach models the human-machine relationship. Rather than seeing the interaction as a set of instructions, it includes describing and debating tasks, experimenting with various ways to complete the task, and collaborating to complete it (Allen et al., 2001).

- *The Rational Agent*: this supplies the function of competence. The agent must have access to an external expertise and common sense database. They can then answer the questions based on their database. Besides the database, the agent should also store information that is context-specific, such that it remembers what the user has said before (Sansonnet et al., 2006).

- *The Embodied Agent*: this gives the user the impression that it talks to a real human. To satisfy this condition, the agent should talk and act like a real human, for example with the usage of language tricks to create personas (Sansonnet et al., 2006).

2.1.1 Recommendation chatbot

Chatbots have multiple use cases. They can be used in customer support and engagement, sales and marketing or help employees (Bika, 2020). A recommendation chatbot provides information to their user according to their preferences (Cerezo, Kubelka, Robbes, & Bergel, 2019). So, for example, if a user tells the chatbot that they like a dress with lots of colours and flowers, then a recommendation chatbot could recommend a new piece of clothing based on that information.

Obtaining input in a conversational recommender system can be divided into two categories: navigation by asking and navigation by suggesting. In navigation by suggesting, the chatbot will propose recommendations to the user and determines based on the feedback whether the user likes the recommendation or not. After multiple propositions, the chatbot will get a good insight on what the user likes and dislikes and it can give a final recommendation. In the navigation by asking, the chatbot will ask the user questions and based on the answers of the user the system will recommend something. This framework proposes an ontology structure and investigates user preferences using question-and-answer dialogues similar to those between professional salespeople and customers.

2.2 Personality

In the past, personality theorists have offered hundreds of proposals and factor analysts sought for decades to bring order to the resulting uncertainty by factoring personality scales. Over the last 30 years, The Myers-Briggs Type Indicator or MBTI (Briggs, 1976) has become an extremely popular personality inventory. The MBTI is a self-report questionnaire developed by Katharina Cook Briggs and her daughter Isabel Briggs Myers to quantify non-psychopathological personality types based on Carl Jung's psychodynamic type theory (Boyle, 1995). Jung's theory explains that people's actions can be predicted to some extent due to their preferences for absorbing information, concentrating attention, making decisions, and organizing their lives (Myers & McCaulley, 1988).

The MBTI is about the central concept preference, which means something you feel most at ease and normal with (Bayne, 1997). The four main scales of the MBTI are (Myers & McCaulley, 1988):

- Extraversion and Introversion: Where does a person gets his energy from?
- Sensing and Intuition: How does a person acquire information and focuses attention?
- Thinking and Feeling: How does a person make decisions?
- Judging and Perceiving: How does a person deal with the external world?

With these four scales it can be decided what the personality of a person is. The four dimensions allow people to be represented by four letters that reflect their personality type (Furnham, 2020). So, the first letter can be an E or an I, standing for **E**xtraversion or **I**ntroversion. The second letter can be an S or N, standing for **S**ensing or i**N**tuition. The third letter can be an T or F, standing for **T**hinking or **F**eeling. The fourth letter can eventually be a J or P, standing for **J**udging or **P**erceiving. There are in total sixteen personality types, all consisting of four letters of the sixteen letter combinations. The 16 possible personalities are INTJ, INTP, ENTJ, ENTP, INFJ, INFP, ENFJ, ENFP, ISTJ, ISFJ, ESTJ, ESFJ, ISTP, ISFP, ESTP, and ESFP. For this research only the three most common personalities are used in experimental conditions, these are ISFJ, ESFJ and ISTJ. ISFJ is found in 13.8% of all persons, ESFJ in 12.3% and ISTJ in 11.6%(*How Rare is Your Myers Briggs Personality Type?*, 2021). These percentages can also be seen in figure 2.2.

In this research we are only interested in the top three personality types since it is too extensive to create a chatbot with 16 different personalities. Each of the three personalities can be explained by the following characteristics:

- *ISFJ*: The person who helps. At their peak, ISFJs are patient, compassionate, and loving individuals. They can be self-centered at their worst, forcing their own desires and thoughts onto others and being controlling. They are reserved, traditional, amicable, and responsible (Thor, 2019).

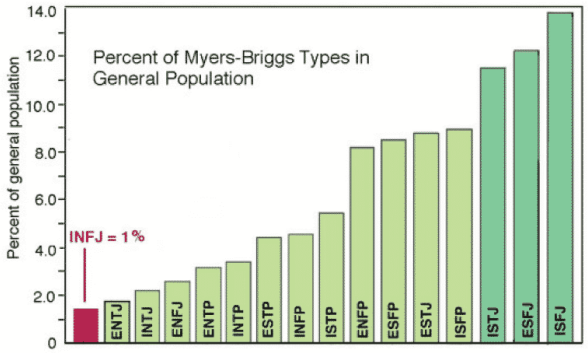

Figure 2.2: Most common personalities (Belyh, 2020)

- *ESFJ*: The Samaritian. In a flow state, ESFJs are positive and see the best in all. However, when they are stressed, they can experience feelings of being exploited or manipulated by others. They are gregarious, traditional, agreeable, and responsible (Thor, 2019).
- *ISTJ*: The organizer. This type is known for inspiring stability and confidence in a flow state. They have the potential to be pillars of their communities. They can act out in times of stress to 'shake off' people who are overly reliant on them. They are reserved, conventional, abrasive, and self-aware (Thor, 2019).

The four scales are used in the MBTI questionnaire, and based on the answers, people are classified as one of 16 personality types. The questionnaire has 90 questions and people need to pick one of the two choices.

2.2.1 Personality in a chatbot

People attribute humanlike characteristics to nonhuman entities, which is called anthropomorphism. Adults make anthropomorphic attributions to a broad variety of targets, including weather patterns, creatures, and moving geometrical figures, according to research by Zlotowski (Złotowski et al., 2018). This implies that humans also attribute humanlike characteristics to chatbots. Since humans also have a personality a person will also assign the chatbot a personality. It has been shown that people usually approach technology in a social way, but when they are specifically debriefed about it, they are hesitant to admit any anthropomorphic attributions (Złotowski et al., 2018). So, the process of anthropomorphizing happens implicitly, people do not recollect anthropomorphizing the nonhuman entity.

Despite their understanding that computers are not humans, people apply social laws, norms, and perceptions to their experiences with computers, and

treat computers as if they were "social actors," according to research focused on the Computers-Are-Social-Actors (CASA) paradigm (J.-E. R. Lee & Nass, 2010). People react to computers in the same way they react to other people, and these reactions can be caused when the computers show certain social cues. That is, if a machine looks, "talks," or acts like a human, people will react to it as though it were a real person, regardless of how minor these cues are (J.-E. R. Lee & Nass, 2010). This paradigm cannot be applied to every technology, the machine needs to have social cues and the machine needs to be presented as an autonomous source (Gambino, Fox, & Ratan, 2020). These requirements are met with respect to a chatbot.

2.2.2 Matching personality

Kaitlin and Graham did a study which purpose was to see if the discloser's emotion and cognitive mechanism terms, as well as interlocutors' language style matching, altered the reappraisal process required to feel better. Positive emotion words and language style matching influenced reported emotional improvement through the technique of cognitive reappraisal, according to the findings (Cannava & Bodie, 2017). There have also been studies that researched whether having personality similarities between spouses or friends has a positive influence on the relationship satisfaction. There have been, however, mixed results in studies that researched whether personality satisfaction is linked to genuine personality similarities between spouses. A few studies, such as the study of Gonzaga et al., have found that there were significant effects of personality similarity on relationships satisfaction (Gonzaga, Campos, & Bradbury, 2007), but others did not find any significant effect. It has also been shown that congruity in personality, and preferably in the personality domain of openness, leads to a more stable relationship (Rammstedt, Spinath, Richter, & Schupp, 2013). Also in friendships there has been research on personality similarities. Consensual validation is used to describe the relationship between attractiveness and similarity of attitudes (Byrne, 1961). To put it another way, an individual is perceived as seeking out people who share his viewpoints and thus provide some evidence for the correctness of his beliefs.

