André Tørlen Lønvik

The Highly Sensitive Person and Chatbots

What personality traits are more important to have in a chatbot, based on a person's HSP value.

Master's thesis in Interaction Design
Supervisor: Frode Volden
June 2022

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Architecture and Design
Department of Design

NTNU
Norwegian University of Science and Technology

André Tørlen Lønvik

The Highly Sensitive Person and Chatbots

What personality traits are more important to have in a chatbot, based on a person's HSP value.

Master's thesis in Interaction Design
Supervisor: Frode Volden
June 2022

Norwegian University of Science and Technology
Faculty of Architecture and Design
Department of Design

NTNU
Norwegian University of
Science and Technology

Abstract

Human-Computer conversations have increasingly become cheaper and more available worldwide through the use of chatbots. It has been established that giving chatbots personality increases usability and the perception of intelligence, while also increasing human trust. This study aims to determine what people who score highly on HSPS want in a chatbot. Specifically, we investigate whether high sensitivity is likely to determine what traits a chatbot should have when conversing with a user. To test the hypothesis that which traits a person wants is correlated with how high they score on the HSPS, an online survey was distributed to the author's network. Responses were analyzed using a correlation. Subsequently, a separate interview process was conducted with students available on the local campus based on the results of the questionnaire. The results showed a high correlation between high HSPS and traits that favored human characteristics, and a high correlation between lower HSPS and traits focused on efficiency. These results suggest that there is a likely need for a chatbot to be able to slightly adapt to the personality of a user. In addition, more research is needed to determine whether the personalities of users are also likely to affect how a user would like a chatbot to function.

Sammendrag

Samtaler mellom mennesker og datamaskiner har i økende grad blitt billigere og mer tilgjengelig over hele verden gjennom bruk av chatbots. Det er fastslått at det å gi chatbots personlighet øker brukervennligheten og oppfatningen av intelligens, samtidig som den øker menneskelig tillit. I denne studien ønsker vi å finne ut hva personer som skårer høyt på HSPS vil ha i en chatbot. Spesifikt undersøker vi om høysensitivitet avgjøre hvilke egenskaper en chatbot burde ha når den snakker med en bruker. For å teste hypotesen om at hvilke egenskaper en person ønsker er korrelert med hvor høyt de skårer på HSPS, ble en online undersøkelse distribuert til forfatterens nettverk. Svarene ble analysert ved hjelp av korrelasjoner. Etter dette ble det gjennomført en egen intervjuprosess med tilgjengelige studenter på den lokale campus basert på funnene i spørreskjemaet. Resultatene viste en høy korrelasjon mellom høy HSPS og egenskaper som favoriserte menneskelige egenskaper, og en høy korrelasjon mellom lavere HSPS og egenskaper fokusert på effektivitet. Disse resultatene tyder på at det er et sannsynlig behov for en chatbot for å kunne tilpasse seg litt til personligheten til en bruker. I tillegg er det behov for å studere videre om personlighetene til brukerne også sannsynligvis vil påvirke hvordan en bruker ønsker at en chatbot skal fungere.

Acknowledgments

I would like to thank the following people for their great help during the course of this thesis project:

My supervisor, Frode Volden, for your excellent guidance and support during this process. Thank you for your keen insight and support in finding multiple articles for me to read on the topic of Highly Sensitive Person (HSP). I am also grateful for all the help with SPSS to analyze the data, which I would not have made it without.

Mom and Dad for your long-standing support, both emotional and material over the five years I have been in Gjøvik.

And finally, the participants for taking time out of their busy schedule to help me with my thesis project.

Contents

Figures

Tables

Code Listings

Acronyms

AES Aesthetic Sensitivity 14, 41

AI Artificial Intelligence 1, 3–6, 9–12, 14, 15, 23, 29, 30, 33, 40–42

CA Conversational Agent 3, 4, 6, 10, 14, 15, 34

CAI Conversational Artificial Intelligence 4, 6

CB1 Chatbot 1 31, 87

CB2 Chatbot 2 31

CNN Conversational Neural Network 6

CRL Continuous Reinforcement Learning 6, 9, 14

EI Emotional Intelligence 13

EOE Ease of Excitation 13, 25–27, 41

HSP Highly Sensitive Person iii, 13, 14, 20, 25–27, 29, 39, 41

HSPS Highly Sensitive Person Scale 13, 25, 39, 58, 60

ISP Internet Service Provider 22

LST Low Sensory Threshold 14, 26, 27

ML Machine Learning 4, 6, 7

NN Neural Network 4, 9–11, 36

RL Reinforcement Learning 9

RNN Recurrent Neural Network x

SPS Sensory-Processing Sensitivity 13

VA Virtual Assistant 6

VH Virtual Human 1

Glossary

Bias A Bias means a deviation from the standard, and is necessary to identify statistical patterns in data used. This is especially apparent within Artificial Intelligence, as without bias, finding and classifying differences between instances would be impossible (Danks and London 2017; Ferrer et al. 2020; Silberg and Manyika 2019). 5, 14

Emotional Quotient Emotional intelligence (also known as Emotional Quotient or EQ). first 14, 42

Generative Pre-trained Transformer 3 (GPT-3) The GPT-3 model created by OpenAI and trained on 45TB of compressed plaintext data (Brown, Mann, et al. 2020, July 22). OpenAI is an AI research and deployment company with a mission to ensure that artificial general intelligence benefits all of humanity (OpenAI n.d.). 10

Model Within the field of ML and AI, a model is defined as a program that has been trained on a set of data (called the training set) to recognize certain types of patterns (Chooch 2020, June 14). 6–8, 10

Natural Language Processing Allows a machine to understand what we say in our natural language, and don't need the exact correct syntax the program to get a meaning behind a string provided to it. A string could be a set of bytes or characters. 1, 4, 5

Seq2Seq Seq2Seq is a ML model which takes as input a sequence of words(sentence or sentences) and generates an output sequence of words (Wadhwa 2018). It is also used in Googles Translation service (Sutskever et al. 2014). The Seq2Seq is based on a Recurrent Neural Network (RNN) which can easily map sequences to sequences whenever the alignment between the inputs the outputs is known ahead of time (Sutskever et al. 2014). For example given a sequence of inputs $(x_1, \cdots, x_T)$, a standard RNN computes a sequence of outputs $(y_1, \cdots, y_T)$. 37

Chapter 1

Introduction

Within the last decade, an increased focus on Artificial Intelligence (AI) has had a rapid increase in popularity, spurring an era of research into Natural Language Processing (NLP), statistics and probability, trying to create a human-like conversations through the assistance of AIs. However, communication between a human and a machine can still be said to be in its infancy (Zhang, Dinan, et al. 2018). Currently, we are within an era of weak AI. A weak AI, sometimes referred to as "narrow AI", can understand certain keywords and produce responses to commands but can't discern the meaning of what was said. Examples of weak AI include Apple's Siri, Amazon's Alexa, or Google Assistant (Lam Research 2021, November 18). To put simply, it is machines behave as if they are intelligent. Bishop (2021) argues that "Weak AI focuses on epistemic issues relating to engineering a simulation of [human] intelligent behavior, whereas strong AI, in seeking to engineer a computational system with all the causal power of a mind, focuses on the ontological". And Searle (1980, p. 417) argues that a weak AI gives us a very powerful tool through the study of the mind.

We expect that within at least a decade or two, we will most likely see the beginning of strong AI and further the research into development of an AI that can function on similar with, if not the same, level of intellect as a human. While development on more advanced AI is ongoing, the progress in the emotional intelligence of chatbots have been increasing. Lucas et al. (2014) have suggested that Virtual Humans (VHs) could likely lead patients to behave more openly during a clinical interview. This indicates to us that there is both a want and a need for improvements in chatbots, such that us humans will more naturally open up and converse with chatbots.

Therefore, we have created a couple of research questions to further the study on Human-Computer interactions, mainly focusing on the human-like aspects of chatbots.

1.1 Research questions

In this thesis we will attempt to answer the following research questions:

1. How can we improve the usability of chatbots by adapting better-defined traits?
2. What traits matter for a user in a chatbot?
    a. What traits would be preferable for users and in which context?
    b. What traits negatively affect the user-experience?
    c. What difference does gender reflect on traits wanted?
3. What will the focus on higher emotional intelligence give users?
    a. Will a chatbot oriented towards the personality and writing style of the user improve the used experience of interfacing with a chatbot?
    b. Will users be more likely to continue using chatbots if it could better "relate" to the users problem and adapt its writing method?

The first research question will be answered through research mentioned in part in Chapter 2 and some in Chapter 5.

1.2 Thesis Outline

This thesis consists of six main chapters.

Chapter 1 This chapter introduces the thesis and structure, show similar studies, and lay the out some of the topics we will discuss in the next chapter.

Chapter 2 In this chapter we will lay the foundation of theory and explain why it is important to focus on, while also show what we are focusing on.

Chapter 3 The methodology chapter lays the foundation for what we are investigating, specific methodologies used and how it was used.

Chapter 4 In this chapter, we show the results from the questionnaire and the interviews, and also attempt to explain explain our findings.

Chapter 5 The discussion chapter is where we discuss our findings, and also show how we interpret some of the data we managed to gather when performing the methods found in the chapter 3. We will also talk about the limitations of our study and what a further study should include.

Chapter 6 And in the last chapter, we conclude and end the thesis.

Chapter 2

Theory, Background, and Existing Literature

In the last two decades, AI has been particularly helpful in transforming the mechanisms and limits of many industries. In order to understand the scope of this study on chatbot and human interaction and what factors to consider when designing chatbot, we must explain. We also have to look at the difference between chatbots and Conversational Agents (CAs), while also discussing the need for specialized robots. In this chapter, we discuss the theoretical basis of CAs and personality, insight into necessary components, and attempt to give the reader a better understanding of the broad scope of teaching AI to function on a large scale.

2.1 Human-Computer Interaction

Human-Computer interactions are increasingly becoming more common, as more and more companies employ AIs as the first line of support with a large population of customers requiring assistance. This is more common in product-based websites that focus on sales (Kindly n.d.[a]; Kindly n.d.[b]; Kindly n.d.[c]; Kindly n.d.[d]), but also in other ways such as banking (Boost.AI n.d.[a]; Boost.AI n.d.[b]), insurance (Boost.AI n.d.[d]), telecommunications (Boost.AI n.d.[c]), and more.

The questions of most visitors who require assistance are often solved quite quickly and without the aid of human customer support agents. And it can even give a 75% reduction in live chat inquiries that require employee interaction (Kindly n.d.[d]). However, as noted in Elliott (n.d.) citing a report by CGS, "Consumers report frustration with chatbots arising from misunderstood questions, irrelevant responses, and poor integration with human service agents". From this we can understand it as most people tend to struggle with being misunderstood, and the questions they ask are not forwarded to the human service agent.

2.1.1 Artificial Intelligence, Machine Learning and Neural Networks

Since 1760, manual work to other products has been off-loaded or even replaced by technological revolutions. Many of which have allowed humans to increase productivity and efficiency in producing new innovations. In the mid-1900s, the first digital computing devices were presented and used (Copeland 2004). From there, the spark of innovation leads to many high-performance electronic devices. With high-performance electronic computers, the possibility of creating a machine whose intellectual capabilities would be close to, if not identical to, that of humans and even surpass human capabilities would be probable. AI is one of those technologies that has spurred during an increased focus on the computerization of tools. Artificial Intelligence is an offshoot branch of technology that allows machines and computer programs to simulate human intelligence to perform various tasks that require human intelligence.

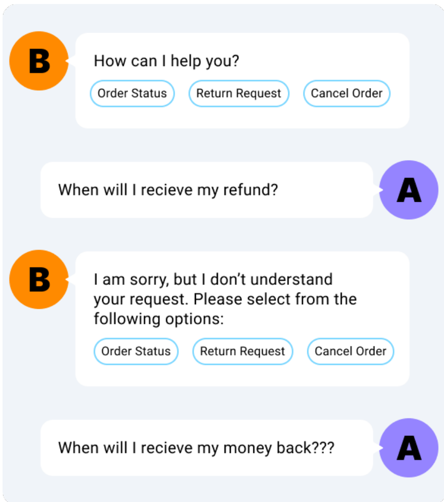
Humans have a fundamental set of basic capacities that are possessed by many other animals, and common sense is one of them. AI often lacks the common sense aspect of human understanding. This has proven to be a topic of interest and is one of the biggest challenges faced in AI (Davis and Marcus 2015).

In order for any Machine Learning (ML) to function correctly and generate a Neural Network (NN) that is usable for the purpose of speech, large datasets are generated to create a list of statements that correlate to other statements. These are often created by crawling websites to generate a large number of datapoints from which to train. There are multiple currently available, each of which contains everything from several thousands to millions of lines of text. These include datasets such as Customer Support on Twitter (Thought Vector and Axelbrooke 2017), Yelp Reviews (Yelp n.d.), Discussion on Reddit (Baumgartner n.d.), Empathetic Dialogues (Jairath 2022; Rashkin et al. 2019), conversations spanning multiple domains and topics (Budzianowski 2022), or topical conversations (Gopalakrishnan et al. 2019) to cite only a few.

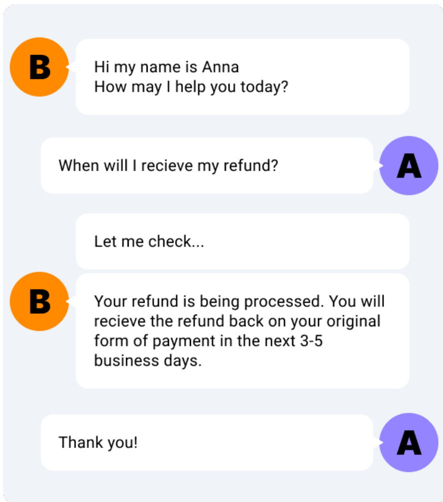
2.1.2 Chatbots and Conversational Agents

Chatbots and CAs are often used interchangeably by those who are not familiar with these terms and how they are created. However, there is a noticeable difference in terms of their functionality and application. A chatbot is a computer program or machine that can respond to your prompts in a “chat”, and the most common way to interact with chatbots is by text. These programs are often based on prewritten questions and answers and do not have the ability to change or modify the answers provided (Figure 2.1a). Although the term chatbot fits within the umbrella term CA, they are classified as a less intelligent form of CA (De Angeli and Brahnam 2008). The chatbot can be traced back to Eliza in 1966, which examined keywords received as input and then triggered the output according to a defined set of rules (Weizenbaum 1966). This method of output generation is still used by several chatbots, although some advancement has been made by adding NLP to the process. Although fully developed CA (or Conversational Artificial Intelligence (CAI)) refers to a host of AI technologies used to allow computers to converse in an "intelligent" manner. Using NLP to allow the agent to process and automatically respond to a request using human language and can collect additional information based on the request from the user (Figure 2.1b). This is of course not limited to text alone (Sciuto et al. 2018), but in the scope of this project we will focus mainly on text. These agents are representative in showing the practicality of computational linguistics and are usually employed as chatbots on the Internet or as assistants on portable devices (DeepAI 2019).

Although a combination of low-intelligence and high-intelligence is used to generate the final output of a customer support agent. Often this is heavily biased towards a certain writing style defined by a set of guidelines[1].

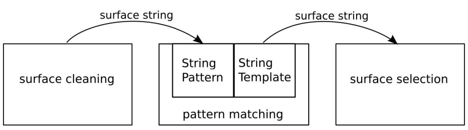


(a) Rule-based Chatbot

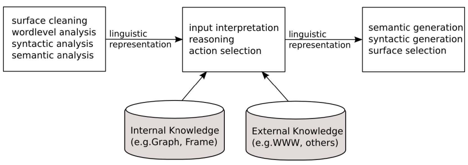


(b) AI-powered Chatbot

Figure 2.1: Examples of two types of chatbots and how they may function. The B used in this image refers to Bot, while A is customer.



(a) The core of a traditional chatbot architecture (Klüwer 2011a, p. 4)



(b) A prototypical dialog system architecture (Klüwer 2011b, p. 5)

Figure 2.2: The basic architecture of both a low intelligence chatbot (a) and conversational agent (b).

[1]These guidelines may be based on how a company's sentence structure and how the company wish to appear. Certain writing styles seem to be preferred for customer support. For example, friendliness (Brown and Sulzer-Azaroff 1994; Engel et al. 2013) and empathy (Clark et al. 2013; Varca 2009; Wieseke et al. 2012) appear to be high on the scale.

The similarity of how the low intelligence chatbot functions in figure 2.1a is easily understood when viewed in context with 2.2a, where we can clearly see how it requires certain words to match a pattern for which it will use to respond. In a similar fashion, Fig. 2.1b could also be a bit easier to understand while viewing in context with Fig. 2.2b.

We will mostly refer to Conversational Agents when speaking of chatbots, except in some cases where we state otherwise. This is to avoid misunderstandings and to simplify the meaning of the many terms that can refer to CAs (e.g. Conversational Neural Network (CNN), CA, CAI, Virtual Assistant (VA)).

2.1.3 Data Drift

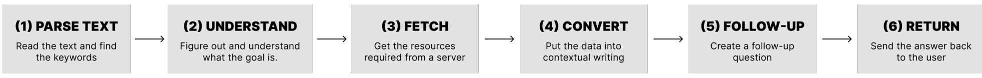
The term “Data drift” refers to the degradation of AIs’ model over time. This is due to a change in input data “For machine learning models, data drift is the change in input data of the model that leads to model performance degradation” (@buchananwp et al. 2021). There are many different types of data drift that can occur; however, the most common is concept drift and refers primarily to the supervised online learning scenario (Gama et al. 2014).

The term concept drift can be described as a type of change that was not foreseeable. Lu et al. (2018) summarized it quite effectively as “Concept drift describes unforeseeable changes in the underlying distribution of streaming data over time”. Simplified to be “the situation when the functional relationship between the model inputs and outputs changes” (Tannor 2021). This effectively means that the data have varied sufficiently from the initial training data for it to be counted as statistically significant, which would lead to the current model not fitting or understanding the input. In most cases, tools based on Continuous Reinforcement Learning (CRL) do not have this problem, as they are gradually adapting based on the input.

2.1.4 Using Big Data for Personality traits

One study has shown that using automated methods to extract and analyze digital footprints in social media, you could effectively predict which of the Big 5 personality traits a person has (Azucar et al. 2018). They argued that “social media-based predictions can then be used for a variety of purposes, including tailoring online services to improve user experience”. This also correlates to what Bleidorn et al. (2017) argued, in which giving unconstrained access to digital traces would allow researchers to detect personality indicators not found through lexical or deductive approaches. Additionally, they argued that ML algorithms can “help answering some of the most puzzling questions in contemporary personality theory involving the boundaries of personality traits, personality development, and cultural influences on personality.”. Interestingly, digital footprints can be quite easily converted into personality traits, Chittaranjan et al. (2011) showed and concluded that smartphone usage could be used in predicting which of the Big-Five personality traits a person has. Therefore, we can state fairly clearly that using a large

(a) Task-Completion Systems



(b) Social Oriented Chatbot

Figure 2.3: Simplified examples of chatbots flow.

population to gather personality traits and their digital footprints, we can easily gather data that can be used to improve and adapt Machine Learning models to specific personality traits.

2.1.5 Social chatbots vs Task-Completion Conversational Systems

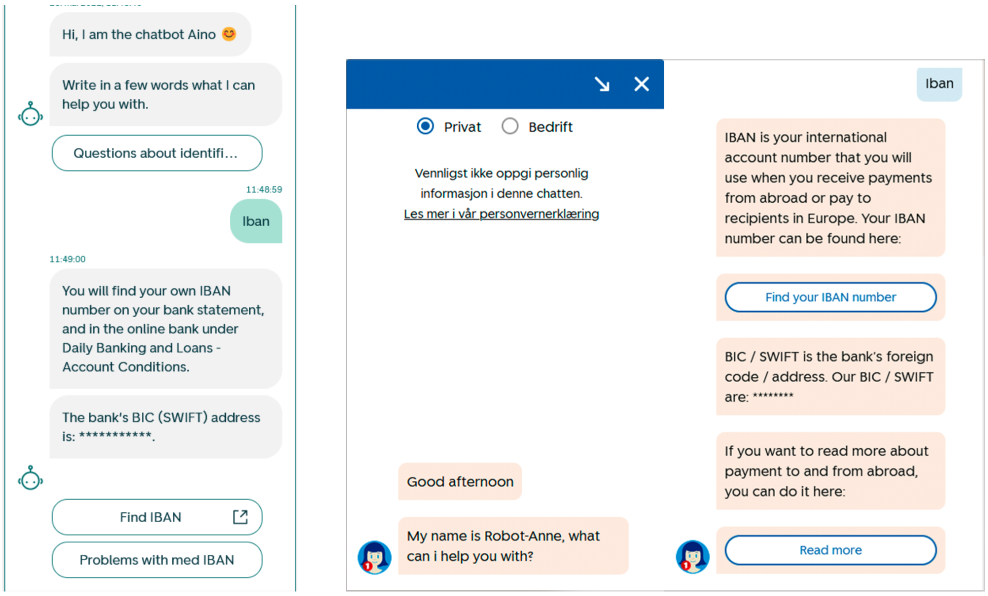
We can often attach a set of predefined categories on which type a chatbot is. This can be done through what sort of topics they can discuss and how they function, mainly if they are open to unknown questions and will attempt to answer them, and if they only respond to certain questions.

