**FINAL REPORT**

**CHATBOT**

**By,**

**Chirag Rohilla 1181325 A02**

**Ananya Mishra 11809064 B49**

**Vaibhav Yelam 11806530 B46**

**Sandu Yeshwanth Ravi 11806514 A33**

Description:

In our project we explore how a Chabot can give information to students about school-related information. In the first iteration of the project we created a Chabot for giving students information about where to get coffee etc. at IFI. One of our hypothesis was that information given by Chabot’s would be useful for new students at IFI, giving them information about things that we consider to be important when you’re a first year students. In the second iteration we wanted to explore the use of Chabot’s through theory and used this in combination with testing to learn more about how a Chabot for this context should be. In the final iteration, iteration three, we improved and changed the Chabot based on the results from the last iteration and made a plan for evaluate the Chabot. The plan was then executed with five participants. In our conclusion we discuss the results from the evaluation in the light of our research question.

Questions: Using a chatbot in a context

We wanted to investigate users' trust in an AI system such as a chatbot. We therefore designed a research questions we wanted to look further into

“How will helpfulness affect trust in chatbot technology for students at If when it comes to school-related information?”

A chatbot needs a purpose, and if we consider that if this purpose is to be helpful, it also needs to gain trust from the users. There is no need to ask a chatbot for help if you don’t trust the information it gives you. With this in mind we consider the first question to be a bit too ambiguous and large for us to investigate in this course. We have therefore used this question as a guideline for what we can actually manage to explore in this course and what we can find on the existing literature in this field. Trust is an important factor for reliance on and implementation of technology (Lee & See, 2004). In relationships trust means being reliable, having confidence in the other person both physically and emotionally (Lewicki & Bunker, 1995). So one can say that trust will also play a role in the interplay between human and machine. The problem with systems taking control is that it’s often hard for people to rely upon it appropriately. Because people respond to technology socially, trust influences dependence in it. So trust will inevitably guide reliance when we are faced with complex and unanticipated situations.we use systems to navigate and make decisions about 3 our health, finances, relationships, and future — they must be trustworthy. In human-technology interaction trust is an example of the important influence of affect and emotions. Emotional feedback in technology is not only important for acceptance, but can also make a fundamental improvement regarding safety and performance (Lee & See, 2004). To make the project more feasible we wanted to explore the following questions:

1. How useful is information given by a chatbot compared to a human counsellor?

2. Does students find information given by a chatbot trustworthy?

By exploring these questions we hoped to get indicators on how students experience interacting with a chatbot contra interacting with a human, and address if the students prefer one communication format over the other. This was done via selected methods in the design process, see chapter 4. Due to time constraints we later in the project had to focus our efforts more on the second question

Background:

Chatbots has emerged as a hot topic in the latest years, and it is used by numerous companies in various areas - help desk tools, automatic telephone answering systems, e-commerce and so on. Even though the technology has been around since the 60’s (Atwell & Shawar, 2007). Why are we suddenly so interested in this technology now? This can likely be explained by the recent year's advancements in messaging applications and AI technology (Brandtzaeg & Følstad, 2017).

In the article Chatbots: Are they really useful? Atwell and Shawar provide real-life examples of different chatbots in different contexts. One of the examples is Sophia, a robot that was developed to assist in mathematics at Harvard by answering students questions. This turned out to be applicable in many other contexts. Living in Norway you have probably noticed “Kommune Kari”. A chatbot that many of the municipality have available on their web-pages. Kari is there to answer “easy” questions like “when will the garbage truck come?” and “where can I find available jobs?”. Kari’s goal and the job is to provide information so that you as a user don’t have to navigate the “massive information flow” (Schibevaag, 2017). This way of using a chatbot is a part of the Question Answering (QA) field which is a combination between AI and information retrieval (Molla & Vicedo, 2007). QA can be defined as:

“... the task whereby an automated machine (such as a computer) answers arbitrary questions formulated in natural language. QA systems are especially useful in situations in which a user needs to know a very specific piece of information and does not have the time—or just does not want—to read all the available documentation related to the search toppic in order to solve the problem at hand”. (Molla & Vicedo, 2007).

