

Traveling with a chatbot

How to adapt chatbot personality to different traveler types

Sven Neuteboom
Utrecht University
Utrecht, The Netherlands
s.y.neuteboom@students.uu.nl

Abstract

Train travel is a popular way of transport in the Netherlands. Multiple tools exist for travelers to help them plan their journey and to look up information about their trip. However, these tools do not account much for the personality and behavior of the traveler, which might have a substantial impact on their preferences and needs. This thesis will focus on the different preferences and needs of different types of travelers when providing input through a chatbot meant for Dutch Railways, such as questions and complaints. The chatbot is meant to adapt based on the type of traveler using it. A Wizard of Oz experiment was conducted to see if traveler type has an impact on the preferences towards chatbot behavior. Some patterns could be observed, especially among the most outspoken traveler types, but further research with more participants is necessary to draw any decisive conclusions.

1. Introduction

Every day in the Netherlands, 1,3 million travels are made on average using the trains and services of the Dutch Railways [38]. Within this huge number of people using the railway system all over the country, there are lots of different kinds of travelers. Some commuters use the train to get to their work every day, while there are also people who mainly use train transport for one-time outings. Besides those, there are also travelers from abroad who might use the public transport system to visit friends or family. These different kinds of travelers might differ in regard of purpose of travel, travel frequency, or the route they take, but also on personality aspects. Some travelers tend to keep to themselves during their journey, while others like to strike up a conversation with their fellow passengers. All these different kinds of people have different needs and might require different information about their travels. A commuter who takes the same route to work five days a week is likely less interested in knowing the departure times- or platform of their train but might benefit more from getting updates about delays or other changes. A tourist, on the other hand, might need way more information about basic procedures, like buying tickets, knowing where to go, and finding the

right train. However, not only the content of the information could differ based on the traveler, but also their preference regarding the way this content is presented.

As of now, to get information about your trip as a traveler, the Dutch Railways offers an app which allows the user to plan their journey, shows the traveler when- and where to go, and informs them of potential delays or roadworks, among other features. Users can also add their regular routes and in turn get a notification in case of a disturbance along their journey. All these features help the travelers by making their journey as smooth and easy as possible. However, the app currently does only account for the ‘hard’ characteristics of the travelers, i.e., characteristics such as travel frequency, ticket usage, business/private purpose of travel, etc. There are more characteristics besides those, which can give a good indication of the behavior and personality of the travelers, which in turn can offer insight into their wishes and needs of how this information is presented. These so-called ‘soft’ characteristics can help to develop and broaden the current understanding of travelers further based on passengers’

intrinsic motives, needs, motivations and values [45].

Besides looking up information, travelers might also have other reasons for interacting with the Dutch Railways, such as filing complaints, or providing suggestions. An increasingly common way deployed by companies to handle these kinds of inputs is the use of chatbots [25]. A chatbot is a program which can simulate human conversation [30, 17]. This can be used to provide quick responses to input of the traveler, such as questions or suggestions. This thesis will focus on the preferences of different kinds of travelers when providing input to the Dutch Railways through such a chatbot. These different kinds of travelers will be segmented based on certain ‘soft’ characteristics. Input can mean any kind of information the traveler provides to the Dutch Railways, such as questions, complaints, and improvement suggestions. Later in this thesis, four concrete types of input which will be focused upon will be described and explained. Since, as mentioned earlier, chatbots are used increasingly more often [2], this thesis will also investigate how willing travelers are to use a chatbot, and what benefits this has compared to other methods, such as writing an email.

To be able to explore these topics, it is necessary to understand the current state of chatbots, to see what works and what does not, and to see how the perceived ‘personality’ of a chatbot influences the user experience. This will be explained in section 2. In this chapter, this thesis will also look at the different kinds of travelers, and how they differ from one another. After this, in chapter 3, the research questions and sub questions will be presented, along with hypotheses. In section 4 and 5, the two methods for answering those questions will be explained and described. Thereafter, in section 6, the results of those methods will be presented, which will be discussed in chapter 7. Chapter 8 will address the limitations of this research, chapter 9 will mention future works, and this thesis will be concluded with chapter 10.

2. Literature Review

In this section existing relevant literature will be covered. The first section is about chatbots, their favorable characteristics and attributes according to users, and their perceived ‘personality’. The personality of a chatbot can be categorized using certain facets and compared to the user’s personality. This section will then check to see if having similar personalities from both the chatbot and the user enhances the user experience by explaining the Similarity-attraction theory. After the section about chatbots, the different kinds of travelers will be explored, their differences highlighted, and a connection will be made between the attributes of the different travelers and the different attributes a chatbot can have. Lastly, the different kinds of inputs the traveler can submit which will be focused upon will be explained, to help give a better idea of the scope and application of the chatbot.

2.1 Chatbots

As mentioned before, chatbots are already used by many companies, in varying degrees of intensity [2], and nearly 80% of today's businesses use or expect to implement chatbots to connect with users 24/7 and solve their issues [25]. Chatbots allow for easy and fast communication from the user towards their desired entity, although they often still also have many shortcomings.

According to the dictionary, a chatbot is “A computer program designed to simulate conversation with human users, especially over the Internet” [31]. It uses Natural Language Processing to communicate in human language by text or oral speech with humans or other chatbots [17]. The big advantage chatbots have and the reason they have become more common is because they reduce service costs and can handle many customers simultaneously [37].

Chatbots have not been as common as they are now for long: in 2017, 84% of Internet users had not used a chatbot yet [13]. Although these numbers are different nowadays, this can still be a challenge, since this could mean that there still

are a lot of users not familiar with the use of a chatbot, and may struggle to interact with one, especially in the beginning. This means that dealing with a chatbot should be as intuitive and natural as possible. Users interacting with chatbots all have a clear goal: talking to one should feel like talking to a fellow human [6, 22]. Developers are moving from app-first design – where each app comes with its own interface, thus incurring a small learning curve – to a chatbot-first model, which uses the already familiar messaging interface [12]. This means that the user does not have to grow accustomed to using the interface of the system. They might have difficulty conveying whatever they want to the chatbot, though, as those are not on human levels yet.

Different research has been done on the favorable characteristics and attributes of chatbots. It appeared that users prefer chatbots which provide either a ‘human-like’ natural language conversation ability, or an engaging experience that exploited the familiar turn-based messaging interface [15]. Chatbot designers should ensure that chatbots understand and sustain conversation context, provide a clear and ongoing indication of the chatbot’s capabilities, engage in small talk, indicate when the chatbot fails to perform a task, and end a conversation gracefully. Users also appreciated a chatbot which can add value to their life by making useful recommendations, endowed with a sense of humor, while being reassuring, empathetic and non-judgmental [27]. Liao et al. [20, 21] studied the deployment of a Human Resource (HR) chatbot in a workplace environment. Apart from functional usage, they found participants getting involved in playful interactions with the chatbot, which are rich signals to infer user satisfaction with the chatbot. These are all just characteristics and attributes of a chatbot generally desired by the user. But can these traits altogether create a ‘personality’ for the chatbot? When people interact with other humans, personalities have a big impact on building connections and forming relationships [25]. When a human user is interacting with, and forming their perception of, a chatbot, they might deduce personality traits from its language and response style [39]. In

commercial contexts, a human representative’s personality can be very important to enhance the user experience [46], and the same is true for chatbots [39]. This is why chatbots need to have traits which correspond with the user’s preferred personality, as it can help enhance the likeability and overall experience of the chatbot.

2.2 *Categorizing Chatbot Personality*

Personality is often used to describe individual differences associated with attitudes, needs, feelings, and behaviors [4, 28]. Many studies about chatbot personality use the well-established five-factor model (FFM), which is considered the most extensive categorization in personality research and incorporates the critical components of personality [26]. The most widely adopted FFM is known as the Big Five, and consists of five characteristics: agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience.

Agreeableness refers to a person’s tendency to be tolerant, compassionate, courteous, and cooperative toward others. *Conscientiousness* is the degree in which a person acts in a planned, organized, or thoughtful manner. *Extraversion* means the extent to how sociable and talkative one is. *Neuroticism* refers to the degree to which a person’s emotions are sensitive to the environment, and includes being nervous, insecure or anxious. Finally, *openness* is the extent to which a person is open to experiencing a variety of activities, is broad minded and independent.

Not every characteristic of the Big Five is as relevant to the evaluation of chatbots. In an article by Von der Pütten et al. (2010) [46], the authors investigated how participant personality affected both their behavior when interacting with a chatbot and their evaluation of said chatbot. They found gender and age did not affect the evaluation, but some personality traits were predictive, including agreeableness and extraversion. Agreeableness had a positive impact on how participants perceived the interaction, and extraversion impacted the participant’s verbal behavior, in particular the number of words they used. This is also

supported in multiple other publications, which suggests that the language use of people in their everyday conversations provides insights into their personality and is stable over time [31, 32, 40]. Out of the Big Five, the focus of this thesis will be on the extraversion-introversion spectrum. The first reason for this is that extraversion, together with agreeableness, impacted the way people interact with chatbots [46]. The second reason is that extraversion – which can be manifested in outgoing, talkative behavior [44] – can be altered easily for chatbots and will make for a noticeable difference. As mentioned before, a higher word count can indicate extraversion, but also the use of less complex words, and social and emotive language is associated with extraversion. Reason for this is that people high in extraversion are more likely to be talkative and expressive and less concerned about using precise language [9, 27, 31].

2.3 Similarity-attraction Theory

Knowing about the difference in language use between users can be useful, but it is not enough to be able to apply the theory and increase the overall experience of the user. It is important to know how to model the chatbot to match with the personality of the user. Does an extraverted user prefer the chatbot's personality to be the same as their own? Research about the similarity-attraction theory focuses mainly on understanding the role of personality and similarity of attitudes to explain why people are attracted to others. According to the similarity-attraction theory, people prefer matching personality types because they like interacting with others who are similar to themselves [5]. This also extends to consumer interactions with computers. As mentioned before, users create a personality for a computer based on certain cues, verbal or non-verbal, in the interface [30]. Users also demonstrate a preference for language congruent to their own personality [30]. Similarly, congruence between user personality and technology interaction style resulted in more positive evaluations of technology [40]. These results suggest that technology responsive to user personality is more likely to

result in higher rates of acceptance and adoption by users.

2.3 Different kinds of travelers

Now that we know that the acceptance and adoption of chatbots can depend on the personality of both the chatbot and the user, it is also important to classify the personality of the traveler, in order to match them up with a compatible chatbot. The most obvious way to categorize different kinds of train travelers would be to split them up based on 'hard' characteristics, such as age and gender, traveling habits (traveling for work or for leisure), and special conditions, such as people with (temporary) disabilities. Although looking at such characteristics already gives valuable information about different train travelers, it is not enough to fully encompass the personality and preferences of the individual travelers. There are other characteristics which also should be considered in order to fully understand and adapt to the personality of the user. The focus of this thesis regarding chatbot personality will be on extraversion. To be able to test the similarity-attraction theory, it is necessary to also categorize the measure of extraversion in travelers. In the article by van Hagen [45], the author tried to find an unambiguous segmentation instrument that would be able to classify different kinds of travelers into different groups, taking into account soft characteristics as well. By better understanding the different target groups, Dutch Railways wanted to

“attune existing and novel products and services as optimally as possible to the wishes and needs of the train passengers' market, as well as gaining an effective and efficient communicative approach of segmented target groups through the correct choice of communication channels, message content and tone-of-voice.” [43, p. 3].

This basically boils down to being able to personalize products, services and ways of communication based on the personality of the traveler.

Ultimately, the segmentation instrument chosen was the NeedScope segmentation that was conducted by TNS NIPO. NeedScope focuses on hidden needs and underlying motivations. It divides consumers' needs into three categories:

- The most accessible category accounts for the functional needs, such as safety, speed and ease. These needs should be fulfilled by the product characteristics of a brand.
- The social identity needs are somewhat more difficult to interpret. People want to belong to a group, and brands are often a manifestation of the social group to which one belongs or wishes to belong.
- The core of all consumer behavior is formed, however, by emotional needs, such as the need for safety or the need for control. This core corresponds with the symbolic meaning of brands, and it is these deep-rooted yet fundamental needs which are difficult to fathom.

NeedScope checks each level of needs and indicates the degree to which a brand does or does not meet the needs.

The research consisted of both a qualitative- and a quantitative phase. In the qualitative phase, interviews were conducted with train passengers to identify their different needs. During the interviews, the researchers identified a difference between the needs with regard to the "travel process" and those with regard to how time was spent, i.e., during the wait at the station or whilst on the train. In the end, a total

of six segments were traced based on the needs of different train travelers. Where a traveler falls within these six segments is mainly determined based on how intro- or extroverted they are, as well as whether they are more individually- or socially minded. How the travelers interact with their fellow passengers plays an important role in what segment they would belong to. The six different segments will be covered below. Besides the hidden needs, the quantitative phase of the research also focused on the more 'hard' characteristics, so that a link could be made between this new form of segmentation and other already existing forms of segmentation.

The six different types of train travelers according to the NeedScope segmentation are summarized in Figure 1 below and will be covered in more detail in Table 1 on the following page.