Task-Completion Systems

These types of systems are designed only to perform specific tasks and are usually operating in constrained domains (Glass et al. 1995; Raux et al. 2005; Wang et al. 2011). The goal of such a system is to (1) parse text, (2) understand what the goal is, (3) retrieve the required data from a server, (4) convert the data into contextual writing, and (5) return an answer (Fig. 2.3a). This sort of system is very useful for personal assistants, productivity tools, and other similar use-cases. Its goal is to quickly solve a problem or find information based on a prompt.

A good example of a task-completion system is a movie ticket reservation, which has been proven to outperform the baselines of modularized dialogue systems for objective and subjective evaluation. However, this comes with the added risk that the model predicts the wrong intention and may mistakenly purchase a ticket for another movie (Li, Chen, Li, et al. 2018, February 11). This might limit its usability if this type of error happens to a significant degree, which might be 2-5 times per 5000 runs or even more often depending on how many users it should service.

This sort of chatbots are often used for question-and-answer type of work, which could include information fetching, service work, guiding users to correct pages on a website. An example of this exact thing can be seen in figure 2.4.



(a) Chatbot used by DNB **(b)** Chatbot used by SMN (1 of 2)

Figure 2.4: Example of a chatbot used by two of the banks in Norway using fetch based search looking for information on IBAN.

Social Chatbots

In contrast to task-completion systems, these types have the main purpose of just talking. The goal they have is not to just answer a question, but rather to keep a conversation going. This type of chatbot is sometimes referred to as "chit-chat" dialogue chatbot. The fundamental overall structure of the flow is similar to that of the task-completion system. However, social chatbots have an additional goal, which is to add a follow-up question related to the topic presented in step 1 (Fig. 2.3b). Although a large amount of research has been done on chit-chat models, extended conversations with these types of models often quickly expose their weaknesses (Lowe et al. 2016; Vinyals and Le 2015). A recent study has shown that with a modified dataset focusing on persona, participants were more engaged than previous datasets (Zhang, Dinan, et al. 2018).

A fairly recent study on whether the addition of a supportive module that focuses on chit-chat tasks would improve task-oriented dialogues showed that it could increase the likelihood that the new model would appear more engaging, interesting, knowledgeable and human-like, without losing out on task performance (Sun et al. 2021). Their proposed model was explicitly trained to predict user goals and to generate contextually relevant chit-chat responses. Which to us indicates that adding a chit-chat module with some personality to the chatbot would likely increase usability; this also correlates with what Smestad and Volden (2019) found about personality matters in chatbots.

2.1.6 Trust in chatbots

Customer service is perhaps the most common use of chatbots in this decade, which has both its benefits and negatives. Human customer support agents are required less and less for answering easy questions that have an answer already on the website, and can focus more on answering questions that are higher priority such as; Faulty software, problems a user is facing, account problems, etc. In general, questions that require more inference and human intuition to answer correctly. A study on human-computer trust Madsen and Gregor (2000) defined trust as “the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid”. This is also what we believe to be an accurate representation of what trust is in Human-Computer conversations, and therefore is the basis for trust factors we use in this thesis.

Følstad et al. (2018) found that the most common factors affecting trust according to his participant. The top two factors he listed was 1) “Quality in interpretation of the user request and advise in response to request” and 2) “Human-likeness. The chatbot’s appearance as human-like, personal, or polite”. He also noted one of the participants stating “having some kind of personal or relational flair to its style of communication to potentially enhance trust”. In opposition to this, another one of his participant argued that “a too humanlike chatbot could even reduce levels of trust”. Interestingly he found that participants “specifically noted that the efficiency in interaction would be decisive for their future use of customer service chatbots”. We interpreted this as a factor which both is related to trust and usability, as user experiences are affected by the tool or program they use, a more accurate software would of course be more trustworthy, and the likelihood of a user returning to this software being greater. This is further corroborated by the findings of Nordheim (2018), which also found that the correct answer and the interpretation of the question were a large factor in determining trust. While there are many reliable reasons for a chatbot, since many chatbots operate on a black-box principle, the opposite is also true. This means that we know what we put in and we know what we get out, but how the machine functions and interprets the input to make the output is unknown.

2.1.7 Dangers, ethics, and problems of Artificial Intelligence

AI is not without its problems. The main problem facing AI and NN is that they are often built using Reinforcement Learning (RL). Although there is no problem with using RL in its own right, it can become a problem once CRL is integrated and used. This is due to the way CRL can be used to improve the chatbot based on the data provided while the program is running. A study that looked at the security difficulty of using chatbots noted that “[. . .] chatbots even ‘learn’ from these data. This means they could be considered to be a security problem” (Hasal et al. 2021). This is true in most cases, unless one specifies to stop training on current and future data. Though this cannot be considered as a future proof system, we can also assume that it is not malleable for continuous use. We can also assume that it requires retraining every once in a while, which would perhaps also mean retraining it fully from scratch. This is so that we can avoid data drift and losing the accuracy of the utilized model.

A separate but still viable problem with AI that arises in cases of use is that if not properly managed and trained, the chatbot could become a failure, which disappoints users and causes less engagement with the chatbot. Fossa and Pisa (2019) argued that “When an AA [CAs] fails to achieve the goal it is programmed to pursue, users ought not to interpret this failure in terms of betrayal, but rather in terms of disappointment. Disappointment refers to functional expectations that are not met and, as such, is the appropriate reaction to reliability issues.” which he continues on stating “repeatedly disappoint their users are unreliable—and ‘untrustworthy’ ”. A study on chatbots and customer service found that “positively disconfirms customers’ normative expectations will lead to greater customer satisfaction, across positive and negative emotions” (Menon and Dubé 2000). In section 2.2.3, we will further discuss this exact topic of anthropomorphism in chatbots.

A further danger, which is one of the more severe problems of AI, is that there is no “filter” that limits speech based on inference and future thinking. This can be exemplified by an experiment performed on a model called GPT-3, where this model implied that a patient should kill themselves (Riera et al. 2020). The nature of AI makes this an extreme possibility of an AI harming a human if not properly managed and trained on how to handle emotions, sarcasm, human nature, and ethical behavior.

Ethical implication of AI

Ethically, the ‘learning’ part of NN is the most troublesome to deal with, as this could lead to unforeseen circumstances. Let us use a banking service as the example, in this case, we train the chatbot to answer questions based on a set of documents and frequently asked questions. This allows us to create a chatbot that can easily respond to general questions and is usable in a general sense. Further training is done, and it is taught to answer specific questions related to a users account. We leave it open for questions, which it will answer as best it can. It gets asked a question related to “Expenses incurred”, which it interprets and looks for the answer in the user’s data. These data are then collected, summarized, and sent to the customer. We have not specified that it should not remember anything it searches for, so it accidentally learned some new information that contains the previous user identification number. Another user asks a similar question and have mostly similar data, which then the bot uses to answer. However, it had collected a set of numbers from the previous user, which was not related to the current user. This is then used to answer the current user with that data added on. Now, is the AI in the wrong for using said data to answer the question? In this exact case, yes, since the data is related to another user and could be misused if detected.

Thinking back to Bryson (2010), where he argued that “We determine their goals and behaviour, either directly or indirectly through specifying their intelligence or how their intelligence is acquired." Are we at fault for providing this option for the chatbot with the ability to learn from data, without thinking of the implication on a global scale?

2.1.8 Benefits of AI

While the dangers and ethical implications of AI provide a stark reminder of what we should be concerned about. We believe that the benefits, if properly maintained, can outweigh the negatives. For example, as stated by Luxton (2020) about chatbots; "[. . .] can easily be replicated (scaled-up) and affordably to meet demand, and, unlike with human professionals, users can access them via the internet at any time and almost anywhere.". Therefore, we can conclude that while we must keep vigilant in how AIs are used, the application of NN systems to build software that improves the lives of users is important. AI does also provide a better option for

In the paper by Huang and Rust (2018), he noted that AI could be separated into multiple stages of learning, mainly the "Four Intelligences" as he puts it. Mechanical, Analytical, Intuitive, and Empathetic. Each of which comes with their own benefits but are still built upon the previous state (E.g. Analytical is the next step of Mechanical). While discussing the multiple uses of AI he noted in the mechanical intelligence section that;

> Mechanical intelligence concerns the ability to automatically perform routine, repeated tasks. [. . .] Mechanical AI has the relative advantage over humans of extreme consistency (e.g., free from human fatigue, and responding to the environment in a very reliable manner).

Since AI does not have the same limitations that a human has, the possibility of user errors is less likely to occur. This makes it highly suitable for repeatable tasks and tasks that require high precision every time. This is probably the largest benefit of any AI, that it can function without the need to sleep. We often discuss human errors within user experience, as we have to design for it to be possible. This is something we also have to consider when designing AIs, as the probability that a human asks a question that an AI does not understand is highly likely. We work around this by adding a slightly higher intelligence to the ai, which could then be said to be "designed to be just intelligent enough to perform the necessary tasks. Intelligent search by Google, Bing, or other search engines are another application [of this]" (Huang and Rust 2018). Other benefits such as an assistive tool for the elderly (Demir et al. 2017), while this often requires a supportive method of text-to-speech, the possibility that this method works in favor of the elderly is high.

Cui et al. (2017) found that they could create a chatbot that could scrape e-Commerce websites to help users find the information they are looking for. This chatbot used both a chit-chat module and a task-oriented (Q&A) type module to create a useful and convenient way to summarize and find information on product pages that contains too much user-generated content. This is a viable and extremely user-friendly method of using chatbots to solve a problems with finding information in a data-heavy environment, such as Amazon's user reviews on products.

Therefore, we can conclude that there are multiple applications of chatbots (and AIs), and that the probability of chatbots becoming a more assistive tool in the future is highly plausible. This also includes the need to improve and adapt chatbots based on the needs and wants of users to better fit their style of information gathering.

2.1.9 Limitations in chatbots

Since in most cases, chatbots are trained on large datasets, with data from many different topics and methods. We know that the efficacy of general-purpose humanoid robots has not yet been proven and is still being researched; in contrast, we have seen that specialized robots have been shown to be highly usable and effective in performing their tasks (Sheridan 2016). This also relates to chatbots, which could be considered a subtype of robot just without the physical aspect of it. Merriam-Webster (n.d.[b]) defines a robot as a device that automatically performs complicated, often repetitive tasks.

Li, Chen, and Chang (2019) found, while analyzing conversations by 1,837 users with a chatbots, that users often failed to progress in a conversation when they requested information rather than when they provided information. They also found that the wast majority of user (71.5%) quit the topic when users did not make progress rather than rephrase or retry the question, which was done by a minority of users (28.5%). This would actually be something that is a large limitation of current generation chatbots. A later more in-depth study focusing on the same information by (Li, Yeh, et al. 2020, April 21), found that most users also still quickly abandoned the chatbot if it misunderstood or did not recognize the question asked by the user. This indicates to us that there is a huge problem with the chatbots not understanding and cannot infer what the user actually wants as well as a human agent would, and this would in fact limit the usability of this type of chatbots to users with higher patience for errors. This would not include people with a heightened emotional state, such as when a problem has occurred and they require assistance as soon as possible.

We could discuss the possibility of having a "Contact Human" button, but that would actually be detrimental, since most users would likely attempt to contact a human first, bypassing the chatbot entirely. This is something we would like to avoid, as chatbots are actually quite useful in sorting out which users have questions that can be found on the website and which users actually require the support of a human agent.

2.2 Emotional Intelligence

The term “Emotional Intelligence (EI)” was first coined in 1990 by Salovey and Mayer (1990), where the model was defined and explaining what EI is. This model came in the form of a set of four major areas; (a) “Percieving emotions”, (b) “Facilitating thought by using emotions”, (c) “understanding emotions”, (d) “managing emotions in oneself and others”. They later revised into a set of seven principles (Mayer et al. 2016), which expanded on, and corrected some problems with, the previous model. The most notable was that they added the second principle “Emotional Intelligence Is Best Measured as an Ability”. This included the statement of “People are poor at estimating their own levels of intelligence — whether it is their general intelligence or their emotional intelligence”. Not that some individuals might have a better understanding of oneself but that we need to establish a baseline for how we can detect whether a person is individually more sensitive to emotions. Therefore, we concluded that a check was needed to understand if an individual is sensitive to emotions. This was done using a set of tests called HSP.

2.2.1 Highly Sensitive Person

Psychologist Elaine Aron, who studied how to identify high sensitivity in people, coined the term Highly Sensitive Person. HSPs are a subset of the human population that ranks high in the personality trait commonly known as Sensory-Processing Sensitivity (SPS). The Highly Sensitive Person Scale (HSPS) was developed during a series of seven studies on SPS. In this study, they identified a set of 27 questions that relate to the unidimensional core variable of high SPS and, in addition, demonstrated its partial independence from social introversion and emotionality, maintaining reliability and discriminant validity (Aron and Aron 1997). The basic personality trait SPS has been linked to a genetic basis (Assary et al. 2021), allowing us to conclude that there are people who would be more likely to respond positively to the emotional intelligence of another subject. This also includes the other end of the spectrum, where the opposite is true. Aron, Aron, and Jagiellowicz (2012) also associated SPS with emotional reactivity, empathy, awareness, self-other relations, reward processing, and reflective thinking. Sensitive people have been shown to easily be overstimulated and affected by the moods of others (Acevedo, Aron, Pospos, et al. 2018; Acevedo, Aron, Aron, et al. 2014).

Therefore, we conclude that HSP directly affects how the user detects emotions in written language and speech. This allows us to base our finding on a set of predefined standards known as HSPS. This also allows us to gather information based on the assumption that the user would prefer how they wish the interaction to function.

A separate study by Smolewska et al. (2006) found that HSPS could be categorized into three factors. These factors were labeled Ease of Excitation (EOE) (the feeling of being mentally overwhelmed by external and internal demands), Aesthetic Sensitivity (AES) (aesthetic awareness), Low Sensory Threshold (LST) (unpleasant sensory arousal to external stimuli) (Smolewska et al. 2006).

Several studies have questioned the reliability of this method (Evans and Rothbart 2008; Rinn et al. 2018), however, we can conclude that, for our purposes, this method is accurate enough and its validity is valid. Therefore, this method will be used to compare our data with and attempt to see a correlation with the wants and needs of chatbots.

A recent study by Trå et al. (2022) found that there is a significant difference in gender and their self reported HSP scores, therefore, we can effectively assume that there is a likelihood of the different genders affecting the choices in wants and needs. This relates to Research Question 2.c found in section 1.1, which we will discuss in further detail in Chapter 5.

2.2.2 Emotional Intelligence, Emotions, and Personality

EQ is the ability to monitor oneself and others' emotions, to discriminate between different emotions and label them appropriately, and to use emotional information to guide thinking and behavior (Colman 2009). The application of a personality to an object allows us to humanize it so that we can form a sort of bond with the object. This can be seen in what is known as *Tamagotchi Effect*.

The *Tamagotchi Effect* is based on a virtual pet with simulated needs, which the user would be required to take care of. If it did not receive sufficient care, the pet would die. "These simple devices did stimulate a range of emotional responses in users [...] The emotional impact is not surprising given the animistic tendency, which some people believed to be a novel effect and labeled the Tamagotchi effect [...] " (Frude and Jandrić 2015). Lawton (2017) showed that "Tamagotchi was first to execute a form of mobile technology designed specifically to elicit an emotional response from the user". This is probably the best example that can show how people can emotionally connect to personified creatures; this is also highly related to how we personify chatbots by anthropomorphizing them through expectations and connecting to them emotionally. One of the benefits of connecting with AI is the ability to let it learn from you, and if it is a CRL algorithm, as Frude and Jandrić (2015) stated perfectly, "the machine will gradually adapt as it comes to appreciate the user's tastes and preferences".

Emotional Intelligence in Conversational Agents

CAs are often defined by the data used during training, and this forms the basis for its personality. By biasing the towards a certain writing style, this allows us to customize how much the chatbot will focus on creating a similar text as the training data. The limitation of intelligence and personality is mainly based on the training data available, which is why large companies often have billions of samples from multiple datasets used to train a chatbot. This also helps alleviate the problem of bias, since with the support of a large sample, we can avoid common training pitfalls.

2.2.3 Humanness and Anthropomorphism

Humanness is a set of characteristics that is defined by an agent's design of acting and appearing human.

> [...] the extent to which an agent is designed to act and appear human [...] encompasses the objectively established human capabilities (having eyes, a face, or the ability to respond politely) of an agent by design and is distinctively different from anthropomorphism, the psychological attributions of human features to non-human agents (Meyer et al. 2016).

Within a similar spectrum, there exists anthropomorphism, which is the attribution of human form, mental states, or personality to objects that are not human (Merriam-Webster n.d.[a]; Waytz, Cacioppo, et al. 2010), such as animals and objects. While this type of attribution is necessary for things to appear human, we wish to invoke humanness over that of anthropomorphism. Since anthropomorphism is related to perception (Waytz, Epley, et al. 2010), humanness will be related to how an agent functions. Thus, since the goal is to appear as human as possible, the term humanness would be the correct term to use in our case. However, we should note that currently we force the chatbot to use words to distance itself from appearing entirely human through sentences such as "Hi I am Robot-Anne" (See Fig. 2.4b). This is due to us attempting to slightly lower the expectation of the human interacting with the chatbot and making sure that they do not think it is a human. Anthropomorphism does not help the chatbot in all cases, as a recent study found that chatbot anthropomorphism has a negative effect on customer satisfaction when customers enter a chatbot-led service interaction in an angry emotional state (Epley et al. 2007). This is due to people expecting adequate help in times of need, and most chatbots have an overblown expectation of efficacy, which is subsequently violated, causing dissatisfaction towards the brand.

In regular chatbot conversations (e-Commerce, help desks, question-and-answer types), we wish that the chatbot maintains the standard set by the company in how it is supposed to talk and not go any further. However, there is the opposite of this, that is, having no restriction that limits the scope of AI. This was applied by a team from Microsoft that applied the principle of human communication to a chatbot, without the limitations of trying to stay within a preset boundary, such as company policy. This chatbot was named XiaoIce and has been the subject of more than 10 billion conversations with humans (Zhou, Gao, et al. 2020) and has shown that it can emotionally connect to the conversation it was having. "XiaoIce can dynamically recognize emotion and engage the user throughout long conversations with appropriate interpersonal responses" (Shum et al. 2018). XiaoIce is to be considered as a chit-chat robot (explained in Sect. 2.1.5), which focuses on conversing for a longer period rather than quickly solving problems.

Contrary to popular studies on the personification of chatbots, Bryson (2010) argues that robots should not be described by personas, nor should CAs be anthropomorphized, as this "invites inappropriate decision such as misassignations of responsibility and misappropriations of resources". He concludes with the statement that "robots are often overly personified", meaning that we often attempt to humanize and anthropomorphize robots to such a great degree that it "can lead to a large range of category errors which can significantly bias decision making" (Bryson 2010). It is not uncommon to expect more from a chatbot than it can actually do, this is due to us thinking it should function more or less the same as regular humans. We can argue that by humanizing a chatbot too much, we will automatically attribute even more human-like traits to the chatbot. For example, we would probably expect more from them, be less tolerant to simple failures, assume that they would know more than they do, and expect interpretation. This is also a problem that chatbots are facing when it comes to the ethical implication of answering the wrong things, without thinking of what it can cause in damage. This is something we currently are not really expecting chatbots to do, as they usually have little to no ethical value in their thinking. According to some studies, this is due to the lack of training data (Lin et al. 2020). This has been somewhat mitigated by additional training and data gathering, but we can assume that a lot of training is required before it is ready for anything empathy-related tasks without a fear of it misspeaking and going something bad.