Sophia and Kari are examples of chatbots that operate in “very specific” domains. This means that if you were to ask Kari about math and Sophia about when the garbage truck comes none of them would know the answer - because the question is outside of their domain. Chatbots have what is called a natural language user interface and therefore communicate with users via natural language ㅡ how a human would talk on a regular basis (Brandtzaeg & Følstad, 2017). Therefore they use what is called natural language processing (NLP) where the chatbot uses computational techniques to analyze text, where the goal is to produce a human-like answer based on a linguistic analysis (Hirschberg & Manning, 2015). For a chatbot to be especially useful to a certain domain some criteria have to be met. Minock (2005) proposes the following criteria for a domain to be successful in answering domain-specific questions: a domain should be circumscribed.complex and practical this is summerized in the table below.

|  |  |
| --- | --- |
| Criteria | Description |
| circumscribed | Clearly defined knowledge sources and comprehensive resources available(a database etc) |
| Complex | If you could develop a simple FAQ then it would not be useful with QA system. There has to be some level of complexity in the domain while still being able to meet the circumscribed criteria. |
| practical | Should be of use to a large group of people in the domain and take into account: how the users will formulate questions, what is commonly asked and how detailed the answers should be. |

Design process and methods:

For the project, we wanted to have a simplified user-centred approach (hereby referred to as UCD). UCD is an iterative design process in which designers focus on the users and their needs in each phase of the design process (Interaction design foundation, unknown). UCD 5 calls for involving users throughout the design process via a variety of research and design techniques so as to create highly usable and accessible products for them. The reason why we wanted to have a UCD design approach is to use the chatbot to explore how the users can, wish and needs to use the chatbot to achieve their goals. Our goal was to facilitate user involvement through interviews and to learn about their context. The interviews was small where we tried to understand people’s opinion about the subject. They were not only a conversation between the us and the participant but we also asked participants to execute some tasks interacting with a chatbot. After we asked them questions from experience.

How the chatbot meets Minock’s three criteria:

Circumscribed - the information given to first year students are usually dispersed on different sites and information channels. The information are usually given in a way where the students have to perform workarounds to retrieve the information. A lot of information is not written and usually learned and retrieved from other older students. This somewhat contradicts the goal of the system being fully circumscribed. Most of the information is found at the UiO webpage which we see as a “circumscribed source “but we also want to include the more verbal information. Complex - the UiO webpage has many versions of FAQ´s but is in our experience sometimes to general. Because of the dispersed information and the different types of information a fully function chatbot in a school context should have, this could not be realised by a simple FAQ. Making a chatbot that is more advanced than a FAQ is not feasible in our project. But is rather a reason for using a chatbot in a school context. such as IFI.

Practical –

Our chatbot is designed to meet the needs of a large group of students at IFI. We believe that it is practical in the sense that it detects short questions like: “I am hungry” and “Food” or “Where is Epsilon?” and “I can’t find my classroom”. Which in turn can reduce the time it takes for the students to locate this information. This can also be used as a way to gather data on the information that students are interested in. 5.2 Persona In the making of the prototype we also formed a persona for the chatbot to make the chatbot consistent in its language. This worked as a guideline in the design of the chatbot and was very helpful since it gave us a common understanding of the chatbots characteristics. We focused on building the chatbot as an engaging partner with a “happy tone” and a sense of humour, including GIFs to make the experience more fun and intriguing.

