



Patrick Steyer, BSc

# **Rebo - A feasibility study on moving from a static dialogue structure for reflection for apprentices to a conversational agent**

## **Master's Thesis**

to achieve the university degree of

Master of Science

Master's degree programme: Computer Science

submitted to

**Graz University of Technology**

Supervisors

Assoc.-Prof. Dr. Viktoria Pammer-Schindler

Irmtraud Wolfbauer

Graz, June 2020



## Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

---

Date

---

Signature



## Danksagung

An dieser Stelle möchte ich mich bei Jedem bedanken, der mich während meines Studiums unterstützt und motiviert hat.

Zuerst gebührt mein Dank Assoc.-Prof. Dr. Viktoria Pammer-Schindler und Irmtraud Wolfbauer, die meine Masterarbeit betreut haben. Für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit möchte ich mich herzlich bedanken.

Ich bedanke mich bei der Arbeiterkammer Steiermark, im speziellen bei Mag. Alexandra Hörmann, Alexander Perissutti und Mag. Karl Snieder für die finanzielle Unterstützung meiner Forschung.

Des Weiteren möchte ich mich bei meinen Kommilitonen für die aufregenden gemeinsamen Jahre bedanken.

Einen besonderen Dank möchte ich meiner Schwester Isabella für das Korrekturlesen der gesamten Masterarbeit aussprechen.

Abschließend möchte ich mich bei meinen Eltern bedanken, die mir mein Studium durch ihre Unterstützung ermöglicht haben und stets ein offenes Ohr für mich hatten.



## Abstract

The goal of the master thesis was, given a pre-defined dialogue structure for reflection, to implement this as a conversational agent using the Bazaar framework. Firstly, in this master thesis we wanted to study the technical feasibility of this, as the Bazaar framework was designed to moderate collaborative discussions, not to lead bilateral conversations. Second, we wanted to study the acceptance of such a chatbot in the context of apprenticeship training and gather feedback from the apprentices.

During the study a chatbot called Rebo was developed. This chatbot is embedded into a website, where the apprentices can chat with Rebo. The data of the conversations is stored in a database, which can be accessed by anyone to analyse the data in the future. The chatbot was tested in a workshop with seven apprentices in the fourth year of apprenticeship training. The apprentices had to fulfill a task and afterwards they were asked to chat with Rebo. Rebo asked them questions, which were supposed to guide the apprentices through a reflection process, so they would reflect on the task fulfilled before.

Even though the chatbot guided the apprentices through the reflective dialogue, only small amounts of reflectiveness could be detected. This result appears to be due to the fact that the apprentices were asked to reflect on a theoretical rather than on a practical task. This also led to more negative feedback by the apprentices than anticipated. On the other hand, most of the conversations were labelled coherent, which indicates that the dialogue structure itself works. Further, it can be said that Rebo acted correctly in all situations where adaptiveness of the dialogue was required.

The results of the study suggest that Rebo should only be used after practical tasks, where reflection is more suited. This assumption should be confirmed in a larger field study. The fact that most of the apprentices' feedback on the chatbot was either negative or neutral, seems to be related to the type of task they had to do rather than to the chatbot itself. This appears to be supported by the feedback provided in the third flash light round. The bazaar framework provided a suitable starting point to implement our chatbot, though some of the assumptions that our added features were based on need to be investigated further in future studies.





# Contents

<b>1. Introduction</b>	<b>17</b>
<b>2. Related Work</b>	<b>19</b>
2.1. Defining the term "Reflection" . . . . .	19
2.2. Conversational Agents . . . . .	19
2.3. Synthesis . . . . .	26
<b>3. Research Plan</b>	<b>29</b>
3.1. Use Case Setting . . . . .	29
3.2. Research Questions . . . . .	29
3.3. Evaluation . . . . .	30
<b>4. Rebo</b>	<b>31</b>
4.1. Preliminary Work . . . . .	31
4.1.1. Framework . . . . .	32
4.1.2. Text Classification Methods . . . . .	34
4.2. Implementation Server . . . . .	36
4.2.1. Set-up of the Server . . . . .	36
4.2.2. Maintenance Features . . . . .	37
4.2.3. Website . . . . .	38
4.3. Implementation Chatbot . . . . .	44
<b>5. Evaluation Method</b>	<b>51</b>
5.1. Preliminary Work . . . . .	51
5.2. Study Procedure . . . . .	52
5.3. Participants . . . . .	55
<b>6. Results</b>	<b>57</b>
6.1. Coherence of Conversations . . . . .	57
6.2. Reflectiveness of Conversations . . . . .	57
6.3. Feedback . . . . .	58
6.4. General Observations . . . . .	60
<b>7. Discussion</b>	<b>65</b>
7.1. Workshop Results . . . . .	65

7.2. Technical Assumptions . . . . .	68
7.3. Future Research . . . . .	73
<b>8. Conclusion</b>	<b>77</b>
<b>References</b>	<b>79</b>
<b>A. Transcript of the Feedback Rounds</b>	<b>85</b>
<b>B. Python Dictionary Splitter</b>	<b>96</b>

## List of Figures

1.	Simple System Overview . . . . .	31
2.	Architecture . . . . .	33
3.	Website by the developers of the Bazaar framework . . . . .	39
4.	Website updated for the use with Rebo . . . . .	40
5.	Rebo Logo . . . . .	40
6.	Website running in essay mode . . . . .	42
7.	Reflection script of Rebo Junior . . . . .	45
8.	Reflective dialogue used by Rebo . . . . .	46
9.	Example conversation utilizing multi message response . . . . .	49
10.	Flow diagram of the workshop . . . . .	53
11.	Conversation with negative outcome regarding reflection and coherence . . . . .	61
12.	Conversation where apprentice insults Rebo . . . . .	62
13.	Coherent conversation during the workshop . . . . .	63
14.	Coherent conversation during the workshop, where a low level of reflection was achieved . . . . .	64



## Listings

1. Shell script for exporting the database . . . . . 38
2. Automatic shell launch script for Rebo . . . . . 43
3. Self-written Python program to split the polart dictionary . . 96



## List of Tables

1.	Message pool of Rebo for each step in the dialogue. . . . .	47
2.	Interaction statistics between the apprentices and Rebo with the focus on coherence. . . . .	57
3.	Conversation statistics with the focus on whether reflection can be recognized. . . . .	58
4.	Feedback collected during the three flash light rounds with the apprentices. . . . .	59





## 1. Introduction

A chatbot can interact with users by using natural language and is often called a conversational agent (Wallace, 2004). Such chatbots are found across different industries and are usually either text or speech based. Nowadays, people often encounter speech based conversational agents like Amazon's Alexa and the text based ones are usually encountered on company websites or social media. In this work we focus on a text based conversational agent, which is a chatbot, combined with reflective learning. In reflective learning, the goal is to learn from past experiences (Fessler et al., 2017). There are many techniques available that can help users to reflect in a more systematic way and that can help to increase their motivation to do so (Fessler et al., 2017).

Conversational agents (CAs) have been studied for decades. Chatbots are well investigated in educational settings for students, customer support and in workplace settings, whereas our target audience for the study are apprentices, which means that we have an edge case for an educational setting. This master thesis is embedded in a research stream, in which conversational agents for reflection are investigated; and builds upon prior work in this research stream on a default dialogue structure for apprentices to reflect on learning tasks (Wolfbauer, Pammer-Schindler, and Rose, 2020).

In this study we use the Bazaar framework to develop a chatbot called Rebo. We tried to find out in how far the Bazaar framework is suitable to develop a conversational agent that guides users through a reflective dialogue. This dialogue changes according to the messages of the user and is called an adaptive dialogue. During the study we set up a server, different bugs were fixed and several features for the chatbot were added. The server is used to host the website with the chatrooms, where the conversations with Rebo are held. Some bugs in the framework were fixed to achieve a higher reliability and stability of the chatbot. Features that were added to the chatbot are for instance the possibility to respond to questions with multiple messages and a reminder message if a user is not taking the conversation seriously.

The field study was carried out with seven apprentices in the fields of electrical engineering, metal and mechatronics. In addition to their apprenticeship training in their respective companies, the apprentices gain knowledge and skills through additional workshops. We conducted our field study in one

of these workshops with seven apprentices in the fourth and final year of apprenticeship training. Ideally, the apprentices should reflect with Rebo upon a task done and learn something from it. By reflecting with Rebo the apprentices should be able to gain insights into their way of working and it should help them realize why some tasks were easy or hard for them. In the conducted field study the apprentices got a task from their instructors and upon finishing it they were asked to interact with the chatbot, which uses an adaptive reflective dialogue to guide the apprentices through a reflection process with the goal of improving the learning experience of apprentices (Wolfbauer, Pammer-Schindler, and Rose, 2020). For this purpose the dialogue guides the user through the reflection process which should lead to develop learning skills that can be useful for the apprentices their whole lives long.

In the next section past research on chatbots and reflection are discussed. After that, the research plan of this study is presented in Section 3, followed by Section 4, which covers the artefact of this thesis, namely the chatbot Rebo. Several aspects of design decisions and implementation are discussed in this section. In Section 5 the evaluation method of this study is explained. Section 6 covers the results of the study, which are then discussed in more detail in Section 7.

## 2. Related Work

The topic of this thesis is the development of a conversational agent for apprentices. The system uses questions to lead the apprentice through the conversation and has the specific goal of leading the apprentice through the process of reflecting on a task that was completed beforehand. This chapter will give an overview of the work which has been conducted in the field of reflection and conversational agents.

### 2.1. Defining the term "Reflection"

Reflection is the conscious activity of exploring past experiences in order to learn something from them (Kocielnik, Avrahami, et al., 2018; Fessler et al., 2017). It is a necessity for everyone to learn from the past (Schön, 1983) and it is also important for success at work (Pammer, Krogstie, and Prilla, 2017). Furthermore reflection is a powerful tool to build up confidence to achieve set goals (Di Stefano et al., 2014). Self-reflection is an acquired skill which can help students to enhance their thinking before starting their working lives (Kovanović et al., 2018). It has been found that the skill can be more easily developed if the process of reflection is supported, for example through prompts (Renner et al., 2016) or mini-dialogues (Kocielnik, Xiao, et al., 2018).

### 2.2. Conversational Agents

The field of conversational agents, also called chatbots, has a wide field of possible applications. Therefore the amount of research on the topic has been increasing rapidly in the past decade. Some of the most popular research fields for the use of said chatbots are customer service, health care, support for workers and also tutoring assistants for students. In customer service, for instance they can be used to relieve human agents by answering repetitive customer questions (Cui et al., 2017). In fields like health care using chatbots can make people more sincere, since the computer will not judge them based on their answers to certain questions (Lucas et al., 2014).

The developed tutoring systems of the past have been quite successful (Graesser, VanLehn, et al., 2001) and through their use significant learning gains could be detected. The user gets a task or question and then has to construct an answer to it. Constructing an elaborate in-depth answer to a question may prove quite a challenge (Graesser, D. S. McNamara, and VanLehn, 2005). In order to help students to construct knowledge AutoTutor (Graesser, VanLehn, et al., 2001) was created. The chatbot asks questions, such as why and how, to lead the student to detailed explanations. To improve the comprehension of texts iStart (D. McNamara, Levinstein, and Boonthum, 2004) supports students with reading strategy training. The explanations of the user are matched to a sample explanation using properties, such as words in common and the length of the message. The results showed that iStart statistically improved the ability to understand texts compared to the control group.