Due to the current intelligence in chatbots, we are closing in with the speed of light on the drop on the uncanny valley effect. We know this since current chatbots are so close to humans in their speech, but we still know that they are robots based on their own admission when we start a chat. This was added to the program so that we did not experience the feeling of "something working like we expect", but there is something slightly off which puts us in a defensive mode.

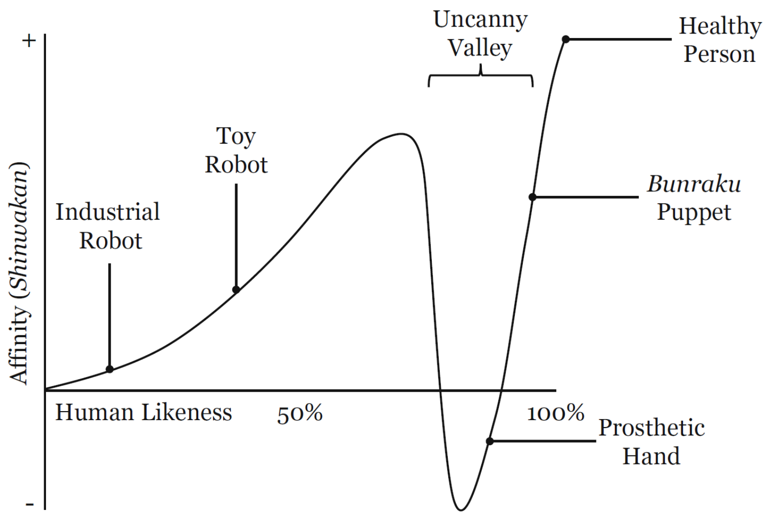
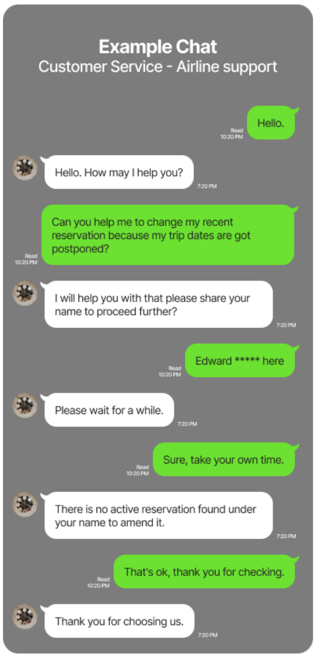
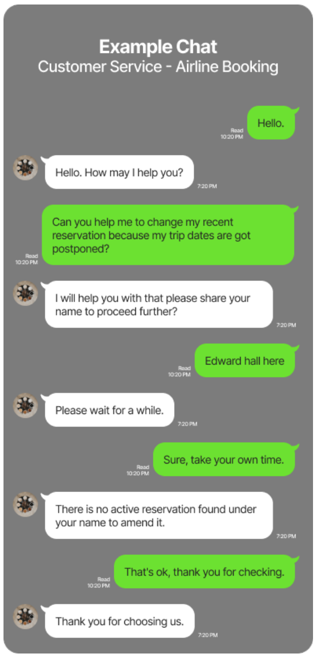


Figure 2.5: The Uncanny Valley (Mori et al. 2012)

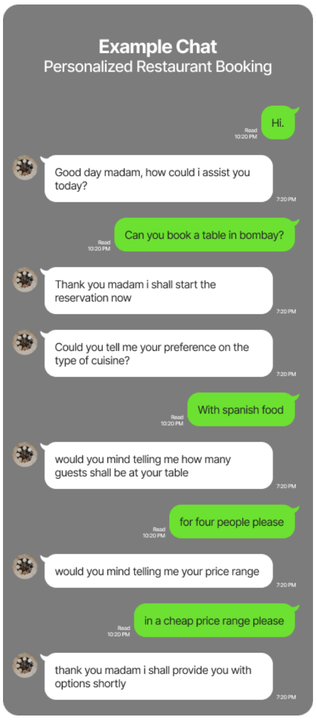
The uncanny valley effect is often defined as a feeling of eeriness and discomfort towards a given medium (Caballar 2019; TechTarget n.d.). Mori et al. (2012) attributed the sense of eerieness as a form of instinct that protects us from sources of danger. These dangers included corpses, different species, and other entities that we can closely approach. We should also note that the literature on the uncanny valley effect is lacking, and in a review of the literature by Zhang, Li, et al. (2020), they stated that “there are some problems in the existing studies, which may lead to inconsistent findings on this topic”. Studies of the Uncanny valley effect in chatbots found that simple chatbots induced less intense psychophysiological reactions than those with more complex animated avatars (Ciechanowski et al. 2019; Skjuve et al. 2019). The feeling of eeriness could lead us to not wanting to interact with the object, which is something we would like to avoid. Thus, we would prefer to avoid this feeling by prefacing our conversations with the statement that “I am a bot” or similar indications.



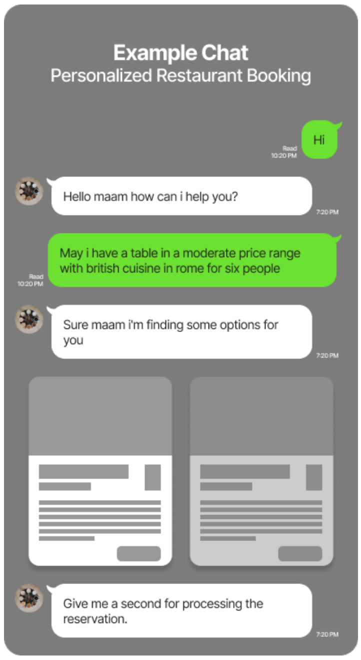
(a) Customer Service - Airline Support



(b) Customer Service - Airline Booking



(c) Personalized Restaurant Booking 1



(d) Personalized Restaurant Booking 2

Figure 2.6: Example of how a chatbot might interact through a chat application through goal related tasks. Fig. a and b uses sample text from (Wei et al. 2018), while fig. c and d uses sample text from (Joshi et al. 2017)

Chapter 3

Methodology

The previous section laid the foundation for the background of chatbots and linked user behavior and humanness. In the following section, we will describe the methodology for this thesis, which consists of several parts. This includes, how we recruited the participants, what payment they received for participation, what tests were done and what tools were used.

3.1 Participants

Participants were recruited by various means, but mainly through the author's own network and easily accessible people at the time of the study. Participants were selected on the basis of volunteerism and freely agreed to participate in this study without any incentives. Participants were required to perform all activities on their own computer, but we did not require them to come to any location for testing. This was due to the need for an authentic experience. An authentic experience in this case means that they were in familiar locations, using their own equipment, while talking on a Web chat service. This was all to simulate how it actually functions currently. The questions posed was mainly focused on use and familiarity to topics such as; Chatbots, smart assistants, personalization, and privacy (See appendix D for questionnaire). The sample size we ended up with was $N = 38$ volunteers, who donated their time to participate in a questionnaire for project and became the main subject behind this study. From this an estimated 28 were from Norway, and 8 from other countries (Unknown as the participants were not asked where they were from). With an additional 12 for an in-depth interview and further study, which were all of Norwegian origin, but would classify themselves as having good to excellent understanding of English. All participants were informed as to the ability of withdrawing from the study at any point (see first page of appendix D).

3.2 Questionnaire

A main portion of the study was the questionnaire, which we used to gather baseline information from participants. We split it into multiple section, which all related to use of smart assistants, preference in chatbots, and the HSP test. This allowed us to get a feeling of what the participants knowledge about smart tools and chatbots were, and also to get the information about the participants sensitivity scale. The questionnaire was hosted on Nettskjema, which provide a secure storage of all data gathered.

3.3 Testing and interview

Prior to anyone being able to participate in the interview, they were informed as to the purpose of the study, their roles as participants, and that they would be able to withdraw from the study at any time (See appendix E). We were allowed to do recordings of voice and video, however, we believe that this would be less relevant and would only be detrimental in our case. The interview was split into five main components;

The introduction Which gave us time to introduce the project, what the goal of the study was, and make sure the participant was able to correctly consent to the study. This interview consent form was sent digitally to the participant, and was filled out on their end and returned. The transfer of documents were done on a secure OneDrive folder managed by NTNU, which was only accessible to the researcher and the participant for the introduction period of the interview and removed from the OneDrive folder to be placed in a encrypted folder on a Linux machine not connected to the internet. This is within reason set by the Data storage guide created for NTNU (n.d.).

General interview Where we gather general knowledge about the different topics we are investigating, and allow us to let the participant relax and talk about their experiences with chatbots. The questions are loosely based on questions stated in appendix F.1.

The HSP test This test is the same as the test used in the questionnaire (Found in the questionnaire, questions 4.1 - 4.27 in appendix D)

Wizard of Oz This technique is described more in detail later on, but is a way for us to quickly build a testing system that can be used to gather accurate testing data.

Follow-up Which consisted of what they think about the chatbot, experiences had during the testing, and if they would liked to try a another chatbot based on the ParlAI product (See 3.3.3 - "Real Chatbot" for more details).

3.3.1 Persona

When developing the Wizard of Oz method we know what information should be available for user to use, this is where the persona tool comes in handy. This character would influence how we would build the chatbot for testing and what we were to focus on. This persona was also added to help guide the users to what state of mind they should use when talking to the chatbot. We created two simple personas that could be used in both of these cases, see appendix H. These personas are adapted on the fly to the participant.

3.3.2 Wizard of Oz

The main test method was to perform a Wizard of Oz test, in which participants were allowed to communicate with a "chatbot" through an online communication program (Rocket.Chat). Although there has been an argument about the ethical implications of using this method (Fraser and Gilbert 1991), we believe that the benefits outweigh the negatives in this case. The purpose behind this selection was based on the need to accurately simulate a chatbot over multiple tests while maintaining consistency in repeated tests. While also making sure that the bot answered correctly. If the participant required some help getting into the mindset of a specific person or goal, we provided a persona from which they could use to get their mind into the correct state, help with questions, and clearly defined what their goal was going to be. The "chatbot" used in this test was for the explicit purpose of letting the user gain some experience to which they could base their subjective answers on. A set of questions (See appendix F.2) relating to how the experience was asked afterwards.

Although we did not create a laboratory experiment, as explained in 3.1, the authentic experience would actually prove to be more useful for our case rather than detrimental. This is due to people being more likely to answer questions in a more relaxed state, and this allows us to see what degree the users were susceptible to changes in methods of writing. As an example, we have the following methods of greeting a user.

#	Message	Happiness	Surprise
1.	Hi <User>! Welcome to <Service>. What can I help you with today?	0.5	0.5
2.	Hi, what can I help you with?	0	1
3.	Hi. Do you require any assistance?	0	0

Table 3.1: This table showing messages used to greet the user, with their respective emotional state extracted with the python library text2emotion

There were created two main chatbots with two main personality traits, which had their own supportive list of answers made before the test. The premade messages used in the trial was created with the help of the dataset MultiWOZ 2.2

created by Zang et al. (2020), and adjusted to be either objective or human-like in response text. The expected questions were created ahead of time, and with a support through a dataset of answers from the MultiWOZ 2.2 dataset which could be used to find other answers if something was missing. The supportive data was created by running the dataset through the code seen in appendix J.2.

3.3.3 The chatbot: Brand, Usage, and Personality

To build the prototype of the chatbot, test the personality, and adjust the output, the chatbot will be based on a real brand and organization (hereafter the brand). There is, however, no collaboration with any brand mentioned in this thesis, and will therefore which brand we will anonymize which was used as baseline. The necessity of using real-life brands to base this prototype on can be explained by the need for the user to have experienced a similar chatbot before to compare ours to. This allowed us to create multiple personas to model the chatbot and personas the user could use to interact with the chatbot. Multiple brands were used to allow a wider range of interviewees as we would like to adapt the chatbot to something the user is familiar with. Note that there is not a significant difference in how the chatbots interact based on the brand, and will not impact the results from the tests and only changes the perceived name of the bot.

The chatbot was produced in two methods, one where the chatbot was trained to respond to commands, another where faked. The faked chatbots were produced for multiple testing, one which were focused on restaurant booking, which was built by getting answers from a dataset produced by (Joshi et al. 2017). And the other was built to be customer support for an Internet Service Provider (ISP).

The fake chatbot used pre-made list of possible answers and keywords that could be put together. This allowed us to fake real answers fairly quickly, without relying on an external service. The goal being to simulate a chatbot which can switch between different writing styles without large adjustments or restarting a program. In order to avoid having the profile picture of the chatbot affect the results in any way, we opted to avoid using any imagery that might link to either a specific chatbot or a specific person.

Real Chatbot

While the real chatbot was trained on the mentioned dataset through ParlAI's training method (Miller et al. n.d.), and only used if the participant wanted to try it. This allowed us to get a better understanding of how well it actually function, while also letting the participants have a little bit of fun. This chatbot was for experimental purpose only, and as such all data found here should only serve as supportive information, and is something we did not put heavy weight on during the analysis.

This was only performed as an after test, and would not affect the results of the assignment, however would give an additional topic to discuss in the interview. This test was performed through the same software as the Wizard of Oz method, which was connected to a separate computer running the AI. The code used for interfacing with the chatbot can be seen in appendix J. The ParlAI chatbot used was the *R2C2 Blenderbot 3Bm*.

3.3.4 The Chatbot

The chatbot was trained and run on a fedora Linux 36 machine running a 12th Gen Intel(R) Core(TM) i7-12700K, with 80GB ram, and NVIDIA 3070 Graphics processor for CUDA cores. This machine was promptly deleted and overwritten to make sure any data was not possible to restore, based on specifications made to the NSD application.

3.3.5 The test server

The test server was created through a DigitalOcean Droplet, running on a basic server with Ubuntu 20.04 using MongoDB (version 5.0.6) as database, and NodeJS (12.18.4) as the backend. SSH hardening was done to protect the server during testing, and quickly erased after test was completed. This was done such that we could easily set up communication between the chatbot and the tester, and which would also allow for remote access with stricter user control. To solve the problem of user registration, the users themselves could not sign up and would need to get a user and password from us.

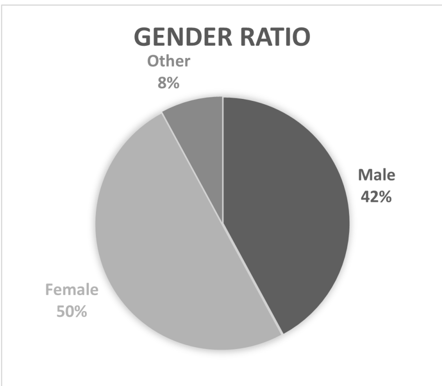
Chapter 4

Results

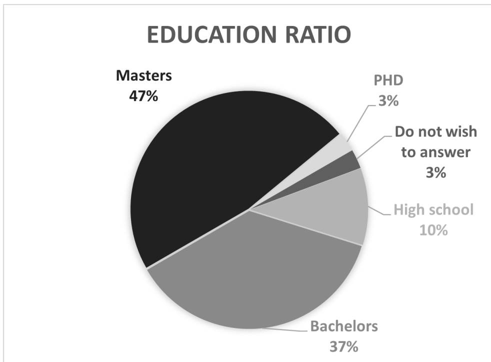
This section discusses the results found during the study. To start the section off, we had a total of 38 respondents to the questionnaire and 12 volunteers for interviews. This allows us to get a general trend of what users actually think, however, more data should be gathered to be sure, which is something we discuss further in the chapter 5.4.

4.1 Sample size

During the collection of data for the questionnaire, we suspected that there would be a higher proportion of women participating, however, this fear was unjust, as we got a decent gender ratio shown here.



(a) Gender ratio of participants



(b) Education level of participants

Figure 4.1: Participant ratios and statistical information

Participants gender ratio (4.1a) showing an overall equal number of participants of each gender, with a small number reporting as the other. The "other" both include those who do not wish to state their gender, and those who identify as other genders. Although the majority of the participants had education similar to that of a master (5 years of higher education, see Fig. 4.1b), it did not affect the results to a significant degree. We believe that the level of education does not reveal any bias in this case, and we can trust that the data are accurate. The only possible problem that we can note here is that the participants were highly educated, which might have affected our results to some extent, which we will discuss later.

Next, we performed a reliability test using Cronback's Alpha on the data available in the questionnaire Fig. 4.1, this was to make sure that the data is valid. This resulted in $N = 35$ of the 38 respondents being complete enough to be added to the test. This resulted in $\alpha = 0.829$, which is considered a sufficient measure of reliability and shows internal consistency in the questionnaire. Cronbach's Alpha usually considers ≈ 0.70 or higher to be valid and sufficient in showing consistency (Cortina 1993). The reliability test was performed with the following questions from the questionnaire (see appendix D): *Q2.2.1—Q2.2.7, Q2.3.1—Q2.3.6, Q3.2.1—Q3.2.6, Q3.4.1—Q3.4.4, Q3.6.1—Q3.6.6, and Q3.7.1—Q3.7.6*.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.829	.846	35

Table 4.1: Reliability test of the questionnaire results. *Listwise deletion based on all variables in the procedure*.

Although our number of participants for this questionnaire is not high enough to gather an exact statistically accurate representation, the number does allow us to infer a general trend.

4.2 Results from Questionnaire

In table 4.2, we can see that, although there is a strong significance in humanness on both the HSP total ($Sig. = 0.002$) and EOE ($Sig. = 0.003$), it only amounts to a moderate positive correlation (HSP total correlation = 0.480, and EOE correlation = 0.472) according to a guide on the strength of the correlation (Evans 1996). Although the correlation is moderate in this case, an increase in sample size would likely also increase the correlation strength.

Both Humanness and Professionalism traits are affected by the EOE factor within HSPS. Where both indicate that users with higher EOE might prefer understanding, friendliness, human-like traits, and other similar traits. We can also see that there is a strong negative correlation between Humanness and Efficiency, indicating that people who want more human-like chatbots (e.g. humor or friendliness) are more likely to not want objective plain-answer without a strong personality. Interestingly, the level of education does not indicate what kind of traits a person would want. But what it tells us is that HSP is negatively affected by it.

		Gender	Education	HSP_total	HSP_EOE	HSP_AES	HSP_LST	Humanness	Efficiency
Education	PC[a]	-.484**	--						
	Sig.[c]	**.002**							
HSP_total	PC[a]	**.504****	**-.502****	--					
	Sig.[c]	**.001**	**.001**						
HSP_EOE	PC[a]	.416**	-.402*	.932**	--				
	Sig.[c]	**.009**	**.012**	<.001					
HSP_AES	PC[a]	.317	-.251	.611**	.386*	--			
	Sig.[c]	.053	.129	<.001	.017				
HSP_LST	PC[a]	**.505****	**-.603****	.894**	.757**	.455**	--		
	Sig.[c]	**.001**	**<.001**	<.001	<.001	.004			
Humanness	PC[a]	.273	-.100	.480**	.472**	.160	.415**	--	
	Sig.[c]	.098	.549	**.002**	**.003**	.336	**.010**		
Efficiency	PC[a]	-.155	.128	-.214	-.190	.060	-.270	**-.807****	--
	Sig.[c]	.352	.444	.197	.254	.722	.101	**<.001**	
Professionalism	PC[a]	-.021	-.004	.301	.352*	.042	.202	.272	-.006
	Sig.[c]	.901	.983	.067	**.030**	.803	.224	.098	.971

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. Pearson Correlation

c. Sig. (2-tailed)

Table 4.2: Correlation of all calculated variables against Gender and Education.

We did not include the type of education in the study, but this could be something to further study if there is a correlation between HSP, traits, and background.

We can infer from this that education affects the probability that a person would prefer less human communication, and rather they answer the question quickly by seeing how education affects the HSP total in Fig. 4.2. We do not exactly know whether the level of education has an actual impact on the HSP score or if those who have a lower HSP score are more likely to have a higher degree. This would be an interesting topic to further investigate. However, if we assume that education can in fact affect how people want their information, we can more easily tailor a chatbot to look for analytical data that shows a person has higher education and then use that to provide a better experience to the user. It is uncertain whether by using a larger sample size we would be likely to see either a higher negative correlation or one closer to 0. However, in this case, we see that education is likely to affect the choice of desired traits. We expect that there is a higher number of design students reached with the questionnaire, as that is what the author is studying and would likely have a higher reach with. We cannot be entirely sure due to none of the questions asking the participants for their education type or field (e.g. design, technology, math, etc.). In table 4.2 we can see that there is at least a medium correlation (≥ 0.5) between gender and both the total and LST. We can also see that there is also a slightly lower correlation on EOE. This indicates that in our results there is a difference in how gender responds to the HSP test. This is not unexpected, and something we assumed from the start. We can also see that the higher the education the person has, the lower they would likely score on the HSP test. This could be speculated to mean either an indication of people with higher education being less sensitive, or that people with higher sensitivity threshold are more likely to participate in higher education.