Looking at recommender systems, there has also been some research on what the influence of the attitude toward the system of the user is if the system uses personality information or not. When compared to systems that do not employ personality information, Hu and Pu (Hu & Pu, 2009) found that personality-based recommender systems are more effective in enhancing user loyalty and minimizing cognitive effort. These studies could explain why having a matching personality between the chatbot and its user would be beneficial for the user experience.

2.3 Recommendation System

Recommendation systems allow mass customization, which is becoming more and more common on the Internet. They are information filtering systems that resolve the problem of information overload by extracting critical information fragments from a vast volume of dynamically generated data based on the user's expectations, interests, or observed actions. Based on the user's profile, a recommender system will predict whether a specific user will like an object or not (Isinkaye, Folajimi, & Ojokoh, 2015). Search engines use recommendation systems to recommend relevant websites or documents based on the keywords that the user gives as an input. Webshops recommend products based on the purchases of other customers. Users' search effort is effectively minimized as a result of such customization. It also offers increased customer satisfaction, increased sales, increased advertisement profits, and the advantage of targeted promotions (Ansari, Essegaier, & Kohli, 2000).

There are multiple options when it comes to filtering out which product will be recommended to the user. Three options will be shortly explained. *Content-based filtering* is an algorithm that analyzes the attributes of items to create a prediction whether the user will like the items or not (Van Meteren & Van Someren, 2000). When it comes to recommending documents like web pages and magazines, the content-based filtering technique is the most effective. The content-based filtering technique allows suggestions based on user profiles and features derived from the content of objects the user has previously evaluated. The consumer is recommended items that are mainly connected to the positively rated items (Isinkaye et al., 2015). *Collaborative filtering* is an algorithm that cannot easily be described by attributes, such as a genre of a movie. So instead, it evaluates items based on opinions of other users (Schafer, Frankowski, Herlocker, & Sen, 2007). The algorithm makes suggestions by matching users with related interests and preferences based on profile similarities. Users that fall into this category form a neighborhood. A user receives recommendations for products that he has not yet rated but that have already received positive feedback from other users in his neighborhood (Isinkaye et al., 2015). *Hybrid filtering* combines both the collaborative and the content-based techniques such that it can optimize the prediction. Using more than one filtering techniques can avoid limitations of a single recommendation technique as another technique will support the system (Isinkaye et al., 2015). There are multiple options when it comes to hybrid filtering but we will only discuss weighted hybridization since it will be used it in this research. In weighted hybridization the algorithm adds up the scores from each source recommender to determine the score of a recommended item. If no source recommender has recommended an object, a score of 0 is assumed (Mueller, 2017).

2.4 Technology Acceptance Model

In 1986, the Technology Acceptance Model was first proposed by Fred D. Davis in his PhD thesis at the Massachusetts Institute of Technology (Davis, 1985). This theoretical model is developed to establish the user acceptance of computer-based information systems as a function of system characteristics. An overview of the original model can be seen in figure 2.3.

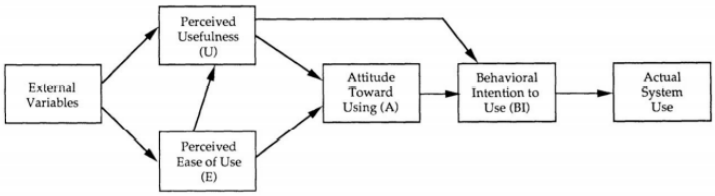

Figure 2.3: Technology Acceptance Model (Legris et al., 2003)

In the model can be seen that the actual system use is determined directly or indirectly on four variables: Perceived Usefulness, Perceived Ease of Use, Attitude Toward Using and Behavioral Intention to Use. Besides this, the variables Perceived Usefulness and Perceived Ease of Use are influenced by the External Variables, which then indirectly also influences the Attitude Toward Using and Behavioral Intention to Use. From the model above, three variables are used and the variable Enjoyment is added between the External Variables and Attitude Toward Using. The Technology Acceptance Model will be used in this research to measure the user experience, which is previously done by J. Portz et al. (Portz et al., 2019) and I. Fedorko et al (Fedorko, Bacik, & Gavurova, 2018).

External Variables The external variables indicate the characteristics of the system, which in this case are the characteristics of the chatbot.

Perceived Usefulness The perceived usefulness is defined as the user's subjective perception of a system's ability to improve job results while performing a task. A system should be viewed as beneficial in order to increase profit and benefits within a company.

Perceived Ease of Use The perceived ease of use is defined as the user's subjective perception of a system's effortlessness. To get the most out of the system, it should be simple and straightforward to use.

Attitude Towards Using The attitude towards using is defined as the attitude the user has towards the system.

2.5 Enjoyment

The enjoyment is defined as the user's degree of enjoying the system. The degree to which using a system is considered to be enjoyable in and of itself, regardless of any potential performance consequences, has been described as enjoyment (Hassanein & Head, 2007). Davis et al. identified enjoyment as an intrinsic motivator for using technology, while the TAM principles of perceived usefulness and ease of use are more extrinsic motivators. Enjoyment has an effect on a user's attitude toward using a device, making it a significant concept to test (Davis, 1985).

2.6 Research model and hypotheses

In this research the Technology Acceptance Model will be used. In combination with the additional concept Enjoyment the model in figure 2.4 is arisen.

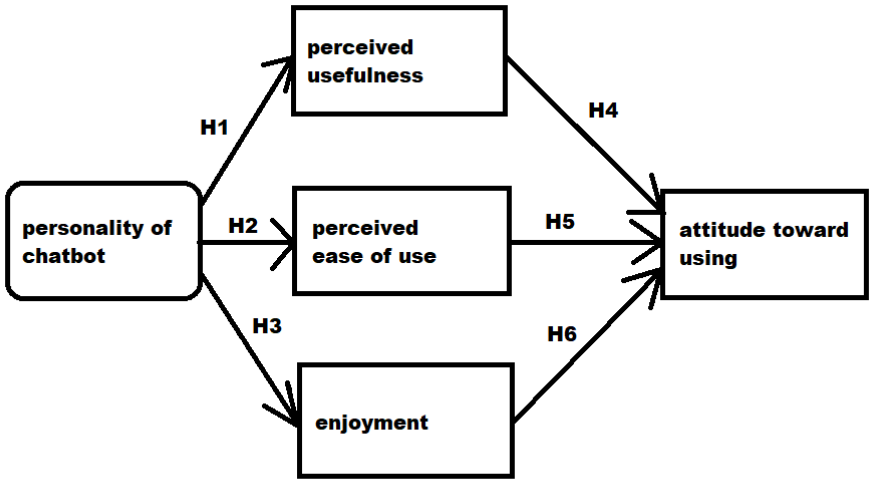

Figure 2.4: Research model

This research model results in six hypotheses:

- H1: A chatbot that has a matching personality with its user is perceived as more useful than a chatbot that does not have a matching personality with its user.
- H2: A chatbot that has a matching personality with its user is perceived as easier to use than a chatbot that does not have a matching personality with its user.