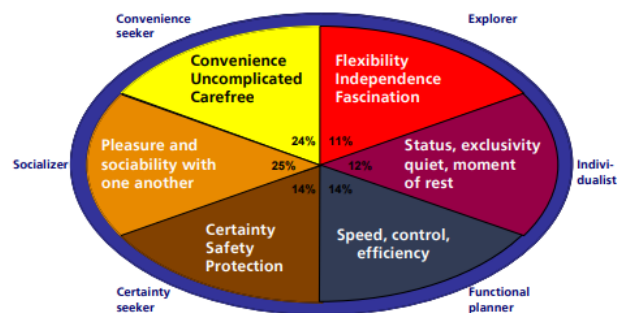


Figure 1: The six different types of train travelers (from van Hagen, 2009)

The Explorer	The Explorer is a young(er), well-educated and employed traveler who enjoys observing their fellow passengers and sees train rides as a source of inspiration. They prepare the moments before departure, if at all, and encountering new experiences and adventures is more important to them than possible dangers, delays, or poor quality facilities like toilets.
The Individualist	The Individualist is an older, often no longer employed traveler, who tends to keep to themselves. When on the train they like to work, read a book or listen to music. They prefer a quiet train, and also travel in first class most of the time. Comfort and luxury is very important to them, which they are willing to pay for. They plan their journey way in advance and always have travel information on hand.
The Functional Planner	The Functional Planner is a young(er), employed traveler who likes to have everything under control and be as organized as possible. Traveling to them is purely functional, and nothing more. This is why they prefer to get from A to B as quickly and efficient as possible. They travel primarily for work or business purposes. They plan their journeys thoroughly, but because they have a lot of knowledge about the routes, this is often not needed.
The Certainty Seeker	The Certainty Seeker is a traveler who is friendly, open and sociable, who wants to feel comfortable, care-free, relaxed, secure and valued. They don't hesitate to ask other passengers for help or advice about their trip when needed. They primarily travel for pleasure, and plan their trip carefully and well in advance.
The Socializer	The Socializer is cheerful, friendly, positive and spontaneous. They enjoy meeting and talking to other passengers, and quickly feel at home on the train. The purpose of their travel is often recreational, and they often travel outside of peak hours. They plan their journey well in advance, and always have travel information by hand, although they also like to ask other travelers for help.
The Convenience Seeker	Lastly, the Convenience Seeker is carefree, relaxed, easy-going, spontaneous and positive. The aspect of convenience is the most important factor for them in choosing to travel by train. They plan their trip at the last moment, or even during their journey. They travel mainly to school or use the train for vacation or trips. They want as much information as needed to help them with their journey, also during the trip.

Table 1: Descriptions of the six types of train travelers (adapted from van Hagen, 2009)

The NeedScope segmentation is not the only segmentation method which can be used to categorize different travelers. There are other publications which also grouped travelers into different groups based on their preferences during traveling, such as the segment profiles developed by Anable [3]. However, the segmentation method by van Hagen is one of the few which focuses solely on train travelers, while the one by Anable focuses on all kinds of travelers, with a focus on car travel. Since this thesis is about train travel only, the segmentation by van Hagen was deemed the more suitable method for this research. Besides just focusing on train travelers, the NeedScope segmentation also places a high importance on extraversion. As mentioned in section 2.2, extraversion is the characteristic which will be focused upon for this paper, so on this aspect van Hagen's segmentation method also suits this thesis very well.

2.5 Different Inputs

These six types of travelers all have different needs and preferences. These preferences might differ in terms of what information is given, and when, but also in how information is presented. This research will focus on the differences between the six groups, and how a system can be adaptive to each user. This will be based not just on 'hard' characteristics, but more on their personality and how that affects the desired feedback from the system. To do so, the final system needs to know in what segment its user belongs, and then act accordingly. To find out what 'accordingly' is, different interviews and tests will need to be conducted to gather information from travelers of each group and find out their preferences and tastes.

As mentioned before, this research will mainly be focusing on the input provided by different kinds of travelers. Reason for this is that this suits the purpose of chatbots well, since they

need- and are designed to deal with input. This input can come in many forms, and for this paper, four concrete types of inputs will be considered. The first three types are complaints, questions and suggestions. The last type of input considered in this research is the inquiry of recommendations. These recommendations can be things like suggestions on where to eat, tips on what to do at the place of destination or giving information about the sights along the route. The first step to get to an adaptive system based on both the 'hard'- and 'soft' characteristics of the passenger is to get a good view of the kinds of input different kinds of travelers like to give.

3. Research Questions

This section will present both research questions, along with the sub questions and their hypotheses. The research questions this thesis will try to answer are:

RQ1: How willing are different kinds of train travelers to use a chatbot to provide input to Dutch Railways instead of the traditional methods?

RQ2: How can a chatbot used to communicate with Dutch Railways be personalized to accommodate different kinds of train travelers and improve the overall user experience?

RQ1 focuses mainly on the attitude of travelers towards using chatbots, while RQ2 focuses on the personalization of information depending on the type of traveler. To create some more structure, both RQ1 & RQ2 will be divided into sub-questions, which will help answer the main research questions.

3.1 Research Question 1:

The first research sub-question of RQ1 is:

RQ1.1: What do travelers consider advantages of using a chatbot over traditional methods?

This sub-question of RQ1 is meant to find advantages about using chatbots, and especially how travelers interpret these advantages; how important and relevant are these benefits to them in the context of public transport?

Expected is that the ease of sending input will be mentioned often, as well as the quick response time. Filing a complaint, for example, now requires the traveler to actively search for the right page on the Dutch Railways website. [18]. Having the option to do all this through a system which is already used for other related matters, such as planning a journey, might lower the threshold for people provide input. Another expected benefit is the quick response time of a chatbot. Especially for inputs where a fast reply is desirable, such as asking questions, a chatbot has an advantage over other method such as mailing, or reaching out through social media, where the response could take a while [10].

RQ1.2: What do travelers consider disadvantages of using a chatbot over traditional methods?

This sub-question is meant to find the drawbacks of using a chatbot over other similar methods. Technical limitations of chatbots, such as having difficulty understanding the user, will not be considered here. This research assumes a perfectly functioning chatbot and is more focused on the potential acceptance and functionality without considering technical issues. Some expected disadvantages are the fact that typing is required. This might not always be the quicker than other methods, such as calling, or using user interface with buttons.

RQ1.3: What do people find most important when wanting to communicate with Dutch Railways?

In other words: do the advantages outweigh the disadvantages according to the travelers? When having a good idea of the benefits and drawbacks of using a chatbot to provide input, it is important to find out whether the travelers prefer this method, or that the potential disadvantages cause most travelers to still favor the other ways to connect with the Dutch Railways.

The hypothesis for this sub-question is that this will depend heavily on the traveler and their preferences, as well as the type of input the traveler wants to provide. It is expected that

younger participants will be more open to the idea of using a chatbot for different kinds of tasks one might want to perform while traveling, since younger people are often more open to learning new methods and technologies [8]. Older people, however, tend to be more fearful and hesitant towards using new technologies and applications [19, 23]. Regarding the different types of input, it is expected that inputs that require immediate feedback, such as asking questions or asking for recommendations, will be more suitable for chatbot usage. One of the advantages chatbots have over other available methods is the quick response time, which will likely be appreciated most when fast feedback is desired.

3.2 Research Question 2:

For RQ2, the first sub-question is:

RQ2.1: To what extent does the someone's traveler type have impact on their preferences regarding providing input to a chatbot?

This sub-question is formulated to find patterns between traveler type and the travelers' preferences. Finding these patterns can help the personalization of the chatbot, since it could check the user's traveler type, and adapt based on this. This makes the system more effective, by knowing which personality traits to focus and place emphasis on.

The hypothesis is that the more extraverted traveler types, such as the Socializer, will also prefer a more 'extraverted' chatbot. This could mean that the chatbot uses a higher word count, makes use of exclamation marks or emojis to convey enthusiasm, or uses any other speech mannerisms in order to feel more extravert. On the other hand, it is expected that the more introverted traveler types, such as the Individualist, will prefer a less outspoken chatbot personality, with a more to-the-point way of speaking. Not every traveler type is as clearly divided regarding being extraverted or introverted, so it is expected that there will also be some differences in preferences within certain groups. Regarding hard characteristics, it is expected age will be an important factor as

well. Older users generally have more difficulty using new technologies and systems [11], and thus might have more trouble using chatbots, rating them lower overall. Someone's age also influences their use of language. Older people tend to use more complex words [35], and thus might prefer the chatbot doing the same, while younger users will likely prefer the opposite.

Finally, it is also expected that the way different traveler types prefer Dutch Railways personnel to behave, will correspond to the way they prefer the chatbot to behave, such as more respectful, or more social.

RQ2.2: To what extent do the preferences of travelers within the same group differ?

This thesis has mentioned so far that traveler's preferences about chatbot behavior will be compared between different traveler types. However, does that determine everything about people's tastes? Not every traveler group has the same, very clear-cut distinctions as to what makes a person a part of this group. It is important to also compare different travelers within the same group to see if there are differences in preferences between them, and what kind of preferences these are. This helps to prevent generalizing the travelers within the same group, which might lead to some users getting a chatbot which does not match with their preferences.

The hypothesis for this sub-question is that while there will probably be many similarities between the travelers within the same group, there will still be slight differences between each individual. Every human is ultimately unique, and so are everyone's preferences and needs. It is expected that most of these slight differences are not important enough to adjust the personality of the chatbot each time. However, there might be some cases where a traveler from one segment might benefit more from a chatbot personality meant for travelers from another group, as some traits and preferences can be more important than others.

RQ2.3: What concrete differences can be made between chatbots to account for different preferences?

The final sub-question or RQ2 is about applying everything learnt about the travelers and their preferences and seeing what the actual differences are between the chatbots based on this.

The hypothesis for this is that language use will mainly be affected. Language use will, as mentioned before, probably depend on age, where formal use of language will likely be more desirable among the older travelers, while younger travelers might prefer more informal language. Besides age, language use might also depend on characteristics such as extraversion or one's preference for efficiency. Expected is that more sociable travelers might prefer more extensive messages with greetings, jokes and other small remarks. People who place importance on efficiency might benefit more from more concise messages.

4. Method 1 – Preliminary Interviews

To answer all questions mentioned before, multiple methods have been used. The first method were preliminary interviews. The purpose of these interviews was to get a general idea about the different travelers and their habits, preferences and needs regarding providing input. The participants' preferences towards chatbots were also explored. Based on these interviews, combined with existing literature, specific scenarios and chatbot personalities have been thought up and designed. The scenarios are different situations in which users could find themselves wanting to use a chatbot. These scenarios were used in a second method, where the goal was to find out the preferences of each traveler type. More on this in chapter 5, which will cover the second method. This chapter will explain the first method, the preliminary interviews, and present its results.

The goal of the interviews was to identify into which traveler segment the interviewee would fit best, and then ask them about their preferences and experiences in the field of

giving input. As mentioned before, the four kinds of inputs accounted for in this research are complaints, questions, suggestions, and recommendations.

4.1 Participants

For the interviews, convenience sampling was used to make sure to get as many participants as possible. The participants needed to have traveled by train before, so this was the only requirement each participant needed to meet. It was also important to make sure to get participants from every traveler segment, but since it is only possible to get an idea of this after the interview, the sampling continued until all groups were sufficiently represented.

In total, a total of 15 people between the age of 21 and 59 were interviewed. Of those 15 people, 9 were male and 6 were female. The interviews were conducted 1 on 1, and recorded, after the interviewee gave their permission.

4.2 Materials

For this method, all questions asked to the participants were written beforehand. These questions can be found in Appendix A. The interviewees also signed an informed consent form, by which they gave permission for their answers and data to be used and analyzed in this research. The consent form can be found in Appendix B.

4.3 Procedure

Before the interviews, a pilot was conducted to assess and revise the questions. After a few revisions to the questions, the interviews were held. This took place in a 1-on-1 setting, where the researcher asked the participants multiple questions in a row. These questions were recorded to be transcribed later. After signing the consent form, the interview started. The first questions meant to divide the participant in one of the six groups of travelers. After this, questions about the four different kinds of inputs were asked. These questions explored the willingness of the interviewee to provide different kinds of input for the NS, such as how likely they are to complain about something, or if and who they would ask something when

having a question. After that, the concept of a chatbot was introduced. For each of the four different kinds of input, the participants were asked if they would be more willing to provide said input when given the option to do it through a chatbot, and why.

4.4 Analysis

After the interviews, the answers were transcribed and qualitatively coded to find the most important suggestions and see if there are any patterns between different groups of travelers. The coding process was inductive, meaning that the keywords are derived from the data, instead of using predefined keywords and scanning the transcriptions for them. The reason for using inductive coding is that these interviews are mainly used for getting a better understanding of the general preferences and wishes of the travelers, without having any predetermined keywords or concepts in mind which could unintentionally influence the results.

By analyzing and coding the results, different ideas have been revised, such as scenarios/situations where using a chatbot might be beneficial over using other, existing methods.

4.5 Results

While the participants were divided into different traveler segments, there were no noticeable patterns between the different traveler types and their preferences regarding their attitudes. However, the interviews gave some good insights about the preferences and behavior of travelers in general. These findings helped create an image and give inspiration to create certain scenarios. The general findings will be discussed here.