Another important aspect of learning is gaining factual knowledge, which can relate to any topic. Ruan et al. wanted to develop an effective way to improve students' factual knowledge. For the sake of having a more engaging learning experience, a dialogue based conversational agent called QuizBot (Ruan et al., 2019) was created. To find out if the system really works, the evaluation was compared to the results gained through the use of a traditional flashcard app and even though the new agent had worse efficiency, students liked it more and also had about 20% more correct answers in the tests (Ruan et al., 2019). Since it is hard for parents to talk with their children about sex, drugs and alcohol Crutzen et al. developed a chatbot for adolescents to gain knowledge about those topics (Crutzen et al., 2011). The system follows the question and answer principle and was investigated in comparison to information hot lines and search engines (Crutzen et al., 2011). Even though the database of the agent was limited it was evaluated more positively than the other two systems and seems to have the potential to reach a large group of people.

In his work Yugo Hayashi focused on how to use a conversational agent in a collaborative activity in order to improve explanations made by the user (Hayashi, 2013). In a collaborative activity two or more people interact with each other over a chatroom. In the study the effect of different utterance output methods of the chatbot like visual, text or audio, on the user was observed. The data showed that audio feedback from the conversational agent

had a positive impact on the text-based interaction of students (Hayashi, 2013).

Conversational agent systems can be enhanced if built as multi agent system, which consists of multiple small agents working together in parallel (Hettige and Karunananda, 2015). Such a multi agent system called *Bazaar* was developed at Carnegie Mellon University <sup>1</sup>. The framework is open source and focuses on collaborative learning. Another study which deals with collaborative learning was conducted by Chaudhuri et al. Its goal was to find out whether the approach of a dialogue based, pointer hint based or a combination of both were most effective as learning support (Chaudhuri et al., 2009). The results showed that the dialogue-based approach was most effective when triggered by the students themselves after they got a pointer hint (Chaudhuri et al., 2009).

Through natural language processing (NLP) it is possible for chatbots to appear human. ALICE (Wallace, 2004) was the first to utilize NLP and had a handwritten database of conversation behaviours by its developer Richard Wallace. The knowledge about conversational patterns is stored in so called AIML files, which are similar to typical XML files (AbuShawar and Atwell, 2015). ALICE was designed as a question answer system for a specific domain, but as long as a human created the required AIML files the system was able to have a conversation about different domains. AbuShawar and Atwell tried to create an automatic approach to generate such AIML files for ALICE (AbuShawar and Atwell, 2015). With their results they showed that ALICE can be easily retrained for different applications and compared to a search engine users preferred to use a chatbot for answering their questions.

Another conversational agent which can easily be used in different domains is OntBot (Al-Zubaide and Issa, 2011). To get around the problem of learning a specific markup language (e.g. AIML) for a single system, the developers of Ontbot created a technique to convert ontologies directly into a relational database. Having knowledge in database tables has advantages over the file system, for instance its simple accessibility through queries or the simple replacing of the ontology to change the domain.

---

<sup>1</sup><https://github.com/DANCEcollaborative/bazaar>

A contrast to traditional teaching agents, which lead their users through a task in order to learn something, are systems in which the corresponding agent receives the instructions from the user (Tanaka and Matsuzoe, 2012). Such a system is called teachable agent and an example for it was developed by Silvervarg and Jönsson with the goal of having students teach the agent a math game through answering multiple choice questions. When the student teaches, the so called on-task dialogues are used and by using additional social conversations, recall of the learning material should be increased (Silvervarg and Jönsson, 2012). The social conversations are implemented in the same way as with ALICE by using Artificial Intelligence Markup Language. The researchers found out that different users and their different attitudes towards the system need to be taken into account when designing such systems (Silvervarg and Jönsson, 2012).

It is hard to draw an exact line between chatbot and virtual agent, but simply spoken a virtual agent is a slightly more advanced version of a chatbot. Virtual agents usually involve not only text based conversation, but also deliver their responses over audio and often have a human like visual representation. To grasp the effect of different interfaces in a human to non-human interaction, Ciechanowski et al. conducted a study with two different systems. The goal was to find out in how far such interfaces would lead to an uncanny effect. One system used was a traditional text based chatbot and the other one had a human like avatar and responded additionally with reading the responses out loud. During the interaction psycho-physiological data was collected from the users (e.g.: EMG, ECG). The results showed that all users had experienced a higher uncanny effect with the animated chatbot and the intensity of reactions was less with the chatbot (Ciechanowski et al., 2019). A real life experiment of a virtual agent was conducted in a museum. The agent called Max provided the visitors of the museum with information about the exhibition via face to face communication. Since this was one of the first field studies of such a system (2004), the study had the goal of evaluating if a system of this kind would be accepted by the visitors (Kopp et al., 2005). Through logs the authors were able to determine that people used human conversation strategies such as greetings, farewell or insults and the log files have also shown that the users were engaged in the interaction (Kopp et al., 2005). Virtual agents can also be applied in universities and can help lecturers and students with

course related activities. For instance, LTKA-Bot (Mulyana, Hakimi, and Hendrawan, 2018) was developed for this purpose. LTKA-Bot supports course related activities from task assignment over group formation up to writing logbooks and marking. This system automates different processes and workflows in the context of lectures. Through splitting the system in many different services, it should be easy to scale and adapt it to more complex tasks in the future (Mulyana, Hakimi, and Hendrawan, 2018).

Since social media use has increased rapidly over the last years, customers have also adapted their way of contacting customer support, which may prove challenging for companies (Xu et al., 2017). Sending a tweet is easier than formulating a good e-mail and companies therefore have to deal with more requests than before. To lower the load of the customer support agents, Rama et al. created a long short term memory (LSTM) network through deep learning. The system which was trained with one million tweets is able to automatically generate responses to user questions (Xu et al., 2017). The input gets converted to a fixed length vector, which then gets fed into the LSTM to generate the response. Through evaluation it could be shown that the developed system is as good as a human for answering emotional questions, but for the other 60% of user requests, which have an informative nature, humans are a better choice. Another conversational agent in the field of customer service is used to answer queries about ID card application and is called UMAIR (Kaleem, O'Shea, and Crockett, 2014). It was designed for the language Urdu, which comes with the challenges of having no online lexical resources and no fixed word order. In order to deal with the issue of having no fixed word order, the WOW algorithm was developed. By using Levenshtein distance and all variations of the word order for a scripted pattern, string similarity is calculated. The results showed that the novel algorithm for string similarity can deal with challenges provided by languages such as Urdu and most of the conversations were able to lead to the goal of answering the user question (Kaleem, O'Shea, and Crockett, 2014). Xue et al. created the conversational agent ISA (Xue et al., 2018) to increase the effectiveness of customer service in call centers. This agent is supposed to assist the human agent with resolving customer issues. For this, the chatbot uses a bi-directional long short term memory (BiLSTM) as network layer and after transforming the sentence into uniform length, the output layer generates the classification labels. It was shown that with the

help of ISA, workers could reduce the wait time for customers. Additionally, ISA was used to function as a knowledge base.

One of the typical use cases for a chatbot is a QA system where a user asks a question and the chatbot has to answer. However, sometimes providing information is not enough, as playful interactions may be expected to achieve a good user experience (Liao et al., 2018). Liao et al. conducted a field study in the sector of Human Resources for new hires. The QA system was designed with the goal to help the new employees to find information and manage administrative tasks. By studying the field deployment the authors tried to find out the interest areas in the workers' conversations. The areas which were found are similar to the conversation areas found with the deployment of the virtual museum guide described above (Kopp et al., 2005) and include feedback giving, playful chit-chat, system inquiry and habitual communicative utterances (Liao et al., 2018). These results may help with the design of CAs and adaptive systems. Another QA system is FarmChat, which was developed to provide farmers in rural areas of India with information needed to better cultivate potatoes (Jain et al., 2018). The system was specifically designed to deal with the problem of the farmers' limited literacy. The system was created with two different systems: one only with audio and the other with audio in combination with text. The results of the study showed that the farmers appreciated the conversational agent and treated the agent as if it was a real human. This study provides developers with valuable insights that can be used for designing CAs for users with low literacy.

Chatbots have been investigated in relation to relieving mental health problems. Problems like depression or anxiety are common among college students and their frequency and severity are steadily increasing (Zivin et al., 2009). In their study Fitzpatrick et al. wanted to find out more about the ability of a text-based conversational agent to deliver cognitive behavioral therapy (CBT) (Fitzpatrick, Darcy, and Vierhile, 2017). For their study the Woebot app was used, a conversation agent specifically designed to deliver CBT through mood tracking and daily conversations. Woebot uses a decision tree with natural language input at certain nodes and suggested responses (Fitzpatrick, Darcy, and Vierhile, 2017). The participants who interacted with the chatbot for two weeks showed reduced signs of depression. On the other hand, the participants who only got the e-book "Depression



in College Students” did not experience any changes in their mental well-being. Another study in this field was carried out by the company Touchkin, which developed a chatbot app called Wysa, which also aims to increase mental well-being. The conversational agent uses self-help practices as CBT in conversations with the user. Inkster, Sarda, and Subramanian tried to determine the effectiveness of Wysa in delivering self-help practices through text-based conversations. The chatbot used machine learning to have conversations with the participants, which were mostly about relationship issues, mental well-being issues and location changes (Inkster, Sarda, and Subramanian, 2018). Results showed that participants who used the app to chat with the chatbot more frequently had a statistically higher improvement of their mood (Inkster, Sarda, and Subramanian, 2018). Another approach for mental well-being focuses on pre-emptive health care. Caring for other human beings increases your own self-compassion (Breines and Chen, 2013). Therefore Lee et al. wanted to find out if the same effect can be achieved when caring for a technological being (Lee et al., 2019). For this purpose, the system Vincent was created, which consists of two chatbots and is a care giving as well as care receiving chatbot. The latter opened up to the user about everyday mistakes, whereas the other guided the user through a dialogue like Woebot. Users were able to show compassion towards another being with care-receiving Vincent. This was done by using suggested responses and sometimes free text input. The results clearly showed that people can gain greater self-compassion by caring for a chatbot than by being cared for by such a system (Lee et al., 2019).

By using wearable devices such as fitness trackers people have gained the ability to easily collect huge amounts of data about their own physical activities (Kocielnik, Xiao, et al., 2018). A key to close the gap between collecting such data and learning from it is reflection (Li, Forlizzi, and Dey, 2010). The conversational agent called Reflection Companion is a system for mobile phones which helps users to reflect on their personal physical activity data collected with Fitbit. Reflection Companion tries to help users reflect by sending them daily mini dialogues and graphs. The goal was to trigger an engaging reflection which should furthermore increase the motivation and should lead to the creation of new behaviours (Kocielnik, Xiao, et al., 2018). The field study was able to show that the used reflective mini dialogues helped participants to reflect on their physical activities

and to motivate them. Further research to integrate external data in conversation dialogues was done by Kocielnik et al. (Kocielnik, Hsieh, and Avrahami, 2018). As with Reflection Companion, they also focused on data about physical activity with the goal to learn and reflect on data and to motivate healthy behaviours (Kocielnik, Hsieh, and Avrahami, 2018). By using external data it will be easier to lead personalized interactions with users, which is important for conversational agents, so they can adapt their dialogues according to the circumstances of the user.

To support the activity of journaling and self-learning in the workplace, the conversational agent Robota was developed (Kocielnik, Avrahami, et al., 2018). It uses text based and voice based activity journaling and should help workers to learn from their daily work through reflection. The chatbot offers a structured way of reflecting on work and uses either a typical chat conversation for the reflection or a voice based conversation. Both kinds have their own benefits and drawbacks. For the chat, the largest benefit appeared to be the ability to reread and to easily respond to questions, whereas the voice based system was found to be more personal and engaging (Kocielnik, Avrahami, et al., 2018). The results showed that workers can benefit from journaling with structured reflection prompts.