- H3: A chatbot that has a matching personality with its user is perceived as more enjoyable than a chatbot that does not have a matching personality with its user.
- H4: The higher the perceived ease of use of a chatbot, the higher people's attitude toward using it.
- H5: The higher the perceived enjoyment of a chatbot, the higher people's attitude toward using it.
- H6: The higher the perceived usefulness of a chatbot, the higher people's attitude toward using it.

Chapter 3

Research methodology

To find out what the answer to the research question should be, an experiment needed to be conducted. To run the experiment, I have created a chatbot that talks with the user about music. In the end of the conversation, the chatbot will recommend a song to the user.

3.1 Personality of the user

The chatbot can have multiple types of personalities. In the beginning of the conversation, the chatbot will ask 16 questions to determine which of the 16 possible personalities the participant has. After these 16 questions the participant will talk to the chatbot. The questions are based on the four dimensions of the MBTI, and each dimension has four statements which need to be answered by "strongly disagree", "disagree", "neutral", "agree" and "strongly agree". Example statements can be found in figure 3.1.

Introvert - Extrovert	You are generally sociable.
Sensors - Intuitives	You are creative and imaginative.
Feelers - Thinkers -	You use your head over your heart.
Judgers - Perceivers	You are organized and structured.

Figure 3.1: Example statement of each dimension of the MBTI

There is an alternation between the view of the statement. So, for example, some statements are from the Introvert view and others are from the Extrovert view. For each personality an ideal participant is made in the back-end. For

example, ISFJ has to "strongly agree" for the introvert statements and "strongly disagree" for the extrovert statements. Strongly disagree has a score of 0 and strongly agree a score of 4. In the end, the participant will then have a list of scores for each statement. This list of scores is then compared to the ideal participant for each personality. This is done with the mean squared error. The personality that has the smallest mean squared error with the list of scores of the participant is set as the personality of that participant.

3.2 Chatbot

3.2.1 Music database

For this chatbot, a dataset by the data scientist Yamac Eren Ay from kaggle is used (Ay, 2021). The data has been taken from Spotify, which is a music streaming service with more than 155 million subscribers worldwide (by Statista Research Department & 25, 2021). This dataset consists of 4 csv files from which 2 are used in this research. First we have the csv file data.csv, this file consists of audio features of all tracks. Second, we have the csv file data_by_artist.csv, which has the audio features of all artists.

3.2.2 Recommendation algorithm

This chatbot needs to recommend a song at the end of the conversation. As mentioned in section 2.3, there are multiple ways to create a recommendation system. For this research weighted hybridization was be used. The recommended song is based on scores. The participant needs to tell the chatbot which genre they like. If a song from the dataset is then of one of the preferred genres, 10 points will be added to the score. This is done because then the recommended song will then most likely be of a genre the user likes.

The user will also give a song they like during the conversation. Of this song, the accousticness, dancability, liveliness, energy, and year will be extracted. There will be looked through all songs in the database and if the difference between the value of the variables accousticness, dancability, liveliness or energy and the value of that variable of the liked song is 0.05 or smaller, 5 points will be added to the score. If the difference is bigger than 0.05 but smaller than 0.1, 2 points will be added and if the value is between 0.1 or 0.2, 1 point will be given. For the years, if the difference between a year of a song in the dataset and the year of the liked song is 5 years or smaller, 5 points are added. If the difference is between 5 and 10 years, 2 points are added and if the difference is between 10 and 15 years, 1 point is added to the score. In the end an amount of points will be added based on the popularity of each song, if the song is very popular a higher amount of points will be added then if the song is not popular. This is done because most people prefer popular songs over less popular songs.

The song with the highest score in the end will be recommended. If there are songs with the same highest score, then the song that is first found in the dataset will be recommended.

3.2.3 Dialogflow

The chatbot's conversation flow is implemented in Dialogflow. Dialogflow is an online platform where people can create agents that can have rich and intuitive customer conversations (Sabharwal & Agrawal, 2020). It uses machine learning to comprehend what the user is saying. The conversation is made up of *intents*. These intents consist of training sentences and a response the chatbot should give. The agent will determine the intention of the user based on the input of the user. If the input matches one of the training sentences of an intent, it will give the response given in that intent.

An example can be found in figure 3.2. In this figure, we can see that this intent is called elderly_age. The first box called *Contexts* contains the input context in which the intent should be called and the output context it will give as output. If another intent has the output context age, and the input of the user matches one of the training phrases, the response stated in the *Responses* box will be given.

In the training phrases can be seen that there are words marked yellow. Those yellow words are *entities*. An entity is a list of words that can be used in the same context. For example, in this case the entity consists of all ages that are considered old. This ensures that it is not necessary to enter the same sentence with one word different as a training phrase. The entities are the parameters of the training phrases. The parameters and action can be found in the box *Action and parameters* and those can be extracted using the back-end.

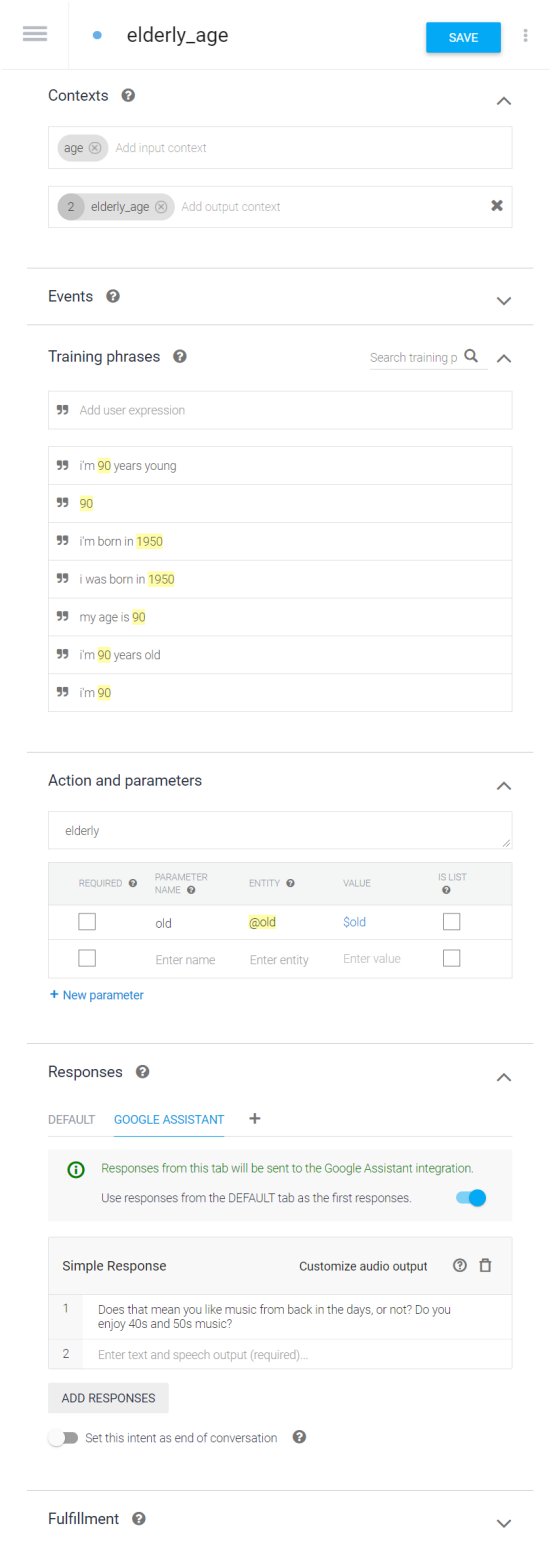

Figure 3.2: elderly_age intent in Dialogflow

3.2.4 Personality of the chatbot

The chatbot can have three different types of personality, and there is one control chatbot that does not have a certain personality. As been said before, they can have the personalities ISTJ, ESFJ, and ISFJ. The personalities are based on the characteristics of the three personality types from MBTI. Next to this, the communication styles of all three personalities have been researched. For example, the ISTJ personality is direct and straightforward, they will talk when it is necessary and when they see a practical need for it (Owens, n.d.). They can tend to be formal and zealous defenders of their social environment's rules and communication norms (*http://www.humanmetrics.com/personality/istj-communication-style*, n.d.). ESFJs, however, are practical, concrete, and at the same time helpful and receptive to the other party's needs. They are confident in their logic and want others to know about it. ESFJs communicate in a delicate but aggressive manner (*http://www.humanmetrics.com/personality/esfj-communication-style*, n.d.). ISFJs strive to be kind and friendly to others, and they are skilled at expressing this vocally. They make an effort to comprehend others and will go to great lengths to ensure that they are aware of everyone's requirements (Moodie, 2018).