Results: Input

Firstly, the vast majority of the interviewees had never provided any input of any kind towards Dutch Railways, be it through email or by addressing an employee in the train or on the station. However, most people interviewed have

had complaints, questions or suggestions. A commonly mentioned argument about why these kinds of input were not acted upon was that the travelers did not consider or forgot about the possibility of filing complaints, suggestions or asking questions. As one interviewee said:

“As soon as I leave the train, I don’t really think about these kinds of things anymore”.

Some other travelers mentioned they would not know how to do so, although this was less common. Most people mentioned they did not know the exact procedure of, for example, filing a complaint, but that they likely would figure it out by looking it up online. Giving travelers the option to provide input through a system which is already used for other related matters, such as planning a journey, will likely improve awareness of the possibility to do so, which will hopefully result in less people forgetting about these possibilities.

Results: Chatbots

Regarding chatbots, most interviewees have used one before, and the ones who had not were at least familiar with them. However, the general opinion about chatbots was quite low, mainly because of previous bad experiences interacting with them. These bad experiences were in all cases related to the technical limitations and difficulties of chatbots, where the chatbot was not able to understand the users correctly or did not give any accurate answers. Because of this, the majority of the participants so far were quite hesitant about using chatbots to provide different kinds of input. However, when mentioning that they needed to imagine a good working chatbot which can accurately help the traveler with all their needs and can communicate just the same as a human would, the vast majority said that they would be open to use one for certain scenarios. As one participant said:

“It would be nice to use a chatbot for questions, as that would mean you wouldn’t have to walk around looking for a staff member”.

While this was the case, there was one more factor that was important for some, regarding when using a chatbot would be beneficial. The type of input that would be given also made a difference for some. For inputs that required immediate feedback, the advantages of using a chatbot were more apparent. Most people understood the benefit of using a chatbot while needing a question answered or wanting some recommendations about their trip, the same could not always be said about the other two types of input. When asking a question or needing a recommendation, it is important to receive an answer as quickly as possible; exactly the advantage a chatbot has over other methods such as sending an email or looking something up. However, for complaints and suggestions, immediate feedback is not always needed. Interviewees also mentioned that sending an email to file a complaint felt more official to them than sending a text message to a robot, and this would give them more of a sense that their complaint/suggestion would actually be read by someone. While this might just be because not many people are used to using chatbots yet, it is a valid point, and definitely something to keep in mind.

5 Method 2 – Chatbot Evaluation

The conducted interviews helped to give some insights and ideas about travelers' preferences, but more quantitative data was necessary to draw any significant conclusions. To obtain this data, the second method was used: the chatbot evaluation. For this, participants would interact with chatbots, each chatbot having a different personality, and rate them on multiple characteristics. Based on the insights gathered from the preliminary interviews, different scenarios were developed, in which users could find themselves wanting to use a chatbot. In these different scenarios, the task the user needed to perform differs. After using the chatbot to complete certain tasks, such as planning a trip or asking a question, the users would evaluate the chatbot by filling in a survey. The scenarios will be effectuated by using Wizard of Oz (WOZ) prototyping. WOZ is a rapid-prototyping method used to simulate certain systems which are very costly or

advanced [48]. A human acts as the system itself and interacts with the user through a real or mock computer interface. In the case of this research paper, the researcher took the role of the chatbot and responded to the queries of the participants, them not knowing they were actually talking to a human instead of a computer. The reason for choosing this method is that while it is very important to keep the test-experience as close and similar to actually chatting with a chatbot, it would be too time-intensive and out of the scope of this project to code and develop an actual chatbot of high enough quality. Just showing screenshots of potential conversations with the chatbot, on the other hand, would be not immersive enough, because typing yourself is a big part of the appeal of chatbots. With this method, the participants believed they were actually talking to a chatbot, while it did not carry the challenges of developing an actual working prototype.

As previously mentioned, this paper focuses on four types of input, being complaints, questions, suggestions and recommendations. However, for this chatbot evaluation, the decision was made to only focus on one of these four input types. Reason for this was to reduce the number of variables in the experiment. With different scenarios and traveler types, there were already quite some variables which could influence each other. While measures were taken to try and minimize this effect, it is still good practice to try and limit the number of variables. This is why for the chatbot evaluation, only questions could be asked.

The reason questions were chosen instead of any of the other three input types is because the results from the preliminary interviews indicated that this is the type of input most travelers would preferably use a chatbot for. A second reason is that because this experiment is mainly meant for users to evaluate different chatbot personalities, it is necessary to have a minimum level of interaction with each chatbot. Because questions require feedback every time and can easily be followed up upon by other questions, this seemed the most suitable out of the four input types to guarantee a minimum level of interaction with the chatbot.

The Ethics and Privacy Quick Scan of the Utrecht University Research Institute of Information and Computing Sciences classified this research as low-risk with no fuller ethics review or privacy assessment required. The Quick Scan results can be found in Appendix G.

5.1 Participants

For this method, convenience sampling was used. The reason for this being that every individual's opinion might add value to this research, as long as they have traveled by train before. That was the only condition for the participants. Besides this, because the total duration of the experiment was about 30 – 45 minutes, and the experiment could not be done remotely, finding a large number of participants proved difficult. Convenience sampling is the easiest and most reliable way to gather a large sample group, which is necessary for a quantitative analysis. It was also important to make sure to get participants from every traveler segment, so the sampling continued until all groups were represented as much as possible.

In total a number of 30 participants partook in the study. Of those 30 participants, 19 were male, and 11 were female. The participants' age ranged between 22 and 89 years old.

5.2 Materials

The experiment was done using Microsoft Teams to emulate the chatbot interface. A document with pre-written answers to potential questions was used to copy and paste those answers into Microsoft Teams. This document (in Dutch) can be found in Appendix C. A smartphone was given to the participants with Microsoft Teams running. The scenarios were written down in a separate document and read to the participants. The scenarios can be found (in Dutch) in Appendix D. The questionnaire was made using Qualtrics and was filled in by the participants on a laptop. The questionnaire can be found in Appendix E. The participant also signed a consent form, which can be found in Appendix F, to indicate they agreed to

participate with the experiment, and that their data could be used for the analysis.

5.3 Procedure

Before the experiments, a pilot study was conducted to assess and revise the pre-written answers, to make most potential questions were accounted for. It also became apparent during the pilot study that the difference between chatbot personalities was not always noticed, as the participants paid more attention to the content of the messages than the chatbot's manner of speaking. This issue has been addressed during the actual experiment by stressing the goal of the experiment more clearly. Finally, the pilot study also helped determine the effectiveness of the Wizard of Oz method, by checking whether the participant actually believed they were chatting with a chatbot. After this was confirmed and some pre-written answers were added, the actual experiment could begin. This proceeded as follows.

Preparation

The participant entered the room with the researcher and the experiment was explained to them. It was stressed to them that the focus of the research lied the chatbot's personality and behavior, and less on the accuracy or content of its answers. Then, the participant was handed the phone with Microsoft Teams open, where a new chat is opened with an account called "NS Chatbot". A screenshot of the interface can be found below. The researcher sat across the participant with a laptop in front of them. On the laptop, in split screen windows, all the pre-written answers to potential questions were opened, as well as a Microsoft Teams chat window, which had been opened on the account the participant will be chatting with. The participants were told the laptop was there for the researcher to take notes, as the participants are required to "think aloud" and say everything that comes up into their minds out loud. This served both the purpose of giving the participant a reasonable explanation as for why the researcher is working on a laptop, as well as being able to hear the thoughts of the

participants, as these could give more insights about the participants' opinions of the chatbot besides the questionnaire which would be used later on. Once the participant signed the consent form and everything else was set up, the experiment could begin.

The experiment

The experiment started by the researcher explaining the different scenarios the participant would have to act out. The participant would get three scenarios where they will have to take the role of a traveler who needs a certain question answered. These three questions were:

- A question about transferring railway services and how this needs to be done using the OV-chipkaart
- A question about what train to take to reach Amsterdam Central Station
- A question about companion discount, how this works, and who applies for it

To get an answer to these three questions, the participants could ask any question they liked, as well as follow-up questions if they deemed them necessary. In fact, the participants were encouraged to follow up with other questions regarding the same topic, as this allowed them to spend more time with the chatbot, which would help them paint a better picture of the chatbot's personality. Answers to a lot of questions were already pre-written, but in the case there was no answer prepared for a certain question, the participant would be answered with a message telling them the chatbot did not understand the question. This question and its corresponding answer were not added to the list of potential questions for further experiments, since this would mean that the chatbot would grow more competent as the number of participants grows, which could give skewed results.

For each different scenario, the personality of the chatbot would also differ. There were three different chatbot personalities in total: a formal one, an informal one, and an informal one which would use emojis in their messages. The formal

one addressed people in a more polite way by using the Dutch pronoun "u", instead of "je", which is more respectful. It also acted less enthusiastic and more reserved, which is done by using a fewer number of words, not including any exclamation marks, and using a vocabulary of mostly written language. The other two chatbots used more words in total, and more words from spoken language. The informal chatbot and the chatbot using emojis used the same kind of language, with the only difference being that the latter chatbot used emojis in nearly every message. As mentioned before, the allocation of the chatbot's personality for each scenario would differ for each participant, to minimize the chance of prejudice based on the type of questions needed to be asked, which might skew the results.

The participants themselves were also allocated to a certain Traveler Type. If the participants had been interviewed before, their traveler type would already be known based on some of the questions during the interview. If not, these questions would be asked after the Wizard of Oz experiment.

As the participants used the phone to type messages to send to the chatbot, the researcher copied and pasted the corresponding answers from the answer list into the chat. This continued until the participant indicated to the chatbot that their question was sufficiently answered. Then, the participant would fill in the questionnaire about the chatbot they just interacted with. This was done on a laptop, and the questionnaire consisted of several questions where the participant would have to evaluate the chatbot personality on different attributes by using a Likert scale.

When this was done, the researcher would tell the participant they would "reset" the chatbot and change its personality. This was done by sending a new greeting message to the participant. Each chatbot personality had its own greeting message, so this indicated to the participant the chatbot has "switched" personalities. The reason for telling the participant that the chatbot would switch personalities each time was to make it easier for

them to distinguish and remember every different personality. This was done until all three scenarios are settled.

After finishing all three scenarios and talking to all three different chatbot personalities, participants were finally asked to rank the three chatbot personalities they talked to, based on which one was their favorite to talk to and which one they would prefer in real life. After asking for a brief explanation about their choice, the participants were thanked and dismissed.



Figure 2: Example of interaction with chatbot on MS Teams

5.4 Analysis

After interacting with each chatbot personality, participants were asked to rate the chatbot on multiple facets. These facets were certain relevant metrics from the user experience perspective section of the paper of Peras (2018) [7]. In this paper, different evaluation metrics for chatbots were examined to be able to evaluate chatbots on different facets. The author proposed to split up the evaluation into five perspectives:

- a) the user experience perspective
- b) information retrieval perspective
- c) linguistic perspective
- d) technology perspective
- e) business perspective.

The user experience perspective is about how the user rates the overall experience while using the chatbot, which uses measures such as ease of use, performance, engagement, and other attributes, which will be explored more into detail later. The information retrieval perspective evaluates how well the information needs of the users are met, such as how accurate the responses of the chatbot are, and how efficient the process is. The linguistic perspective analyzes the level of the linguistic accuracy of chatbots and ability to return appropriate response. The technology perspective measures how well the chatbot can imitate human behavior based on its responses to input. Finally, the business perspective measures the business value of a chatbot, which is the difference between the effectiveness and the costs of the chatbot.

Not every chatbot needs to be evaluated on all five perspectives, as the evaluation process and the combination of different perspectives should align with different areas of chatbot application and with the needs of users, as well as the scope of the research. For this paper, the focus lies on the evaluation of the social aspects of the chatbot's behavior, and thus the chatbot was only evaluated on the user experience perspective. The reason for the other four perspectives to not be included in this evaluation is that these focus on other facets than the (perceived) personality of the chatbot. Also, the evaluation will be conducted after a Wizard of Oz experiment, which means there will not be an actual chatbot which looks up information itself, so evaluating it on points like efficiency and accuracy would be pointless. The user experience perspective, however, does match with the goals of this research. This is why (some of) the metrics of this perspective will be used in the evaluation after the Wizard of Oz experiment.

User experience perspective

The user experience perspective consists of four categories:

- a) usability
- b) performance
- c) affect
- d) satisfaction

First off, usability refers to efficiency and effectiveness, and is related to the ease of using and the time needed to complete the task. Metrics will be assigned to evaluate the level of the task completion and the amount of effort required from users to complete the tasks. Because these metrics rely on the user's personality, experience and ideas, they are mostly qualitative.

Performance refers to completion of a certain task in terms of completeness, promptness and appropriateness. Like usability, performance is affected by users and their goals, so the assigned evaluation metrics are also qualitative by nature. The category of affect is used to evaluate the experience and emotions of chatbot users. Those attributes are hard to express numerically, and they are usually measured by surveys or questionnaires.

Finally, satisfaction refers to user's pleasure arising from the comparison of their expectations and chatbot performance. Again, those feelings are hard to express numerically.