As you can see in Table 4.2, the general separation between the genders in selecting which trait (humanness vs. efficiency) they would want is low, which indicates to us that there is actually no discernible difference in how each gender wants a trait. However, since we can see that both total and LST are affected by gender, if we were to compare the value of HSP with the traits, then we would get a different story. This does not actually tell us that there is a difference in how people want it, since education is actually also a factor in this case, but it does tell us that there is a possibility for gender to play a role in what traits they want. Especially in terms of professionalism and humanness, which correlates with EOE and LST.

An interesting fact found was that those who selected *Humanness* traits were less likely to select the efficiency traits and vice versa, this is shown with a correlation of $-.807$ and a significance of $< .001$. This shows that there is significance in which traits a person would like when conversing with chatbots, and this is also something that we can further correlate with each of the HSP scores to show which traits a person would prefer based on their own self-reported sensitivity.

Each of the tables (4.3, 4.4, 4.5) indicating the correlation between HSP and the traits desired by users, where we can see that EOE has the greatest impact on what users would like in terms of traits (table 4.3). We can also see that there is a significance in the linearity between EOE and LST with *Humanness*, and EOE with *Professionalism*. Meaning that there is a significant change in what people want based on their own emotional sensitivity. This is interesting, since we do not have an exact explanation as to why that is, but we can speculate that there is a high probability that the person with higher sensitivity would pick up on small changes in tone.

*ANOVA Table | HSP * Humanness*

			Sum of Squares	df	Mean Square	F	Sig.
HSP_EOE * Humanness	Between Groups	(Combined)	10.133	3	3.378	4.366	**.011**
		Linearity	8.114	1	8.114	10.489	**.003**
		Deviation from Linearity	2.019	2	1.010	1.305	.284
	Within Groups		26.301	34	.774		
	Total		36.434	37			
HSP_AES * Humanness	Between Groups	(Combined)	1.467	3	.489	.836	.484
		Linearity	.549	1	.549	.939	.339
		Deviation from Linearity	.917	2	.459	.784	.465
	Within Groups		19.885	34	.585		
	Total		21.351	37			
HSP_LST * Humanness	Between Groups	(Combined)	13.614	3	4.538	3.735	**.020**
		Linearity	9.453	1	9.453	7.780	**.009**
		Deviation from Linearity	4.161	2	2.081	1.712	.196
	Within Groups		41.311	34	1.215		
	Total		54.926	37			

Table 4.3: HSP correlated with Humanness

*ANOVA Table | HSP * Efficiency*

			Sum of Squares	df	Mean Square	F	Sig.
HSP_EOE * Efficiency	Between Groups	(Combined)	6.154	3	2.051	2.303	.095
		Linearity	1.311	1	1.311	1.472	.233
		Deviation from Linearity	4.843	2	2.421	2.719	.080
	Within Groups		30.280	34	.891		
	Total		36.434	37			
HSP_AES * Efficiency	Between Groups	(Combined)	2.474	3	.825	1.486	.236
		Linearity	.076	1	.076	.137	.713
		Deviation from Linearity	2.398	2	1.199	2.160	.131
	Within Groups		18.877	34	.555		
	Total		21.351	37			
HSP_LST * Efficiency	Between Groups	(Combined)	9.596	3	3.199	2.399	.085
		Linearity	4.000	1	4.000	3.000	.092
		Deviation from Linearity	5.596	2	2.798	2.099	.138
	Within Groups		45.329	34	1.333		
	Total		54.926	37			

Table 4.4: HSP correlated with Efficiency

*ANOVA Table | HSP * Professionalism*

			Sum of Squares	df	Mean Square	F	Sig.
HSP_EOE * Professionalism	Between Groups	(Combined)	8.545	3	2.848	3.472	**.027**
		Linearity	4.510	1	4.510	5.498	**.025**
		Deviation from Linearity	4.035	2	2.017	2.459	.101
	Within Groups		27.889	34	.820		
	Total		36.434	37			
HSP_AES * Professionalism	Between Groups	(Combined)	.546	3	.182	.297	.827
		Linearity	.038	1	.038	.061	.806
		Deviation from Linearity	.508	2	.254	.415	.663
	Within Groups		20.805	34	.612		
	Total		21.351	37			
HSP_LST * Professionalism	Between Groups	(Combined)	6.742	3	2.247	1.586	.211
		Linearity	2.237	1	2.237	1.579	.217
		Deviation from Linearity	4.505	2	2.252	1.589	.219
	Within Groups		48.183	34	1.417		
	Total		54.926	37			

Table 4.5: HSP correlated with Professionalism

4.3 Results from interviews

We will first begin by summarizing the main findings of the interviews, then go further into detail on the topics discussed with the participants. After this is done, we will show our findings from the testing and link them with the previous part.

4.3.1 Sample size

The interviewees were heavily biased toward people with greater compassion, as the majority of participants were students studying nursing. These were selected at random by the available people on campus in Gjøvik and Trondheim, who volunteered to participate in this study, in addition to some who mentioned their willingness to participate in the study through the questionnaire. A total of $N = 12$ participated in this interview and completed the study. Of which 7 were women and 5 were men, which may slightly bias the following results toward humanness traits. This is due to the fact that women are more likely to have higher HSP value, which is corroborated by a recent study on HSP using a gender matched study by Trå et al. (2022), where it was shown that women were more likely to have a higher HSP values than men. This is also something we also found during our study of the questionnaire data which can be seen in table 4.2. Furthermore, this is also corroborated by the participants' HSP values seen in appendix B.1.

4.3.2 General Findings

Smart Assistants

We find that people associate mainly AI with chatbots or smart assistants. All participants said that they were most familiar with the "Siri" and the "Google" assistants. This is to be expected since most people have access to these assistance through either phones (iPhones and Android based phones), their MacBook, and Google Home Assistants. They were less familiar with Amazon Alexa but had heard of it.

Two of the participants stated that they used smart assistants at least once a week, while the rest did not use smart assistants at all. The two who used the smart assistants used the Google Assistant. This does not indicate anything in particular, rather it was interesting to hear that this was the main tool. However, this might indicate that there is slightly more ease of use on the one created by Google. When asking questions about how it felt like to ask questions to a smart assistants, all participants stated that they preferred searching for data themselves rather than asking a question to a smart assistant. This might indicate a lack of trust in the system, while also indicate that they feel more in control when searching for themselves.

Chatbots

Initially, of the participants, only 5 knew exactly what a chatbot was without any explanation or visual reminder. However, they all knew what it was when we showed examples (Examples were shown from banking, e-Commerce, and customer support. However, none of them directly indicated any knowledge about how chatbots work, other than being a AI. A recurring trend was that customer support and banks were the most used by participants when talking about chatbots. Participants stated that in most cases, the questions were solved quickly, which is usually related to fetch-based questions, which are similar to the demonstrative in Figure 2.4.

Although around half did mention that they felt annoyance when the chatbot would misinterpret or not understand the question, which would lead to either attempt to bypass the chatbot by reaching a human agent, or quit the conversation to solve it themselves. Any question that required a more advanced understanding and inference, account help, and other difficult problems required human intervention. Three of the participants stated that they had no particular feelings after using the chatbot, two of them were male. This is not to state that there is a probable correlation between gender and being more affected by text, but rather an observation found in these interviews.

"*How do you know you're talking to a chatbot as opposed to a human?*" was probably the most interesting question, since the participants would have to consider whether the agent is a real human. Interestingly, most people though that it would be easy to distinguish between a chatbot and a human. This is true to some extent as a chatbot would not be able to understand everything the user is asking, nor would it be able to infer what the user is asking (to a certain extent it could, but in general not). Two of the participants did note that they would not be able to distinguish between them, but noted that with extended use they would perhaps be able to do so.

"*Do you prefer human customer support? Why, Why not?*". All participants liked to talk to humans because they would be more understanding. A statement one participant had, which summed up most of what people though was that:

> "*I prefer it over chatbots, because at least when it is a human they tend to ask better questions and can more easily diagnose the problem, rather than having to answer multiple questions before being sent to a human to do it all again*".

We believe that openness in chatbots seems to be an important trait, as most said that they wanted to feel like the chatbot *wants* to talk to the users. Positive human traits that affect how users perceive the chatbot are important for its continued use. While it is important that we coincide that chatbots are still an early stage low intelligence, we experience this and believe that it might be that a chatbot is not smart enough *yet* to do the exact tasks a human would.

Wizard Of Oz Chatbot

This test was split into two sections, the objective and the human-like chatbot test. Which chatbot started first was selected by the user themselves through selecting a number (1 and 2). This was to pseudo-randomize which chatbot started such that there was no way for the user to know which was the first.

Chatbot 1 (CB1): The chatbot which focuses on Human-like traits and being friendly.
Chatbot 2 (CB2): This chatbot focused on being quick, efficient, and objective.

This test lasted an average of 8 minutes per chatbot. Their goal being what is shown in the personas found in appendix H. When they found their answer, depending on how long it took to find all the answers, they had some additional time to find more information on additional topics such as “other entertainment” and “other transportation options”. This was to extend the time and allow them to infer more on the personality of the chatbot.

We did not include any negative words in our tests, which negates the question of whether the chatbot was negative (appendix F.2).

As expected, CB1 were defined as likable, sociable, friendliness, and intelligent. This is due to its method of writing being formulated to directly speak to the person rather than the assignment it was tasked with. Almost all of the participants stated that they felt understood while communicating with it, comments such as “I felt like it wanted to talk to me” and “It felt like it tried to give me good suggestions” were not uncommon. An excerpt from the best conversation can be found in appendix I. This allows us to infer that having a personality matters when the conversation is goal-oriented. At one point, some of the participants found that it was *too* chatty, and it blurred out what the answer actually was and what it wanted to converse about. This may be due to our own fault in writing, or it could be that they expected something else.

CB2 were defined as attentive and neutral. The users felt like it answered the questions exactly like they expected; this does not mean that this was the best way of answering, but rather a subjective observation from the participants. It did not feel as if it was communicating with them, but rather just giving the answer they were asking for. In general, this chatbot was defined as neutral and less intelligent than CB1. This may be because it did not specifically interact *directly* with the user, but rather simply responded with the exact answer. As one user stated “This is not different than simply searching”. Thus, we can actually conclude that having *some* personality is actually important for the user experience.

The important finding is that having traits that are too human-like is important to appear intelligent, while being too human-like would cross over into the *Uncanny Valley*.

"Real Chatbot"

Of the participants, five of the participants had time and would like to try this chatbot. We told the participant that they had a total of 10 minutes to converse. We found that having a time limit would only benefit the test, as this was not the main purpose of the test. Our goal was to allow the user to experience a slightly more advanced chatbot, which has a bigger personality and which is able to be conversed with.

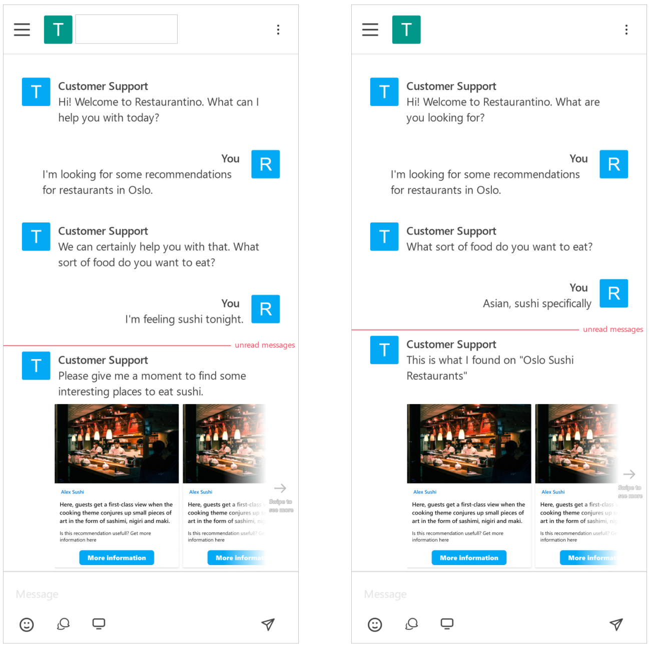
The participants stated that the chatbot felt very human-like, but also robotic in its writing method, meaning the method was too structured for how a human might answer. This indicates to us that there is a possible gap in chit-chat robots where they are too friendly from the start, or staying too consistent in their writing method. When asked about the intelligence and how it attempted to get to know the user, the participants found that it was intelligent but a bit too inquisitive for a regular human. This might be due to how it was trained, but the chatbot would ask questions and lead the conversation to topics ranging from games to sports to personal questions. The following excerpt is from the beginning of a conversation between a user and the chatbot.

> **User:** Hi! My name is *****. Who are you?
>
> **Bot:** *Hi *****, how are you today? I'm good, just watching some soccer. What do you do for fun?*
>
> **User:** I'm doing good. I love playing games for fun. Who's playing right now?
>
> **Bot:** *I'm not sure who's playing, but I love watching soccer. Do you have a favorite team?*
>
> **User:** Yea, my local team ***, who's your favourite team?
>
> **Bot:** **** is my favorite as well. I'd love to see them play in London one day.*

And finally, it was found to be a highly social and somewhat likable chatbot that had some attentiveness. Overall, we found that most people, while they enjoyed testing the conversation, found it too structured, which indicates that more training on less structured sentences was needed. However, we note that this was not our main purpose and so this should be considered a side note. Therefore, we can say that the conversation is not natural, as there is no discussion around the topic once it started.

4.3.3 Interview vs questionnaire

During the interview process, we found a couple of important facts related to the questionnaire. The main findings were that the majority of participants appeared to prefer a more human-like chatbot when interacting with it. All participants appeared to prefer Fig. 4.2a over Fig. 4.2b, and citing it as "appeared more open and inviting to ask questions".



(a) Human-like traits **(b)** Efficiency traits

Figure 4.2: Excerpt from testing showing difference between human-like traits and efficiency traits

4.3.4 Other findings

An interesting finding is that most of the participants did not think of search engines (such as Google, Bing, or similar) as AI, but rather as a search function. Meaning a lack of knowledge of what AI can function as. This allows us to state that the interviewees reflect a layman's view of chatbots and AIs.

Chapter 5

Discussion

In this section we will discuss the findings and correlate it with what we have shown in the theory. We will begin with discussing chatbots and findings related to this thesis, then move on to discussing the limitations on the study, and after that conclude the thesis.

5.1 Chatbots

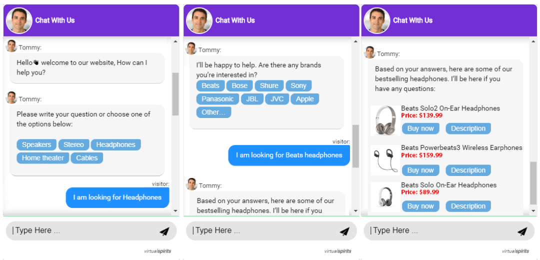
Some people noted that when they used chatbot, a portion of the responded did answer that they used they mostly used the chatbots when it was necessary. This usually was things such as; Customer support, finding information, and Those who mentioned customer support also sometimes said that they had to quarrel with a chatbot to reach a human agent, this was due to the CAs not being able to give an answer to the questions. We will note that there is a possibility of the question from the participant to the chatbot was not formulated in such a way that the chatbot understood the message. However, this is something that is or should be included in any chatbot since fault tolerance is a minimum for any program, not to mention chatbot.

5.1.1 Chatbot Personality Traits

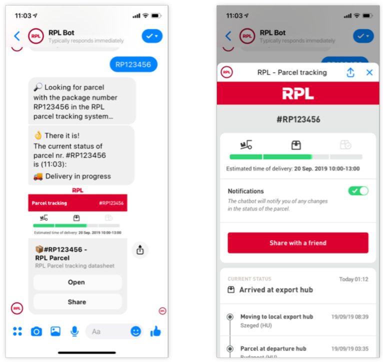
Numerous respondents listed the traits “Simple” and “Efficient” as the most important trait to have in their chatbot. In this context, “Simple” would relate to “simplicity of use”, and “Efficient” would relate to how the chatbot interprets the questions and suggests answers without an extensive explanation. This is because chatbots that require too much know-how to use would be useless for an agent that is only used for simple “question and answer” type of work.

The second most important trait is “Intelligent”, which is related to the ability to think and understand what is asked. The importance of this trait is understated, as most of the chatbots are not intelligent and limited in functionality because of limited datasets to train on. Although this is usually enough for chatbots used on websites focusing on e-Commerce. This is due to them not requiring as much understanding of complex language to find the information the user is seeking. One thing to note is that this is more prevalent in chatbots working in the Norwegian language than it is for English ones. We know this because training data sets are more extensive and more widespread as more people use the English language (Kaggle n.d.).

Most respondents reasoned the need for quickness, intelligence and simplicity as wanting their questions answered quickly and not wanting a conversational partner. While this would be preferable on chatbots built for specific purpose, anything that requires more reasoning would struggle by only using this. This includes searching database of information and fetching topics related to this subject. This could be thing such as e-commerce (fig. 5.1a) and getting user specific data, such as tracking packages (fig. 5.1b).



(a) Chatbot used for e-Commerce, figure by virtualspirits (n.d.)



(b) Chatbot used for tracking packages, figure by Hajjar (2021)

Figure 5.1: Specific use-cases for efficiency chatbots

Some noted that for a chatbot to be usable, it should be able to answer at least a complex question. If we compare it with today's chatbots, we notice that if it cannot understand parts of the questions, it will not attempt to answer the rest and respond with "I don't know" or a similar statement and tell the user to contact a human agent. This is not a poor solution per se. It would be preferable to at least attempt to answer the parts that it can understand and suggest that they could pass the conversation to a human agent for more complicated parts.

There were also some who stated that they only needed answers and thought that additional development in the personality was a waste of development time. Although we know that personality is an important factor in increase user satisfaction (Jakobsen 2021; Smestad and Volden 2019), although this primarily relates to customer support chatbots.

5.1.2 Selected Traits and High Sensitivity Persons

We expected that the "personality" a person would want for a chatbot would differ based on their technical knowledge, since this is because a person with higher skill will often only require specifics rather than converse around the topic. In the questionnaire, we found that there seems to be a correlation between how sensitive a person is and what traits they select. People with higher sensitivity towards emotions pick traits that affect the personality of the chatbot, while those with lower sensitivity pick traits that focus on efficiency and speed. This tells us that people who often have higher technical competences would rather have their questions answered quickly, with little to no conversations "around" the current topic of conversation. One study looking at differences in "Web-Use Skill" and how they searched for information tells us that "skilled Web users effectively filtered information according to search intentions and data sources" (Feufel and Stahl 2012). This correlates to the findings in terms of technically inclined people are more efficient in how they want information, and in terms of speed and pure information-based searching.

In both the general questionnaire and questions answered by interviewees (both through interview and just as a single question asked passersby), the most common answers were that they wanted efficiency rather than a conversational partner. Where the common factor in all answers were that people expect the correct answer quickly, when they require more detailed help, a human is necessary.

We do expect that more helpful chatbots are being developed at this moment, with more advanced abilities such as; inference based chatbots (Finch et al. 2021), relatable chatbots with empathy (de Gennaro et al. 2020; Lahoz-Beltra and López 2021), chatbots focusing on user experience methodology such as design thinking (Bittner and Shoury 2019). This also correlates with the argumentation of chatbots (robots) being slaves mentioned by Bryson (2010), a tool rather than a conversational partner. We can see this in answers provided to both the questionnaire and the interviews, in which the following is a good excerpt from all of them.