3.2.5 Research conditions

There are four types of chatbots implemented, one control chatbot with no personality and three chatbots with the personalities ISFJ, ESFJ and ISTJ. In the figures below a difference can be seen between the four chatbots, the ISFJ chatbot (figure 3.3), the ESFJ chatbot (figure 3.4), the ISTJ chatbot (figure 3.5), and the control chatbot (figure 3.6).

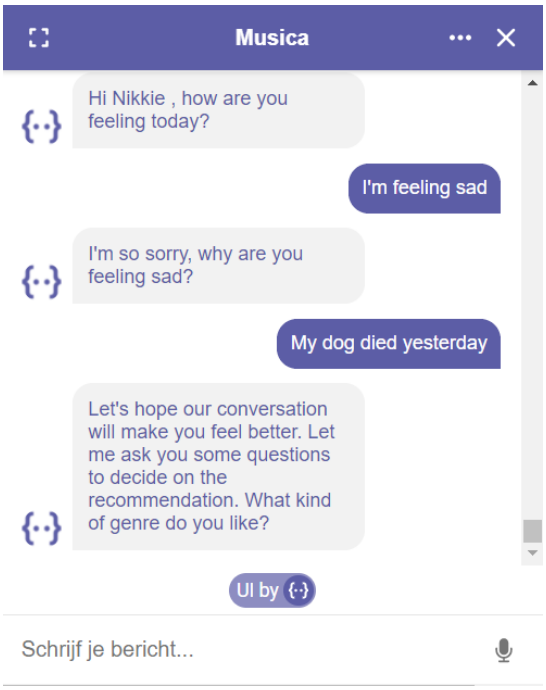

Figure 3.3: Conversation with a chatbot with ISFJ personality

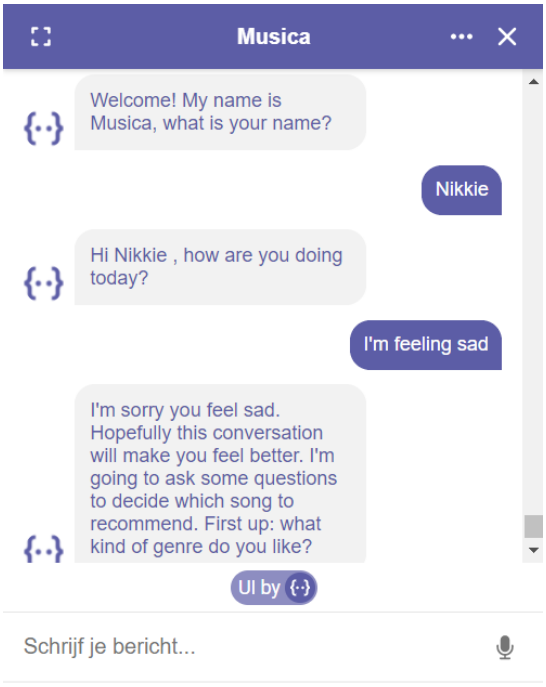

Figure 3.4: Conversation with a chatbot with ESFJ personality

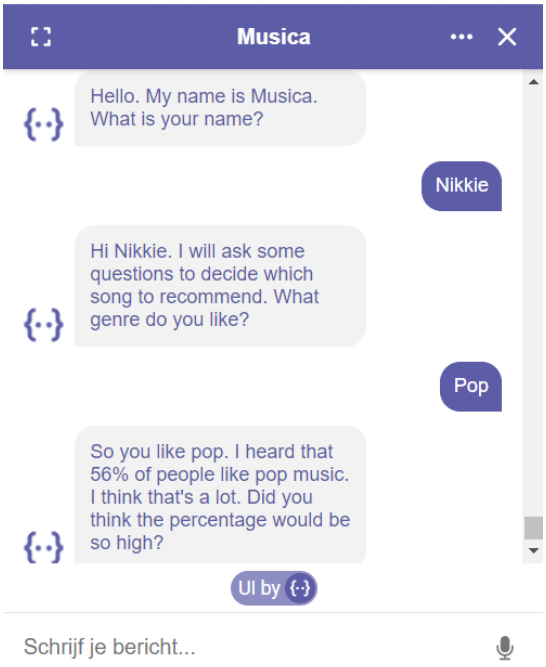

Figure 3.5: Conversation with a chatbot with ISTJ personality

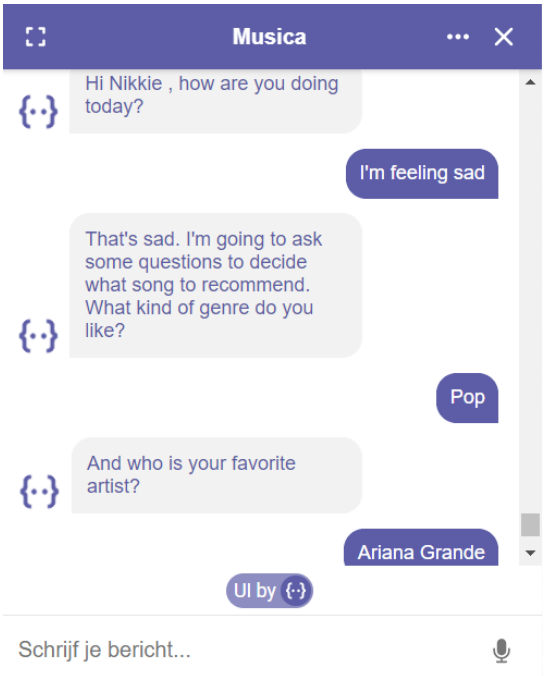

Figure 3.6: Conversation with a chatbot with no personality

In these conversations it can be seen that there is a difference in the way the chatbot communicates. The personalities ISFJ and ESFJ both like to please other people, they want to be helpful. In the conversation can be seen that they both express that they hope the conversation will make the participant feel better. The personality ISTJ on the other hand is quite insensitive, and is not interested in the feelings of the other person. In the control condition the chatbot will acknowledge the feelings of the participant but is not going deeply into it.

In general, the difference between extrovert and introvert can be seen the most. The extrovert will be more outgoing and enthusiastic. The extrovert uses for example exclamation marks to stress their enthusiasm, while the introvert is more reserved. A table with a few communication characteristics of each personality can be seen in figure 3.7.

ISTJ	ISFJ	ESFJ
Dislike small talk	Dislike small talk	Like to talk in general
Reserved	Reserved	Gregarious
Insensitive	Sensitive	Sensitive
Withdrawn	Withdrawn	Open
Practical	Talk about feelings	Talk about feelings
Straightforward	Want to be liked	Want to be liked

Figure 3.7: Personality characteristics in the conversation

3.3 Questionnaire

To measure the concepts representing the user experience, a questionnaire was used. The questionnaire was created with Qualtrics, which is an online questionnaire builder. The questionnaire begins with some demographic questions as what is the participants age, education etc. After this, questions based on the Technology Acceptance Model will be asked. The questionnaire has 15 statements that need to be answered on a 5-Point likert scale. Participants can give the answers "strongly agree", "agree", "neutral", "disagree" and "strongly disagree". After each 4 or 3 statements, an open question is present where participants can explain one of their answers. Example statements are given in figure 3.8.