### 2.3. Synthesis

As can be seen, chatbots have been used and studied in various fields. Extensive research on chatbots has been conducted in educational settings and workplace settings. Educational settings research has so far mostly been related to high school and university students, whereas the workplace settings research has mainly been about customer support. The research connected to Rebo has the potential to complement past studies, as it focuses on a new target group, namely that of apprentices. It appears that so far, no other research in the field has been conducted on this type of subjects.

Secondly, there is some research that has been done on the use of chatbots for reflective learning specifically. However, said research has mainly focused on the fields of physical activities (Kocielnik, Xiao, et al., 2018; Kocielnik, Hsieh, and Avrahami, 2018). Only one study that used chatbots for reflection

in the workplace could be found (Kocielnik, Avrahami, et al., 2018). The present study tries to fill this gap in research by focusing on the new target group of apprentices in relation to reflecting with chatbots. An additional benefit of this study is the fact that if we try to train people to reflect on their fulfillment of tasks even before they enter the working world, they may make use of their acquired new skill right from the start of their working lives.



## 3. Research Plan

This chapter provides a quick overview of the research plan of this master thesis.

### 3.1. Use Case Setting

The use case of this work lies in apprenticeship training. In particular, this means that the study was carried out in cooperation with several Styrian companies that have joined forces to create a workshop in which their apprentices are taught the fundamental knowledge needed in their workplaces. The apprentices that were part of the study are active in the fields of electrical engineering, metal and mechatronics. In the workshops, apprentices first acquire theoretical knowledge. Then they get practical tasks to solve individually, which they have to document digitally on the ABV Digital Moodle platform upon completion. Once this step has been performed, the work of Rebo that is the subject of this study starts. Rebo steps forward to lead the apprentices through a reflection dialogue, which is supposed to help the apprentices reflect on the task completed beforehand. By doing so it should be possible for them to realize why certain tasks were easier or harder for them. Additionally, it should be possible for the apprentices to find out why the results were or were not as expected.

### 3.2. Research Questions

The goals of this master thesis were to study the technical feasibility of using Bazaar for a conversational agent that leads through reflective dialogues; and to study the agent's acceptance by apprentices in the context of apprenticeship training and gather first feedback as a step towards a full field study.

### **3.3. Evaluation**

Another goal of the present study was to evaluate Rebo's success in the apprentices' workshop. We wanted to achieve high user acceptance, in the sense of an emotionally positive reaction, ease of use and a perceived benefit from the interaction. The method used to evaluate whether these goals were met were flashlight feedback rounds conducted in a workshop with the apprentices.

## 4. Rebo

In this section the artefacts of this thesis, i.e. the chatbot Rebo and the server with the running system, will be presented. Firstly, the work done before the start of the thesis are explained. Then the main features and changes are presented. In Figure 1, a simplified overview of the system components can be seen. The files for the finished chatbot can be found at [https://github.com/Tot333/MA\\_Rebo](https://github.com/Tot333/MA_Rebo) and the files for the corresponding webserver can be found at [https://github.com/Tot333/WebServer\\_Bazaar](https://github.com/Tot333/WebServer_Bazaar).

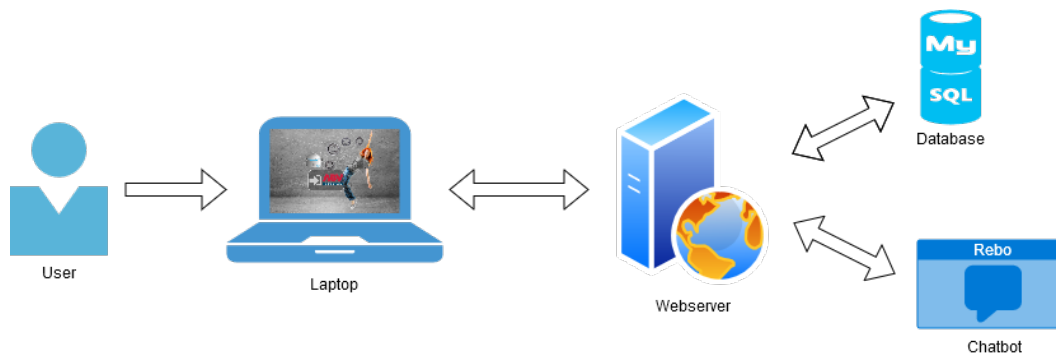


Figure 1: A simple system overview showing the main components of a conversation. A user accesses the website through an end device like a laptop. The server holds all the files for the website, database and chatbot, whereas the chatbot leads the conversation and the database holds the conversation data.

### 4.1. Preliminary Work

In this first sub section the main focus lies on the theoretical work done for this thesis and the decisions that go with it. The main points covered are the decisions on the framework as well as those for the text classification methods.

#### 4.1.1. Framework

The decision for the used framework fell onto the publicly available Bazaar framework. It was developed at Carnegie Mellon University and is a framework for collaborative learning. There were two main reasons why this framework was chosen for this thesis. First, it was clear that a pre-existing framework for the typical chatroom implementation should be used. The reason for this is quite simple. Even though many solutions for typical chatroom systems exist, programming them from scratch takes up a lot of time. Therefore, a working framework was used in order not to generate an overwhelming additional workload for this thesis. The reason why the framework for collaborative learning of Carnegie Mellon University was used is that the supervisor of this master thesis had previously worked with that particular system and its developers. Therefore, a communication system with the developers of the system was already established and available when problems occurred. Additionally, the system supports dialogue based conversations, which were used in this study. On the whole, it can be said that Bazaar was chosen in order to have a solid base construct.

The Bazaar system is a so called conversational agent and was designed as a moderator for group discussions that consists of multiple behavioural components. The moderator uses these components to interact during ongoing discussions, which should provide support for the users in a group learning task. Since the framework was specifically designed to support group learning and not to actively lead a full conversation, it had to be adapted to be suitable for the apprentices. In order to do this the first goal was to understand the system and the multiple components it comes with. Figure 2 shows an overview of these components. After analysing the code base, reflections with the supervisor and getting to know the system, it became relatively clear that the system is suitable for our use case, but that some of the complexity is not necessary. This basically means that many of the behavioural components of the framework were not directly relevant for the use case of this thesis. However, this was to be expected since many of the components are meant to encourage discussions between multiple users. Also, it was obvious that the so called TutorAgent component, which is basically the chatbot functionality, would be the most interesting component for this work.



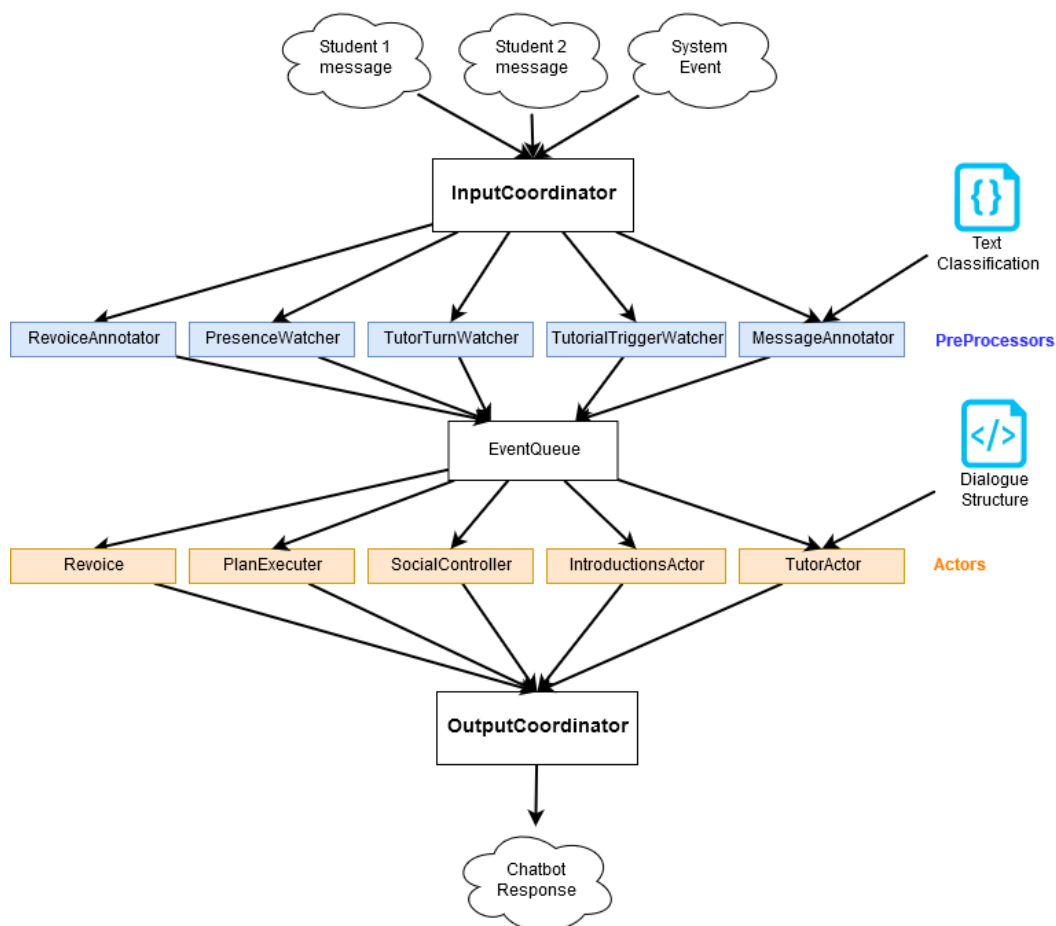


Figure 2: Architecture overview of the Bazaar framework. System events and user messages are handled by the *InputCoordinator*, which gives the received information to the *PreProcessors*. In this figure we can see a few of the *PreProcessors* and *Actors* implemented in the framework. The *MessageAnnotator* is one of the most important *PreProcessors*, since it uses the implemented text classification to label the received messages. These labels are used in the *TutorActor*, which uses them to follow different paths in the dialogue. After the pre-processing step is finished, the event is put into the *EventQueue*. Subsequently all the actors process the event and propose a response. The *OutputCoordinator* then accepts a proposal and sends the response to the user.

Since the TutorAgent component is at the core of the chatbot, it was kept in the framework and it was examined closely to fully grasp how it works. The TutorAgent follows a structured dialogue, which is scripted in an XML file and either starts automatically or is triggered through a keyword of the user. When such a trigger word is used in the discussion, the TutorAgent asks for confirmation to start the dialogue. In the dialogue file different branches can be created by setting conditions through dictionary files. These files are used for pattern matching and the message of the user gets analysed if certain words of the dictionary are used whenever a branching in the script is reached. For pattern matching a typical regex is used. While looking at the code base in combination with testing the chatbot, problems while matching could be identified.

The last part of the preliminary work was to decide for each component whether it should remain in the framework or should be removed. Unfortunately, there was one big bug in the system, which was that occasionally the whole system would get stuck during a conversation. This was one of the main reasons for removing some of the components. The bug could be traced back to a component which was not connected to the chatbot and not necessary for the use case. Therefore, a decision was made to reduce the supporting components for the conversation which were not too interconnected with one another. To achieve this, they had to be decoupled from the rest of the system. Once they were no longer entangled in the rest of the system, it was possible to remove them. Nevertheless, some components seemed to be too interrelated with other core parts of the system, which is why these components were kept in the code base.