Earl testing and Findings: -

In the beginning of our project we wanted to test the first version of our chatbot (from appendix 1) on first year students. This was late in the fall and most of the first year students were familiar with a lot of the answers our chatbot could provide. We therefore developed a scenario to help the participants imagine the context of use (see figure 2). We wanted to test this early version of the prototype to get input on what the chatbot could and could not answer in the future. After the test was completed we had a short interview with the participants. The main purpose for this test was to see how the participants interacted with the prototype and find out if a chatbot could be suitable to find the information they needed. Before the testing we also carried out a pilot test to find immediate flaws in the plan



Results from the first testing:

The first participant enjoyed talking to the bot, but stressed the fact that you had to “talk like “a dummy” for it to understand what you were asking. The participant pointed out that this really would have come in handy in his first weeks at the university, as he didn’t always know who to ask - especially if he was in a hurry. He pointed out that the prototype needs to get more features like tell you exam dates, or “ifi life-hacks, like get your coffee before all of the students have their break”. The second participant was a bit frustrated that the chatbot wasn’t flexible enough (Fig.3). “I don’t like having to guess what questions to ask”. He would like more instructions to know how to get more out of the chatbot.

The third participant had also problems with understanding what the chatbot could do. When given a hint for what the chatbot could do, the chatbot did not function properly. Here we tried to restart the system and then the chatbot displayed it´s welcome message一 what it could do. Afterwards it was more clear what the participant could ask it, but the chatbot did not always give the response that the participant wanted.

Evaluating:

the chatbot We wanted to evaluate the prototype in the right context, which for the IFI chatbot was at IFI. As mentioned before, most of the new students are more or less ‘integrated’ per now we could not test on “real potential users”. However we consider IFI-students as a good substitute since they have been in the situation before and a group that we easily can make contact with.

The evaluation was formed as a formative usability test where the goal is to look at metrics that are more qualitative than quantitative (Lazar et. al. 2017). In the evaluation we wanted to combine small semi-structured interviews with the users executing tasks because this could give us more information about the experience beyond the metrics.

|  |  |
| --- | --- |
| Set up | Candidates:  Five randomely picked evaluators. The only criteria is that they have to be students from IFI.  Context:  In the institute for the informatics building. |
| Warming up | -Have you talked with a chatbot before? If yes: What type of chatbot?  - How do feel about getting information from a chatbot? Do you consider the information as more or less reliable? |
| Task’s | Scenario:  Imagine you are a new student. Use the chatbot and try to figure out when your next lecture starts, which room it is in and where is it located? Later you are feeling thirsty and are interested in a cup of coffee near the university.  Tasks:  Use the chatbot to find out:  Where is the room named ‘Normarc’?  Where can you buy coffee at ifi?  Have a chat with the chatbot |
| Control Questions | - Did you feel like the chatbot gave you a good answer?  - Do you think that the answer from the chatbot was trustworthy?  - Do you feel a need to ‘double ch eck’ the answers you got from the chatbot?  - If you were to rate this chatbot from 1-6 where six is the best, what would you rate it?  - If low: What improvements does it need to get a six? |
|  |  |

Findings from the evaluation:

All of our participants reported that they had interacted with chatbots before, but had very little knowledge about how they worked. They found the chatbot to be nice to interact with and enjoyed that it had a friendly and casual tone. One of the participants said that she did not want a chatbot that felt too ‘human-like’, and that the prototype did not feel ‘human-like’ at all. This became clear when the same errorr message appears several times during the test.

They found it hard to get the right answer but when they did they were very satisfied with the answers. “It was a good answer when I finally got the right one.”. It was pointed out that the chatbot was not a smart chatbot, but that it provided the most necessary information sparing them from precious time spent on ‘Google’