Measured concept	Amount of statements	Example statement	Source
Ease of Use	4	The interaction with the chatbot was clear and understandable	F. Davis (Davis, 1989)
Usability	4	The chatbot would be useful in my life	F. Davis (Davis, 1989)
Enjoyment	4	It was enjoyable to share a conversation with the chatbot	Lee et. al. (J. Lee, Kim, & Choi, 2019)
Attitude toward Using	3	I have a generally favorable attitude towards the chatbot	M. Masrom (Masrom, 2007)

Figure 3.8: Example statement of each measured concept

3.4 Design

This experiment has a between subject design. To see whether a matching personality of the chatbot and its user has an influence on the user experience, three conditions are used. In the first condition, participants will have to talk to a chatbot with a matching personality. In the second condition, participants will have to talk to a chatbot with a different personality. In the control condition, participants will talk to a chatbot with no personality. The chatbot will ask some questions in the beginning of the conversation to determine the personality of the participant. If the personality of the participant is one of the implemented personalities of the chatbot, the participant will get the matching personality condition. If the participant has not got a personality that is one of the implemented personalities, the participant will get a fifty-fifty chance of getting the control condition or the condition with a non-matching personality.

3.5 Procedure

At the beginning of the experiment, the participant reads about the experiment on the website, what the terms and conditions are and what is expected from them. After this, the participant needs to click on the experiment tab of the website to start the experiment. First, they need to agree with the terms and conditions, and then they will be able to talk to the chatbot. In the conversation, the chatbot will first ask some questions to determine the personality of the participant. After the personality is determined by the back-end, the conversation with the correct condition will start. The participant will at the same time get a color code which they need to fill in in the questionnaire. This ensures that it is clear for the results which research condition the participant participated in. The participant will talk to the chatbot for a while about music and in the end the chatbot will give a recommendation based on the answers of the participant. After the conversation ends, the participant gets a password which can be filled in in the evaluation tab. This ensures that the participant completes the conversation with the chatbot. In the evaluation tab, the participant fills in the password and then it will bring them to the link for the evaluation. Here they fill in a questionnaire about their conversation and after

the questionnaire an explanation of the experiment can be found. Then the experiment is finished and the participant can close the tab.

Chapter 4

Results

4.1 Demographics

This experiment was done by mostly people from my social network. I asked friends and family to do the experiment online and a few of them also asked their friends to do the experiment. In the end there were 34 people who did the experiment. 10 of them had the matching personality condition, 13 had the non matching personality condition and 11 had the control condition.

Most participants were females, there were 20 female participants and 14 male participants. In the matching personality condition, 7 participants were female and 3 were male. In the non matching personality condition, 8 participants were male and 5 were female. In the control condition, 3 participants were male and 8 were female. The majority of females was probably due to the fact that most of my friends are females. The majority of people were between 19 and 25, but there were also quite some older participants. The oldest participant was 55. The distribution of all ages can be seen below in figure 4.1.

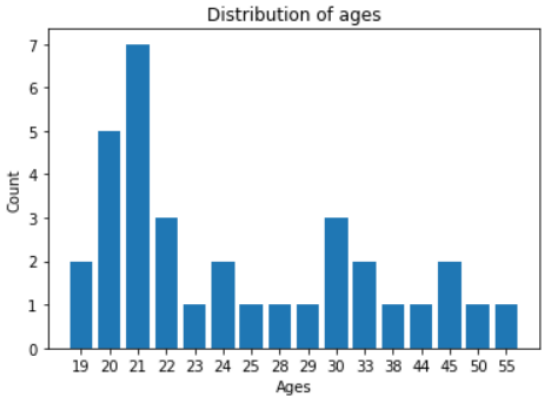

Figure 4.1: The distribution of ages

From all participants, 27 people had used a chatbot before, only 4 never used a chatbot before and 3 did not know whether they have used a chatbot before. Most people were highly educated, there were 18 people that have done academic education, and 10 people that have done applied university. The rest of the participants only did secondary school and 1 participant chose other as highest level of education. Most participants study or work in the Artificial Intelligence field or in an unspecified field. There were 2 participants who work or study in the computing science field, 2 participants who work or study in the psychology field and 1 participant who works or studies in the communication science field.

4.2 Quantitative results

In this research, the goal was to define whether having a matching personality with the chatbot has a positive impact on the user experience. The user experience was measured by 4 different variables: Ease of Use, Usefulness, Enjoyment and Attitude Toward Using. We can look at the difference between the matching personality condition and the control condition for these variables to determine what the influence of a matching personality has on the user experience.

Before we can do the analysis, we need to do some preprocessing. First, there is checked whether there are participants which answers can be considered as outliers. The z-score is calculated and no participants need to be removed. All z-scores are below 3 and above -3, which is seen as the boundary for a participant to be an outlier or not (Frost, 2021). After this, we will check whether the data is normally distributed. This is done by a test based on D'Agostino and Pearson's, which combines skew and kurtosis to create an all-encompassing normalcy test (Strangman, 2002). From this test we can conclude that our data is not normally distributed. This can be due to the fact that we have a quite a small dataset and the use of a 5-point likert scale. This, however, has no influence on the ANOVA test we are about to perform. We use ANOVA to see whether there is a significant difference between the control condition and the matching personality condition. The one-way ANOVA is regarded as a reliable test against the assumption of normality. This means it tolerates deviations from its normalcy assumption fairly well. As regards the normality of group data, the one-way ANOVA can tolerate data that is non-normal with only a small effect on the Type I error rate (*One-way ANOVA*, n.d.). The results of ANOVA can be interpreted in the following way. The P-value is the chance that means of two groups are significantly different. If the P-value is below 0.05, there is less than 5% chance that the results are random and therefore the means of the two groups are the same. The descriptive statistics of the conditions can be found in figure 4.2 and 4.3.

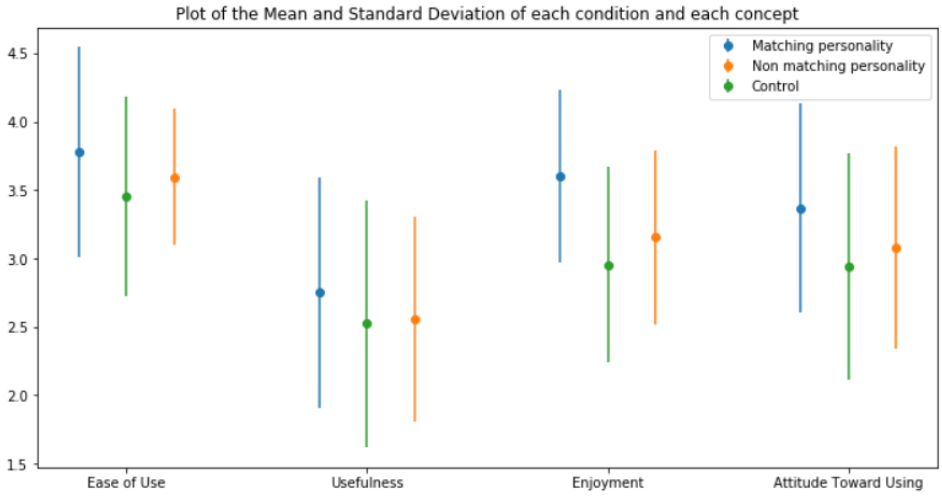

Figure 4.2: Plot of the mean and standard deviation of all conditions and concepts