#### **4.1.2. Text Classification Methods**

In order to create an adaptive dialogue, the messages of the user need to be analysed so it is possible to react to them in various different ways. One of the methods used for this is called text classification. It is one of the most well-known tasks of natural language processing and has a wide variety of applications. By using classification, a text or message can get assigned to previously defined classes (Suppala and Rao, 2019). One of the best known applications for text classification is sentiment analysis, where a text gets

classified according to its sentimental orientation (Leung and Chan, 2008). The system needs to automatically classify messages if we want to make sure it is adaptive and there exist many different ways to do this. For this thesis the decision had to be made between one rule based and one machine learning approach.

In the machine learning approach, the system uses past observations to build a model to classify new data. To build a model which can differentiate between the different classes, the system needs to get trained with pre-labelled data. To verify that the model works, a part of the data is used to test the created model. For this thesis the text classification algorithm Naive Bayes was used. There are many different algorithms available. Naive Bayes seemed particularly suitable, because it needs relatively little sample data to achieve good results. It is based on the Bayes Theorem and uses the probability of occurrence for each word in a category to compute the likelihood that the message is from said category. When trained properly, such a system can classify the messages sent in this use case very reliably. Through the work of Irmtraud Wolfbauer and her chatbot Rebo Junior about 175 conversations are available for the use case with apprentices. However, to train a Naive Bayes system, at least a thousand data points in the specific domain are necessary to achieve a reliable model. Therefore, using a machine learning approach was not suitable for this study.

The second approach which was considered are rule based systems. They use handcrafted rules in order to classify texts or messages. With the help of these specific rules the system tries to identify relevant elements which are used for classification. One of the typical ways to do this is pattern matching. To create such a system, domain knowledge is needed. Generating rules can be quite time consuming, but on the positive side, they can be improved over time. Techniques in rule based systems that are typically used are for instance stemming, tokenization and dictionaries. The reason to use such a rule based approach for this thesis was mainly based on its simplicity and expandability for the future. The system uses the straight forward lexicon or dictionary technique for matching. In order for this to work, as many dictionaries as required classes had to be created and manually filled with words corresponding to the specific categories. For the technique of stemming, the plan was to use the simple porter stemmer, but it was not

implemented in the scope of this thesis and therefore could be an extension for future work.

## 4.2. Implementation Server

In this subsection the focus lies on the server implementation and its three main components, which are the set-up of the server itself, the maintenance features and the webserver. For the maintenance features the focus is on the automatic database backup and the automatic server restart scripts, whereas for the webserver the focus lies on the website. This includes the created URLs, the different modes for the website and most importantly, the automatic start script for the chatbot.

### 4.2.1. Set-up of the Server

The first work that needed to be done was to set up a server. For this master thesis the Know Center provided a Linux server instance. The server was used as webserver to host the website for the chatrooms. These chatrooms were used for the conversation between the apprentices and Rebo. The developers of the Bazaar framework provided their used server files and a few instructions on how to set up a webserver which can be used with the framework. Basically, the webserver consists of multiple docker containers, of which one is for the website, one for the database and one only used as proxy. After setting up the server, the first step was to set up the docker images to host the chatrooms. Since the docker files contained some problems, they were corrected and subsequently the three images were created. During the practical work of this master thesis the server twice had unexpected reboots on Sundays and because of that the docker images stopped running. To prevent that from happening again, the decision was to configure the docker images so that they always get restarted when an error has occurred or a server restart has been performed.

## 4.2.2. Maintenance Features

Over the next years Rebo will be used in the apprentices' workshops and it is important that the server runs with as little human maintenance cost as possible. To achieve this two automatic scripts were created.

First, to restart the server periodically without human intervention, the automatic restart feature for the server was implemented. It is configured in a way that the server restarts every Saturday at 2am. To achieve this behaviour the pre-installed crontab command from Linux was used. Crontab is a process which can be used to schedule activities. This means it executes commands and scripts at specific dates or times. For this master thesis the restart command is configured as a reoccurring task. The command is located in the crontab file (command `"0 2 * * 6 /sbin/shutdown -r now"`). To achieve the periodic behaviour a crontab command generator website <sup>2</sup> was used. The command makes sure that on every Saturday at 2:00am the simple shutdown command of the server gets executed. The main reason for adding this feature was to get rid of manually restarting the server, which is otherwise necessary because of stuck processes. Stuck processes occurred regularly at the beginning of working on this thesis through a bug in the framework. To tackle this problem the simplest solution was chosen, which was to automatically restart the server periodically to kill stuck processes and free the RAM. Even though this problem could be resolved, the restart feature was kept to clear all remnants of the chatbot processes or other processes which may accumulate.

The second feature that needed to be incorporated is the automatic database backup. In order to achieve the automatic database export from the mysql database running in the docker image again a command in crontab was used. The command this time (`"0 1 * * 6 sh /home/DBBackups/backup.sh"`) was created with the crontab command generator website <sup>2</sup>. The command executes a self-written shell script, which exports the database into an SQL file. The used shell script can be seen in Listing 1 and shows how the database dump is created. As with the command for the automatic restart, this script gets executed every Saturday with the only difference that the database backup occurs one hour before the server restart, i.e. at 1:00am.

---

<sup>2</sup><https://crontab.guru/>

The reason to save the database before restarting the server is to prevent any potential data loss occurring through the restart. Considering that backups of the database have to be made periodically, this feature ensures that this is done and by being done automatically, the time for human maintenance is decreased.

Listing 1: Shell script for exporting the used database, running in the docker image "nodechat" into a file on the server. The created backup file is an sql file which can be loaded to restore the database.

---

```
docker exec nodechat /usr/bin/mysqldump -u root
--password=***** nodechat
> /home/DBBackups/$(date +%Y-%m-%d_%H:%M).sql
```

---

### 4.2.3. Website

In order to be able to use Rebo, a website for the chatrooms was created. The developers of the Bazaar framework provided their files for the website, but in order to have a visually more pleasing experience the visuals had to be changed. In Figure 3 the website before it was changed can be seen and Figure 4 shows the updated version which was created for this thesis. To create the new version, the first step was to get rid of all the elements not needed, which was done by hiding them through an css file. Next, dark colours were replaced with lighter ones, which match with the colours of Rebo. Additionally, the Rebo logo was added next to the conversation, which can be seen in Figure 5. As said before, the reason for changing the layout and visuals of the website was to give the user a better experience when having a conversation with Rebo. By hiding unused elements and only keeping relevant ones, the interface was simplified. With the decision to simply hide the other elements, but not get rid of them, the possibility of re-enabling them in the future was kept. The reason for the change from dark colours to lighter ones was to make the website friendlier and more pleasant to use. Furthermore, to create the feeling of a "normal" conversation like over Whatsapp, the logo of Rebo was added so the apprentice can see with whom the conversation is conducted.

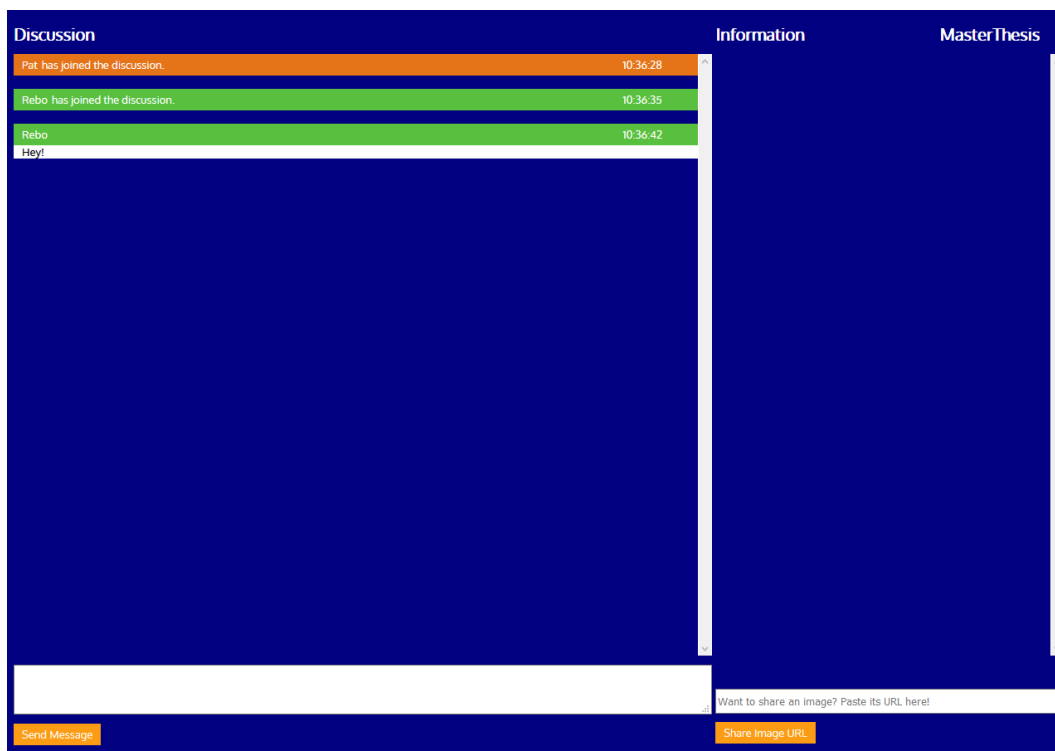


Figure 3: This figure shows the website used by the developers of the Bazaar framework. It uses a dark blue background and has two different chat windows. One window is for the text messages and one for image URLs.

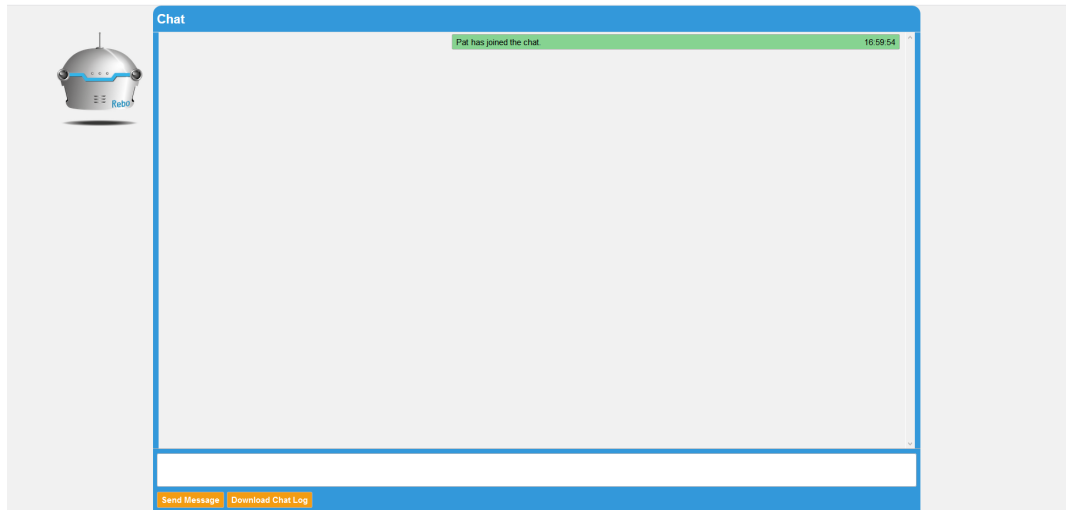


Figure 4: Here the updated website for the apprentices can be seen. In the upper left corner the Rebo Logo is shown, the layout has been redesigned and through the lighter colours the website looks friendlier.



Figure 5: This is the final design for Rebo, the reflective guidance chatbot. It is based on the fields of work of the apprentices, which are electrical engineering, metal and mechatronics. [Image created during the Rebo Junior study of Irmtraud Wolfbauer.]