The metrics are mostly quantitative, although some papers are indicating the relation of satisfaction with the duration and frequency of chat. [42]. After some pilot tests, however, it was deemed pointless to also use metrics such as the total time spent interacting with the chatbot or the number of messages sent, since they fluctuated way too much depending on the participant, the scenario, or other peripheral matters. The data gathered is both quantitative as qualitative, where the quantitative data are the Likert-scale scores, and the qualitative data are the explanations and remarks given by the participants.

6 Results

This section consists of the quantitative- and qualitative results. The quantitative results will look at the scores given in the questionnaire and examine any effects of the independent variables on these scores. The qualitative results will look at the preferred chatbots for each traveler as indicated by themselves, as well as the reasons for this. These reasons were also given by the participants themselves.

Out of the total of 30 participants, 6 were Explorers, 4 were Individualists, 3 were Functional Planners, 4 were Certainty Seekers, 4 were Socializers, and 9 were Convenience Seekers.

6.1 Quantitative Results

Total score

In order to answer RQ2 and its sub-questions, a Kruskal-Wallis H test was conducted to analyze the effect of different traveler groups and different chatbot types on the total score given by the participant. There was a significant effect of traveler type on the total score ($\chi^2(5) = 18.36$, $p = .003$). This means that the scores given by each participant do depend on the traveler type. There was, however, no significant effect of chatbot type on the total score ($\chi^2(2) = 3.04$, $p = .219$). When splitting the output per traveler type, there was also no significant effect of chatbot type on total score, for none of the different traveler groups. The effect closest to significance was for the biggest group, the Convenience Seekers ($\chi^2(2) = 5.32$, $p = .070$). As for the other groups, the p-value was higher, and there was no significant effect of chatbot type on total score for the Explorers ($\chi^2(2) = 1.28$, $p = .529$), Individualists, ($\chi^2(2) = .88$, $p = .957$), Functional Planners ($\chi^2(2) = .04$, $p = .955$), Certainty Seekers ($\chi^2(2) = 2.01$, $p = .365$) and Socializers ($\chi^2(2) = 3.77$, $p = .152$).

Affect score

Now that the total score has been analyzed, it is also important to look at the various aspects of what the chatbots have been rated on. The affect score takes into account the added value of 3 of

the 10 questions from the questionnaire: the questions about personality, entertainment, and engagement. With this score, the same Kruskal-Wallis H test was conducted to analyze the effect of traveler type and chatbot type on their assessment of the chatbot's personality.

Again, there was a significant effect of traveler type on the affect score ($\chi^2(5)= 13.01, p = .023$). As for chatbot type ($\chi^2(2)= 3.84, p = .146$), there was no significant effect on the affect score. When splitting the output per traveler type, there was a significant effect of chatbot type on the affect score for the Socializers ($\chi^2(2)= 7.98, p = .018$), and for the Convenience Seekers ($\chi^2(2)= 6.30, p = .043$). There was no significant effect of chatbot type on the affect score for the Explorers ($\chi^2(2)= 1.32, p = .517$), Individualists ($\chi^2(2)= .07, p = .966$), Functional Planners ($\chi^2(2)= .82, p = .663$), and the Certainty Seekers ($\chi^2(2)= .88, p = .644$).

Usability score

The usability score takes into account the value of 1 of the 10 questions from the questionnaire: the question about the ease of use of the chatbot.

There was a significant effect of traveler type on the usability score ($\chi^2(5)= 22.25, p < .001$). As for chatbot type ($\chi^2(2)= .84, p = .656$), there was no significant effect on the usability score. When splitting the output per traveler type, there was no significant effect of chatbot type on the usability score for any of the traveler groups: Explorers ($\chi^2(2)= 0, p = 1$), Individualists ($\chi^2(2)= .58, p = .747$), Functional Planners ($\chi^2(2)= .31, p = .859$), Certainty Seekers ($\chi^2(2)= .14, p = .934$), Socializers ($\chi^2(2)= 1.10, p = .577$), and the Convenience Seekers ($\chi^2(2)= 1.18, p = .554$).

Performance score

The performance score takes into account the added value of 2 of the 10 questions from the questionnaire: the questions about helpfulness and accuracy.

There was a significant effect of traveler type on the performance score ($\chi^2(5)= 17.72, p = .003$).

As for chatbot type ($\chi^2(2)= .11, p = .946$), there was no significant effect on the performance score. When splitting the output per traveler type, there was no significant effect of chatbot type on the performance score for any of the traveler groups: Explorers ($\chi^2(2)= 2.09, p = .352$), Individualists ($\chi^2(2)= .81, p = .669$), Functional Planners ($\chi^2(2)= .07, p = .966$), Certainty Seekers ($\chi^2(2)= .93, p = .627$), Socializers ($\chi^2(2)= 1.10, p = .577$), and the Convenience Seekers ($\chi^2(2)= 1.19, p = .909$).

Satisfaction score

The satisfaction score takes into account the added value of 2 of the 10 questions from the questionnaire: the questions about expectation and satisfaction.

There was a significant effect of traveler type on the satisfaction score ($\chi^2(5)= 17.65, p = .003$). As for chatbot type ($\chi^2(2)= .89, p = .640$), there was no significant effect on the satisfaction score. When splitting the output per traveler type, there was no significant effect of chatbot type on the satisfaction score for any of the traveler groups: Explorers ($\chi^2(2)= 1.48, p = .492$), Individualists ($\chi^2(2)= .84, p = .658$), Functional Planners ($\chi^2(2)= .41, p = .816$), Certainty Seekers ($\chi^2(2)= .94, p = .625$), Socializers ($\chi^2(2)= 2.00, p = .368$), and the Convenience Seekers ($\chi^2(2)= .33, p = .847$).

Trust score

Additionally, a trust score was also measured and analyzed. The trust score is a measure of how trustworthy people deemed the chatbot, and how willing they are to provide personal data to the chatbot. The trust score takes into account the added value of 2 of the 10 questions from the questionnaire: the questions about trustworthiness and privacy.

There was a significant effect of traveler type on the trust score ($\chi^2(2)= 17.87, p = .003$). There was also a significant effect of chatbot type on the affect score ($\chi^2(2)= 10.15, p = .006$). When splitting the output per traveler type, however, there was no significant effect of chatbot type on the performance score for any of the traveler

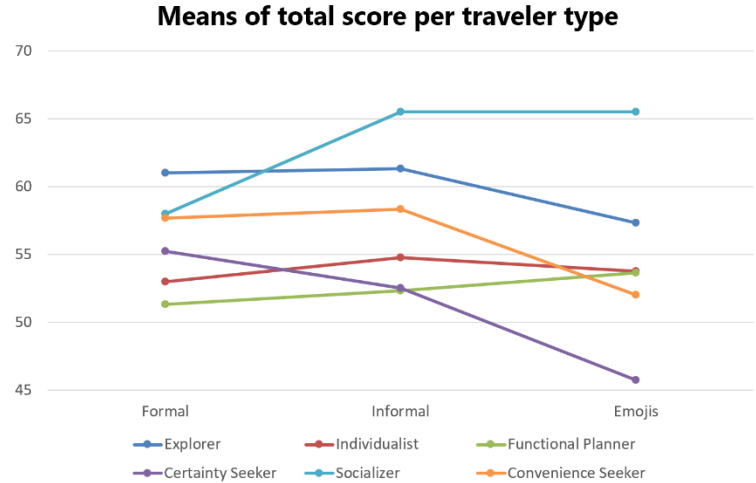
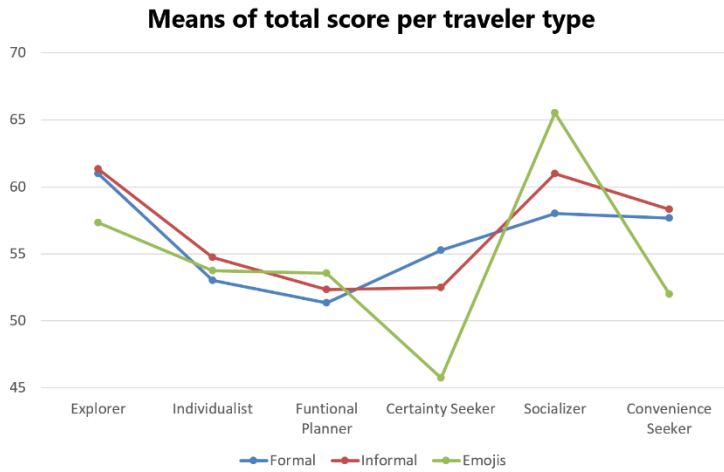


Figure 3: Graphs of the mean total scores of each chatbot type, per traveler type

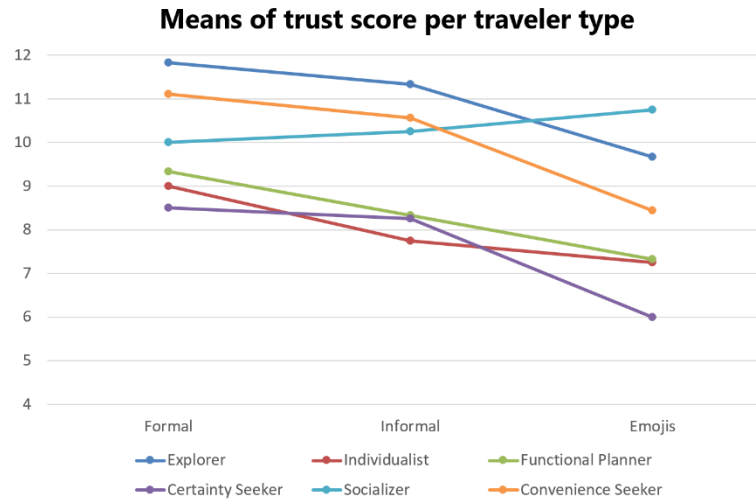
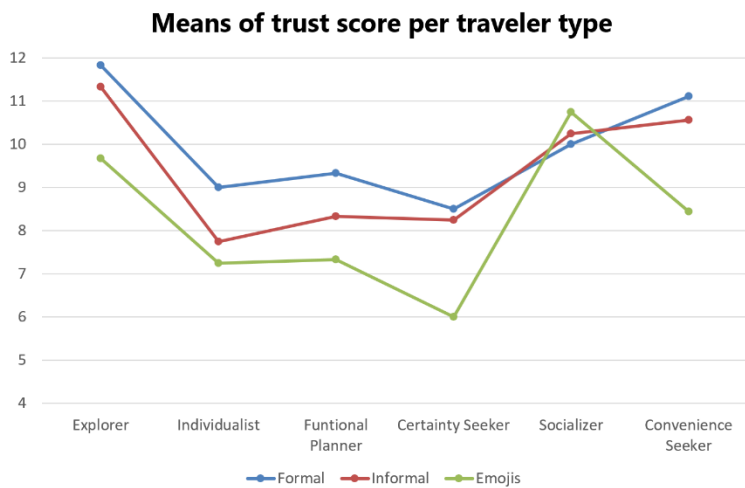


Figure 4: Graphs of the mean trust scores of each chatbot type, per traveler type

Traveler type	Chatbot type	Mean score	Std. Deviation	N
Explorer	Formal	61.00	3.69	6
	Informal	61.33	3.93	6
	Emojis	57.33	8.24	6
	Total	59.89	5.66	18
Individualist	Formal	53.00	7.53	4
	Informal	54.75	6.70	4
	Emojis	53.75	3.77	4
	Total	53.83	5.67	12
Functional Planner	Formal	51.33	11.72	3
	Informal	52.33	6.35	3
	Emojis	53.67	4.62	3
	Total	52.44	7.13	9

Certainty Seeker	Formal	55.25	9.36	4
	Informal	52.50	9.95	4
	Emojis	45.75	12.09	4
	Total	51.17	10.40	12
Socializer	Formal	58.00	5.83	4
	Informal	61.00	2.94	4
	Total	61.50	5.18	12
Convenience Seeker	Formal	57.67	3.87	9
	Informal	58.33	6.75	9
	Total	56.00	6.20	27

Table 2: Mean total scores of each chatbot type, per traveler type

groups: Explorers ($\chi^2(2)= 2.96, p = .228$), Individualists ($\chi^2(2)= 3.36, p = .187$), Functional Planners ($\chi^2(2)= 1.33, p = .513$), Certainty Seekers ($\chi^2(2)= 2.47, p = .291$), Socializers ($\chi^2(2)= .126, p = .939$), and the Convenience Seekers ($\chi^2(2)= 4.87, p = .088$).

6.2 Qualitative Results

This section will look at what the participants' indicated to be their favorite- and least favorite chatbot, and the reasons given by them as to why. This will be done for each group of traveler types, to see if there are any noteworthy patterns, similarities or differences in- and between groups.

Explorers

For the Explorers, four out of six participants indicated they favored the informal chatbot over the other two. The most mentioned reason for this, as indicated by them, was that they liked the friendliness compared to the formal chatbot. This group's overall least favorite chatbot was, however, the emoji chatbot. It was often mentioned that using emojis felt less professional and weird. One participant also mentioned that the use of emojis made it feel like they were taken less seriously and gave them the impression of a "company trying to be hip". In total, four of the six participants indicated the emoji chatbot to be their least favorite.