Variable	Condition	n	Mean	Standard deviation
Enjoyment	Control	11	2.95455	0.713802
	Matching personality	10	3.6	0.632456
	Non matching personality	13	3.15385	0.638153
Ease of use	Control	11	3.45455	0.729911
	Matching personality	10	3.775	0.76753
	Non matching personality	13	3.59615	0.495454
Usefulness	Control	11	2.52273	0.901901
	Matching personality	10	2.75	0.839719
	Non matching personality	13	2.55769	0.75182
Attitude toward using	Control	11	2.93939	0.826869
	Matching personality	10	3.36667	0.76489
	Non matching personality	13	3.07692	0.739296

Figure 4.3: Descriptive statistics of all conditions and concepts

Let us first look at the results of ANOVA with only the experimental condition with a matching personality and the control condition. We can see that only the condition with the matching personality had a significant impact on the enjoyment of the user. We can namely see that the P-value is smaller than 0.05, and therefore we know that there is less than 5% probability that the results are random. The results can be seen in the table 4.4.

Variable	F statistic	P-value
Ease of use	3.845	0.053
Enjoyment	19.083	3.634e -5
Usefulness	1.420	0.237
Attitude Toward Using	4.505	0.038

Figure 4.4: ANOVA results of the experimental condition with a matching personality versus control condition

So, based on the results of ANOVA, we can conclude that we can reject H1 and H2, and we can accept H3.

To test whether we can reject or accept H4, H5 and H6, Pearson correlation coefficient was used. The results of this can be seen in the table 4.5.

	Enjoyment	Ease of Use	Usefulness	Attitude toward using
Enjoyment	1	0.5113855716177522	0.5643183583505933	0.603739
Ease of Use		1	0.4790673886742736	0.448581
Usefulness			1	0.790335
Attitude toward using				1

Figure 4.5: Correlation between all concepts

All correlations have a P-value below 0.05 and are therefore significant. In the table we can see that there is a strong positive correlation between usefulness and attitude toward using ($r = 0.790$), which makes us accept hypothesis H6, and between enjoyment and attitude toward using ($r = 0.604$), which also makes us accept hypothesis H5. The correlation between ease of use and attitude toward using is moderate ($r = 0.449$), so therefore we can accept hypothesis H4 as well.

Besides the analysis between the experimental condition with a matching personality and the control condition, there has also been researched whether having a random personality as chatbot has an influence on the user experience. The results in figure 4.6 concluded from the ANOVA one way analysis.

Variable	F statistic	P-value
Ease of use	1.268	0.263
Enjoyment	2.085	0.152
Usefulness	0.043	0.836
Attitude Toward Using	0.555	0.459

Figure 4.6: ANOVA results of the experimental condition with a random personality versus control condition

Here can be seen that no significant P-value and therefore can be concluded that the chatbot having a random personality does not have an influence on the user experience.

4.3 Qualitative results

In the questionnaire were also a few open ended questions for the participants to explain why they agree or disagree with a certain statement.

For the ease of use questions, most people claimed the chatbot had a clear dialog. However, there was also said that their answers were not always understood by the chatbot and it was therefore sometimes confusing what the chatbot actually wanted. This might have influenced the results.

In the part about enjoyment, it can be seen that people did think the conversation was more engaging when the chatbot had a matching personality. In the control condition the mean is 2.95, while in the matching personality condition the mean is 3.6. People therefore more often agreed with the fact that the chatbot was more enjoyable in the case of a matching personality. The control chatbot was sometimes said to be "just some sort of question and answering", people thought the chatbot just asked a plain question and the participant answered the question. They also thought that the chatbot did not really interact with the user or that the conversation was not engaging.

For usefulness, most people agreed that they did not know why they would need a song recommendation chatbot in their life. The mean of usefulness is in all conditions between 2.5 and 2.75, which meant that they mostly disagree with the fact that this chatbot would be useful in their life. People also mentioned that even though they liked the conversation, the conversation was too long for it to be efficient to find a new song.

Since people enjoyed the conversation more in the matching personality condition, we can see that the attitude toward using is a bit higher in the experimental condition compared to the control condition. The mean in the

matching personality condition is 3.37, while the mean in the control condition was only 2.94. People mentioned that it might be fun to have a chatbot that recommends songs but that it needed to be quicker and that for example spotify is way better in recommending a song.

Chapter 5

Discussion

In the beginning of this paper I have explained the problem that users find chatbots often merely unintelligent or pointless, and that this has a negative impact on the user experience. According to a study by Shumanov & Johnsen, a matching customer personality with congruent chatbot personality has a positive effect on customer engagement with chatbots (Shumanov & Johnson, 2021). In this research I applied this knowledge to a song recommendation chatbot in order to improve the user experience. The research questions was therefore: "What is the influence of a matching personality of a chatbot and its user on the user experience?". I created four different chatbots, one chatbot with no personality and three chatbots with all different personalities. Depending on the personality of the user, the user had to chat with the control chatbot, a chatbot with congruent personality or a chatbot with a different personality. Looking at the results, we can see that the mean of all variables is indeed higher in the condition of the congruent personality. The difference between experimental and control condition is, however, only significant with the variable enjoyment, all other variables did not have a significant difference. A matching personality with the chatbot therefore increased the enjoyment of the user.

Why a matching personality did not have any influence on the other variables could have different causes. First of all, the Myers-Briggs Type Indicator is not very reliable. In a major review study done by the National Academy of Sciences it appeared that just 47% of respondents kept their initial type designations after a 5-week period (Nowack, 1997). This disadvantage could have resulted in the participants to be assigned a wrong personality, and therefore the personality did not match the chatbot. Besides this, I have also decided to adapt the questionnaire to only have 16 questions, instead of the original 90 questions. This of course also influences the reliability of the MBTI and could result in an incorrect assignment of a personality. The MBTI also only represents stereotypes of personalities, it does not describe individuals and it puts you in a box that prevents you from combining your preferences (*https://www.teamtechnology.co.uk/myers-briggs-criticisms.html*, n.d.).

These disadvantages of the MBTI could have lead to an unreliable result.

In the Technology Acceptance Model was explained that both perceived usefulness and ease of use directly influence the attitude toward using. In the results can be seen that indeed perceived usefulness has a strong positive correlation ($r = 0.790$) with attitude toward using, and ease of use has a moderate positive correlation ($r = 0.449$) with attitude toward using. In the model is also mentioned that perceived ease of use is also directly influencing perceived usefulness, which is also supported by the results ($r = 0.479$). Enjoyment was cited by Davis et al. as an inherent motivator for system use (Davis, 1985). Looking at the results it can be said that enjoyment indeed did have a strong correlation with attitude toward using ($r = 0.604$)

This research has also some limitations. First of all, the chatbot was not perfect. It could not understand every variation of a response and it therefore asked a few times whether a person could rephrase their response in order for the chatbot to understand what the user meant. Next to this, sometimes when the chatbot did not understand the response, it just asked the question again. These issues of the conversation could have caused the results of this study no to be entirely reliable. A study by Folstad and Brandtzaeg researched which attributes of a chatbot lead to a better or worse user experience. About 11% of the participants mentioned that poor user experience often concerned pragmatic attributes (Følstad & Brandtzaeg, 2020). This mostly contained interpretation issues. The apparent difficulty of getting the chatbot to understand what the user was trying to tell was mentioned in a large proportion of participant reports (Følstad & Brandtzaeg, 2020).