The next feature that was implemented is the possibility to access the website through three different modes. The first mode is for the chat with Rebo, the second for writing a small essay and the last for a group discussion. For the chatbot and the group mode the website looks and behaves the same, with the only exception that in the group mode the discussion is not led by a chatbot and multiple users have the possibility to chat with one another. In the essay mode the website is slightly modified, so that an essay can easily be written. Therefore, this mode cannot be used for a typical conversation. Figure 6 shows the website in essay mode with one of its modifications, which is a confirmation window for sending the essay. The URL to access the chatrooms was previously constructed as follows: `<domain>/bazaar/chat/<chatroom>/1/<name>/1/`. However, since the parts of the URL with 1 are not used, it was decided to change them. To achieve these modes, the mode parameter in the URL was created which replaced an unused parameter and the last part of the URL was simply removed since only one new parameter was needed. The URL to access the chatrooms is now constructed as follows: `<domain>/bazaar/chat/<chatroom>/<chatbot|essay|group>/<name>/`, the `<chatroom>` placeholder gets automatically replaced with a unique string for each conversation and the placeholder `<chatbot | essay| group>` shows the three modes which can be used. Furthermore, small changes in the client and server files were necessary to achieve the different behaviours and visual representations. The reason for adding different modes was to have more options for future use of the system. Through these changes it is easily possible to either have a chatroom to chat with the chatbot for a group discussion or to write an essay and the differences are simply triggered by using another mode in the URL.

In order to access the data of the conversations more easily, another feature in the form of special URLs was created. The URL [http://rebo.know-center.tugraz.at/bazaar/room\\_status\\_all](http://rebo.know-center.tugraz.at/bazaar/room_status_all) lists all used chatroom names in a small table, the URL [http://rebo.know-center.tugraz.at/bazaar/data/<room\\_name>](http://rebo.know-center.tugraz.at/bazaar/data/<room_name>) automatically exports the data of the desired room as csv file and the URL <http://rebo.know-center.tugraz.at/bazaar/AllData> automatically exports the data of all chatrooms. The exported data can then easily be imported into another program, such as Excel to analyze it. All URLs were implemented by routing the corresponding GET requests

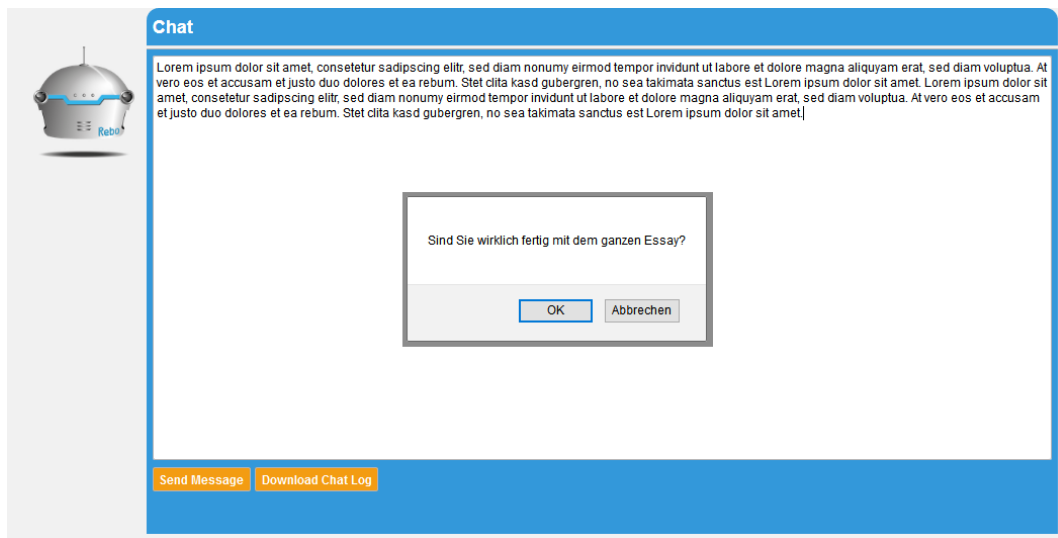


Figure 6: The website running in essay mode. The chat window is hidden while writing and upon sending the message the confirmation window is displayed.

and responding with the respective data. The response can be either a simple table with the room names or a download prompt for the csv file. The URL which shows the room status was changed to support the current database structure and lists all chatrooms. The other two URLs were not supported by the server files provided by the framework developers and had to be implemented additionally. The goal of this was to simplify the process of retrieving the data of the stored conversations. One URL is used if only the data of one specific conversation is desired, whereas a different URL is used to extract all available data into a csv file. These URLs were created to help researchers who want to analyse the collected data in the future. The most important feature implemented on the server is the automatic launch script for the chatbot. Whenever a user connects to a chatroom in chatbot mode, the script gets executed and starts up Rebo. This only happens if it is the first time a user connects to the chatroom. This is checked by the server javascript file on each connection with a simple database query. Whenever this check is true and the used mode is chatbot, a shell script to launch the chatbot is executed. This script can be seen in Listing 2 and uses the room name to which the chatbot has to connect as argument. As can be seen in the Listing 2 the executable jar file for Rebo is called "ReboAgent.jar" and is

located in the same folder on the server as the shell script. Since the Bazaar framework was designed to run with a graphical user interface, it was not possible to run the chatbot on the server without finding a workaround. The workaround used is to make the chatbot believe that the server is a normal desktop system. By pretending that a desktop is available, the jar file with the chatbot can be executed. This addition was a necessity for this master thesis, because it is unreasonable to have an instructor start the chatbot each time a conversation is held. The practice of always starting the chatbot manually was used by the developers of the Bazaar framework, since they did not have the need for an automatic start of the chatbot. With the implemented change, the chatbot launch is automated and therefore the conversation can be held at any time and place. This has the additional benefit that the apprentices have the possibility to do the reflections from home or any other place, in case they were not able to do so in the workshop or have the urge to reflect over something again or additionally.

Listing 2: Shell script for automatically launching the chatbot. The room name is given as argument to the script and the jar file of the chatbot has to be located in the same folder as the script. By pretending that a display is available the jar field can be executed.

---

```
#!/bin/bash
cd "$(dirname "$0")"

roomname=$1;

jar_name="ReboAgent.jar"

echo $jar_name
echo $roomname

DISPLAY=:19 java -jar $jar_name --room=
    $roomname --launch
```

---

### 4.3. Implementation Chatbot

The main part of the chatbot is the dialogue structure, which the chatbot follows during the conversation. As described before it is crucial to have a well thought out dialogue structure to achieve the desired learning effects. As starting point the non-adaptive structure of Rebo Junior was chosen and developed through an iterative process by Irmtraud Wolfbauer. In Figure 7 this non-adaptive reflection script of Rebo Junior is shown. In this script the questions asked are always the same and are not dependent on the answers given by the user. Figure 8 shows a flow diagram of the final dialogue structure. Depending on the answers of the user, the chatbot follows different paths in the dialogue structure. Furthermore, for every step in the diagram, the chatbot selects a question from a pool of phrases, which makes sure identical questions are phrased in different ways. In Table 1 some sample phrases for each step are shown. For each reflection level used with Rebo the corresponding step in the flow diagram can be seen. Furthermore, for each step of the dialogue up to three messages of the message pool are shown. After each iteration, the dialogue structure was implemented in an XML file. Since the goal of this master thesis was to implement an adaptive chatbot, it was necessary to create a dialogue structure with different paths. Furthermore, to achieve the desired reflection, well thought-out questions had to be constructed. Additionally, to prevent always getting the same questions, a question pool had to be built for each step in the flow diagram.

As explained before, the branching in the dialogue is achieved through text classification. As for dictionaries, it was decided to use polart dictionary (Klenner, Fahrni, and Petrakis, 2009) by Klenner, Fahrni, and Petrakis as a starting point. This dictionary is a German dictionary with over 9,000 words and their respective sentiments. To make it usable with the Bazaar framework, a small python program was written to pre-process the dictionary. This program can be seen in Appendix B. After pre-processing, the dictionary had to be scanned manually for unnecessary words in this use case. Additionally, the data from the Rebo Junior conversations was analyzed and the dictionary was expanded with more domain specific words.

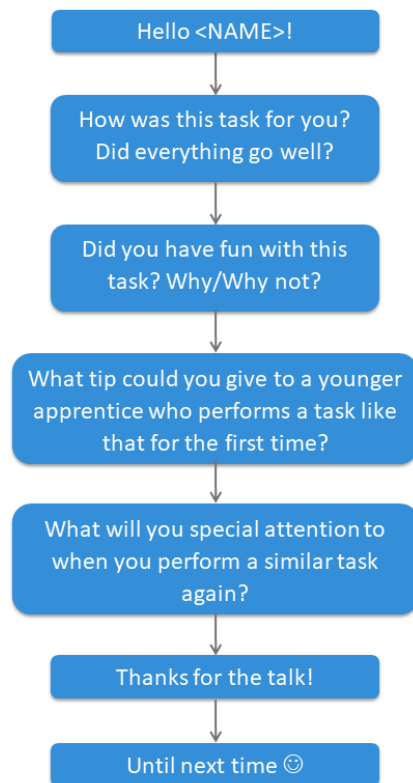
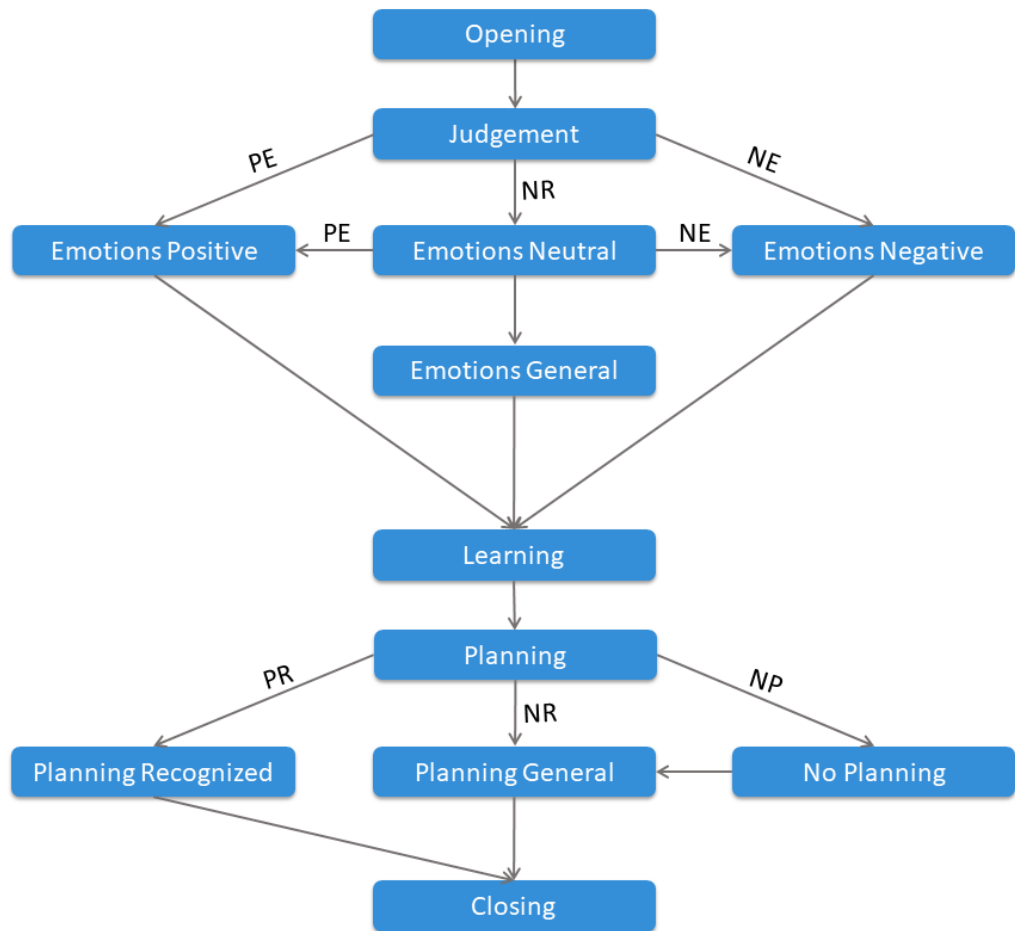


Figure 7: Non-adaptive reflection script used by Rebo junior. Every time a user reflects on a task, he or she is confronted with the same questions.