- "When I'm using a chatbot outside of the context of entertainment (ex. Cleverbot) I'm looking for an answer to a simple question, so I'd prefer for it to not waste time."
- "I want an answer quickly to my question, not having to always look through help pages or search for information."
- "[A] Chatbot is for me a quick way to get answers to something I need or wonder [about]. It can effectively replace other inquiries such as my own searches on websites, email, phone, Service Requester"

The chatbot named "Cleverbot" is a NN which is based on 3 billion conversational interactions (Existor 2016). However, a modern chatbot are often better optimized through better curated datasets, which would allow the bot to better understand meanings of questions and respond with accurate information back. This was shown by Vinyals and Le (2015), where a set of 200 questions were asked to two chatbots; Cleverbot and their own chatbot based on the Seq2Seq framework. From this they had a human evaluation of the answers both bots gave, of which Seq2Seq managed to get the human evaluators to like 97 out of the 200 answers it gave, while Cleverbot only managed 60. This shows us how different older chatbots, even with more training data, is lacking compared to more recent ones.

5.1.3 Chatbots, Smart Assistants and the Elderly

During a period of data gathering, a participant made a comment about the application of smart assistants and their use for the elderly, stating that "she [their grandmother] though it was an actual person". Which is a problem relating the anthropomorphism and thinking that the object is an actual person. Not that it is without benefits, as another comment was "she [their grandmother] did find that the information it gave her was what she was usually looking for". While we might consider this conjecture, the problem of being too human might also be something that could lead to problems further down the line. There is not enough difference that separates a computer and a human for those with, for example, visual deficiencies. Basically, the difference between auditory and text-based "Chatbots" is not that far apart, as the main difference is the conversion of text into speech. The process is mostly the same, using similar methods to solve the problem. While most chatbots today use the Task Completion System (Section 2.1.5).

One study have suggested that using smart assistants to help elderly may improve living conditions, as they have (Barros and Seabra 2020) While others concluded that by polishing existing tools, there will most likely be a possible increase in the quality of life (Masina et al. 2020). We believe that chatbots can actually be a very helpful tool as to the elderly, especially if you have human-like traits added to them. This would allow the elderly to be less likely to get cognitive impaired over a longer time. A study found that there were a noticeable improvements in the communication abilities for 90% of the residents (Tulsulkar et al. 2021), this allows us to speculate that the use-case for chatbots in the homes of elderly is actually quite high.

In general, elderly at this moment are not as familiar with these types of tools, but we suspect as the current generation gets older the use of smart tools such as this would become more common. Most interview participants did not have elderly with smart tools, however, some of them could see that there could be some benefit in using such tools.

5.2 General findings

We found that $\approx$ 85% of respondents uses or have used a smart assistant, and of those only $\approx$ 6% did not use it daily at all. This does indicate a lack of usefulness in the product itself, we can speculate that this is because those under 40 are more likely to prefer written over spoken language, this might not indicate anything by itself, but it might tell us something about the perceived usefulness of this tool (see table A.2). The majority tended to not use it for shopping list, which is usually its selling point and used a lot in advertisements.

In table A.2 we can see that at least 45% of respondents want to use a smart assistant at least occasionally, and a further 44% agree that it is something that they want to use. How often is something to speculate, and could be a further research point later on. Interestingly, in the same table, we also see that 50% of respondents think that a smart assistant gives the correct answer occasionally, which to us indicates that it misunderstands the question approximately between 40% and 60% of the time. This is not good enough to be useful in a daily driver situation. This also further goes back into the same chatbot topic, as a voice assistant does not truly understand the "voice" of the user, it converts the audio recording into text through automatic speech recognition (Këpuska and Bohouta 2018), which it then uses to search and understand the meaning behind the sentence. The diagram seen in fig 5.2 show how the basics of how a general dialogue system functions. There is a slight difference in how a pure chatbot and a smart assistant should function, the main difference being that a chatbot should only respond with the information required, while a smart assistant should include some personality. This is due to how we expect language to be spoken, and often prefer to not only have the information spoken to us.

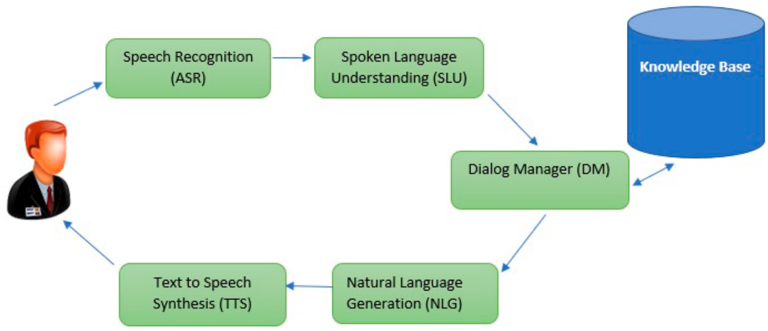


Figure 5.2: The structure of general dialogue system (Figure provided by Këpuska and Bohouta 2018)

We place chatbots and smart assistants in a similar toolset, as both function in a similar fashion *input > understand > get data > build sentence > return sentence*, the main difference being that it needs to convert audio to text, and slightly translate spoken language to text. This was similarly shown previously in fig J.1, which is part of the dialog manager shown in fig. 5.2.

The Google Smart assistant responds fairly directly with "According to Wikipedia, [explanation of topic]". This is perhaps because it is the first search it came up with using Googles own search engine. Sometimes it answers slightly differently with "This is the information I found on [Topic]. [Information]", but that has not happened often. Asking for recipes has shown it give a slight different types of answers, but it is still the same sentence structure. "*Personality (I the robot) > found this (recipe or answer) > read answer*" This is actually quite suited to the HSP persons who want's chatbots to be efficient, since there is little talk around the subject and quickly shows what the user wants instead.

5.3 Key Findings

There were a couple of key points that were repeated in multiple answers, both in the questionnaire and in interviews. These key points were summarized into two general must-haves, and two HSPS related points:

Fault tolerant The chatbot should answer my questions, even if they are spelled wrong, use another word for the same thing.

Answers correctly Gives the correct answer on the first try (highly linked with efficiency). That also included with the addition of not having to give additional follow-up explanation of what the users actually wants to see.

And depending on which sensitivity type, either:

Quick and efficient To get quick answer to questions one may be wondering on. They are there to reduce time spent on searching yourself and answer the question asked with information related to the business. A common theme in the sensitivity types with the efficiency trait was that; when they used a chatbot it was usually looking for information and did not need a conversational partner.

Human-like Want it to feel like you are not talking to a robot, would still like some efficiency, but they would rather a chatbot talking like a regular human. Using human-like language, inserting little bit of humor, be able to answer the user back intelligently, validate the users feelings, and they want to feel understood.

While we did expect a certain difference in how people want their information, we did not expect that there would be a clear difference in how people would want the chatbot types. Though we do not specifically mention the feeling of "wanting to be understood" on the "quick and efficient", we do not discount that the user would like to feel understood in this case too, it is just a different type of understood. Mostly the sentence participants used to describe "not being understood" was "to quarrel with the chatbot", meaning that they would have to largely struggle through using a chatbot to reach a human agent. This mainly related to customer support, and often the questions and help requests they had would fall on deaf ears when talking to the chatbot available.

While most chatbots at this moment are quite fault tolerant by nature, in the Norwegian chatbot market, there is still a bit more to be worked on. In most cases, it is only the customer support that has this problem, as fetch-based chatbots (similar to task-completion chatbots, see chapter 2.1.5).

Both the fault tolerant and answers correctly are actually points that relates to usability, while the efficiency and human-like would be more towards personalization. This however does not mean that you should pick either of them to focus on, as in most cases the smartest would be to combine all of them into a single multi functional service which can be utilized in specific cases.

There were some discussions with the interviewees as to what they would pick if they got to see both types of chatbots (see fig. 4.2), and since we know that the majority in this case would rather have the human-like traits. This might actually be because we are more prone to like things which appear human, as it is easier to trust things that are like us. Subtle changes in personality of a chatbot might not mean much, but as we found in section 4.3.2 under “real chatbot”. Being too friendly or too distant would cause the user to feel either that the conversation is unnatural or that the chatbot does not care about you.

We know now that, based on how sensitive the users are, some want to feel listened too and understood, while others are more likely to just want a more direct answer quickly.

5.4 Limitations and Further Study

While the goal was to gain understanding through a broad network of people. We expected that the data would be slightly biased toward students, as the main sample of people participating was based on network and available in close proximity (as seen in fig. 4.1). There were also an additional smaller sample that was gathered from external volunteering users from France, Germany and some others. This was initially not intended, however, by studying the effect the additional population had on the data, we concluded that this did not alter the outcome to any significant degree. Therefore, we can conclude that results from the participants represent a general overview. We can therefore conclude that another study of same or similar nature should be performed as to the validity globally, since a majority of participants were Norwegian in this study, which also has an increased sample size.

Additionally we would also recommend that a chatbot be trained on multitudes more data and tested against more people, as the probability of training AIs to be the exact same is low. This is due to a number of different factors which can include everything from the algorithm used, training data, Evaluation Procedure, and Platform (Brownlee 2020). We acknowledge that a model should be included in this thesis, but was not able to extract one.

In this study, we did not focus on specific types of personality that should be attached to a chatbot, this is something that could be further studied. The questionnaire did utilize the possibility of ranking the personality traits rather than selecting four of them, which might limit the data we gathered from it (See appendix D, question 3.5). This is something that should have been changed, as it would give a better overall result in terms of to what degree people would like the traits. We should note that the results were not wrong, even with this method, as they had the ability of selecting four of them, which we used to create a score (see appendix C.1) that could be correlated with the other scores such as the HSP score. We did also not include which area of focus and study the participants were, this limits our knowledge of how technically advanced they are. This is not to suggest that people with education that is not programming or heavy science based are less technically advanced, we believe that the method of searching for information might correlate with what field of study they partake in. As a result, a further study would be necessary in which this parameter is added to the calculations. Due to the authors own education within the field of design, this also includes the possibility of current results having higher bias towards the design field, which could influence the number of people with higher AES or EOE results. We would also like in further study to add the "empathic" option for selecting traits, as this might be relevant in terms of how much empathy should matter into the chatbot (or any AI for that matter).

The HSP score used in this study is based entirely on self reported measures, and is something that can be affected by many different factors. This includes things like internal factors (current state of mind, tiredness, emotional state, stress, mindfulness, and others) and external factors (noise level, location, lighting, and others). This however does not invalidate the results, but is rather a possible factor which we decided was unnecessary to focus on for this trial. A future study should perhaps include a more controlled environment, in which you could simulate multiple different environments over a multi-day trial.

There were, in addition an additional limitation in terms of the interview process, as there were a slight bias towards the humanitarian side of personality. This did impact the data found in the study, but overall did not change the findings to a significant degree. We also did not focus on the participants actual technical knowledge, but would perhaps be something that could be further studied.

For the actual chatbot trained and used during the interview process, were not used as any basis, as the time it takes to train a fully functional and adjustable chatbot is far too expensive and long for this study. It might have been of interest in seeing and analyzing how people wrote to a chatbot which has different traits utilized, however, this would be too large of a scope for this thesis, and require many more months to test the chatbot on multitudes more participants.

Further effort could have been done in trying to reach participants, perhaps even offering some sort of "gift" for participating might have increased the participation ratio. We believe that this could have been detrimental, as people would quickly go through the questions without thinking of the responses. However, the benefits would be that there would most likely have been increased enthusiasm in participating.

Another trait that could have been included in the Humanness calculation is T5 (Intelligent), as this generally only relates to how intelligent the chatbot appears. This was not done in our case, which slightly limits the number of participants who scored higher on Humanness. This should be changed in future studies and perhaps split into two main types of intelligence such as emotional and general.

The real chatbot was run on a local machine, which has limitations as to what it can do. This affected the results in the trial period. A larger server running better equipment would likely show better results in the conversations, but is something we are limited by at this moment.

5.5 Similar Studies

There have been several studies into the personality (Smestad and Volden 2019) and how adding specific personalities to a chatbot improves user experience, the effect of humanness on user experience (Jakobsen 2021), and the role of emotions in personalization (Nguyen 2019). Our goal was not to connect between personality and chatbots, nor was it on personalization and EQ. While many of them are of similar nature, here we focus on linking the sensitivity of people to what type of traits they require a better user experience, while also trying to expand on the multitude of. Our study focused on the sensitivity of people, and would

Other studies have shown that what is important for trust and what is important for establishing trust in chatbots (Følstad et al. 2018; Nordheim 2018), while some have shown to focus on creating AIs that has distinct personalities and used in real-world events (Zhou, Mark, et al. 2019). We do not focus on this part in this case, since it is off less importance to our study.

Some studies are focusing on personalizing chatbots for (Shumanov and Johnson 2021) where they demonstrated that “personality can be predicted during contextual interactions” and that “consumer personality with congruent chatbot personality had a positive impact on consumer engagement with chatbots”. This is also within a similar spectrum, but another part of the personality based studies.

Chapter 6

Conclusion

In this thesis, we answered the research questions listed in Section 1.1. We found that focusing on human-like traits and efficiency traits depending on the HSP score can effectively improve the usability of specific users. We found that the top 3 traits for Efficiency was efficient(63%), simple (61%), and quick (50%), while the top 3 traits for Humanness was Understanding (29%), Friendly (18%), and Human-like (11%). Additional traits do not necessarily need to be produced but should still be studied in terms of effectiveness on usability.

A general rule of thumb would be that; Human-like traits would be more effective in chatbots that require emotional investments (e.g. healthcare), while efficiency traits should be focused on chatbots-related e-Commerce. This is similar to how it functions at this moment; however, an increase in the focus on to what degree you need to apply each trait should be studied.

We have also discussed how higher emotional intelligence might allow a chatbot to be better suited towards healthcare, which requires general understanding and the ability to show empathy towards any being.

There is a noticeable difference in how people with high sensitivity vs. low sensitivity want their chatbots to function, this indicates that there is a reason to invest in figuring out which personality based on analytic data, which could be further utilized to tailor a chatbot to a specific user.

Our findings indicate that gender is not the determining factor for which trait a user wants, but gender has *some* effect on the HSP values of a person. This in turn can affect the user's choices. However, our finding does not indicate any correlation between gender and selected traits.

Chatbots that are tailored more towards a user's own personality and information seeking will increase the likelihood of them using it further, however, you would require more or less all traits for a chatbot to be usable. We will note that having the ability to adjust the *efficiency* and *humanness* settings of a chatbot based on the user's *profile* could likely make the chatbot more user friendly, and the possibility of a user continuing to interact with it increases.

Bibliography

Acevedo, B., Aron, E., Pospos, S., & Jessen, D. (2018). The functional highly sensitive brain: A review of the brain circuits underlying sensory processing sensitivity and seemingly related disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1744), 20170161. https://doi.org/10.1098/rstb.2017.0161

Acevedo, B. P., Aron, E. N., Aron, A., Sangster, M.-D., Collins, N., & Brown, L. L. (2014). The highly sensitive brain: An fMRI study of sensory processing sensitivity and response to others' emotions. *Brain and Behavior*, *4*(4), 580–594. https://doi.org/10.1002/brb3.242

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of personality and social psychology*, *73*(2), 345. https://doi.org/10.1037/0022-3514.73.2.345

Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory Processing Sensitivity: A Review in the Light of the Evolution of Biological Responsivity. *Personality and Social Psychology Review*, *16*(3), 262–282. https://doi.org/10.1177/1088868311434213

Assary, E., Zavos, H. M. S., Krapohl, E., Keers, R., & Pluess, M. (2021). Genetic architecture of Environmental Sensitivity reflects multiple heritable components: A twin study with adolescents. *Molecular Psychiatry*, *26*(9), 4896–4904. https://doi.org/10.1038/s41380-020-0783-8

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–159. https://doi.org/10.1016/j.paid.2017.12.018

Barros, T. C. G. M., & Seabra, R. D. (2020). Usability Assessment of Google Assistant and Siri Virtual Assistants Focusing on Elderly Users. In S. Latifi (Ed.), *17th International Conference on Information Technology–New Generations (ITNG 2020)* (pp. 653–657). Springer International Publishing. https://doi.org/10.1007/978-3-030-43020-7_88

Baumgartner, J. (n.d.). *Directory Contents*. Pushshift. Retrieved May 23, 2022, from https://files.pushshift.io/reddit/

Bishop, J. M. (2021). Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.513474

Bittner, E., & Shoury, O. (2019). Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method. https://doi.org/10.24251/HICSS.2019.029

Bleidorn, W., Hopwood, C. J., & Wright, A. G. (2017). Using big data to advance personality theory. *Current opinion in behavioral sciences*, *18*, 79–82. https://doi.org/10.1016/j.cobeha.2017.08.004

Boost.AI. (n.d.[a]). *How conversational AI is pioneering the digitization of banks in the Nordics*. Retrieved April 22, 2022, from https://www.boost.ai/case-studies/how-conversational-ai-is-pioneering-the-digitization-of-banks-in-the-nordics

Boost.AI. (n.d.[b]). *How Norway's biggest bank automated 51% of its online chat traffic with AI*. Retrieved April 22, 2022, from https://www.boost.ai/case-studies/how-norways-biggest-bank-automated-51-of-its-online-chat-traffic-with-ai

Boost.AI. (n.d.[c]). *How Scandinavia's largest telco is enhancing customer experience with conversational AI*. Retrieved April 22, 2022, from https://www.boost.ai/case-studies/how-scandinavias-largest-telco-is-enhancing-customer-experience-with-conversational-ai

Boost.AI. (n.d.[d]). *Tryg uses conversational AI to improve customer experience and operational efficiency across three unique markets*. Retrieved April 22, 2022, from https://www.boost.ai/case-studies/tryg-case-study-conversational-ai

Brown, C. S., & Sulzer-Azaroff, B. (1994). An Assessment of the Relationship Between Customer Satisfaction and Service Friendliness. *Journal of Organizational Behavior Management*, *14*(2), 55–76. https://doi.org/10.1300/J075v14n02_05

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020, July 22). *Language Models are Few-Shot Learners*. Retrieved May 11, 2022, from http://arxiv.org/abs/2005.14165

Brownlee, J. (2020). *Why Do I Get Different Results Each Time in Machine Learning?* Machine Learning Mastery, 17. August. Retrieved May 19, 2022, from https://machinelearningmastery.com/different-results-each-time-in-machine-learning/

Bryson, J. J. (2010). Robots should be slaves. *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues, 8,* 63–74. https://doi.org/10.1075/nlp.8.11bry

@buchananwp, @alexbuckgit, @cjgronlund, @nibaccam, @lostmygithubaccount, @lbsong, @luisquintanilla, @yuanstanley, @DCtheGeek, @sdgilley, @j-martens, @PeterCLu, & @YutongTie-MSFT. (2021). *Detect data drift on datasets (preview) - Azure Machine Learning*. Retrieved May 7, 2022, from https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets

Budzianowski, P. (2022). *Multi-Domain Wizard-of-Oz dataset*. Retrieved May 23, 2022, from https://github.com/budzianowski/multiwoz

Budzianowski, P., Wen, T.-H., Tseng, B.-H., Casanueva, I., Stefan, U., Osman, R., & Gašić, M. (2018). MultiWOZ - A large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Caballar, R. D. (2019). *What Is the Uncanny Valley?* IEEE Spectrum. November 6. Retrieved April 21, 2022, from https://spectrum.ieee.org/what-is-the-uncanny-valley

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Mining large-scale smartphone data for personality studies. *Personal and ubiquitous computing*, *17*(3), 433–450. https://doi.org/10.1007/s00779-011-0490-1

Chooch. (2020, June 14). *What is an AI Model?* Chooch. Retrieved May 7, 2022, from https://chooch.ai/computer-vision/what-is-an-ai-model/

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future generation computer systems*, *92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055

Clark, C. M., Murfett, U. M., Rogers, P. S., & Ang, S. (2013). Is Empathy Effective for Customer Service? Evidence From Call Center Interactions. *Journal of Business and Technical Communication*, *27*(2), 123–153. https://doi.org/10.1177/1050651912468887

Colman, A. M. (2009). Emotional intelligence. *A Dictionary of Psychology* (3rd ed.). Oxford University Press. Retrieved May 6, 2022, from https://www.oxfordreference.com/view/10.1093/acref/9780199534067.001.0001/acref-9780199534067-e-2728

Copeland, B. J. (2004). Colossus: Its origins and originators. *IEEE Annals of the History of Computing*, *26*(4), 38–45. https://doi.org/10.1109/MAHC.2004.26

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). Superagent: A customer service chatbot for e-commerce websites. *Proceedings of ACL 2017, System Demonstrations*, 97–102.