Individualists

For the Individualists, all participants preferred the formal chatbot over the other two. An often-mentioned reason was that they wanted the information as quick and concise as possible, and had no need for anything more, like emojis or enthusiastic greetings. One participant also said that they actually prefer chatbots to act like robots, instead of trying and emulate human behavior. So they prefer a chatbot acting dry, without adding any emotion. Regarding their least favorite chatbots, the group of Individualists was equally split between the informal chatbot and the emoji chatbot. The reasons why they disliked those chatbots

compared to the others were the same across the board; they added unnecessary stuff to the conversation, such as the use of popular language or distracting emojis.

Functional Planners

The Functional Planners were very evenly spread out regarding their favorites and least favorites. Each of the three members of this group favored- and disliked a different chatbot. Reason for this could be that this was the smallest group, so it is more difficult to accurately portray the overall preferences of this traveler group, with only a few participants. The reasons they gave for preferring one chatbot over the other were different for each Functional Planner, since they each favored a different chatbot. However, those reasons were very similar to those of other traveler types preferring the same chatbots. The reasons for the participant favoring the formal chatbot were him liking how concise and direct it was. As he said:

"The formal one was to the point, no nonsense, and gave quick information. I don't need anything else."

The reason for the participant preferring the informal chatbot was mainly that the use of language was relatable, which they liked. Lastly the participant favoring the emoji chatbot indicated this was because the emojis made it seem like the chatbot was more attentive.

Convenience Seekers

In contrast to this, the group of Convenience Seekers was the largest out of all traveler types. Both the informal chatbot and the emoji chatbot scored high in this group based on the ranking provided by the participants. Six out of nine Convenience Seekers indicated that their favorite chatbot was the informal one, and the other three said the emoji-using one was their preferred choice. None of the participants picked the formal chatbot as their number one. The most mentioned reason participants preferred the informal chatbot was that it felt more personal, friendly and pleasant compared

to the formal one, while not being as casual as the emoji chatbot. The emoji chatbot was considered too casual by those Convenience Seekers. The Convenience Seekers preferring the emoji chatbot, however, liked the level of casualness, indicating that the emojis made the conversation feel more casual, personal, and that it was closest as to how they would type to friends. This again corresponds with the similarity-attraction theory. As for this group's least favorite chatbot personality, an interesting detail is that all five people who indicated their favorite personality was the informal one, plus one more participant, also said that their least favorite personality was the emoji-using chatbot. This means there is a distinction in this group based on whether the travelers like the use of emojis or not. Reasons given for this ranking were that the emojis felt annoying, unprofessional, and forced. As one participant said:

"I'm not talking to my mother here".

As for the other four Convenience Seekers, three indicated that their least favorite chatbot was the formal one, as they did not like it to be addressed as politely and distant as the chatbot did. One Convenience Seeker mentioned that the informal chatbot was their least favorite one, because its personality was in their regard not similar enough to their own. Therefore, they preferred no personality at all in the formal chatbot instead of a clashing personality in the informal chatbot.

Socializers

As for the Socializers, a clear pattern could be seen. All participants from this group preferred the emoji-using chatbot over the other two. This is because the emojis made it look like the chatbot was more friendly, enthusiastic and attentive. They liked the playfulness of the emojis, the fact that it made the chatbot feel less rigid and more human-like, and they mentioned that it made chatting with the chatbot more entertaining. One participant mentioned:

"Emojis made it look friendly and enthusiastic, and it made me feel sorry for him when he was sad he couldn't help me".

The Socializers' least favorite chatbot was the formal chatbot, with two participants indicating that one as their least favorite. The other Socializer appointed the informal chatbot as their least favorite. Reasons of the participants for disliking the formal one were that it was less engaging and that it felt impersonal. One participant said:

"Formal use of language made it seem less engaging, like it wanted to get my question over with and leave".

Both participants also mentioned they preferred not to be addressed as formal. The other Socializer who preferred the formal chatbot over the informal one mentioned that this was because the informal one felt less trustworthy than the formal chatbot, but also was not as engaging and personal as the emoji using one. They said they prefer a personality which is one or the other, and not a personality in-between.

Certainty Seekers

Lastly, the preferences of the Certainty Seeker group were relatively evenly distributed, with the slight majority having a preference for the formal chatbot. Reasons given for this were that it felt more professional and reliable, and more accessible since it's shorter. Of the other two members of this group, one chose the informal one being their favorite, while the other participant chose the emoji chatbot. The reason for preferring the informal chatbot was that it felt professional enough, but also human-like enough to be able to identify with. As the participant said:

"It feels like it really has attention for you".

The informal chatbot was chosen as people's least favorite chatbot two times, as well as the emoji-using chatbot, with two more people indicating that one to be their least favorite. It was mentioned that the informal one was the least appealing of the two since the emoji one

was funny and the formal one felt reliable, but this one had no real distinguishing features. Reasons for people to choose the emoji chatbot as their least favorite was that the emojis felt unprofessional and distracting.

In conclusion, there were some groups where a clear pattern in preference could be seen, such as the Socializers, the Individualists, and to a certain extent also the Convenience Seekers. The positives and negatives about each chatbot did not differ much between groups. However, their preferences did, so that could mean every group has different priorities regarding aspects of interacting with the chatbot they find more important. On the next page, Table 2 shows an overview of the different arguments in favor- or against each chatbot type, divided by traveler-type group. Table 3 summarizes how many participants of each group chose which chatbot type as their number one, -two, or -three.

7 Discussion

Interpreting the data

When looking at the quantitative results, it turned out that for the Total Score, the only variable that had a significant effect on the total score was the traveler type, but not the chatbot type or the interaction between the two. This indicates that certain groups rate chatbots significantly different, but not based on the chatbot type. When looking at the graphs, however, there seem to be some patterns, at least for certain groups, which correspond with the qualitative analysis. For example, for the Explorer group, the informal chatbot was rated highest, only slightly above the formal chatbot. The emoji chatbot is clearly in last place, with quite a margin. This matches the verbal statements of participants during the qualitative analyses quite well, where the favorite chatbot of most Explorers was the informal one, followed by the formal chatbot. Four out of six people said the emoji chatbot was their least favorite. A reason as for why these differences were not significant might be the low number of participants. In total, thirty people evaluated the

three chatbots, and those thirty people were also split up into the six traveler segments, making the sample size for each segment even smaller. As mentioned earlier, even within certain traveler groups, people's preferences still differ. These differences within groups might be smaller than the differences between groups, but with this small a number of participants, those differences will make the data less consistent. This could be a reason as for why a lot of the effects of chatbot type on different score metrics were not significant.

The p-value was often lowest when looking at the traveler segment with the largest number of people: the Convenience Seekers. In contrast, it was often one of highest for the Functional Planners, the smallest group. This further supports the theory that the small number of participants could be a reason for many effects to not be significant.

Another reason for the fact that the effects were not significant could be that the similarity-attraction theory might not always apply for chatbots. According to an article by Kornfield et al. about automated text messaging tools meant for mental health support, people respond very differently to messages from computers than messages from people [19]. To some, an overly human-like tone and the use of emojis did not come across as believable and might instead evoke more negative emotions. This could mean that the similarity-attraction theory does not always apply to human-chatbot interaction the same way it does to human-human interaction.

However, when combining the quantitative-with the qualitative data, some patterns can still be seen, which could be interpreted as some conclusions. Firstly, it is obvious from the data that the Socializer group rated the emoji chatbot the highest overall. This matches the similarity-attraction theory that more extraverted travelers prefer a matching, extraverted chatbot personality [5]. The qualitative analysis also confirmed this, with every Socializer indicating

	Formal		Informal		Emoji	
	Positives	Negatives	Positives	Negatives	Positives	Negatives
Explorer	- Professional - Trustworthy	- Static x2 - Too formal	- Friendly x2 - Relatable			- Annoying - Unprofessional - Weird - Less serious - Forced
Individualist	- Robot-like - Trustworthy x2 - Professional - No nonsense - To the point			- Language use - Redundant	- Robot-like - Forced	- Unnecessary x2 - Unprofessional
Functional Planner	- To the point - Quick information - No redundancy	- Static - Associate with bad chatbots - Too formal - Dry	- Casual - Relatable		- Attentive - Fun - Funny	- Unnecessary - Too many
Certainty Seeker	- Professional - Reliable x2 - Accessible - Proper	- Too formal	- Professional enough - Human-like - Attentive		- Entertaining x2 - Funny x2	- Unprofessional - Unpleasant - Messy - Less serious
Socializer	- More trustworthy	- Less engaging - Impersonal	- Friendly		- Friendly - Enthusiastic - Playful - Funny - More real - Cute - More feeling	
Convenience Seeker	- Clear - Straight to the point - No personality clash	- Impersonal - Non-human like - Too formal	- Human-like x2 - Clear - Personal	- Personality clash	- Personal - Matches my personality - Low-key & casual	- Too much - Unnecessary - Forced - Fake

Table 3: Positives and negatives mentioned about each chatbot type, separated by traveler group

		Explorer	Individualist	Functional Planner	Certainty Seeker	Socializer	Convenience Seeker
#1	Formal chatbot	1	4	1	2		
	Informal chatbot	4		1	1		6
	Emoji chatbot	1		1	1	4	3
#2	Formal chatbot	3		1	2	1	6
	Informal chatbot	2	2	1	1	3	2
	Emoji chatbot	1	2	1	1		1
#3	Formal chatbot	2		1		3	3
	Informal chatbot		2	1	2	1	1
	Emoji chatbot	4	2	1	2		5

Table 4: Division of chatbot preferences, separated by traveler group

that they favored the emoji chatbot over the other two. The reasons given for this also match the similarity-attraction theory, with the chatbot being human-like, and the chatbot's perceived "social qualities". When looking at arguably the traveler type most different from the Socializer, the Individualist, something else is noticeable. The quantitative analysis and the qualitative analysis do not match. When looking at the quantitative analysis, the mean total score of the formal chatbot is lowest overall. At the same time, every Individualist indicated this to be their favorite chatbot. This seems contradictory, but a reason for this could be that Individualists valued different aspects of the evaluation more than others. This could be a reason for the lower total score of this chatbot, even though all Individualists indicated the formal chatbot to be their favorite. For example, on Trust Score, the formal chatbot scored the highest among the Individualists. If they place higher importance on trust than all other aspects, the Total Score does not always correspond with someone's favorite chatbot since trust is only a part of the total score. When looking at the qualitative analysis, it indeed seems like Individualists place a higher importance on trust since terms like these were mentioned multiple times among the reasons why they picked the formal chatbot as their favorite.

This all indicates that the Total Score is not an unconditional, perfect way to display how pleased everyone is with the chatbot, especially with a low sample size like in this study, where you have higher chances of having extreme or non-representative groups.

Overall though, the combination of the quantitative- and qualitative results show that there does seem to be a link between the personality of the traveler and their desired chatbot personality. More extraverted travelers generally preferred the more extraverted chatbots, and the less extraverted travelers generally preferred the less extraverted chatbots. There were obviously also differences in preference between travelers within the same group, but these were overall less obvious than those between groups. A notable difference in preference within groups was the preference for

emojis. Some people, from across all segments, had a strong aversion to them, while others liked the added value they brought.

Summary and research questions

To answer RQ2, it seems likely that for traveler groups which are either quite extravert or - introvert, the similarity-attraction theory holds up, and they prefer a chatbot with a matching level of expression as themselves. With other, less defined groups, personal preferences play a bigger role than the group they belong to. These personal preferences can be things like acceptance of emojis, preference of how a chatbot should act, and how somebody would like to be addressed. However, further research with more participants is needed to confirm this.

To answer RQ1, the results of the preliminary interviews will mainly be used. Looking at all the answers given during the interviews, it became clear that the main advantage of using chatbots for most people was the reaction speed. Besides this, the ease of use was also mentioned a couple of times. There are also some disadvantages when using chatbots compared to other methods, however. By far the most mentioned disadvantage of using chatbots was that they often do not work correctly. Other than those issues, there were also more personal problems with using chatbots. Some people indicated that they prefer calling over texting, simply because they can express themselves better using their voice. Others mentioned that for certain things, using a chatbot felt too casual, and that they would prefer a more "official" way to contact customer service. It was in some cases also dependent on the type of input they would need to provide to the Dutch Railways. When an immediate response was needed, some people would prefer using a chatbot, but when this was not the case, they indicated that they would prefer to do this in a different way, since the main benefit of a chatbot would be otiose at that point.

To conclude, it seems that people have no problem using chatbots to communicate with Dutch Railways, as long as they work as

expected, bring benefits that are required for the situation, and is easy to use.

8 Limitations

As mentioned before, one of the main limitations of this research was the number of participants. Because the methods had to be done in person, the maximum achievable number of participants turned out to be 30. This number could likely have been a reason as for why some effects were not significant, especially because these 30 participants were divided further into six groups. This limitation has been addressed to a certain extent by including the qualitative analysis, which helps interpret and support the patterns found in the data. This makes the assumptions and takeaways more credible. Besides this, the variety of participants could also have been better. The different traveler segments were not all represented equally in terms of numbers. This made the interpretation of patterns in certain groups more debatable than others.