PE ... Positive Emotion  
 NE ... Negative Emotion

PR ... Planning Recognized  
 NP ... No Planning Recognized

NR ... Not Recognized

Figure 8: Reflective dialogue used by Rebo the reflective guidance chatbot. Depending on the answers of the user the chatbot follows different paths during the conversation. For this thesis the chatbot has the possibility to react to positive/negative emotions twice and once the possibility to recognize if a participant makes plans for their future behavior during the conversation.

Reflection Level	Flow Diagram Ref.	Example Messages Rebo
Opening	Opening	Hey [USERNAME]! Hallo [USERNAME] Grüß dich [USERNAME]
Judgement	Judgement	Wie ist es dir bei dieser Aufgabe gegangen? Alles gut gelaufen? Bist du mit dem Ergebnis der Aufgabe zufrieden? Warum/Warum nicht? Wie gut erfüllt dein Ergebnis die gestellte Aufgabe?
Emotions	Emotions Positive	Toll zu hören! — Was macht diese Aufgabe für dich interessant? Was hat dir dabei am besten gefallen? Worauf bist du bei der Lösung der Aufgabe besonders stolz?
	Emotions Negative	Oje! — Was war denn das Problem? Wobei hattest du die meisten Schwierigkeiten? Was hat dir nicht so viel Spaß bei der Aufgabe gemacht? Warum?
	Emotions Neutral	Was gefällt dir an deinem Ergebnis und was nicht? Wie gut erfüllt dein Ergebnis die gestellte Aufgabe?
	Emotions General	Wie hast du dich während der Durchführung der Aufgaben gefühlt? Hat dich die Arbeit an der Aufgabe gefesselt? Hattest du Spaß bei der Aufgabe? Warum oder warum nicht?
Learning	Learning	Was hast du für dich aus der Aufgabe gelernt? Was hat dich bei der Aufgabe besonders gefordert? Wie gut hast du die Aufgabe verstanden?
Planning	Planning	Worauf wirst du beim nächsten Mal besonders achten, wenn du einen ähnlichen Auftrag bekommst? Wo kannst du das Erlernte noch einsetzen? Was wirst du das nächste Mal besser machen?
	Planning Recognized	Das klingt nach einem Plan! Großartig! Klingt gut!
	No Planning	War nicht vielleicht doch etwas Interessantes für dich dabei?
	Planning General	Was hast du für die Zukunft gelernt? Wenn du die gleiche Aufgabe morgen nochmals machen würdest, was wäre dann anders als heute?
Closing	Closing	Danke, dass du das mit mir besprochen hast! Danke! Das war's für heute :D Danke das war's für heute, viel Spaß bei deiner nächsten Aufgabe!

Table 1: Message pool of Rebo for each step in the dialogue.

An improvement compared to the Bazaar framework is the possibility to have different introductions, which may be used in combination with the username. Before the chatbot starts the conversation an introduction phrase is sent, which the user needs to respond to. In this thesis this introduction step was incorporated into the dialogue structure through the opening step. In this opening step the user is greeted and then the chatbot expects a greeting in return. After receiving a greeting in response, Rebo starts to follow the dialogue structure. The framework was adapted to randomly pick one of the provided phrases for the introduction. Additionally, the possibility to use the username of the apprentice in the introductory step was implemented. These changes make it possible to use a wider variety of phrases for the introductory step, which should counteract the feeling of the greeting being monotonous. Furthermore, by using the username in these introduction phrases, a more personal greeting is created.

The dialogue structure used consists of multiple different paths, which the chatbot can follow through analysing the messages sent by the user. Whenever there is a branching in the structure, Rebo uses pattern matching to analyse the message and checks if the condition for the individual paths are met. To get rid of the problem where partial words were matched, the regex was changed in a way so that only whole words are matched. The chatbot is used with German words and since in German, words can be easily put together to build a new word, this change was quite important. This small regex change prevents false matches when checking the conditions for the individual paths of the dialogue.

Another extension for the chatbot is the possibility to answer Rebo through multiple messages. In a conversation Rebo originally responded after the first message of the user. In order to allow multiple responses by the apprentices, a trigger was built in to wait for further messages if necessary. To signal Rebo that the current message consists of multiple messages, three dots at the end of the message have to be added. These three dots are the trigger for Rebo to wait for another message before continuing the conversation. This trigger can be used repeatedly in different messages to chain an unlimited amount of messages if necessary. The messages are then put together and are treated as one message when analysing it. This change was made so the conversation seems more like that of a typical chat program such as WhatsApp or Telegram, where people can respond with



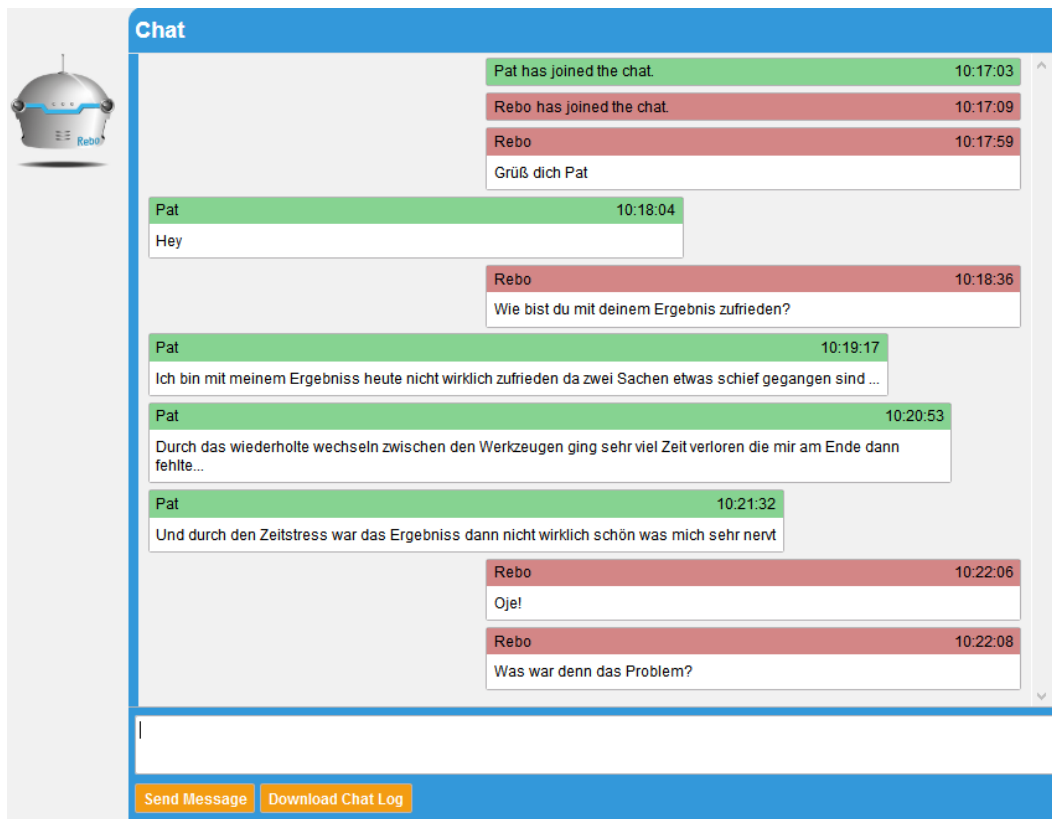


Figure 9: Example conversation taken from a testing conversation which utilizes the multi message response trigger, in our case the three dots at the end of the message.

multiple messages to a question. An example of a conversation with the use of the multi message trigger which was taken from a testing conversation can be seen in Figure 9.

The next addition to the system was a reminder message. This message is sent by the chatbot to remind the user to take the conversation seriously. Since the goal of the dialogue is to lead the user to reflect on a task, it is expected that the user will answer with at least one sentence to each question asked by Rebo. For this purpose the chatbot calculates the average used words per answer and if the amount is lower than a threshold, the chatbot sends a reminder message before asking the next question. A special message, which is excluded from this calculation is the greeting message,

as it is typically quite short. Rebo waits for at least 3 answered questions before calculating the average amount of words and potentially sending the reminder. This is supposed to decrease the possibility to send the reminder, even though the user has answered the questions sincerely. This extension of the chatbot was implemented after analysing the Rebo Junior data. When analysing said data, it could be seen that some users tricked the system, which means that users skipped through the whole conversation by spamming "ok" or similar words. The goal of the reflection with Rebo is to achieve a learning effect, which cannot be achieved if users behave in such a way. Therefore, the reminder extension for the chatbot was implemented, which triggers whenever such behaviour is suspected.

## 5. Evaluation Method

In this section we cover the evaluation method used in this thesis, which should help us answer our research questions. As mentioned before, these are to study the technical feasibility to create a conversational agent to lead reflective dialogues by using Bazaar, to study the acceptance of Rebo by apprentices and to gather feedback as a step towards a full field study. The last two research questions should be answered through a field study in a workshop with apprentices.

### 5.1. Preliminary Work

This section focuses on the evaluation of the chatbot and on how it was conducted. The basic plan for evaluating the artefact of this master thesis was to conduct a workshop with the apprentices. Before specifically planning how to do this, Irmtraud Wolfbauer was accompanied at two of her workshops for Rebo Junior. The idea behind this was to get to know some of the apprentices, get in contact with the instructors in the workshop and most importantly, get an understanding of how such a workshop is set up.

In order to understand why Rebo was used in the way it was, it is important to know that there are plans to embed Rebo into the moodle platform of the apprentices in the future. This was not yet done for the planned workshop and therefore the unique URLs to access the required chatrooms were handed out to the apprentices. To ensure data privacy, the chatroom names were a concatenation of apprenticeship year and a random three digit number. Additionally, the apprentices were told that they could use any username they wanted. This step should also prevent the tracking back of chat messages to a specific user through his or her name. Once everyone finished reflecting, it was decided to gather some feedback from the apprentices. In the Rebo Junior workshops two flash light rounds were conducted and since these worked quite well, it was decided to also conduct two flash light rounds in this study. The first question of the flash light round was about the apprentices' thoughts on the chatbot itself. This should provide conclusions on whether the apprentices generally liked the interaction with

Rebo or not. In the second question, the apprentices were asked whether they had gained anything from the interaction with Rebo or if they might benefit from it in the future.

## 5.2. Study Procedure

In order to conduct the study with the apprentices the physical presence of the author of this thesis was mandatory. The reasons for this were that in case of problems solutions can be worked out immediately, feedback from the apprentices can be collected and most importantly, before the start of the study, the project can be presented to them together with an explanation why reflections using Rebo could be useful to them. Figure 10 shows the flow diagram of the study procedure with additional information on who was responsible for each step. Irmtraud Wolfbauer mainly accompanied the author of this thesis to help with the flash light rounds to collect the feedback of the apprentices. Of course the two workshop instructors were also present to hand out the workshop task to the apprentices. The workshop was conducted like any other workshop the apprentices had frequented before with the exceptions that the Rebo study was explained to the apprentices at the beginning of their lesson and that after fulfilling their workshop task, the apprentices were asked to take part in the study. The workshop was done in the "Binder Lernwerk"<sup>3</sup>, which is located in Voitsberg, Austria.