Danks, D., & London, A. J. (2017). Algorithmic Bias in Autonomous Systems. *IJCAI*, *17*, 4691–4697. https://doi.org/10.24963/ijcai.2017/654

Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, *58*(9), 92–103. https://doi.org/10.1145/2701413

De Angeli, A., & Brahnam, S. (2008). I hate you! Disinhibition with virtual partners. *Interacting with Computers*, *20*(3), 302–310. https://doi.org/10.1016/j.intcom.2008.02.004

DeepAI. (2019). *Conversational Agent*. DeepAI. Retrieved April 22, 2022, from https://deepai.org/machine-learning-glossary-and-terms/conversational-agent

de Gennaro, M., Krumhuber, E. G., & Lucas, G. (2020). Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.03061

Demir, E., Köseoğlu, E., Sokullu, R., & Şeker, B. (2017). Smart Home Assistant for Ambient Assisted Living of Elderly People with Dementia. *Procedia Computer Science, 113*, 609–614. https://doi.org/10.1016/j.procs.2017.08.302

Elliott, C. (n.d.). *Chatbots Are Killing Customer Service. Here's Why.* Forbes. Retrieved April 21, 2022, from https://www.forbes.com/sites/christopherelliott/2018/08/27/chatbots-are-killing-customer-service-heres-why/

Engel, J., Tran, C., Pavlek, N., Blankenberg, N., & Meyer, A. (2013). The Impact of Friendliness on Brand Perception. *Marketing Review St. Gallen*, *30*(6), 82–95. https://doi.org/10.1365/s11621-013-0302-2

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864. https://doi.org/10.1037/0033-295X.114.4.864

Eric, M., Goel, R., Paul, S., Sethi, A., Agarwal, S., Gao, S., & Hakkani-Tur, D. (2019). *MultiWOZ 2.1: Multi-domain dialogue state corrections and state tracking baselines*.

Evans, D. E., & Rothbart, M. K. (2008). Temperamental sensitivity: Two constructs or one? *Personality and Individual Differences*, *44*(1), 108–118. https://doi.org/10.1016/j.paid.2007.07.016

Evans, J. D. (1996). *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Publishing Co.

Existor. (2016). *Cleverbot Data for Machine Learning*. existor.com, 15 January. Retrieved May 23, 2022, from https://www.existor.com/products/cleverbot-data-for-machine-learning/

Ferrer, X., van Nuenen, T., Such, J. M., & Criado, N. (2020). Discovering and categorising language biases in reddit. *International AAAI Conference on Web and Social Media (ICWSM 2021)(Forthcoming)*.

Feufel, M. A., & Stahl, S. F. (2012). What do web-use skill differences imply for online health information searches? *Journal of medical Internet research*, *14*(3), E87–331. https://doi.org/10.2196/jmir.2051

Finch, S. E., Finch, J. D., Huryn, D., Hutsell, W., Huang, X., He, H., & Choi, J. D. (2021). *An Approach to Inference-Driven Dialogue Management within a Social Chatbot* (arXiv:2111.00570). https://doi.org/10.48550/arXiv.2111.00570

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In S. S. Bodrunova (Ed.), *Internet Science* (pp. 194–208). Springer International Publishing. https://doi.org/10.1007/978-3-030-01437-7_16

Fossa, F., & Pisa, I. T. (2019). I don't trust you, you faker. *On trust, reliance, and artificial agency. Teoria, 1*, 63–80. https://doi.org/10.4454/teoria.v39i1.57

Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language, 5*(1), 81–99. https://doi.org/10.1016/0885-2308(91)90019-M

Frude, N., & Jandrić, P. (2015). The Intimate Machine – 30 years on. *E-Learning and Digital Media*, *12*(3-4), 410–424. https://doi.org/10.1177/2042753015571830

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), 44:1–44:37. https://doi.org/10.1145/2523813

Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., Sakai, S., Seneff, S., & Zue, V. (1995). Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communication, 17*(1), 1–18. https://doi.org/10.1016/0167-6393(95)00008-C

Gopalakrishnan, K., Hedayatnia, B., Chen, Q., Gottardi, A., Kwatra, S., Venkatesh, A., Gabriel, R., & Hakkani-Tür, D. (2019). Topical-chat: Towards knowledge-grounded open-domain conversations. *Proc. Interspeech 2019*, 1891–1895. https://doi.org/10.21437/Interspeech.2019-3079

Hajjar, A. J. (2021). *Retail Chatbots: Top 12 Use Cases & Examples in 2022*. AIMultiple. 11 May. Retrieved May 16, 2022, from https://research.aimultiple.com/chatbot-in-retail/

Hasal, M., Nowaková, J., Ahmed Saghair, K., Abdulla, H., Snášel, V., & Ogiela, L. (2021). Chatbots: Security, privacy, data protection, and social aspects. *Concurrency and Computation: Practice and Experience, 33*(19). https://doi.org/10.1002/cpe.6426

Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, *21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Jairath, A. (2022). *Empathetic Dialogues (Facebook AI) 25k*. Kaggle. Retrieved May 23, 2022, from https://www.kaggle.com/atharvjairath/empathetic-dialogues-facebook-ai

Jakobsen, A. (2021). *Humanlike Customer Service Chatbots: The Effect of Humanness on User Experience and User Behavior* (Master's thesis). University of Oslo. Oslo. http://urn.nb.no/URN:NBN:no-90268

Joshi, C. K., Mi, F., & Faltings, B. (2017). *Personalization in Goal-Oriented Dialog* (arXiv:1706.07503). https://doi.org/10.48550/arXiv.1706.07503

Kaggle. (N.d.). *Find Open Datasets and Machine Learning Projects*. Retrieved May 7, 2022, from https://www.kaggle.com/datasets

Këpuska, V., & Bohouta, G. (2018). Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 99–103. https://doi.org/10.1109/CCWC.2018.8301638

Kindly. (n.d.[a]). *AI Sales Chatbot Case Study*. Retrieved April 22, 2022, from https://www.kindly.ai/case-studies/ai-sales-chatbot-kahoot

Kindly. (n.d.[b]). *Airlines Using Chatbots Norwegian Airlines Case Study*. Retrieved April 22, 2022, from https://www.kindly.ai/case-studies/norwegian-airlines-using-chatbots

Kindly. (n.d.[c]). *E-commerce Chatbot Case study*. Retrieved April 22, 2022, from https://www.kindly.ai/case-studies/ecommerce-chatbot-elkjop

Kindly. (n.d.[d]). *Recruitment Chatbot Case Study*. Retrieved April 22, 2022, from https://www.kindly.ai/case-studies/recruitment-chatbot-adecco

Klüwer, T. (2011a). *Figure 1. The core of a traditional chatbot architecture*. https://doi.org/10.4018/978-1-60960-617-6.ch001

Klüwer, T. (2011b). *Figure 2. A prototypical dialog system architecture*. https://doi.org/10.4018/978-1-60960-617-6.ch001

Lahoz-Beltra, R., & López, C. C. (2021). LENNA (Learning Emotions Neural Network Assisted): An Empathic Chatbot Designed to Study the Simulation of Emotions in a Bot and Their Analysis in a Conversation. *Computers*, *10*(12), 170. https://doi.org/10.3390/computers10120170

Lam Research. (2021, November 18). *The era of artificial intelligence*. Lam Research, 18 November. Retrieved May 13, 2022, from https://blog.lamresearch.com/the-era-of-artificial-intelligence/

Lawton, L. (2017). Taken by the Tamagotchi: How a Toy Changed the Perspective on Mobile Technology. *The iJournal: Student Journal of the Faculty of Information*, *2*(2). Retrieved May 25, 2022, from https://theijournal.ca/index.php/ijournal/article/view/28127

Li, C.-H., Chen, K., & Chang, Y.-J. (2019). When there is no progress with a task-oriented chatbot: A conversation analysis. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. https://doi.org/10.1145/3338286.3344407

Li, C.-H., Yeh, S.-F., Chang, T.-J., Tsai, M.-H., Chen, K., & Chang, Y.-J. (2020, April 21). A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery. Retrieved May 24, 2022, from https://doi.org/10.1145/3313831.3376209

Li, X., Chen, Y.-N., Li, L., Gao, J., & Celikyilmaz, A. (2018, February 11). *End-to-End Task-Completion Neural Dialogue Systems*. https://doi.org/10.48550/arXiv.1703.01008

Lin, Z., Xu, P., Winata, G. I., Siddique, F. B., Liu, Z., Shin, J., & Fung, P. (2020). Caire: An end-to-end empathetic chatbot. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(09), 13622–13623. https://doi.org/10.1609/aaai.v34i09.7098

Lowe, R., Pow, N., Serban, I., & Pineau, J. (2016). *The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems* (arXiv:1506.08909). https://doi.org/10.48550/arXiv.1506.08909

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2018). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, *31*(12), 2346–2363.

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043
Labeled:Random

Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, *98*(4), 285–287. https://doi.org/10.2471/BLT.19.237636

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. *11th Australasian Conference on Information Systems*, *53*, 6–8.

Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., Mapelli, D., Spagnolli, A., & Gamberini, L. (2020). Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study. *Journal of Medical Internet Research*, *22*(9), e18431. https://doi.org/10.2196/18431

Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The Ability Model of Emotional Intelligence: Principles and Updates. *Emotion Review*, *8*(4), 290–300. https://doi.org/10.1177/1754073916639667

Menon, K., & Dubé, L. (2000). Ensuring greater satisfaction by engineering salesperson response to customer emotions. *Journal of Retailing*, *76*(3), 285–307. https://doi.org/10.1016/S0022-4359(00)00034-8

Merriam-Webster. (N.d.[a]). Anthropomorphize. *Merriam-Webster.com dictionary*. Retrieved April 21, 2022, from https://www.merriam-webster.com/dictionary/anthropomorphize

Merriam-Webster. (N.d.[b]). Robot. *Merriam-Webster.com dictionary*. Retrieved May 24, 2022, from https://www.merriam-webster.com/dictionary/robot

Meyer, J., Miller, C., Hancock, P., de Visser, E. J., & Dorneich, M. (2016). Politeness in Machine-Human and Human-Human Interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 279–283. https://doi.org/10.1177/1541931213601064

Miller, A. H., Li, M., & Weston, J. (N.d.). *Intro to ParlAI — ParlAI Documentation*. ParlAI. Retrieved March 15, 2022, from https://parl.ai/docs/tutorial_basic.html?highlight=training#training-and-evaluating-existing-agents

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Nguyen, M. T. (2019). The Role of Emotions in Personalization: Detecting Emotions in Real-Time as Context Information. Retrieved May 4, 2022, from https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2617727

Nordheim, C. B. (2018). *Trust in chatbots for customer service – findings from a questionnaire study* (Master's thesis). University of Oslo. Oslo. Retrieved

May 25, 2022, from http://urn.nb.no/URN:NBN:no-66052
Accepted: 2018-08-21T22:03:17Z

NTNU. (n.d.). *Data storage guide*. Kunnskapsbasen. Retrieved May 24, 2022, from https://i.ntnu.no/wiki/-/wiki/English/Data+storage+guide

OpenAI. (n.d.). *About OpenAI*. Retrieved May 12, 2022, from https://openai.com/about/

Paweł Budzianowski. (n.d.). *Multiwoz/data/MultiWOZ_2.2*. GitHub. Retrieved May 25, 2022, from https://github.com/budzianowski/multiwoz

Ramadan, O., Budzianowski, P., & Gasic, M. (2018). Large-scale multi-domain belief tracking with knowledge sharing. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, *2*, 432–437.

Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2019). *Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset* (arXiv: 1811.00207). https://doi.org/10.48550/arXiv.1811.00207

Raux, A., Langner, B., Bohus, D., Black, A. W., & Eskenazi, M. (2005). Let's go public! taking a spoken dialog system to the real world. *In Proc. of Interspeech 2005*, 885–888. https://doi.org/10.21437/Interspeech.2005-399

Riera, K., Rousseau, A.-L., & Baudelaire, C. (2020). *Doctor GPT-3: Hype or reality?* Nabla, 27. October. Retrieved May 11, 2022, from https://www.nabla.dev/blog/gpt-3/

Rinn, A. N., Mullet, D. R., Jett, N., & Nyikos, T. (2018). Sensory Processing Sensitivity Among High-Ability Individuals: A Psychometric Evaluation of the Highly Sensitive Person Scale. *Roeper review*, *40*(3), 166–175. https://doi.org/10.1080/02783193.2018.1466840

Rocket.Chat. (n.d.). *Rocket.Chat: Communications Platform You Can Fully Trust*. Retrieved May 9, 2022, from https://rocket.chat/

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*(3), 185–211. https://doi.org/10.2190/DUGG-P24E-52WK-6CDG

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868. https://doi.org/10.1145/3196709.3196772

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Sheridan, T. B. (2016). Human–Robot Interaction: Status and Challenges. *Human Factors*, *58*(4), 525–532. https://doi.org/10.1177/0018720816644364

Shum, H.-y., He, X.-d., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/FITEE.1700826

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

Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). *McKinsey Global Institute*, 1–6. https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. B. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human technology*, *15*(1), 30–54. https://doi.org/10.17011/ht/urn.201902201607

Smestad, T. L., & Volden, F. (2019). Chatbot personalities matters: Improving the user experience of chatbot interfaces. *Internet science* (pp. 170–181). Springer International Publishing.

Smolewska, K. A., McCabe, S. B., & Woody, E. Z. (2006). A psychometric evaluation of the Highly Sensitive Person Scale: The components of sensory-processing sensitivity and their relation to the BIS/BAS and "Big Five". *Personality and Individual Differences*, *40*(6), 1269–1279. https://doi.org/10.1016/j.paid.2005.09.022

Sun, K., Moon, S., Crook, P., Roller, S., Silvert, B., Liu, B., Wang, Z., Liu, H., Cho, E., & Cardie, C. (2021). *Adding Chit-Chat to Enhance Task-Oriented Dialogues* (arXiv:2010.12757). https://doi.org/10.48550/arXiv.2010.12757

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf

Tannor, P. (2021). *Data Drift vs. Concept Drift: What Are the Main Differences?* Deepchecks, 06. October. Retrieved May 10, 2022, from https://deepchecks.com/data-drift-vs-concept-drift-what-are-the-main-differences/

TechTarget. (n.d.). *What is uncanny valley? - Definition from WhatIs.com*. WhatIs.com. Retrieved April 21, 2022, from https://www.techtarget.com/whatis/definition/uncanny-valley

Thought Vector, & Axelbrooke, S. (2017). *Customer Support on Twitter*. Kaggle. Retrieved May 23, 2022, from https://www.kaggle.com/thoughtvector/customer-support-on-twitter

Trå, H. V., Volden, F., & Watten, R. G. (2022). High Sensitivity: Factor Structure of the Highly Sensitive Person Scale and Personality Traits in a High and Low Sensitivity Group. Two Gender-Matched Studies.

Tulsulkar, G., Mishra, N., Thalmann, N. M., Lim, H. E., Lee, M. P., & Cheng, S. K. (2021). Can a humanoid social robot stimulate the interactivity of cognitively impaired elderly? A thorough study based on computer vision methods. *The Visual Computer*, *37*(12), 3019–3038. https://doi.org/10.1007/s00371-021-02242-y

Varca, P. E. (2009). Emotional empathy and front line employees: Does it make sense to care about the customer? *Journal of Services Marketing*, *23*(1), 51–56. https://doi.org/10.1108/08876040910933093

Vinyals, O., & Le, Q. (2015). *A Neural Conversational Model* (arXiv:1506.05869). arXiv. https://doi.org/10.48550/arXiv.1506.05869

virtualspirits. (n.d.). *Ecommerce Chatbot – Use Chat bots as Your Shopping Assistants*. virtualspirits. Retrieved May 16, 2022, from https://www.virtualspirits.com/ecommerce-chatbot-use-chat-bots-as-your-shopping-assistants.aspx

Wadhwa, M. (2018). *Seq2seq model in Machine Learning*. GeeksforGeeks, 29. September. Retrieved May 23, 2022, from https://www.geeksforgeeks.org/seq2seq-model-in-machine-learning/

Wang, Y.-Y., Deng, L., & Acero, A. (2011). Semantic Frame-Based Spoken Language Understanding. *Spoken Language Understanding* (pp. 41–91). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119992691.ch3

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on psychological science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social Cognition Unbound: Insights Into Anthropomorphism and Dehumanization. *Current Directions in Psychological Science*, *19*(1), 58–62. https://doi.org/10.1177/0963721409359302

Wei, W., Le, Q., Dai, A., & Li, J. (2018). AirDialogue: An environment for goal-oriented dialogue research. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3844–3854. https://doi.org/10.18653/v1/D18-1419

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Wieseke, J., Geigenmüller, A., & Kraus, F. (2012). On the Role of Empathy in Customer-Employee Interactions. *Journal of Service Research*, *15*(3), 316–331. https://doi.org/10.1177/1094670512439743

Yelp. (N.d.). *Yelp Dataset*. Retrieved May 23, 2022, from https://www.yelp.com/dataset

Zang, X., Rastogi, A., Chen, J., Sunkara, S., Gupta, R., Zhang, J., & Chen, J. (2020). MultiWOZ 2.2 : A Dialogue Dataset with Additional Annotation Corrections and State Tracking Baselines. *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, 109–117. https://doi.org/10.18653/v1/2020.nlp4convai-1.13

Zhang, J., Li, S., Zhang, J.-Y., Du, F., Qi, Y., & Liu, X. (2020). A Literature Review of the Research on the Uncanny Valley. In P.-L. P. Rau (Ed.), *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments* (pp. 255–268). Springer International Publishing. https://doi.org/10.1007/978-3-030-49788-0_19

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). *Personalizing Dialogue Agents: I have a dog, do you have pets too?* (arXiv:1801.07243). https://doi.org/10.48550/arXiv.1801.07243

Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *Computational Linguistics*, *46*(1), 53–93. https://doi.org/10.1162/coli_a_00368

Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting Virtual Agents: The Effect of Personality. *ACM Transactions on Interactive Intelligent Systems*, *9*(2-3), 10:1–10:36. https://doi.org/10.1145/3232077

Appendix A

Additional Findings

Use of smart assistants	Not at all	Very rarely	Rarely	Occasionally	Often	Very often
Daily	6%	38%	13%	19%	16%	9%
Before I go to work / school	34%	25%	22%	6%	3%	9%
For music	22%	9%	22%	22%	19%	6%
as an Alarm (wake up and timers)	31%	19%	9%	16%	13%	13%
Get information (eg weather, facts, news)	16%	28%	19%	22%	3%	13%
For the Shopping List	59%	31%	3%	6%	0%	0%
Control smart homes (eg switch on / off lights)	44%	19%	3%	9%	9%	16%
For other things	25%	44%	16%	9%	3%	3%

Table A.1: Use of smart assistants

I think smart assistants	Strongly disagree	Disagree	Occasionally	Agree	Strongly Agree
Is Easy to use	0%	11%	43%	41%	5%
Always understand what I'm asking for	15%	38%	29%	18%	0%
Always gives me the answer I'm looking for	8%	28%	50%	14%	0%
Is helpful and allows me to do several things at once	8%	17%	31%	36%	8%
Is a fun toy	3%	6%	17%	54%	20%
Is something I want to use	0%	14%	42%	33%	11%

Table A.2: How people view smart assistants

I think it's okay for a chatbot to store and use ...	Don't know	Strongly disagree	Disagree	Both agree and disagree	Agree	Strongly Agree
My IP address	8%	21%	21%	26%	18%	5%
Previous conversations I have had with the chatbot	3%	8%	11%	27%	32%	19%
Logs (When you sent the message, length of the conversation, etc. No personal info)	3%	0%	16%	18%	37%	26%
My email	5%	13%	16%	37%	21%	8%
My name (Part or all of the name)	3%	16%	32%	13%	26%	11%
My Location (GPS / City)	0%	47%	13%	21%	11%	8%

Table A.3: What data users don’t mind being stored when only used in relations to the user

I think it's okay for a chatbot to store and use ...	Don't know	Strongly disagree	Disagree	Both agree and disagree	Agree	Strongly Agree
My IP address	8%	53%	18%	11%	8%	3%
Previous conversations I have had with the chatbot	3%	8%	8%	42%	32%	8%
Logs (When you sent the message, length of the conversation, etc. No personal info)	0%	8%	3%	26%	47%	16%
My email	0%	42%	29%	18%	3%	8%
My name (Part or all of the name)	0%	34%	37%	18%	3%	8%
My Location (GPS / City)	0%	45%	26%	18%	8%	3%

Table A.4: What data users don't mind being stored when used for any purpose and further training of the chatbot

Trait	Percent selected
Neutral	13%
Understanding	29%
Friendly (Tone)	18%
Technically proficient	37%
Intelligent	42%
Human-like	11%
Professional	32%
Objective	8%
Humorous	5%
Quick	50%
Simple	61%
Analytical	11%
Efficient	63%
Other	13%

Table A.5: Traits selected in the questionnaire showing what percent people most likely wants. Users were able to select a total of 4 traits.