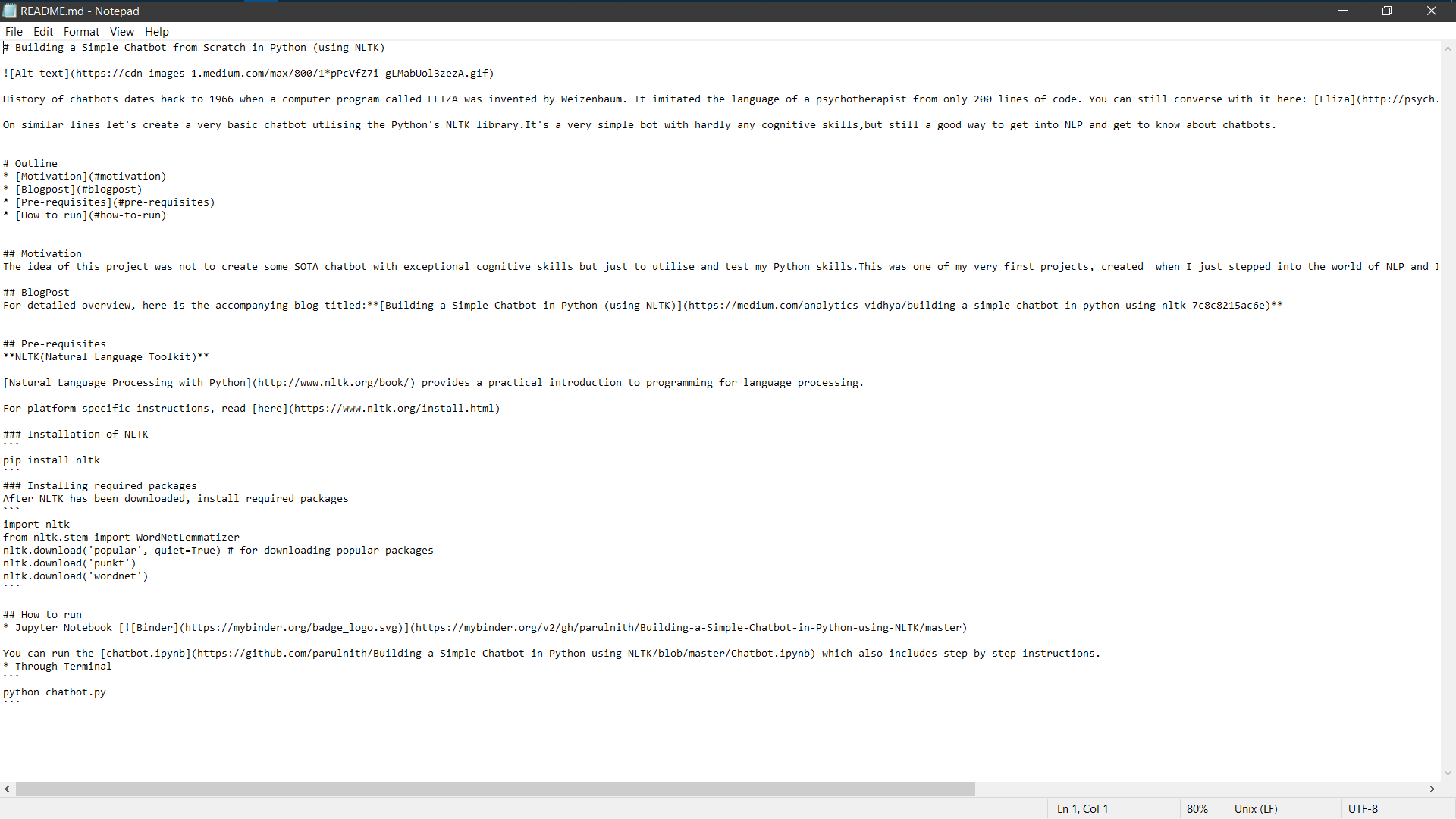
Discussion and conclusion:

When testing the last prototype we got findings suggesting that the participants did not have a problem with getting information from a chatbot instead of a human. The information that they got was not seen as less trustworthy, this could be supported by the fact that the chatbot provided a source for the information it gave. It has been interesting to investigate how the participants interacted with the chatbot and how they reported on it afterwards. Our findings have some indicators leading towards that a chatbot could be a good alternative for acting as a helpful friend for freshmans at a new school. Still we have to stress the fact that the chatbot was not very intelligent and that the evaluators had to adjust their language to match the chatbots.