Besides this, the sample size used for this research is quite small. An optimal sample size is about 10% of the population, as long as it does not exceed the 1000 (Bullen, 2021). This research is applicable to all persons above 18 speaking English, so the sample size should be at least 1000. This is of course not achievable in the time and resources available for this thesis research. This study only had a sample size of 34, so this is way too little for the results to be completely reliable. Secondly, there can be questioned whether the participants of this research are representative for the entire population. The population consists of all people from ages above 18, while this research mostly had participants between 19 and 25. Is also only consisted of males and females while there are also quite some nonbinary people. The results of this research could therefore be subject to bias, certain groups may be over-represented, resulting in their views being amplified, while others may be under-represented (Austin, 2014).

A last limitation is that the experiment is conducted in an uncontrolled environment. It could not be controlled whether the participants did the experiment serious and also filled in their actual findings in the evaluation form. Participants could have rushed the experiment or might have just filled in random things. These participants of course influence the results of the research.

Another reason for the results to be not significant could be that a matching personality is improving the user experience in other types of chatbots, but not in song recommendation chatbots. The conversation the participants had was not a in depth conversation. The questions to the participants were shallow and their answers were therefore also shallow. Research has shown that congruity in personality, and preferably in the personality domain of openness, lead to a more stable relationship(Rammstedt et al., 2013). This could mean that a more in depth conversation, for example a conversation with a therapist, could benefit more when it comes to having a matching personality.

Chapter 6

Conclusions

In this research I studied whether a matching personality between a chatbot and its user has a positive effect on the user experience. The findings suggest that a congruent personality between chatbot and its user only has a positive influence on the enjoyment of the user, but not on the other concepts of user experience. The research has shown that having a chatbot with the same personality as the user does not have any positive influence on the usefulness, ease of use or attitude toward using compared to a chatbot that does not have any personality. There also has not been found any significant influence on the user experience when the user had a conversation with a chatbot with a random personality. The findings are not entirely in line with previous work, there it has been shown that a congruent personality had a positive impact on consumer engagement and buying outcomes. The difference in conclusions could be caused by my choice of chatbot, the limited sample size, the interpretation issues of the chatbot or the uncontrolled environment. Future research should be done to determine whether the results can be applied in general.

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, 100006.

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Toward conversational human-computer interaction. *AI magazine*, *22*(4), 27–27.

Ansari, A., Essegaier, S., & Kohli, R. (2000). *Internet recommendation systems.* SAGE Publications Sage CA: Los Angeles, CA.

Austin, G. (2014, Jul). Retrieved from `https://austinresearch.co.uk/the-importance-of-representative-samples-and-how-to-get-them/`

Ay, Y. E. (2021, Apr). *Spotify dataset 1922-2021, 600k tracks.* Retrieved from `https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks?select=data_by_genres_o.csv`

Bayne, R. (1997). The myers-briggs type indicator: A critical review and practical guide.

Belyh, A. (2020, Sep). *Myers-briggs: This is the rarest personality type in the world.* Retrieved from `https://www.cleverism.com/myers-briggs-this-is-the-rarest-personality-type-in-the-world/`

Bika, N. (2020, Sep). *Chatbot use cases: What bots can do per industry and function.* Retrieved from `https://acquire.io/chatbot-use-cases/`

Boyle, G. J. (1995). Myers-briggs type indicator (mbti): some psychometric limitations. *Australian Psychologist*, *30*(1), 71–74.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In *International conference on internet science* (pp. 377–392).

Briggs, K. C. (1976). *Myers-briggs type indicator*. Consulting Psychologists Press Palo Alto, CA.

Bullen, P. B. (2021, May). *How to choose a sample size (for the statistically challenged).* Retrieved from `https://www.tools4dev.org/resources/how-to-choose-a-sample-size/#:~:text=A\%20good\%20maximum\%20sample\%20size,\%2C\%2010\%25\%20would\%20be\%2020\%2C000.`

Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology*, *62*(3), 713.

by Statista Research Department, P., & 25, M. (2021, Mar). *Spotify users - subscribers in 2020.* Retrieved from https://www.statista.com/statistics/244995/number-of-paying-spotify-subscribers/

Cannava, K., & Bodie, G. D. (2017). Language use and style matching in supportive conversations between strangers and friends. *Journal of Social and Personal Relationships*, *34*(4), 467–485.

Cerezo, J., Kubelka, J., Robbes, R., & Bergel, A. (2019). Building an expert recommender chatbot. In *2019 ieee/acm 1st international workshop on bots in software engineering (botse)* (pp. 59–63).

Copeland, B. J. (2000). The turing test. *Minds and Machines*, *10*(4), 519–539.

Dahiya, M. (2017, 05). A tool of conversation: Chatbot. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING*, *5*, 158-161.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340.

Fedorko, I., Bacik, R., & Gavurova, B. (2018). Technology acceptance model in e-commerce segment. *Management & Marketing*, *13*(4), 1242–1256.

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience*, *5*(1), 1–14.

Frost, J. (2021, May). *5 ways to find outliers in your data.* Retrieved from https://statisticsbyjim.com/basics/outliers/

Furnham, A. (2020). Myers-briggs type indicator (mbti). *Encyclopedia of personality and individual differences*, 3059–3062.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger casa: extending the computers are social actors paradigm. *Human-Machine Communication*, *1*(1), 5.

Gonzaga, G. C., Campos, B., & Bradbury, T. (2007). Similarity, convergence, and relationship satisfaction in dating and married couples. *Journal of personality and social psychology*, *93*(1), 34.

Hassanein, K., & Head, M. (2007). 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.

How rare is your myers briggs personality type? (2021, Apr). Retrieved from https://personalitymax.com/personality-types/population-gender/

(n.d.). Retrieved from https://www.teamtechnology.co.uk/myers-briggs-criticisms.html

(n.d.). Human Metrics. Retrieved from http://www.humanmetrics.com/personality/esfj-communication-style

(n.d.). Human Metrics. Retrieved from http://www.humanmetrics.com/personality/istj-communication-style

Hu, R., & Pu, P. (2009). Acceptance issues of personality-based recommender systems. In *Proceedings of the third acm conference on recommender systems* (pp. 221–224).

Isinkaye, F. O., Folajimi, Y., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261–273.

Koufaris, M., Kambil, A., & LaBarbera, P. A. (2001). Consumer behavior in web-based commerce: an empirical study. *International journal of electronic commerce*, *6*(2), 115–138.

Lee, J., Kim, J., & Choi, J. Y. (2019). The adoption of virtual reality devices: The technology acceptance model integrating enjoyment, social interaction, and strength of the social ties. *Telematics and Informatics*, *39*, 37-48. Retrieved from https://www.sciencedirect.com/science/article/pii/S0736585318306415 doi: https://doi.org/10.1016/j.tele.2018.12.006

Lee, J.-E. R., & Nass, C. I. (2010). Trust in computers: The computers-are-social-actors (casa) paradigm and trustworthiness perception in human-computer communication. In *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives* (pp. 1–15). IGI Global.

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? a critical review of the technology acceptance model. *Information & management*, *40*(3), 191–204.

Masrom, M. (2007). Technology acceptance model and e-learning. *Technology*, *21*(24), 81.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175–215.

Moodie, K. (2018, Sep). *What your natural communication skills are like, based on your personality type.* Retrieved from https://personalitygrowth.com/what-your-natural-communication-skills-are-like-based-on-your-personality-type/

Mueller, J. (2017). Combining aspects of genetic algorithms with weighted recommender hybridization. In *Proceedings of the 19th international conference on information integration and web-based applications & services* (pp. 13–22).

Myers, I. B., & McCaulley, M. H. (1988). *Myers-briggs type indicator: Mbti.* Consulting Psychologists Press Palo Alto.