Next, the structure of how the study was carried out with the apprentices needs to be explained. First, the apprentices got general information about the chatbot, its goals and the study procedure. It was clarified that Rebo is a computer program, which has the purpose to guide them through a reflection process related to a previously fulfilled task. The website is designed to be accessed with a laptop, since the apprentices who will work with Rebo in the future all get laptops from their corresponding companies. The apprentices in this study were not equipped with laptops and therefore were informed that they could alternatively use their phones for reflections with Rebo. They were also told that the visuals on the website may look

---

<sup>3</sup><http://www.binder-lernwerk.at/>

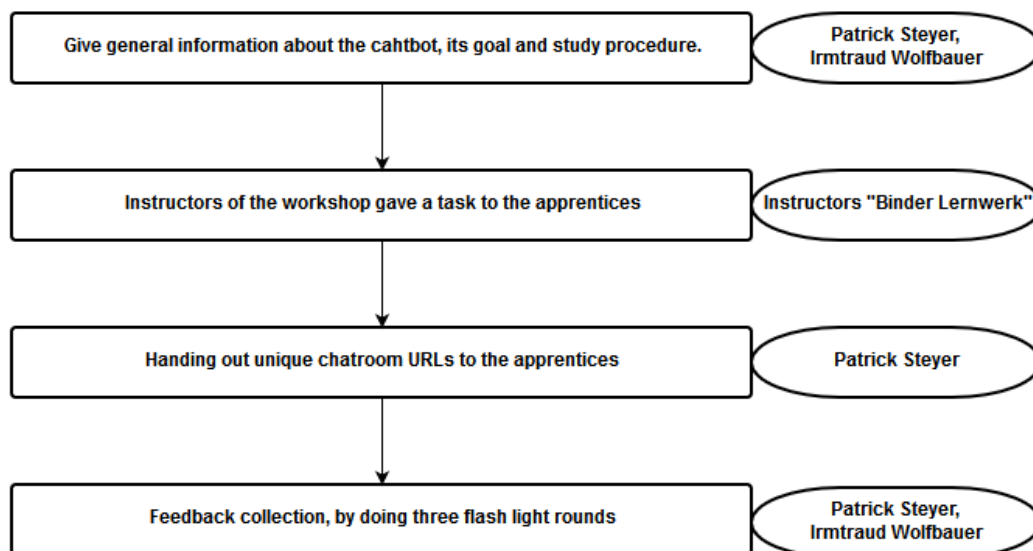


Figure 10: Flow diagram of the study procedure for the conducted workshop. The right side shows who was involved in which step.

strange on some mobile phones and were asked not to be put off by this. Furthermore, they were asked that if any problems or questions arose they should not hesitate to ask, so they could get help immediately.

After this first introduction to Rebo, the apprentices got a task from the instructors of the workshop. The instructors are not related to the study and are employed by the "Binder Lernwerk". The instructors are responsible to educate the apprentices in the workshops. Before the field study was conducted it was decided, by the author of this thesis, that the apprentices should get a small task from their instructor, whose workload should cover roughly an hour. Unfortunately, it was not clear before carrying out the study that the apprentices in the fourth year mostly use the workshop for studying, in contrast to usual workshops which are used for trying out practical tasks. This fact was not known to the author of this thesis or Irmtraud Wolfbauer. The workshops connected to the Rebo Junior study which we had attended before had suggested otherwise. The apprentices had the task to individually study an hour for their upcoming final exam (Lehrabschlussprüfung LAP). They were asked by their instructor to pay

special attention to the chapter they feel least comfortable with and they had to prepare a few theoretical questions they might be asked on that topic in their final exam and to prepare the corresponding answers.

Once all the apprentices had finished the task, each apprentice got a unique URL to a chatroom. As mentioned before, to ensure data privacy the URLs were constructed by incorporating a random number. Additionally, the apprentices were told that they could insert any username they wanted for this study, which ensures that no information used in this thesis can be tracked back to a specific user. In order not to put them under pressure, they were told to take their time to think and reflect about the task properly. They also had the possibility to ask questions, if any arose. All apprentices were able to connect to the chatrooms without problems and Rebo started the conversation with them. Rebo guided them through the reflection process by asking questions according to the messages of the apprentices. During the conversations there was not a single apprentice who ran into a problem.

Finally, feedback from the participants was collected in form of three flash light rounds. Initially, only two flash light rounds were planned, but based on the wishes of the participants of the study, a third flash light round was added. In each of these rounds every apprentice had to give a short oral answer on the asked question. This was done together with the rest of the group in a plenary setting. The questions were asked by Irmtraud Wolfbauer. The first question asked was "How did you like the interaction with Rebo and what are your thoughts about it?" and the second "Do you think you have benefitted or could benefit from the interaction with the chatbot in the future?". In response to a question during the second round Irmtraud Wolfbauer mentioned that Rebo Junior was only used for practical tasks, which initiated positive responses from the apprentices. Therefore, the third question arose, whether they think it would be more beneficial to use Rebo after a theoretical or a practical task. The apprentices were eager to discuss this question, which is why this question was added to the flash light rounds and is referred to in this thesis as the third flash light round. The complete transcript of the flash light rounds can be found in [Appendix A](#).

### 5.3. Participants

The implemented system is specifically designed to be used in workshops for apprenticeship training. Therefore, it seemed sensible to conduct the field study in one of these workshops. The apprentices, in the workshop were in the fourth year of their apprenticeship and the field study was carried out shortly before their final apprenticeship examination. Furthermore, it is important to mention that none of the participants of this study had participated in the Rebo Junior study. At the time when the study was carried out ten apprentices were in the fourth year of the apprenticeship training and seven of them were present in the workshop on the day of the study. All of the participants were male, which was probably due to the fact that the rate of female apprentices in the fields of electrical engineering, metal and mechatronics is quite low.





## 6. Results

In this section the results of the study are presented. The conversations of the workshop are analysed and the feedback of the apprentices is summarised.

### 6.1. Coherence of Conversations

Firstly, the focus lies on the coherence of the conversations. To achieve a learning effect through reflection, it is indispensable to have a coherent conversation between the apprentice and Rebo. Hence, when analysing the data, in a first step each conversation was either assigned a "yes" (coherent) or a "no" (not coherent). Table 2 shows the conversations labelled in terms of coherence. It can be seen that two conversations were not completely coherent, which are roughly 28,6% of the conversations. The other five conversations with Rebo were coherent, which are about 71,4%. Of these five conversations three are coherent and in the other two, one response does not directly answer the question asked.

Coherence of the Conversations		
yes	71,4%	(5)
no	28,6%	(2)

Table 2: Interaction statistics between the apprentices and Rebo with the focus on coherence.

### 6.2. Reflectiveness of Conversations

It is the goal of Rebo to lead the user through a reflection process. Therefore, in a second step each conversation was analysed in terms of achieving reflection and was then either labelled "yes" (reflection recognizable) or "no" (no reflection recognizable). Table 3 shows the statistics of the labelled conversations. In two of the seven conversations, which are about 28,6% of the conversations, reflection could be recognized. On the other hand 71,4% were labelled with "no", which means that no significant amount of reflection could be identified.

Reflection in Conversation Recognizable		
yes	28,6%	(2)
no	71,4%	(5)

Table 3: Conversation statistics with the focus on whether reflection can be recognized.

### 6.3. Feedback

As a third step, the feedback from the flashlight round was analysed. The first flash light round focused on whether the apprentices liked the interaction with the chatbot and on what their thoughts about Rebo were. When analysing the feedback of this round, the feedback from each apprentice was assigned to one of three classes. Either it was positive, negative or neutral feedback. Table 4 shows the feedback results from round one with respect to each question. With a share of 57,1%, most of the feedback in this flash light round was neutrals. The feedback from two out of seven apprentices was rather negative, which accounted for 28,5% of all feedback. These apprentices found that interacting with Rebo "felt strange and they could not see how they should benefit from it"<sup>4</sup>. Furthermore, one apprentice gave a very positive feedback, accounting for roughly 14,3% of all feedback. This apprentice found that "interacting with Rebo was cool"<sup>5</sup> and that "it can be a great help for future apprentices"<sup>6</sup>.

In the second question, the apprentices were asked whether they felt they had benefitted from the interaction with the chatbot or if they think that they could benefit from such an interaction in the future. As for the first flash light round, the feedback from each apprentice was assigned to one of the three classes positive, negative or neutral. The class *neutral*, as well as the class *negative* each have a share of 44,9% of the feedback. For each of those two classes three of the seven instances of feedback were assigned to them. Several apprentices mentioned that they do not think that they personally can benefit from Rebo<sup>7</sup>. The class *positive* was only assigned to the feedback

<sup>4</sup>Verbatim quote: "Komisch eindeutig... ich was net in was für einer Hinsicht wie weit des helfen soll"

<sup>5</sup>Verbatim quote: "Die Interaktion mit dem war ganz cool."

<sup>6</sup>Verbatim quote: "wird sicher a große Hilfe für andere Lehrlinge"

<sup>7</sup>Verbatim quote: "Mir persönlich bringt der sicher nix"

Interactions with Rebo		
Total number: 7		
Did you like the interaction with Rebo?		
positive:	14,3%	(1)
negative:	28,6%	(2)
neutral:	57,1%	(4)
Do you think you could benefit from the interaction with the chatbot?		
positive:	14,3%	(1)
negative:	42,9%	(3)
neutral:	42,9%	(3)
Do you think it would help more to reflect upon a practical task?		
yes	100%	(7)
no	0%	(0)

Table 4: Feedback collected during the three flash light rounds with the apprentices.

of one apprentice, who mentioned that they feel someone could benefit from Rebo if they were motivated<sup>8</sup>.

As stated before, the third flash light round arose spontaneously during the second round and deals with the question whether the apprentices believe using Rebo for reflection after fulfilling a practical task might be more useful than for reflecting on a theoretical task. Again all feedback was assigned to a class, but unlike in the previous rounds, feedback from this round was assigned to one of two classes, namely "yes" or "no". In this question all seven apprentices gave feedback which was assigned to the class "yes". This means that 100% of the participants believe that using a chatbot for reflecting makes more sense for practical tasks than for theoretical ones. One apprentice stated that he thinks they could benefit from it, because they would reflect on the whole process of fulfilling the task again<sup>9</sup>.

<sup>8</sup>Verbatim quote: "Ja wenn du motiviert darauf zugehst"

<sup>9</sup>Verbatim quote: "I denk scho, dass es was bringen würd, weil ma zum Nachdenken angeregt wird"

## 6.4. General Observations

First, it is important to say that it appears that two apprentices out of the seven apprentices that took part in the study did not take the conversation seriously. One of them started to spam the conversation using just one word after the second question. This conversation can be seen in Figure 11. The other apprentice did not even try to have a normal conversation and insulted the chatbot almost throughout the whole conversation. This can be seen in Figure 12. These conversations are a perfect example of what a negative outcome regarding reflection and coherence looks like. Figure 13, on the other hand, shows a conversation where another apprentice had a coherent conversation during the workshop and in Figure 14, a conversation with a low level of reflection can be seen.

The implemented reminder message to take the conversation seriously was sent in three of the seven conversations. Two of these messages were sent to the apprentices who did not take the conversation seriously. A final interesting observation is that the newly implemented multi message response was not used a single time in this workshop.