The Total Score, the metric intended to measure the overall appreciation for each chatbot type, might also not be the most suitable in every case. This metric did not correspond with the qualitative results for some traveler types, indicating that other factors might also influence someone's appreciation for certain chatbot types. These factors could be the importance placed by travelers on certain characteristics over others, such as attentiveness or trustworthiness. The Total Score was also a subjective metric, with each participant supplying the score themselves through a Likert scale. This could have caused issues regarding consistency, as certain participants might have interpreted the scoring system differently than others.

Finally, the classification of travelers into certain groups was also not as clear-cut in every case. Certain participants could have belonged in multiple groups, since their preferences, personality and habits were not necessarily exclusive to one specific group. The division was not always as clear as desired, which could also have impacted the data.

9 Future work

Further research is necessary to prove some of the inferences made based on the data. Most importantly, a follow-up research with a much higher number of participants would add a lot of value in terms of credibility. The other limitations mentioned before could also be addressed to get more reliable results. For the Total Score, by allowing participants to place a certain weight to different characteristics, the Total Score could be more representative of the preferences of the participant. The sometimes ambiguous classification of different traveler types could also be addressed, for instance by focusing more on one specific characteristic, instead at multiple characteristics combined. This could help prevent the overlapping of certain segments.

This thesis was solely focused on train travelers, but the topics covered, and methods used are also applicable to other contexts. Chatbots are used in nearly every industry, not just train travel [2]. The exact specifics, such as participant segmentation and the content discussed with the chatbots would differ, but the essence would be the same.

The same goes for the characteristics focused upon. This thesis mainly focused on extraversion, but other characteristics could also be explored, such as agreeableness from the Big Five, which also impacted the evaluation of chatbots [46]. Other characteristics entirely could also be used to check the similarity-attraction theory in different contexts.

Chatbots are becoming increasingly more customary each day [25], and thus further research about this topic in different contexts could always prove beneficial.

10 Conclusion

This paper aimed to find out if and when people would be willing to use a chatbot to provide input to the Dutch Railways and if the soft characteristics of travelers could be used to optimize their experience using this chatbot. As

for the first objective, a lot of people are open to this idea, as long as it works as intended, and offers an advantage over using more traditional methods. By splitting travelers into six traveler types, this thesis tried to determine if the different soft characteristics of each group had an impact on how people prefer a chatbot to behave and talk. While there was no significant effect proven of the interaction between traveler type chatbot type on the rating participants gave the chatbot, some patterns could be seen which could be further investigated and proven significant with a bigger sample size. These effects were that very extraverted or - introverted groups preferred matching chatbot

types. The use of emojis was very divisive among participants, with some loving it and others hating them. While there seems to be a slight pattern in each traveler group regarding preferences like this, further research is needed to prove whether these effects are significant.

References

1. Abu Shawar, B., & Atwell, E. (2007). Chatbots: Are they Really Useful? *Journal for Language Technology and Computational Linguistics*, 22(1), 29–49. <https://doi.org/10.21248/jlcl.22.2007.88>
2. Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. *IFIP Advances in Information and Communication Technology*, 373–383. https://doi.org/10.1007/978-3-030-49186-4_31
3. Anable, J. (2005). ‘Complacent Car Addicts’ or ‘Aspiring Environmentalists’? Identifying travel behaviour segments using attitude theory. *Transport Policy*, 12(1), 65–78. <https://doi.org/10.1016/j.tranpol.2004.11.004>
4. Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, 87(1), 43–51. <https://doi.org/10.1037/0021-9010.87.1.43>
5. Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology*, 62(3), 713–715. <https://doi.org/10.1037/h0044721>
6. Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM*, 43(4), 70–78. <https://doi.org/10.1145/332051.332075>
7. Chatbot evaluation metrics. (2018). In D. Peras (Red.), *Economic and Social Development*.
8. Chimbo, B., Gelderblom, J., & Villiers, D. (2011). A comparative analysis of the meaning of “learnability” for child and adult users. *Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction Rome*.
9. Cohen, A. S., Minor, K. S., Baillie, L. E., & Dahir, A. M. (2008). Clarifying the Linguistic Signature: Measuring Personality From Natural Speech. *Journal of Personality Assessment*, 90(6), 559–563. <https://doi.org/10.1080/00223890802388459>
10. *Contact / NS*. (z.d.). Nederlandse Spoorwegen. <https://www.ns.nl/klantenservice/contact>
11. Czaja, S. J., Sharit, J., Ownby, R., Roth, D. L., & Nair, S. (2001). Examining age differences in performance of a complex information search and retrieval task. *Psychology and Aging*, 16(4), 564–579. <https://doi.org/10.1037/0882-7974.16.4.564>
12. Elimeliah, C. V. (2016, 3 augustus). *Why chatbots are replacing apps*. VentureBeat. <https://venturebeat.com/business/why-chatbots-are-replacing-apps/>
13. *First time bot users deserve good bots*. (2017). Orange Hive. Geraadpleegd op 3 mei 2022, van <https://unfiltered.orangehive.de/first-time-bot-users-deserve-good-bots/>
14. Forbes, P. J., Gabrielli, S., Maimone, R., Masthoff, J., Wells, S., & Jylhä, A. (2014). Towards using Segmentation-based Techniques to Personalize Mobility Behavior Interventions. *ICST*

- Transactions on Ambient Systems*, 1(4), e4.
<https://doi.org/10.4108/amsys.1.4.e4>
15. Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. *Proceedings of the 2018 Designing Interactive Systems Conference*.
<https://doi.org/10.1145/3196709.3196735>
 16. John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. *University of California, Berkeley*.
 17. Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Pradeepkumar, B., & Das, T. (2015). A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence. *International Journal of u- and e-Service, Science and Technology*, 8(7), 277–284.
<https://doi.org/10.14257/ijunesst.2015.8.7.28>
 18. *Klacht over NS-dienstverlening / NS*. (z.d.). Nederlandse Spoorwegen.
<https://www.ns.nl/klantenservice/overig/klachtenafhandeling.html>
 19. Kornfield, R., Meyerhoff, J., Studd, H., Bhattacharjee, A., Williams, J. J., Reddy, M., & Mohr, D. C. (2022). Meeting Users Where They Are: User-centered Design of an Automated Text Messaging Tool to Support the Mental Health of Young Adults. *CHI Conference on Human Factors in Computing Systems*.
<https://doi.org/10.1145/3491102.3502046>
 20. Leung, R., Tang, C., Haddad, S., McGrenere, J., Graf, P., & Ingriany, V. (2012). How Older Adults Learn to Use Mobile Devices. *ACM Transactions on Accessible Computing*, 4(3), 1–33.
<https://doi.org/10.1145/2399193.2399195>
 21. Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What Can You Do? *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*.
<https://doi.org/10.1145/2901790.2901842>
 22. Liao, Q. V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M. P., Wang, D., Muller, M., Shami, N. S., & Geyer, W. (2018). All Work and No Play? *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.
<https://doi.org/10.1145/3173574.3173577>
 23. Luger, E., & Sellen, A. (2016). “Like Having a Really Bad PA”. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.
<https://doi.org/10.1145/2858036.2858288>
 24. Mahmud, S., Alvina, J., Chilana, P. K., Bunt, A., & McGrenere, J. (2020). Learning Through Exploration: How Children, Adults, and Older Adults Interact with a New Feature-Rich Application. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.
<https://doi.org/10.1145/3313831.3376414>
 25. Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457–500.
<https://doi.org/10.1613/jair.2349>
 26. McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509–516.

- <https://doi.org/10.1037/0003-066x.52.5.509>
27. Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., & O'Neill, J. (2017). How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. *Human-Computer Interaction - INTERACT 2017*, 441–459. https://doi.org/10.1007/978-3-319-67744-6_28
 28. Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5), 862–877. <https://doi.org/10.1037/0022-3514.90.5.862>
 29. Moon, Y. (2002). Personalization and Personality: Some Effects of Customizing Message Style Based on Consumer Personality. *Journal of Consumer Psychology*, 12(4), 313–325. <https://doi.org/10.1207/15327660260382351>
 30. Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
 31. Oxford University Press (OUP). (2021). Dictionary.com Is The World's Favorite Online Dictionary. In *Lexico.com*. <https://www.dictionary.com/>
 32. Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6), 1296–1312. <https://doi.org/10.1037/0022-3514.77.6.1296>
 33. Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003a). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
 34. Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003b). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
 35. Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, 85(2), 291–301. <https://doi.org/10.1037/0022-3514.85.2.291>
 36. Ramasamy, A. (2019, 18 januari). *How Businesses Can Begin Using Chatbots The Right Way*. Forbes. <https://www.forbes.com/sites/forbestechcouncil/2019/01/18/how-businesses-can-begin-using-chatbots-the-right-way/?sh=7e2ca4802183>
 37. Ranoliya, B. R., Raghuvanshi, N., & Singh, S. (2017). Chatbot for university related FAQs. 2017 *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. <https://doi.org/10.1109/icacci.2017.8126057>
 38. *Reizigersgedrag | NS Dashboard*. (z.d.). NSJaarverslag.nl. Geraadpleegd op 18 oktober 2022, van <https://dashboards.nsjaarverslag.nl/reizigersgedrag>
 39. Ruane, E., Farrell, S., & Ventresque, A. (2021). User Perception of Text-Based Chatbot Personality. *Chatbot Research and Design*, 32–47. https://doi.org/10.1007/978-3-030-68288-0_3

40. Ruijten, P. A. M. (2020). The similarity-attraction paradigm in persuasive technology: effects of system and user personality on evaluations and persuasiveness of an interactive system. *Behaviour & Information Technology*, 40(8), 734–746. <https://doi.org/10.1080/0144929x.2020.1723701>
41. Schnurr, P. P., Rosenberg, S. D., Oxman, T. E., & Tucker, G. J. (1986). A Methodological Note on Content Analysis: Estimates of Reliability. *Journal of Personality Assessment*, 50(4), 601–609. https://doi.org/10.1207/s15327752jpa5004_7
42. Schumaker, R. P., Ginsburg, M., Chen, H., & Liu, Y. (2007). An evaluation of the chat and knowledge delivery components of a low-level dialog system: The AZ-ALICE experiment. *Decision Support Systems*, 42(4), 2236–2246. <https://doi.org/10.1016/j.dss.2006.07.001>
43. Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, 117, 106627. <https://doi.org/10.1016/j.chb.2020.106627>
44. Thompson, E. R. (2008). Development and Validation of an International English Big-Five Mini-Markers. *Personality and Individual Differences*, 45(6), 542–548. <https://doi.org/10.1016/j.paid.2008.06.013>
45. van Hagen, M. (2009). *How to meet the needs of train passengers? A successful customer segmentation model for public transport*. European Transport Congress.
46. Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters. *Journal of Computer-Mediated Communication*, 19(3), 529–545. <https://doi.org/10.1111/jcc4.12066>
47. von der Pütten, A. M., Krämer, N. C., & Gratch, J. (2010). How Our Personality Shapes Our Interactions with Virtual Characters - Implications for Research and Development. *Intelligent Virtual Agents*, 208–221. https://doi.org/10.1007/978-3-642-15892-6_23
48. Wilson, J., & Rosenberg, D. (1988). Rapid Prototyping for User Interface Design. *Handbook of Human-Computer Interaction*, 859–875. <https://doi.org/10.1016/b978-0-444-70536-5.50044-0>

Appendix A

Interview Questions

General questions

- Name
- Age
- Gender
- Occupation
- Location -> city name / city size / province

Questions about traveling habits

- How often do you take the train?
- Why do you use the train mostly?
- When do you use the train mostly?
- What class do you usually travel by when in the train?
- Do you often take the same route (same train/bus/etc.) when using public transport?
- When taking the same route, do you have multiple carriers and/or transfers?
- What are some of the reasons you take the train?
- What are some of the reasons you don't take the train?
- What do you do when on the train?
- How do you prepare for a journey?
- Do you like to observe your surroundings and other passengers while in the train?
- Do you keep to yourself in the train?
- How would you like the railway employees to behave?
 - Respectful/polite
 - Friendly/sociable
 - Professional/distant
- Do you see traveling by train as entertaining itself as well, or just as a means to transport?
- Do you think speed and efficiency is more important, or feeling safe and being well prepared?
- Do you enjoy meeting and talking to fellow passengers on the train?
- How important do you think convenience and simple planning is?

Questions about input:

Complaints:

- Have you ever had (or still have) any complaints about traveling by train?
- Have you ever filed this/a complaint?
 - Why not -> do you know how to file a complaint?
 - OR: What happened with your complaint?
- Would you be more willing to file a complaint if it were made easier?
 - How would it be easier for you?
 - Tips/suggestions?
- Have you ever used chatbots before? Are you aware of the chatbot on the NS site?
- Would you like to file complaints using a chatbot? Or prefer a real person over mail?

Questions:

- Have you ever had (or still have) any questions about traveling by train?
- Have you ever asked this question to an NS employee? (on station, through e-mail or website, etc.)

- Why not -> do you know who to go to?
 - OR: What happened with your question?
- Would you be more willing to ask a question if it were made easier?
 - How would it be easier for you?
 - Tips/suggestions?
- Have you ever used chatbots before? Are you aware of the chatbot on the NS site?
- Would you like to ask questions using a chatbot? Or prefer a real person over mail/in real life?