Nowack, K. (1997). *Personality inventories: The next generation.* Retrieved from http://www.opd.net/abstracts5.html

One-way anova. (n.d.). Laerd Statistics. Retrieved from https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-3.php

Owens, M. (n.d.). *Istj relationships & compatibility with other personality types.* Retrieved from https://www.truity.com/personality-type/ISTJ/relationships#:~:text=ISTJs\%20communicate\%20in\%20a\%20direct,to\%20move\%20a\%20task\%20along.

Portz, J. D., Bayliss, E. A., Bull, S., Boxer, R. S., Bekelman, D. B., Gleason, K., & Czaja, S. (2019). Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. *Journal of medical Internet research*, *21*(4), e11604.

Rammstedt, B., Spinath, F. M., Richter, D., & Schupp, J. (2013). Partnership longevity and personality congruence in couples. *Personality and Individual Differences*, *54*(7), 832–835.

Sabharwal, N., & Agrawal, A. (2020). Introduction to google dialogflow. In *Cognitive virtual assistants using google dialogflow* (pp. 13–54). Springer.

Salecha, M. (2020, Jun). *Story of eliza, the first chatbot developed in 1966.* Retrieved from https://analyticsindiamag.com/story-eliza-first-chatbot-developed-1966/

Sansonnet, J.-P., Leray, D., & Martin, J.-C. (2006). Architecture of a framework for generic assisting conversational agents. In *International workshop on intelligent virtual agents* (pp. 145–156).

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web* (pp. 291–324). Springer.

Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful? In *Ldv forum* (Vol. 22, pp. 29–49).

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, *117*, 106627.

Smestad, T. L., & Volden, F. (2018). Chatbot personalities matters. In *International conference on internet science* (pp. 170–181).

Strangman, G. (2002). *scipy.stats.normaltest.* Retrieved from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html

Thor, E. (2019). The big 5/ocean model and the 16 personalities. Retrieved from https://www.erikthor.com/2019/10/10/connections-between-big-5-and-the-mbti/

Van Meteren, R., & Van Someren, M. (2000). Using content-based filtering for recommendation. In *Proceedings of the machine learning in the new information age: MLnet/ecml2000 workshop* (Vol. 30, pp. 47–56).

Weizenbaum, J. (1983). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *26*(1), 23–28.

Złotowski, J., Sumioka, H., Eyssel, F., Nishio, S., Bartneck, C., & Ishiguro, H. (2018). Model of dual anthropomorphism: the relationship between

the media equation effect and implicit anthropomorphism. *International Journal of Social Robotics*, *10*(5), 701–714.

Appendix A

Appendix

A.1 Personality questions

1. You are generally sociable
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
2. You get energy by spending time alone
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
3. You talk a lot and like to start conversations
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
4. You have a few deep friendships instead of many friends
    A. Strongly disagree
    B. Disagree

C. Neutral
D. Agree
E. Strongly agree

5. You pay attention to the details
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
6. You focus on what is possible (in the future)
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
7. You like to make things
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
8. You are creative and imaginative
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
9. You use your head over your heart
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree

10. You are interested in people and their emotions
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
11. You treat people according to their situation
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
12. You make decisions based on logic
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
13. You are organized and structured
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
14. You prefer to go with the flow instead of making plans
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree
15. You like to be in control of your life
    A. Strongly disagree
    B. Disagree

C. Neutral
D. Agree
E. Strongly agree

16. You are able to change and adapt quickly
    A. Strongly disagree
    B. Disagree
    C. Neutral
    D. Agree
    E. Strongly agree

A.2 User experience questionnaire

Statement	strongly disagree	disagree	neutral	agree	strongly agree
The interaction with the chatbot was clear and understandable	○	○	○	○	○
Interaction with the chatbot does not require a lot of mental effort	○	○	○	○	○
I find the chatbot easy to use	○	○	○	○	○
Learning how to use the chatbot was easy for me	○	○	○	○	○

Why do you think the chatbot was (not) easy to use?

Statement	strongly disagree	disagree	neutral	agree	strongly agree
Is was enjoyable to share a conversation with the chatbot	○	○	○	○	○
The conversation with the chatbot was exciting	○	○	○	○	○
The conversation with the chatbot was interesting	○	○	○	○	○
It was pleasant to share a conversation with the chatbot	○	○	○	○	○

Why do you (not) enjoy the conversation with the chatbot?

Statement	strongly disagree	disagree	neutral	agree	strongly agree
By using the chatbot I can decide more quickly which song to listen to than in the past	○	○	○	○	○
By using the chatbot I can better decide which song to listen to than in the past	○	○	○	○	○
The chatbot would be useful in my life	○	○	○	○	○
Using the chatbot would enhance the effectiveness of finding a new song	○	○	○	○	○

Why do you think the chatbot would (not) be useful in your life?

Statement	strongly disagree	disagree	neutral	agree	strongly agree
I have a generally favorable attitude towards the chatbot	○	○	○	○	○
I would like to use the chatbot again in the future to find new songs	○	○	○	○	○
I believe it's a good idea to use the chatbot to find a new song	○	○	○	○	○

Why do you think it is (not) a good idea to use the chatbot to find a new song?

A.3 Chatbot dialog

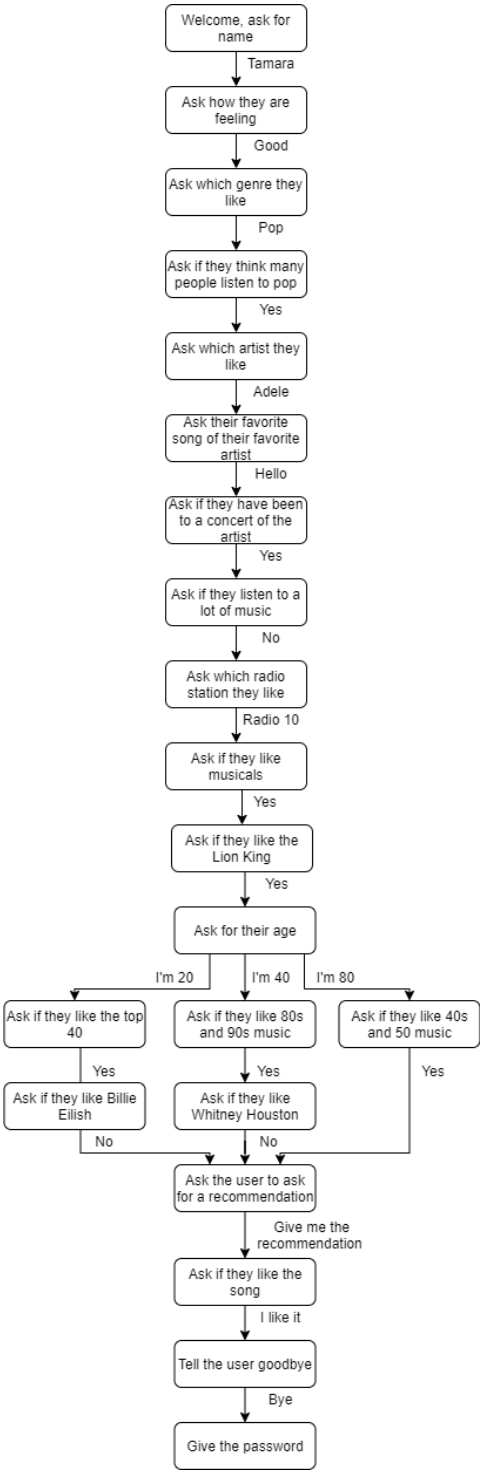

Figure A.1: Dialog