Appendix B

Interview Data

B.1 HSP values from interview

B.1.1 Internal Consistency

N of Items refers to HSP variables seen in table C.2 found in appendix C.

HSP EOE

Cronbach's Alpha	N of Items
.934	12

(a) HSP EOE

HSP AES

Cronbach's Alpha	N of Items
.665	7

(b) HSP AES

HSP LST

Cronbach's Alpha	N of Items
.912	6

(c) HSP LST

Table B.1: The internal consistency of all HSP tests from interview ($N = 12$).

Appendix C

Variables and SPSS calculation

C.1 Chatbot types

Variable	SPSS Variable
Humanness	T2, T3, T6, T7, T9, T12
Efficiency	T1, T4, T5, T8, T10, T11, T13
Professionalism	T2, T4, T5, T7, T8, T12

Table C.1: Variables used to calculate the three main traits of this study

Variable and their meaning

T1=Neutral, T2=Understanding, T3=Friendly (Tone), T4=Technically proficient, T5=Intelligent, T6=Human-like, T7=Professional, T8=Objective, T9=Humorous, T10=Quick, T11=Simple, T12=Analytical, T13=Efficient, T14=Other

These variables are related to question 3.5, which you can see in appendix D. The prefix T refers to the word "Trait".

In this study, the number is calculated by adding up the variables as seen above, which then can be used to check against the HSPS variables. The numbers are represented by 0's and 1's, as they either picked the trait or they did not. In a future study, a expansion of this exact trait selection should be extended to be on a scale rather than a true or false (e.g. from 0 to 7).

C.2 HSP Variables

Variable	**SPSS Variable**
HSP_total	1 — 27
EOE	3, 4, 13, 14, 16, 17, 20, 21, 23, 24, 26, 27
AES	2, 5, 8, 10, 12, 15, 22
LST	6, 7, 9, 18, 19, 25

Table C.2: Variables used to calculate HSP scores

The numbers listed in table C.2 refers to the questions found in appendix D (questions 4.1 — 4.27).

C.2.1 Internal Consistency of all HSP tests

N of Items refers to HSP variables seen in table C.2 found in appendix C.

HSP EOE

Cronbach's Alpha	N of Items
.873	12

(a) HSP EOE

HSP AES

Cronbach's Alpha	N of Items
.563	7

(b) HSP AES

HSP LST

Cronbach's Alpha	N of Items
.850	6

(c) HSP LST

Table C.3: The internal consistency of all HSP tests done ($N = 50$).

C.2.2 Internal Consistency of Questionnaire HSP tests

HSP EOE

Cronbach's Alpha	N of Items
.840	12

(a) HSP EOE

HSP AES

Cronbach's Alpha	N of Items
.494	7

(b) HSP AES

HSP LST

Cronbach's Alpha	N of Items
.819	6

(c) HSP LST

Table C.4: The internal consistency of HSP tests from questinnaire ($N = 38$).

Appendix D

Questionnaire

Questions used in this questionnaire was selected based on the need for figuring out the usage of tools such as smart assistants (Topic 2) and chatbots (Topic 3). We also wished to get a overview of to what degree they believed in personalization and privacy (questions 3.6—3.7). And lastly, we wanted to get to know their HSPS score (topic 4) to which we can correlate with what they wish to use.

[redacted]

Do you want to participate in the research project?

This is a question for you to participate in a research project where the purpose is to gain insight into how emotional intelligence in artificial intelligence can affect how users experience chatbots. In this letter, we give you information about the goals of the project and what participation will mean for you.

Purpose

Artificial intelligent agents are increasingly available on websites around the world, which focus on providing you with the best customer support agent quickly. In this study, we want to gain deeper knowledge about how call agents are experienced from the users' point of view and let us improve and suggest changes that will affect chatbots in the future. The purpose is to create insight into how artificial intelligence can be experienced more humanely through a greater focus on emotional intelligence and how this affects the user's experience and how this increases the likelihood that they do not need the help of human customer support to solve their problem.

This project is a master's project at the Norwegian University of Science and Technology in Gjøvik

It is voluntary to participate in the project. If you choose to participate, you can withdraw your consent at any time without giving any reason. All your personal information will then be deleted. It will not have any negative consequences for you if you do not want to participate or later choose to withdraw.

Your privacy - how we store and use your information

We will only use the information about you for the purposes we have described in this article. We treat the information confidentially and in accordance with the privacy regulations. Only the supervisor at NTNU and a student have access to the information that is submitted. Web forms are used to collect the information and are provided by the University of Oslo. All information is anonymized through grouping and aggregation.

Where can I find out more?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:
Norges teknisk-naturvitenskapelige universitet ved. André Tørlen Lønvik på andretl@stud.ntnu.no

If you agree to participate, click "Next Page".

You you would not like to participate, you can simply close the tab. We will not save any information until you click "Send".

Sideskift

1. General information

1.1 - Age

- under 18
- 18-23
- 24-29
- 30-35
- 36-40
- 40 and above
- Do not want to answer

1.2 - Gender

- Male
- Female
- Other / Don't wish to answer

1.3 - Level of education

All of these apply to ongoing or completed education.

- High school
- 3-year college or university (Equivalent to bachelor or engineering degree)
- 5 year college or university (Masters or similar education)
- Over 5 years at college or university (Doctorate or similar)
- Do not want to answer

Sideskift

2. Artificial Intelligence

2.1 - Do you use or have used a smart assistant?

Smart assistants include Google Home, Alexa, Siri and other.
Can also be known as virtual assistant.

- Yes
- No

2.2 - I use my smart assistant

Here comes a set of statements related to your use of smart assistants.

	Not at all	Very rarely	Rarely	Occasionally	Often	Very often
Daily	○	○	○	○	○	○
Before I go to work / school	○	○	○	○	○	○
For music	○	○	○	○	○	○
as an Alarm (wake up and timers)	○	○	○	○	○	○
Get information (eg weather, facts, news)	○	○	○	○	○	○
For the Shopping List	○	○	○	○	○	○
Control smart homes (eg switch on / off lights)	○	○	○	○	○	○
For other things	○	○	○	○	○	○

2.2.1 - Do you use smart assistants for anything else?

Feel free to describe what else you use the assistant for (Not mandatory).

2.3 - I think smart assistants

Here are some statements about what you think about smart assistants.

	Strongly dis-agree	Disagree	Occasionally	Agree	Strongly Agree
Is Easy to use	○	○	○	○	○
Always understand what I'm asking for	○	○	○	○	○
Always gives me the answer I'm looking for	○	○	○	○	○
Is helpful and allows me to do seve-ral things at once	○	○	○	○	○
Is a fun toy	○	○	○	○	○
Is something I want to use	○	○	○	○	○

Sideskift

3. Chatbots and personalization

3.1 - I know what personalization is.

○ Yes

○ Yes, but want to see a description

○ No / Don't know

Personalization is about adapting a product or experience to a specific person.

In this case, it's about adapting what you use to your personality and interests.

Personalization can also be linked to all the "Cookies" the banners you get when you visit a website for the first time.

3.2 - I think it's acceptable that a website

Here are some claims about personalization. Choose the one that suits you best. If you are not sure or have no shape, choose "do not know".

	Don't know	Strongly disagree	disagree	Both agree and disagree	Agree	Strongly agree
Remember I have visited it previously	○	○	○	○	○	○
Gives me a welcome message when I visit the site	○	○	○	○	○	○
Is adapted to my interests	○	○	○	○	○	○
Uses my information to provide me with better customized content	○	○	○	○	○	○
Shows me advertising tailored to my interests	○	○	○	○	○	○
Collect my information	○	○	○	○	○	○

3.3 - I know what a chatbot is

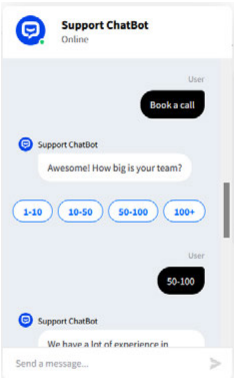
Please answer honestly

- ○ Yes
- ○ Yes, but want to see a description
- ○ No / Don't know

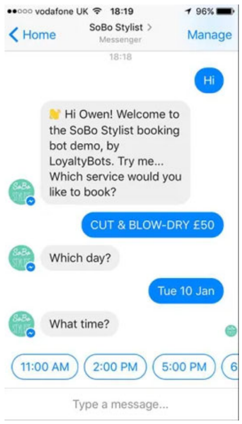
A chatbot is a type of software that can help customers by automating conversations and interacting with them through messaging platforms.

Examples of chatbots: Support ChatBot, SoBo Stylist.

Support Chatbot



SoBo Stylist



3.4 - I am satisfied with

Here are some claims about previous conversations you have had with a chatbot. Choose the one that suits you best. If you are not sure, select "do not know".

	Don't Know	Strongly dis-agree	Disagree	Both agree and dis-agree	Agree	Strongly agree
previous conversations with chat-bots	○	○	○	○	○	○
how quickly it answered my ques-tions	○	○	○	○	○	○
the answer I got (it was the correct answer to what I asked)	○	○	○	○	○	○
the way it answered me (language / spelling)	○	○	○	○	○	○

3.5 - I want a chatbot to be:

Select up to 4 types of characteristics that you want a chatbot to have.

- ☐ Neutral
- ☐ Understanding
- ☐ Friendly (Tone)
- ☐ Technically proficient
- ☐ Intelligent
- ☐ Human-like
- ☐ Professional
- ☐ Objective
- ☐ Humorous
- ☐ Quick
- ☐ Simple
- ☐ Analytical
- ☐ Efficient
- ☐ Other

3.5.1 - What other qualities could you imagine? (Not obligatory)

3.5.2 - Why did you choose these four? (Not obligatory)

Feel free to explain why you chose these four, and why you think it is important.

3.6 - I think it's okay for a chatbot to store and use ...

Here we are basically talking about only conversations with you, which will mean that **it uses what you send to it ONLY when it talks to YOU**. (Active conversations and future ones)

	Don't know	Strongly dis-agree	Disagree	Both agree and dis-agree	Agree	Strongly Agree
My IP address	○	○	○	○	○	○
Previous conversations I have had with the chatbot	○	○	○	○	○	○
Logs (When you sent the message, length of the conversation, etc. No personal info)	○	○	○	○	○	○
My email	○	○	○	○	○	○
My name (Part or all of the name)	○	○	○	○	○	○
My Location (GPS / City)	○	○	○	○	○	○

3.7 - I think it's okay for a chatbot to store and use

In this case, **EVERYTHING you submit is used to improve the chatbot** for everyone who talks to it.

	Don't know	Strongly dis-agree	Disagree	Both agree and dis-agree	Agree	Strongly agree
My IP address	○	○	○	○	○	○
Previous conversations I have had with the chatbot	○	○	○	○	○	○
Logs (When you sent the message, length of the conversation, etc. No personal info)	○	○	○	○	○	○
My email	○	○	○	○	○	○
My name (Part or all of the name)	○	○	○	○	○	○
My Location (GPS / City)	○	○	○	○	○	○

Sideskift

4. Here are a number of characteristics that people may have. How well does each of these characteristics fit you?

Do not dwell on each characteristic. When in doubt, the first answer that comes to mind will often be the best.

Answer on a scale from 1 to 7, where 1 equals «Not at all» and 7 equals «Extremely».

	1	2	3	4	5	6	7
4.1. Are you easily overwhelmed by strong sensory input?	○	○	○	○	○	○	○
4.2. Do you seem to be aware of subtleties in your environment?	○	○	○	○	○	○	○
4.3. Do other people's moods affect you?	○	○	○	○	○	○	○
4.4. Do you tend to be more sensitive to pain?	○	○	○	○	○	○	○
4.5. Do you find yourself needing to withdraw during busy days, into bed or into a darkened room or any place where you can have some privacy and relief from stimulation?	○	○	○	○	○	○	○
4.6. Are you particularly sensitive to the effects of caffeine?	○	○	○	○	○	○	○
4.7. Are you easily overwhelmed by things like bright lights, strong smells, coarse fabrics, or sirens close by?	○	○	○	○	○	○	○
4.8. Do you have a rich, complex inner life?	○	○	○	○	○	○	○
4.9. Are you made uncomfortable by loud noises?	○	○	○	○	○	○	○
4.10. Are you deeply moved by the arts or music?	○	○	○	○	○	○	○
4.11. Does your nervous system sometimes feel so frazzled that you just have to go off by yourself?	○	○	○	○	○	○	○
4.12. Are you conscientious?	○	○	○	○	○	○	○
4.13. Do you startle easily?	○	○	○	○	○	○	○

4.14. Do you get rattled when you have a lot to do in a short amount of time?	○	○	○	○	○	○	○
4.15. When people are uncomfortable in a physical environment do you tend to know what needs to be done to make it more comfortable (like changing the lighting or the seating)?	○	○	○	○	○	○	○
4.16. Are you annoyed when people try to get you to do too many things at once?	○	○	○	○	○	○	○
4.17. Do you try hard to avoid making mistakes or forgetting things?	○	○	○	○	○	○	○
4.18. Do you make a point to avoid violent movies and TV shows?	○	○	○	○	○	○	○
4.19. Do you become unpleasantly aroused when a lot is going on around you?	○	○	○	○	○	○	○
4.20. Does being very hungry create a strong reaction in you, disrupting your concentration or mood?	○	○	○	○	○	○	○
4.21. Do changes in your life shake you up?	○	○	○	○	○	○	○
4.22. Do you notice and enjoy delicate or fine scents, tastes, sounds, works of art?	○	○	○	○	○	○	○
4.23. Do you find it unpleasant to have a lot going on at once?	○	○	○	○	○	○	○
4.24. Do you make it a high priority to arrange your life to avoid upsetting or overwhelming situations?	○	○	○	○	○	○	○
4.25. Are you bothered by intense stimuli, like loud noises or chaotic scenes?	○	○	○	○	○	○	○
4.26. When you must compete or be observed while performing a task, do you become so nervous or shaky that you do much worse than you would otherwise?	○	○	○	○	○	○	○
4.27. When you were a child, did parents or teachers seem to see you as sensitive or shy?	○	○	○	○	○	○	○

Appendix E

Interview Consent Form

Norwegian Only: Please note that the title used in this document is not reflective of the actual purpose, but rather a guideline and was made at the start of this project. This document was only created in Norwegian, as the participants in the interviews were all speaking Norwegian and did not require an English translation.

Note: No video or audio recording was done during this project which voids any part about video or audio recordings in this document.

Vil du delta i forskningsprosjektet

”Can conversational agents provide a better user experience over traditional chatbots”?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å *skaffe innsikt inn i hvordan emosjonell intelligens i kunstig intelligens kan påvirke hvordan brukere opplever chatbots*. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål
Kunstige intelligente agenter er i økende grad tilgjengelig på nettsteder over hele verden, som fokuserer på å gi deg den beste kundestøtteagenten raskt. I denne studien ønsker vi å få dypere kunnskap om hvordan samtaleagenter oppleves fra brukernes ståsted og la oss forbedre og foreslå endringer som vil påvirke chatbots i fremtiden. Formålet er å skape innsikt i hvordan kunstig intelligens kan oppleves mer menneskelig ved hjelp av større fokus på emosjonell intelligens og hvordan dette påvirker opplevelsen til brukeren samt hvordan dette øker sannsynligheten for at de ikke trenger hjelp av menneskelig kundestøtte for å løse problemet deres.
Dette prosjektet er et masterprosjekt ved Norges teknisk-naturvitenskapelige universitet i Gjøvik.

Hvem er ansvarlig for forskningsprosjektet?
Norges teknisk-naturvitenskapelige universitet er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?
Utvalget er trukket ut ifra villige deltakere som selv velger å sende inn spørreskjema, dette gjøres gjennom at du selv valgte å trykke på lenken delt. Du er i alderen mellom 18-40 og har gått på eller går på høyere utdanning.

Hva innebærer det for deg å delta?
Hvis du velger å delta i prosjektet, innebærer det at du er med på et intervju og/eller eksperiment. Det vil ta deg ca. 20-40 minutter. Intervjuet inneholder spørsmål om din erfaring med chatbots og hvordan du ønsker at den skal oppføre seg. Dine svar blir vil bli tatt opp ved hjelp av et digitalt opptaksmiddel, og transkribert etter endt opptak.

Hvis du velger å delta i eksperimentet innebærer det at du er med på et liveforsøk, hvor du gjennomfører et sett med test av emosjonell hypersensitivitet, for så å prate med en chatbot. Dette blir tatt opp (ved din godkjenning) med lyd og video for å følge å følge dine emosjonelle reiser i etterkant. Deretter blir det gjennomført et sett med spørsmål angående hvordan du opplevde denne samtalen.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.
Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.
Bare veileder ved NTNU og en student har tilgang til informasjonen som blir tatt opp.

Etter endt intervju og eksperiment, vil all data bli anonymisert ved transkribering av lyd filer. Og all video vil bli lagret på en separat harddisk som ikke er koblet til internett.
Etter endt prosjekt eller at vi finner ut at videoen ikke er nødvendig lenger, vil all personlig identifiserende data slettes. Dette inkluderer alle opptak.

Video filen vil bare bli benyttet under undersøkelse av data, og bare vist lokalt til forsker André og veileder Frode. Det vi leter etter er å se sammenhengen med svaret du svarer på undersøkelsen og hvordan du opplever det i person.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?
Opplysningene anonymiseres når prosjektet avsluttes/oppgaven er godkjent, noe som etter planen er 01.08.2022. Alt som inneholder personopplysninger, blir enten slettet eller anonymisert ved å fjerne kodenøkler og sletting av data.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.
På oppdrag fra Norges teknisk-naturvitenskapelige universitet har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:
Norges teknisk-naturvitenskapelige universitet ved. André Tørlen Lønvik på andretl@stud.ntnu.no eller Frode Volden på frodv@ntnu.no .
Personvernombud: Thomas Helgesen, thomas.helgesen@ntnu.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:
- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Frode Volden	André Tørlen Lønvik
Veileder	Forsker

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Can conversational agents provide a better user experience over traditional chatbots*, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i intervju
- ☐ å delta i eksperiment

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

---

(Signert av prosjektdeltaker, dato)

Appendix F

Interview

F.1 Interview Guide

The interview guide served as a topic guideline for what was asked. It was slightly modified to be online only, but the topics and questions did not change. These questions were also slightly changed to reflect data from the questionnaire (See appendix D)

Interview guide

Can conversational agents provide a better user experience over traditional chatbots?

Purpose

Artificial intelligent agents are increasingly available on websites around the world, which focus on providing you with the best customer support agent quickly. In this study, we want to gain deeper knowledge about how call agents are experienced from the users' point of view and let us improve and suggest changes that will affect chatbots in the future. The purpose is to create insight into how artificial intelligence can be experienced more humanely through a greater focus on emotional intelligence and how this affects the user's experience and how this increases the likelihood that they do not need the help of human customer support to solve their problem.

Aim:

The aim of these interviews is to investigate in depth details about experiences had with chatbots. This information will be used to go further into detail on how the emotional language of chatbots can affect the users experience. We also investigate how close to humanness we want the chatbots to end up, and if we actually would like to do so.

Date and Location: February – March 2022- Gjøvik
Time: 20-40 minutes
Equipment: Consent form, printed interview guide, recorder, pen, and paper.

Interview

Here we begin the interview.

1. **Warmup / Introduction (5 minutes)**
   Welcome the participants, complete the consent form, and introduce the project.

2. **Key questions (20 minutes)**

   General Questions
   - Age
   - Gender
   - Education level

General questions about AI (Artificial Intelligence)

- What do you associate with “Artificial Intelligence”?
- What AI are you familiar with and how often do you use it?
- Do you know what a chatbot is?
- Have you talked to a chatbot?
    - Where did you talk with it?
    - How long ago was it
    - What was the conversation about?
    - Did it answer your questions? Why / Not
    - Did you have to have human help?
    - How did you feel after talking to it?
- Where have you talked with chatbots and how have you perceived them?
- Do you prefer human customer support? Why, Why not?
- How do you know you’re talking to a chatbot as opposed to a human?
    - Are you ok with talking with them?
    - Do you feel understood while talking to it
- In your opinion, how do you feel about conversing with chatbots?