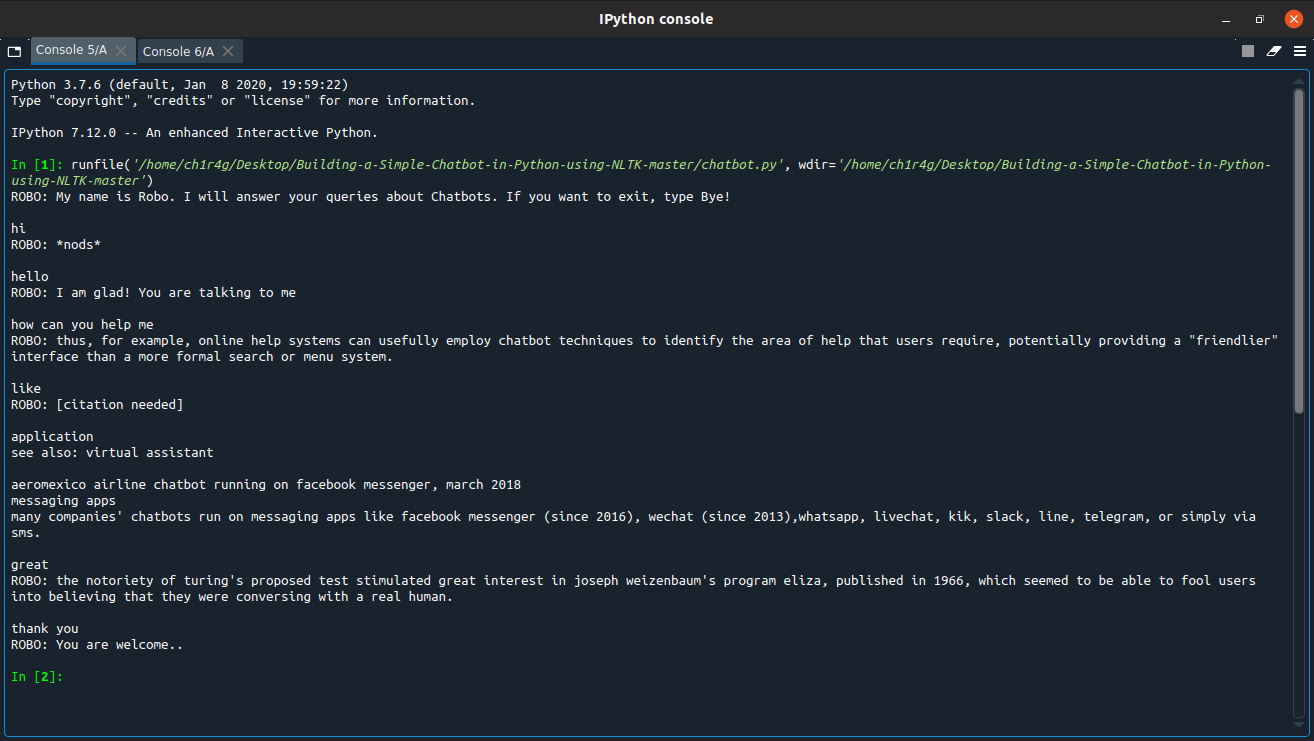
Because of the scope of the project we did not have time to conduct as much user testing and re-design to the chatbot as we would have liked. This has an impact on the validity of our research. Through the project we have touched on some theory when making the chatbot, but this should also have a larger focus for higher validity. Even though the participants trusted the information given in this project we cannot say that people trusts a chatbot as much as they trust a human being. There are also biases in our project, one of them is that all the students that we included in the project already knew a lot of the answer the prototype could provide. Another bias is that the information the chatbot provides could be seen as “casual” and are not crucial and/or vital This could have had an impact on the results regarding trustworthiness. With that being said we also think that some of our findings could give some insights into how a very small group of people think about using a chatbot to gain information in a school context. Some of the characteristics of our chatbot was viewed as appropriate for the given context, like “casualness” and links to where the information was gathered. If the IFI chatbot is to be further developed. This could be something to draw upon.

**Samples: -**

**Code: -**

****

**Output: -**

****