Suggestions:

- Have you ever had (or still have) any suggestions about traveling by train? (can be about stations, train carts, seats, routes, etc.)
- Have you ever suggested this to an NS employee? (on station, through e-mail or website, etc.)
 - Why not -> do you know who to go to?
 - OR: What happened with your suggestion?
- Would you be more willing to suggest something if it were made easier?
 - How would it be easier for you?
 - Tips/suggestions?
- Have you ever used chatbots before? Are you aware of the chatbot on the NS site?
- Would you like to suggest things using a chatbot? Or prefer a real person over mail/in real life?

Recommendation:

- Have you ever wanted any recommendation about traveling by train? (can be about stations, restaurants, sights along the way, places to visit at your destination, etc.)
- Have you ever asked for this to an NS employee? (on station, through the website, etc.)
 - Why not -> do you know who to go to?
 - OR: What kind of recommendation did you get?
- Would you be more willing to ask for recommendations if it were made easier?
 - How would it be easier for you?
 - Tips/suggestions?
- Have you ever used chatbots before? Are you aware of the chatbot on the NS site?
- Would you like to ask for recommendations using a chatbot? Or prefer a real person over mail/in real life?

Appendix B

Informed Consent Form – Method 1

Thank you for agreeing to participate in this study. The purpose of this study is to gain insight into the attitude of different kinds of people towards providing input to the Dutch Railways, as well as their opinions and thoughts about chatbots. It is conducted by a student of Utrecht University as thesis of the Masters Programme “Human Computer Interaction”.

The procedure will take about 30 minutes. You will be answering questions about your travel behavior, your preferences regarding providing input, and your preferences regarding using chatbots. Your answers will be audio recorded and transcribed at a later point for the purpose of analyzing its contents. The recording will be deleted once the derived data is processed. The transcript of the recording will be stored on a hard drive for three months, until the analysis is complete. In accordance with the General Data Protection Regulation (GDPR) you can have access to your recordings and can request them to be deleted at any time during this period.

All data collected during this study will be anonymized, and will not have any identifiable information about participants. Participating is completely voluntary and you have the right to withdraw from the study at any time during the procedure. In this case you may ask for the recording to be stopped and deleted, and well as your given answers. Consequently, your data will not be used in the final paper.

Insights gathered from your answers will be used in writing a qualitative research report, which will be publicly available to read. Direct quotes may be used in the report, but your name and other identifying information will be kept anonymous.

By signing below, you agree that you have read and understood the above information and would be interested in participating in this study. If you have any questions regarding the research, you can contact Sven Neuteboom, s.y.neuteboom@students.uu.nl.

Signature: _____

Date: _____

Appendix C

Pre-written questions and answers for Wizard of Oz evaluation

Algemene antwoorden (formeel):

<i>Begroeting</i>	Goedendag! Waarmee kan ik u van dienst zijn?
"Ik heb een vraag"	Daarmee help ik u graag. Wat is uw vraag?
Ongereleerde dingen	Het spijt mij, maar dit begrijp ik niet. Kunt u het wellicht iets anders verwoorden?
Start	Goedendag. Ik ben Jens, de virtuele assistent van NS, en ik help u graag met uw vragen of opmerkingen over onze diensten.
Afsluiting	Ik hoop u voldoende te hebben geholpen. Ik wens u verder een fijne dag.
	Ik wens u verder een fijne dag.

Algemene antwoorden (informeel):


<i>Begroeting</i>	Hey! Kan ik je ergens mee helpen?
"Ik heb een vraag"	Daar ben ik voor! Wat is je vraag?
Ongereleerde dingen	Sorry, maar dit snap ik niet. Kun je het misschien op een andere manier zeggen?
Start	Hallo! Ik ben Jens, de virtuele assistent van NS en ik help je graag met je vragen of opmerkingen over onze diensten.
Afsluiting	Ik hoop je voldoende te hebben geholpen! Fijne dag en goede reis!
	In dat geval wens ik je een fijne dag en goede reis!

Algemene antwoorden (emojis):

<i>Begroeting</i>	Hey! Kan ik je verder nog ergens mee helpen? (happyface)
"Ik heb een vraag"	Daar ben ik voor! Wat is je vraag? :)
Ongereleerde dingen	Sorry, maar dit snap ik niet. (sadbutrelievedface) Kun je het misschien op een andere manier zeggen?
Start	Hallo! Ik ben Jens, de virtuele assistent van NS en ik help je graag met je vragen of opmerkingen over onze diensten. (happyface)
Afsluiting	Ik hoop je voldoende te hebben geholpen! Fijne dag en goede reis! (satisfied)
	Fijne dag en goede reis! (satisfied)

Vraag stellen (formeel)


Opdracht: vraag stellen over overstappen

<i>Input</i>	<i>Output</i>
Hoe werkt overstappen?	De OV-chipkaart kunt u bij elke aanbieder van openbaar vervoer gebruiken in Nederland. Wanneer u overstapt op een andere vervoerder, checkt u uit bij uw vorige vervoerder en checkt u in bij uw nieuwe. Dit kan bij de overstapzuilen op het perron.
Hoe vind ik die overstapzuilen?	Deze staan op het perron waar uw trein vandaan vertrekt. De zuilen zien er als volgt uit: 
Hoe werken deze overstapzuilen?	U checkt eerst uit bij de vervoerder waar u net mee heeft gereisd. Vervolgens checkt u weer in bij de vervoerder waar u mee gaat reizen. In uw geval checkt u dus uit bij de NS zuil, en vervolgens weer in bij de Valleilijn.
Ah oke dat moet lukken	Fijn om te horen. Heeft u nog verdere vragen?
Nee	In dat geval wens ik u een fijne dag en een goede reis!
Ja	Wat wilt u nog meer weten?
Oke	Is dat zo duidelijk?

Vraag stellen (informeel)

Opdracht: vraag stellen over overstappen

<i>Input</i>	<i>Output</i>
Hoe stap ik over?	Overstappen naar een andere vervoerder is erg simpel! Op het perron check je bij de overstapzuilen uit bij je vorige vervoerder en vervolgens check je in bij de volgende vervoerder. In dit geval dus de Valleilijn.
Hoe vind ik die overstapzuilen?	Deze staan op het perron waar je trein vandaan vertrekt. Deze zien er zo uit:

	
Hoe werken deze overstapzuilen?	Je checkt met jouw OV chipkaart eerst uit bij de vervoerder waar je net mee hebt gereisd, in jouw geval de NS. Daarna check je weer in bij de zuil van de volgende vervoerder, de Valleilijn dus.
Ah oke dat moet lukken	Top! Heb je verder nog vragen voor me?
Nee	Okay! In dat geval wens ik je een goede reis!
Ja	Okay, wat wil je nog vragen?
Oke	Is dat duidelijk zo? Dit doe je met jouw OV-chipkaart.

Vraag stellen (emojis)

Opdracht: vraag stellen over overstappen

<i>Input</i>	<i>Output</i>
Ik stap zo over op de valleilijn, werkt mijn OV-chipkaart dan ook	Overstappen naar een andere vervoerder is erg simpel! (smileeyes) Op het perron check je bij de overstapzuilen uit bij je vorige vervoerder en vervolgens check je in bij de volgende vervoerder. In dit geval dus de Valleilijn.
Hoe vind ik die overstapzuilen?	Deze staan op het perron waar je trein vandaan vertrekt. Deze zien er zo uit: 
Hoe werken deze overstapzuilen?	Je checkt eerst uit bij de vervoerder waar je net mee hebt gereisd, in jouw geval de NS. Daarna check je weer

	in bij de zuil van de volgende vervoerder, de Valleilijn dus ;)
Ah oke dat moet lukken	Top! Heb je verder nog vragen voor me? (facewithmonocle)
Nee	Okay! In dat geval wens ik je een goede reis! (huggingface)
Ja	Okay, wat wil je nog vragen?
Oke	Is dat duidelijk zo? (nerd)

Vraag stellen (formeel)

Opricht: vraag stellen over reis naar Amsterdam

<i>Input</i>	<i>Output</i>
Ik moet naar Amsterdam, welke trein moet ik dan hebben	Op welk station bevindt u zich nu?
Ik ben nu op Utrecht en ik moet naar Amsterdam, welke trein moet ik dan hebben?	Vanaf Utrecht Centraal vertrekt er over 14 minuten (19:09) een Intercity richting Den Helder die stopt op Amsterdam Centraal. Deze vertrekt vanaf spoor 5.
Ah oke	Is deze optie goed voor u, of heeft u liever een trein op een ander tijdstip?
Ah oke thanks	Graag gedaan. Is deze optie goed voor u, of heeft u liever een trein op een ander tijdstip?
Nee is prima zo	Dat is goed om te horen. Indien u nog andere vragen heeft hoor ik dat uiteraard graag.
Kan ik er eentje later krijgen?	Uiteraard. Over 28 minuten (19:23) vertrekt er een Intercity richting Enkhuizen die stopt op Amsterdam Centraal vanaf spoor 7.
Later	Over 28 minuten (19:23) vertrekt er een Intercity richting Enkhuizen die stopt op Amsterdam Centraal vanaf spoor 7.
Kan ik er eentje eerder krijgen?	Uiteraard. Over 4 minuten (18:59) vertrekt er een Intercity richting Alkmaar die stopt op Amsterdam Centraal vanaf spoor 7.
Eerder	Over 4 minuten (18:59) vertrekt er een Intercity richting Alkmaar die stopt op Amsterdam Centraal vanaf spoor 7.
Op welk spoor komt deze trein aan?	Deze trein komt op Amsterdam Centraal aan op spoor 12.

Vraag stellen (informeel)

Opricht: vraag stellen over reis naar Amsterdam

<i>Input</i>	<i>Output</i>
Ik moet naar Amsterdam, welke trein moet ik dan hebben	Op welk station ben je nu?
Ik ben nu op Utrecht en ik moet naar Amsterdam, welke trein moet ik dan hebben?	Vanaf Utrecht Centraal gaat er over 14 minuten (19:09) een Intercity richting Den Helder die stopt op Amsterdam Centraal. Deze vertrekt vanaf spoor 5.
Ah oke	Is deze optie goed, of heb je liever een trein later?

Ah oke thanks	Geen probleem! Is deze optie goed, of heb je liever een trein later?
Nee is prima zo	Mooi! Als je nog meer vragen hebt hoor ik dat wel!
Kan ik er eentje later krijgen?	Zeker! Over 28 minuten (19:23) gaat er een Intercity richting Enkhuizen die stopt op Amsterdam Centraal vanaf spoor 7.
Later	Over 28 minuten (19:23) gaat er een Intercity richting Enkhuizen die stopt op Amsterdam Centraal vanaf spoor 7.
Kan ik er eentje eerder krijgen?	Zeker! Over 4 minuten (18:59) gaat er een Intercity richting Alkmaar die stopt op Amsterdam Centraal vanaf spoor 7.
Eerder	Over 4 minuten (18:59) gaat er een Intercity richting Alkmaar die stopt op Amsterdam Centraal vanaf spoor 7.
Op welk spoor komt deze trein aan?	Je komt op Amsterdam Centraal aan op spoor 12!