Personality

- In your own words, why do you think you like writing/talking with some people compared to others?
- Sentences have meaning
- Questionnaire ()
- Arron Hypersensitivity Test (Perform test)

3. Rounding off and finishing

Summarize the interview and thank the participant for participating in the interview.

F.2 Interview guide Additional questions

This questionnaire was performed orally and the questions were changed to open ended questions, where the goal was to get more subjective data on how the user was feeling towards certain chatbots and the Wizard of Oz Chatbot.

Parts of these questions were extracted from Jakobsen (2021) found in his Appendix A.

Experience with the chatbot

In your opinion, to what degree does the words below describe the

1. Likeable

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very Accurately

2. Sociable

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very Accurately

3. Friendly

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very Accurately

4. Personal

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very Accurately

5. Attentive

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very Accurately

6. Positive

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	○	○	○	○	○	○	○	○	○	○	Very Accurately

7. Negative

	1	2	3	4	5	6	7	8	9	10	
Very Poorly	○	○	○	○	○	○	○	○	○	○	Very Accurately

While interacting with the agent, to what degree did you feel...

8. ... as if it was an intelligent being?

	1	2	3	4	5	6	7	8	9	10	
Not at all	○	○	○	○	○	○	○	○	○	○	Extremely

9. ... as if it was a social being?

	1	2	3	4	5	6	7	8	9	10	
Not at all	○	○	○	○	○	○	○	○	○	○	Extremely

10. ... as if it was communicating with you?

	1	2	3	4	5	6	7	8	9	10	
Not at all	○	○	○	○	○	○	○	○	○	○	Extremely

11. ... as if it was paying attention to you?

	1	2	3	4	5	6	7	8	9	10	
Not at all	○	○	○	○	○	○	○	○	○	○	Extremely

12. ... it was involved with you

Not at all	1	2	3	4	5	6	7	8	9	10	Extremely

13. ... it was understanding you?

Not at all	1	2	3	4	5	6	7	8	9	10	Extremely

Fluency

14. Fluency

1	2	3	4	5	6	7	8	9	10

15. Engagingness

1	2	3	4	5	6	7	8	9	10

16. Consistency

1	2	3	4	5	6	7	8	9	10

Appendix G

NSD Assessment

Norwegian Only:
Results of the assessments by Norwegian Centre for Research Data were that this project was ruled as legal based on storage method and grouping of data.

Vurdering

Referansenummer

190660

Prosjekttittel

Can conversational agents provide a better user experience over traditional chatbots

Behandlingsansvarlig institusjon

Norges teknisk-naturvitenskapelige universitet / Fakultet for arkitektur og design (AD) / Institutt for design

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Frode Volden, frodv@ntnu.no, tlf: 93227262

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

André Tørlen Lønvik, andretl@stud.ntnu.no, tlf: 90990065

Prosjektperiode

01.01.2022 - 01.08.2022

Vurdering (1)

16.03.2022 - Vurdert

OM VURDERINGEN
Personverntjenester har en avtale med institusjonen du forsker eller studerer ved. Denne avtalen innebærer at vi skal gi deg råd slik at behandlingen av personopplysninger i prosjektet ditt er lovlig etter personvernregelverket.

Personverntjenester har nå vurdert den planlagte behandlingen av personopplysninger. Vår vurdering er at behandlingen er lovlig, hvis den gjennomføres slik den er beskrevet i meldeskjemaet med dialog og vedlegg.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til den datoen som er oppgitt i meldeskjemaet.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
Personverntjenester vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:
• lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
• formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
• dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
• lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), og dataportabilitet (art. 20).

Personverntjenester vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
Personverntjenester legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

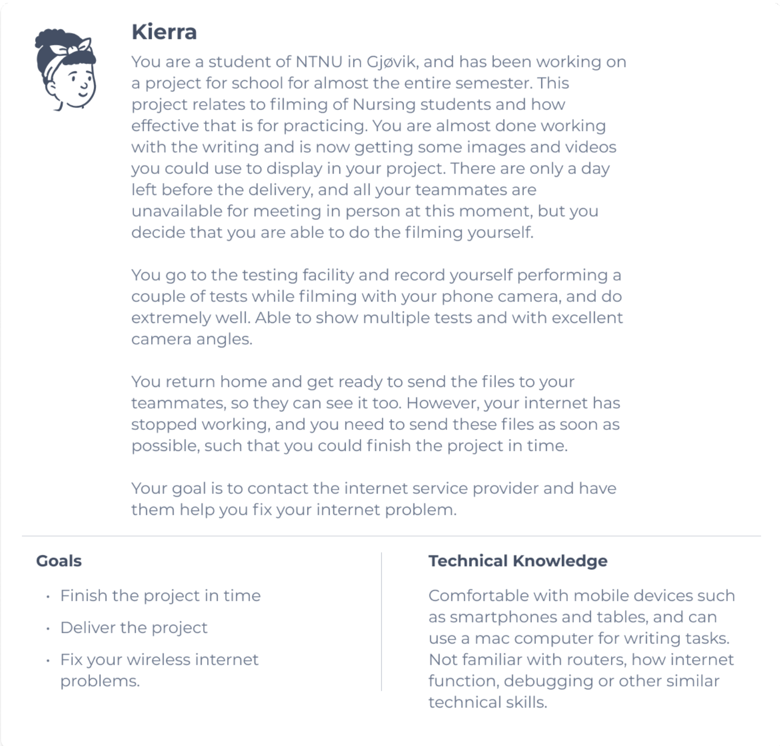
MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til oss ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema
Du må vente på svar fra oss før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET
Personverntjenester vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Appendix H

Personas



Kierra

You are a student of NTNU in Gjøvik, and has been working on a project for school for almost the entire semester. This project relates to filming of Nursing students and how effective that is for practicing. You are almost done working with the writing and is now getting some images and videos you could use to display in your project. There are only a day left before the delivery, and all your teammates are unavailable for meeting in person at this moment, but you decide that you are able to do the filming yourself.

You go to the testing facility and record yourself performing a couple of tests while filming with your phone camera, and do extremely well. Able to show multiple tests and with excellent camera angles.

You return home and get ready to send the files to your teammates, so they can see it too. However, your internet has stopped working, and you need to send these files as soon as possible, such that you could finish the project in time.

Your goal is to contact the internet service provider and have them help you fix your internet problem.

Goals

- Finish the project in time
- Deliver the project
- Fix your wireless internet problems.

Technical Knowledge

Comfortable with mobile devices such as smartphones and tables, and can use a mac computer for writing tasks. Not familiar with routers, how internet function, debugging or other similar technical skills.

Figure H.1: Persona 1 used as a baseline for development of the chatbot, and also used for guiding the participants if necessary

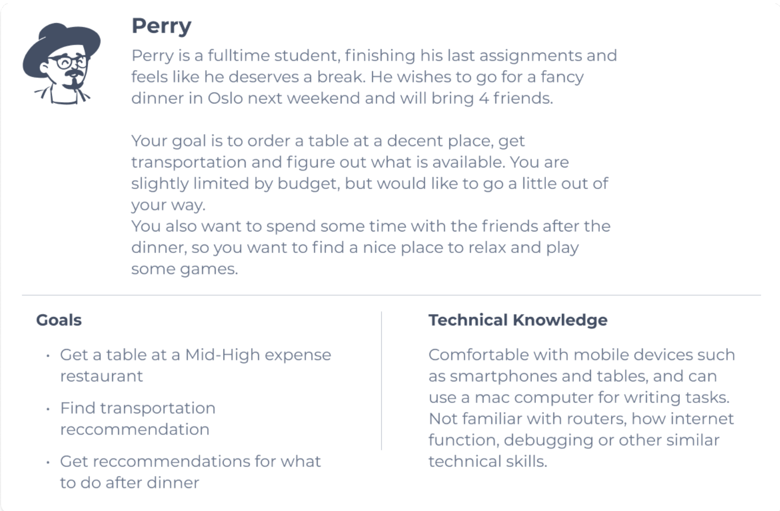


Figure H.2: Persona 2 used as a baseline for development of the chatbot, and also used for guiding the participants if necessary

Appendix I

Chat Example

Excerpt from the best chat with CB1.

```
USER: I'm celebrating finishing my thesis for my bachelors and want
to find a great place to eat.
CB1: Congratulations on completing the thesis! Oslo offers a
variety of options for food, do you have any specific price range
or food requests?
USER: Thank you, I would like a medium to high priced restaurant
serving sushi.
CB1: I have several options for you. But first, how many people
are you bringing with you?
USER: I am bringing four friends.
CB1: Great! When do you want to celebrate?
USER: This weekend would be good around 19 would be great. Are there
any available tables?
CB1: I found one restaurants that fits that criteria named
Alexanders Sushi. To sum up, A table for five at 19:30, would you
like me to book for you?
USER: Yes, please!
CB1: Ok, your restaurant table at Alexanders Sushi is booked for
19:30 this Saturday, Reference number FKRO2HOW . Do you need help
with anything else?
USER: I would like to know about any place to go for a drink after
dinner.
CB1: I can help with that. What sort of bar are you looking for?
USER: I would like a relaxing bar, but it should have a dance floor.
CB1: I found a good bar with nice reviews located in the centre area
of town which has both a calm bar and a dance floor close to the
restaurant.
USER: What is the address, please.
CB1: The address is the "Kirkegata 10". Did you want any more
information about the bar?
```

```
USER: Yes, I need the entrance fee.
CB1: The bar did not list any entrance fee, but I found several
reviews saying it costs 150 kroners.
[...]
USER: Thanks so much. That is all I need for today. Bye.
CB1: Have a nice stay!
```

Appendix J

Code

The code shown in code J.1 was made to cooperate (middleware) between several services, such as a chatbot running on a web socket and a chat service such as Rocket.Chat. With this middleware to handle the data, we could ensure that the data were properly stored in a file for analysis when a session was performed (J.1 lines 64-75). Rocket.Chat is a self-hosted chat application secured by E2E encryption (Rocket.Chat n.d.). This software was selected for privacy reasons and to keep the data from testing safe. It was hosted on a Digital Ocean droplet (Linux server), which was quickly deleted after the test was completed. We could also ensure that the data was not stored for any purpose other than our own. This allowed us to utilize hardware from other places without compromising the security of the hardware or the user. In addition, this allowed us to give users access through the internet. All data transported was done through a Secure Sockets Layer on all endpoints, both user and chatbot hardware.

The purpose of using middleware to solve the connection problems, was researched in a course *IDG3006 Web of Things*, where a middleware server was created to handle data in large quantities. This showed an effective method of building software that can handle both trafic in and out. While the course IDG3006 focused mainly on connecting devices together through Node environment, it formed the basis for handling content through internet connected devices.

J.1 Middelware

The code J.1 utilizes both the RocketChatAPI and RocketChat PyPi packages, this is due to some functionality of these packages that were missing. RocketChatAPI handles the selection of rooms and the fetching of data from rooms, while RocketChat handles the sending of data to rooms.

The following code was connected to a websocket available in the chatbot server, which allowed for bi-directional sending of messages.

Code listing J.1: Python code for running a middleware that handles data transfer between the server and the software that runs the chatbot

```python
import os
import time
import thread
import datetime
import websocket
import multiprocessing as mp
from requests import sessions
from dotenv import load_dotenv
from rocketchat.api import RocketChatAPI
from rocketchat_API.rocketchat import RocketChat



load_dotenv()

api = RocketChatAPI(settings={'token': os.environ.get("api-token"),
                              'user_id': os.environ.get("user-id"),
                              'domain': os.environ.get("domain")})
rocket = RocketChat(os.environ.get("user"),
                    os.environ.get("password"),
                    server_url=os.environ.get("domain"))


# Initialize the neccessary variables
selectedRoom = None
UnreadChatlog = []      # List of unread chatlogs, used for sending to server.
Chatlog = []            # Entire log of chat
me = None               # The user who is logged in
MessagesToSend = []     # Messages to send to the rocketchat server


# This function loads all required data and gets information from the
# Rocket.Chat server.
def bootstrap():
    global selectedRoom
    global me

    try:
        me = rocket.me().json()['name']
        if not me:
            raise Exception ("Could␣not␣get␣user␣id")
    except Exception as e:
        print(e)
        exit()
        return
```

```

    # find rooms available
    rooms = api.get_private_rooms()
    # list and enumerate rooms
    for i, room in enumerate(rooms):
        print(f'{i} - {room["name"]}')

    while True:
        try:
            print('Select a room between 0 and {}'.format(len(rooms) - 1))
            roomnum = int(input("\nSelect a room: "))
            # check if roomnum is smaller than length of rooms and if not negative
            if roomnum < len(rooms) and roomnum >= 0:
                selectedRoom = rooms[roomnum]['id']
                break
            else:
                raise ValueError("Invalid room number")
        except ValueError as e:
            print(e)
    print("Bootstrap done. Moving on to main loop.")


def writeChatlogToFile():
    global Chatlog
    # Get current timestamp
    timestamp = datetime.datetime.now().strftime("%Y-%m-%d__%H.%M.%S")
    # Create a new file inside logs folder with the timestamp as the name
    try:
        # write all data in chatlogs and unreadchatlogs to a csv file
        with open(f'logs/{timestamp}.csv', 'w') as f:
            f.write('ts|username|msg\n')
            f.writelines(f'{i["ts"]}|{i["username"]}|{i["msg"]}\n' for i in Chatlog)
    except Exception as e:
        print(e)



def getAllMessages(room, cl, ucl):
    global Chatlog
    global UnreadChatlog

    messages = api.get_private_room_history(room, count=10)["messages"]
    for message in messages:
        # Check if a message containing an id does not already exists in
        # Chatlog OR UnreadChatlog
        if message["_id"] not in [i["_id"] for i in cl] and\
                message["_id"] not in [i["_id"] for i in ucl]:
            ucl.append(
                {'_id': str(message["_id"]),
                 'ts': str(message["ts"]),
                 'username': str(message["u"]["username"]),
                 'msg': str(message["msg"])})
            # print("Added message to UnreadChatlog")


def sendMessage(msg):
    global selectedRoom
    # Send message to selected room
    rocket.chat_post_message(msg, room_id=selectedRoom)

```

```python


def receive_messages(ws):
    # Loop forever
    while True:
        # Wait for a message from the websocket
        message = ws.recv()
        # Add the message to UnreadChatlog
        UnreadChatlog.append(message)
        # Print the message to the console
        print(message)


def mainLoop(cl,ucl, room):
    try:
        while True:
            getAllMessages(room, cl, ucl)
            for message in ucl:
                # when printed, add message to hasRead and remove from Chatlog
                if not message:
                    continue
                else:
                    cl.append(message)
                    ucl.remove(message)
            time.sleep(0.5)
    except KeyboardInterrupt:
        print("\nExiting...")
    except Exception as e:
        print(e)



def on_message(ws, message):
    global MessagesToSend
    # Add the message to MessagesToSend
    MessagesToSend.append(message)

def on_error(ws, error):
    print(error)

def on_close(ws):
    print("###␣closed␣###")

def on_open(ws):
    def run(*args):
        while True:
            for message in UnreadChatlog:
                # Send the message to the AI
                ws.send(message)
                # Remove the message from UnreadChatlog
                UnreadChatlog.remove(message)
            time.sleep(0.5)

    thread.start_new_thread(run, ())



def soc():
    ws = websocket.WebSocketApp(os.environ["AISERVER"],
                                on_message=on_message,
```

```
                                   on_error=on_error,
                                   on_close=on_close)
    ws.on_open = on_open

    ws.run_forever()




if __name__ == '__main__':
    # Starts the required modules and loads the project variables.
    bootstrap()

    # Create a manager to share the Chatlog and UnreadChatlog between processes
    manager = mp.Manager()
    Chatlog = manager.list()
    UnreadChatlog = manager.list()


    # Create a new process for the main loop
    p = mp.Process(target=mainLoop, args=(Chatlog, UnreadChatlog, selectedRoom), )
    p.start()


    # on keyboard interrupt write chatlog to file
    try:
        soc()
        while True:
            time.sleep(1)
            if len(MessagesToSend) > 0:
                for message in MessagesToSend:
                    sendMessage(message)
                MessagesToSend.clear()

    except KeyboardInterrupt:
        print("\n")
        print("Keyboar␣Interrupt␣(Exiting)")
    except Exception as e:
        print(e)
        print("\n")
        print("Error␣(Exiting)")
        exit()
    finally:
        p.join()
        p.terminate()
        writeChatlogToFile()
        print(Chatlog)

        print("Exited")
        exit()
```

J.1.1 Software Flowchat

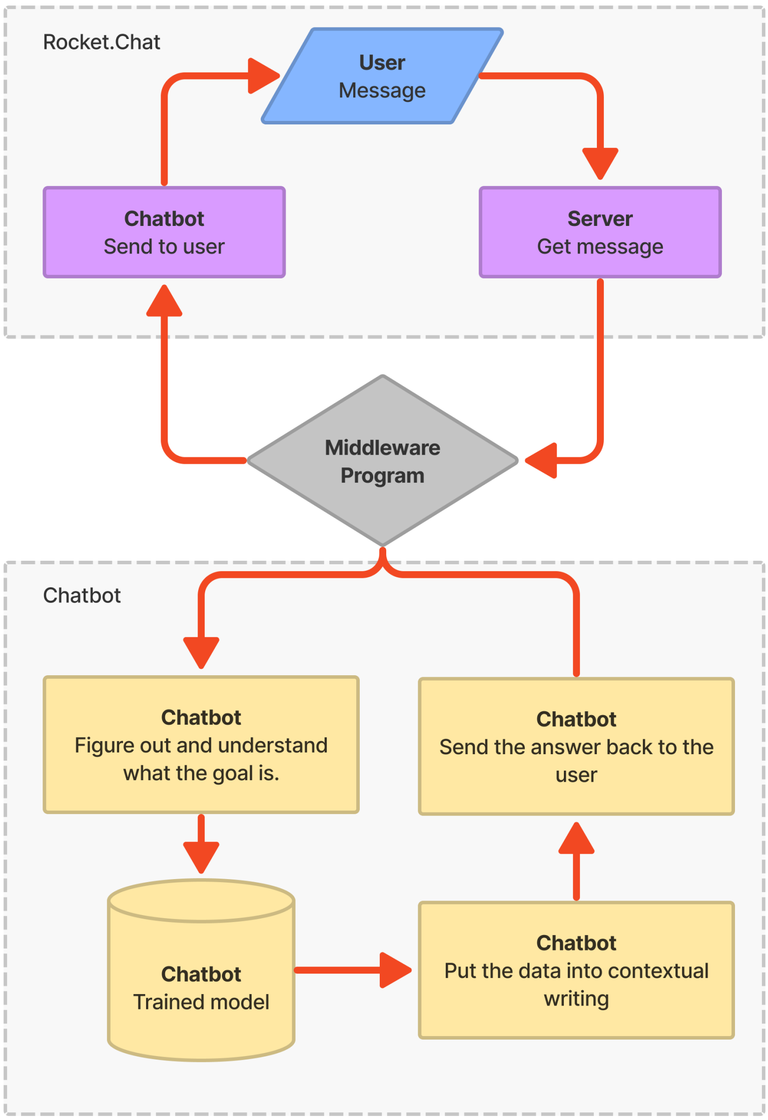


Figure J.1: How the program functioned and connected.

J.2 MultiWoZ Dataset Exporter

This code was used to export the data used for the Wizard of Oz method from the dataset used (Paweł Budzianowski n.d.)

Code listing J.2: Python code for exporting the conversation between User and system

```python
import json

# user input the filename of the json file
filename = input("Please enter the filename of the json file: ")

with open(filename) as f:
    data = json.load(f)
    text = []
    # Get loop through turns
    for cturn in data:
        text.append('\nID: ' + cturn['dialogue_id'])
        turns = cturn['turns']

        for turn in turns:
            # get speaker
            speaker = turn['speaker']
            # get utterance
            utterance = turn['utterance']

            # append speaker and utterance to text using following format:
            # speaker: utterance
            text.append(speaker + ': ' + utterance)

    # Write text to file
    with open('dataset.txt', 'w') as d:
        for line in text:
            d.write(line + '\n')
        d.close()
    f.close()

print('Done!')
```

André Tørlen Lønvik

The Highly Sensitive Person and Chatbots

NTNU
Norwegian University of
Science and Technology