Vraag stellen (emoji)

Opdracht: vraag stellen over reis naar Amsterdam

<i>Input</i>	<i>Output</i>
Ik moet naar Amsterdam, welke trein moet ik dan hebben	Op welk station ben je nu? (facewithmonocle)
Ik ben nu op Utrecht en ik moet naar Amsterdam, welke trein moet ik dan hebben?	Vanaf Utrecht Centraal gaat er over 14 minuten (19:09) een Intercity richting Den Helder die stopt op Amsterdam Centraal. Deze vertrekt vanaf spoor 5 (highspeedtrain)
Ah oke	Is deze optie goed, of heb je liever een trein later? (facewithmonocle)
Ah oke thanks	Geen probleem! Is deze optie goed, of heb je liever een trein later?
Nee is prima zo	Mooi! Als je nog meer vragen hebt hoor ik dat wel!
Kan ik er eentje later krijgen?	Zeker! Over 28 minuten (19:23) gaat er een Intercity richting Enkhuizen die stopt op Amsterdam Centraal vanaf spoor 7 (threeoclock)
Later	Over 28 minuten (19:23) gaat er een Intercity richting Enkhuizen die stopt op Amsterdam Centraal vanaf spoor 7 (threeoclock)
Kan ik er eentje eerder krijgen?	Zeker! Over 4 minuten (18:59) gaat er een Intercity richting Alkmaar die stopt op Amsterdam Centraal vanaf spoor 7 (threeoclock)
Eerder	Over 4 minuten (18:59) gaat er een Intercity richting Alkmaar die stopt op Amsterdam Centraal vanaf spoor 7 (threeoclock)
Op welk spoor komt deze trein aan?	Je komt op Amsterdam Centraal aan op spoor 12! :)

Vraag stellen (formeel)

Opdracht: vraag stellen over met korting reizen

<i>Input</i>	<i>Output</i>
Hoe werkt meereiskorting precies?	Samenreiskorting is 40% korting op uw ritprijs wanneer u in de daluren met iemand meereist die een NS-abonnement heeft (uitgezonderd NS Flex Basis en Kids Vrij), of een Studentenreisproduct.
Wat zijn de daluren?	Van maandag tot en met vrijdag vóór 06.30 uur, van 9.00 - 16.00 uur en na 18.30 uur. In de weekenden en op feestdagen is meereiskorting altijd geldig.
Hoeveel mensen kan ik meenemen met meereiskorting	Als u een NS-abonnement of Studentenreisproduct heeft kunt u maximaal 3 personen mee laten reizen.
Hoe activeer ik meereiskorting	U kunt samenreiskorting aanzetten in de NS-app, in Mijn NS, of u kunt het op uw OV-chipkaart laden via de NS-kaartautomaat.
Moet de persoon met het abonnement erbij zijn?	Er moet een persoon meereizen met een NS-abonnement of Studentenreisproduct. Stapt deze persoon eerder uit of reist deze persoon niet mee terug? Zet dan samenreiskorting uit via de NS-App of Mijn NS. Als u uitcheckt wordt de samenreiskorting automatisch van uw OV-chipkaart gehaald.
Geld samenreiskorting bij alle vervoerders?	Uw korting is geldig bij NS en alle andere treinvervoerders. Als u in de spits overstapt naar een andere treinvervoerder, dan geldt uw korting niet meer.
Geldt samenreiskorting ook in het buitenland?	Reist u naar het buitenland? Dan kunt u samenreiskorting aanzetten voor het gedeelte dat u in Nederland reist. U kiest bij 'Kortings- en loyaliteitskaarten' voor de optie 'Samenreiskorting' tijdens uw bestelling op Nsinternational.com .
Ik heb geen OV-chipkaart, kan ik dan nog steeds samenreiskorting gebruiken?	Jazeker, u kunt ook zonder OV-chipkaart met iemand meereizen. Dan koopt u een Samenreisticket in plaats van een regulier ticket.

Vraag stellen (informeel)

Opdracht: vraag stellen over met korting reizen

<i>Input</i>	<i>Output</i>
Hoe werkt meereiskorting precies?	Bij samenreiskorting krijg je 40% korting op je ritprijs als je in de daluren samen met iemand reist die een NS-abonnement of Studentenreisproduct heeft (uitgezonderd NS Flex Basis en Kids Vrij).
Wat zijn de daluren?	Van maandag tot en met vrijdag vóór 06.30 uur, van 9.00 - 16.00 uur en na 18.30 uur. In de weekenden en op feestdagen is samenreiskorting altijd geldig.
Hoeveel mensen kan ik meenemen met meereiskorting	Als je een NS-abonnement of Studentenreisproduct heeft mag je maximaal 3 personen mee laten reizen.
Hoe activeer ik meereiskorting	Dat kan op 3 manieren! Je kunt samenreiskorting aanzetten in de NS-app, in Mijn NS, of je kan het

	eenmalig op je OV-chipkaart zetten via de kaartautomaat.
Moet de persoon met het abonnement erbij zijn?	Yes. Er moet een persoon meereizen met een NS-abonnement of Studentenreisproduct. Stapt deze persoon eerder uit of reist deze persoon niet mee terug? Zet dan samenreiskorting uit via de NS-App of Mijn NS. Als je uitcheckt wordt de samenreiskorting automatisch van je OV-chipkaart gehaald.
Geld samenreiskorting bij alle vervoerders?	Jazeker! Je korting is geldig bij NS en alle andere treinvervoerders. Let wel op! Als je in de spits overstapt naar een andere treinvervoerder, dan geldt je korting niet meer.
Geldt samenreiskorting ook in het buitenland?	Helaas niet. Reis je naar het buitenland? Dan kunt je wel de samenreiskorting aanzetten voor het gedeelte dat je in Nederland reist. Om dit te doen kiest je bij 'Kortings- en loyaliteitskaarten' voor de optie 'Samenreiskorting' tijdens je bestelling op Nsinternational.com .
Ik heb geen OV-chipkaart, kan ik dan nog steeds samenreiskorting gebruiken?	Jazeker! Je kunt een Samenreisticket kopen in plaats van een normaal kaartje.

Vraag stellen (emoji)

Opdracht: vraag stellen over met korting reizen

<i>Input</i>	<i>Output</i>
Hoe werkt meereiskorting precies?	Bij samenreiskorting krijg je 40% korting op je ritprijs als je in de daluren samen met iemand reist die een NS-abonnement of Studentenreisproduct heeft (uitgezonderd NS Flex Basis en Kids Vrij) (moneymouthface)
Wat zijn de daluren?	De daluren zijn van maandag tot en met vrijdag vóór 06.30 uur, van 9.00 - 16.00 uur en na 18.30 uur. In de weekenden en op feestdagen is meereiskorting altijd geldig (sevenoclock)
Hoeveel mensen kan ik meenemen met meereiskorting	Als je een NS-abonnement of Studentenreisproduct hebt mag je maximaal 3 personen mee laten reizen (yn)
Hoe activeer ik meereiskorting	Dat kan op 3 manieren! Je kunt samenreiskorting aanzetten in de NS-app, in Mijn NS, of je kan het eenmalig op je OV-chipkaart zetten via de kaartautomaat (relieved)
Moet de persoon met het abonnement erbij zijn?	Yes. Er moet een persoon meereizen met een NS-abonnement of Studentenreisproduct. Stapt deze persoon eerder uit of reist deze persoon niet mee terug? Zet dan samenreiskorting uit via de NS-App of Mijn NS. Als je uitcheckt wordt de samenreiskorting automatisch van je OV-chipkaart gehaald.
Geld samenreiskorting bij alle vervoerders?	Jazeker! (happyface) Je korting is geldig bij NS en alle andere treinvervoerders. Let wel op! Als je in de spits overstapt naar een andere treinvervoerder, dan geldt je korting niet meer.

Geldt samenreiskorting ook in het buitenland?	Helaas niet :(Reis je naar het buitenland? Dan kunt je wel de samenreiskorting aanzetten voor het gedeelte dat je in Nederland reist. Om dit te doen kiest je bij 'Kortings- en loyaliteitskaarten' voor de optie 'Samenreiskorting' tijdens je bestelling op Nsinternational.com .
Ik heb geen OV-chipkaart, kan ik dan nog steeds samenreiskorting gebruiken?	Jazeker! Je kunt een Samenreisticket kopen in plaats van een normaal kaartje :D

Appendix D

Chatbot Scenarios

Scenario 1: Overstappen

Je bent onderweg van Utrecht Centraal naar Barneveld Centrum. In de reisplanner zie je dat je eerst de intercity moet hebben naar Amersfoort Centraal, en daar moet overstappen op de Valleilijn stoptrein richting Barneveld Zuid. Er staat aangegeven dat je ook met je OV chipkaart moet overstappen van de NS naar de Valleilijn. Je snapt echter niet zo goed hoe dit werkt. Probeer hier uit te komen door vragen aan de chatbot te stellen. Het wordt aangemoedigd om zoveel mogelijk door te vragen, om zeker te weten dat je het écht goed snapt.

Scenario 2: Route vragen

Je wilt de trein pakken naar Amsterdam Centraal, maar weet niet op welk tijdstip de trein gaat of vanaf welk perron deze vertrekt. Je bent momenteel op Utrecht Centraal. Probeer uit te vinden waar je moet zijn, en op welk tijdstip, door vragen aan de chatbot te stellen. Het wordt aangemoedigd om zoveel mogelijk door te vragen, om zeker te weten dat je de goede trein hebt op het goede tijdstip.

Scenario 3: Meereiskorting

Je hebt gehoord van een vriend dat hij gebruik kan maken van samenreiskorting, en iemand kan meenemen die dan met korting mag reizen. Je had hier nog nooit eerder van gehoord en wil er graag meer over weten, zoals of jij hiervoor in aanmerking komt, hoeveel de korting is, en wat de voorwaarden zijn. Probeer zoveel mogelijk over samenreiskorting uit te vinden door vragen te stellen aan de chatbot. Het wordt aangemoedigd om zoveel mogelijk door te vragen, om zo zoveel mogelijk te weten te komen over de voorwaarden van samenreiskorting.

Appendix E

Questionnaire

On a scale of 1-7, how much do you agree with the following statements?

- The chatbot helped me well with my questions
- The process of getting an answer to my questions was very easy
- The chatbot accurately answered my questions
- I found the personality of the chatbot pleasant
- I found it entertaining to talk to the chatbot
- I found that the chatbot came across as attentive
- I found that the chatbot came across as trustworthy
- I would feel comfortable if I needed to give any private information to the chatbot
- I expected the chatbot to function well
- I thought the chatbot functioned well after I used it

Appendix F

Informed Consent Form – Method 2

Thank you for agreeing to participate in this study. The purpose of this study is to gain insight into the attitude of different kinds of people towards chatbots with varying personalities. It is conducted by a student of Utrecht University as thesis of the Masters Programme “Human Computer Interaction”.

The procedure will take about 45 minutes. You will be testing a chatbot by interacting with it, and fill in multiple questionnaires afterwards. The testing of the chatbot will be audio recorded and transcribed at a later point for the purpose of analyzing its contents. The recording will be deleted once the derived data is processed. The transcript of the recording will be stored on a hard drive for three months, until the analysis is complete. In accordance with the General Data Protection Regulation (GDPR) you can have access to your recordings and can request them to be deleted at any time during this period.

The results of the questionnaire will be stored through Qualtrics on the Utrecht University server until the evaluation phase is completed, and later will be exported onto the hard drive of the researcher for three months, until the analysis is complete.

All data collected during this study will be anonymized, and will not have any identifiable information about participants. Participating is completely voluntary and you have the right to withdraw from the study at any time during the procedure. In this case you may ask for the recording to be stopped and deleted, and well as the results from the questionnaire. Consequently, your data will not be used in the final paper.

Insights gathered from the questionnaire and recording will be used in writing a qualitative & quantitative research report, which will be publicly available to read. Direct quotes may be used in the report, but your name and other identifying information will be kept anonymous.

By signing below, you agree that you have read and understood the above information and would be interested in participating in this study. If you have any questions regarding the research, you can contact Sven Neuteboom, s.y.neuteboom@students.uu.nl.

Signature: _____

Date: _____

Appendix G

Quick Scan Results

Response Summary:

Section 1. Research projects involving human participants

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from Twitter, Reddit) without directly recruiting participants, please answer no.

- Yes

Recruitment

P2. Does your project involve participants younger than 18 years of age?

- No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

- No

P4. Is your project likely to involve participants engaging in illegal activities?

- No

P5. Does your project involve patients?

- No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

- No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

- No

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)

- Yes

PC2. Will you tell participants that their participation is voluntary?

- Yes

PC3. Will you obtain explicit consent for participation?

- Yes

PC4. Will you obtain explicit consent for any sensor readings, eye tracking, photos, audio, and/or video recordings?

- Yes

PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?

- Yes

PC6. Will you give potential participants time to consider participation?

- Yes

PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?

- Yes

PC8. Does your project involve concealment or deliberate misleading of participants?

- No

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of **personal data** (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?

- No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

H1. Does your project give rise to a realistic risk to the national security of any country?

- No

H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?

- No

H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)

- No

H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)

- No

H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?

- No

H6. Does your project give rise to a realistic risk of harm to the researchers?

- No

H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?

- No

H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?

- No

H9. Is there a realistic risk of other types of negative externalities?

- No

Section 4. Conflicts of interest

C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?

- No

C2. Is there a direct hierarchical relationship between researchers and participants?

- No

Section 5. Your information.

This last section collects data about you and your project so that we can register that you completed the Ethics and Privacy Quick Scan, sent you (and your supervisor/course coordinator) a summary of what you filled out, and follow up where a fuller ethics review and/or privacy assessment is needed. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information. Please see the guidance on the ICS Ethics and Privacy website on what happens on submission.

Z0. Which is your main department?

- Information and Computing Science

Z1. Your full name:

Sven Neuteboom

Z2. Your email address:

s.y.neuteboom@students.uu.nl

Z3. In what context will you conduct this research?

- As a student for my master thesis, supervised by:
Marloes Vredenburg

Z5. Master programme for which you are doing the thesis

- Human-Computer Interaction

Z6. Email of the course coordinator or supervisor (so that we can inform them that you filled this out and provide them with a summary):

m.t.r.vredenburg@uu.nl

Z7. Email of the moderator (as provided by the coordinator of your thesis project):

j.f.m.masthoff@uu.nl

Z8. Title of the research project/study for which you filled out this Quick Scan:

Traveling with a chatbot: How can a personalized chatbot make traveling easier & more comfortable?

Z9. Summary of what you intend to investigate and how you will investigate this (200 words max):

This research aims to investigate whether the personality type of a train traveler influences their preference regarding the personality of the chatbot they'll be using to ask questions about their journey. This will be done through a Wizard of Oz experiment, where the participants have to rate three different chatbot personalities on different aspects. These ratings will be compared to find out if there are patterns to be found.

Z10. In case you encountered warnings in the survey, does supervisor already have ethical approval for a research line that fully covers your project?

- Not applicable

Scoring

- Privacy: 0
- Ethics: 0