**Chat**

**Rebo** 10:13:02  
Grüß dich Rebo

**SCHUKO** 10:13:23  
Hallo

**Rebo** 10:13:24  
Welches Wissen konntest du heute gut anwenden?

**SCHUKO** 10:14:12  
Über Hartmetalle

**Rebo** 10:14:16  
Was gefällt dir an deinem Ergebnis und was nicht?

**SCHUKO** 10:15:28  
das man alle sachen verbinden kann und nicht das man hohe presskraft braucht

**Rebo** 10:15:34  
Was hat dich an der Aufgabe interessiert?

**SCHUKO** 10:16:39  
nix

**Rebo** 10:16:44  
Hat dich die Arbeit an der Aufgabe gefesselt?

**SCHUKO** 10:16:50  
na

**Rebo** 10:16:56  
Was war interessant an der Aufgabe?

**SCHUKO** 10:17:10  
nix

**Rebo** 10:17:16  
Heute wieder einmal kurz angebunden.

**Rebo** 10:17:16  
Was würdest du beim nächsten Mal wieder gleich machen? Warum?

**SCHUKO** 10:17:31  
nix unötig

**Rebo** 10:17:36  
Wenn du die gleiche Aufgabe morgen nochmals machen würdest, was wäre dann anders als heute?

**SCHUKO** 10:17:42  
nix

**Rebo** 10:17:46  
Danke für dein Feedback.

**SCHUKO** 10:17:52  
bitte

Send Message Download Chat Log

Figure 11: Example conversation where the apprentice was not able to achieve the targeted reflection and coherence.

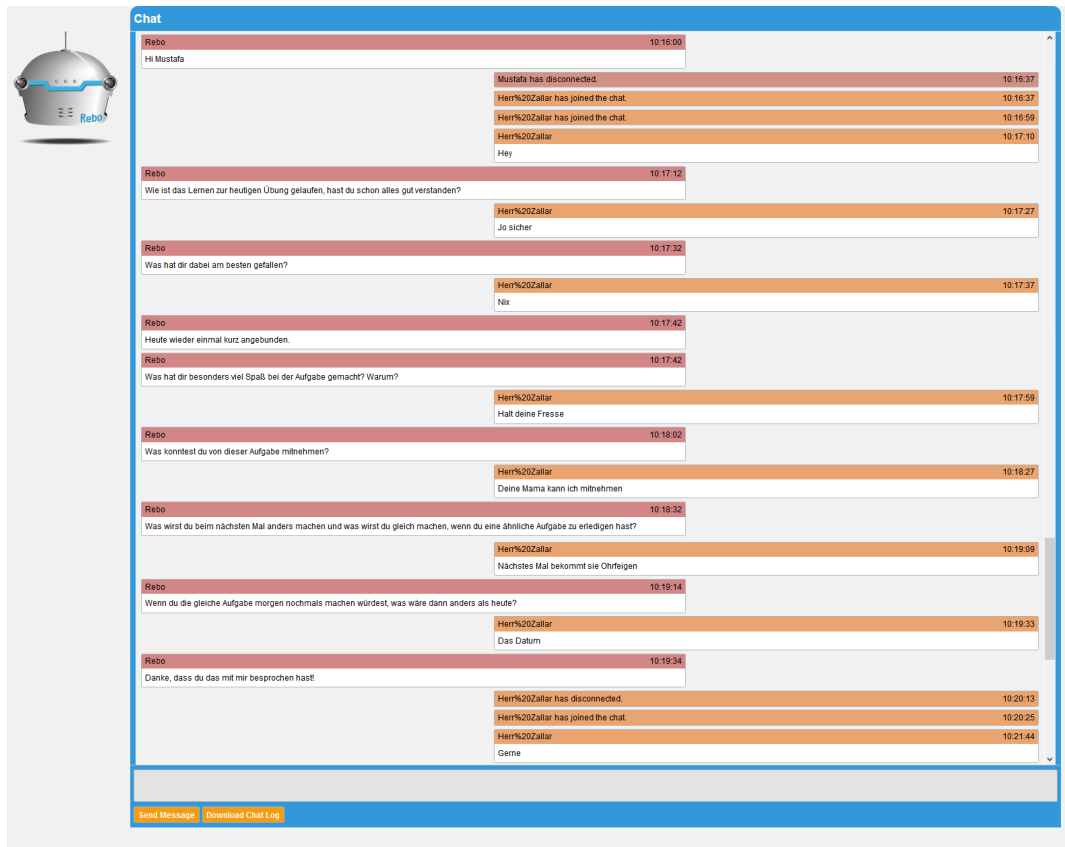


Figure 12: Example conversation where the apprentice insulted the chatbot almost through the whole conversation.

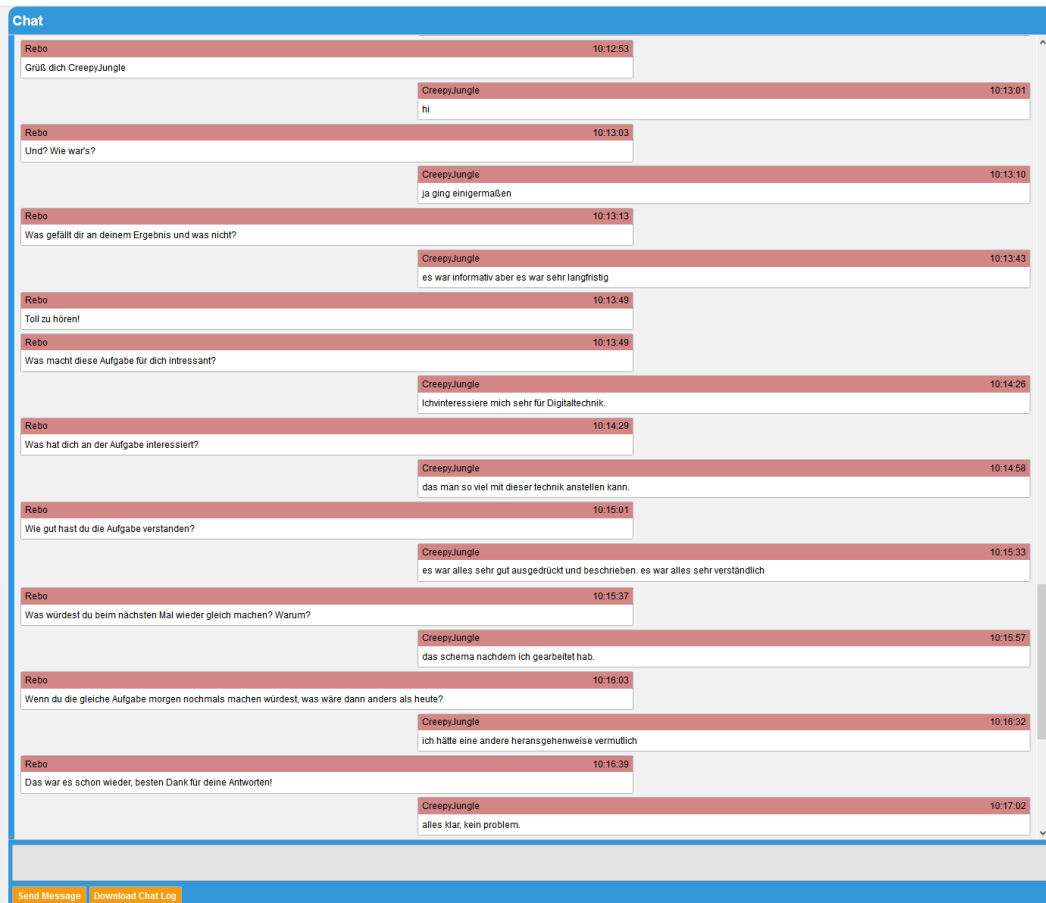


Figure 13: Example conversation where the apprentice was able to have a coherent conversation with Rebo.

**Chat**

**Rebo** 10:14:14  
Hallo Long-dong-silver

**Long-dong-silver** 10:14:38  
Hallo :)

**Rebo** 10:14:38  
Welches Wissen konntest du heute gut anwenden?

**Long-dong-silver** 10:16:56  
Ich weiß viel über Fehlerstromschutzschalter

**Rebo** 10:17:02  
Was gefällt dir an deinem Ergebnis und was nicht?

**Long-dong-silver** 10:20:45  
Mir gefällt es, dass ich viel über den Aufbau und die Funktionen von einem Fehlerstromschutzschalter weiß.

**Rebo** 10:20:46  
Was hat dich an der Aufgabe interessiert?

**Long-dong-silver** 10:21:53  
Alles, was mit diesem Thema zu tun hat, abgesehen von den Berechnungen. :)

**Rebo** 10:21:58  
Was hat dir nicht so viel Spaß bei der Aufgabe gemacht? Warum?

**Long-dong-silver** 10:22:46  
Die Berechnungen, weil dass in der Berufsschule nicht gut erklärt wurde.

**Rebo** 10:22:50  
Was hast du für dich aus der Aufgabe gelernt?

**Long-dong-silver** 10:23:34  
Die Funktion eines FI - Schutzschalters

**Rebo** 10:23:36  
Wo kannst du das Erlernete noch einsetzen?

**Long-dong-silver** 10:23:57  
Privat, und auf meinem Arbeitsplatz

**Rebo** 10:24:02  
Wenn du die gleiche Aufgabe morgen nochmals machen würdest, was wäre dann anders als heute?

**Long-dong-silver** 10:24:32  
Es wäre ein Tag später :)

**Rebo** 10:24:38  
Hast du noch Fragen oder Anregungen an mich?

**Long-dong-silver** 10:25:08  
Im Moment nicht, aber falls ich welche habe, melde ich mich :)

**Rebo** 10:25:14  
Bis bald!

**Rebo** has disconnected. 10:25:15

**Long-dong-silver** 10:25:29  
Auf wiedersehen

[Send Message](#) [Download Chat Log](#)

Figure 14: Example conversation where the apprentice was able to achieve reflection on at least a low level.



## 7. Discussion

In this chapter of the master thesis the results presented in the previous chapter are discussed. Furthermore, ideas for future research are provided.

### 7.1. Workshop Results

This section covers the discussion of the workshop results. On the one hand the focus lies on the results of the conversations with Rebo and on the other hand, the results of the feedback rounds with the apprentices are discussed.

The use of a dialogue structure with different paths and multiple different phrases, led us to two assumptions at the start of the study. The assumption that the use of an adaptive system ensures that the user does not get bored that easily in the long term use cannot be confirmed or disproved in this study, since each apprentice interacted with Rebo only once. Obviously, more data would be needed to see how apprentices behave in further conversations with Rebo. As for the second assumption, namely that conversations remain coherent after switching from a static dialogue to an adaptive system, it can be said that this assumption can be confirmed, as the two non-coherent conversations were from the apprentices who tricked the system and all remaining conversations were coherent.

In the first feedback round we wanted to find out if the apprentices liked the interaction with the chatbot and what their thoughts about Rebo were. In the Rebo Junior study the feedback on this question was throughout very positive (Wolfbauer, Pammer-Schindler, and Rose, 2020) and therefore it was assumed that it would be the same for the more intelligent version Rebo. Surprisingly, the results of the feedback in this study were mostly neutral or negative. By analysing the individual statements of the apprentices, it can be seen that many of the apprentices stated that Rebo did not help them with studying and that it would help more, if it would ask specific questions on the topic they had been studying. It seems that the apprentices had the misconception that Rebo should help them directly with studying. This probably occurred due to the fact that they got a theoretical task in the

workshop, which involved studying. If they would have got a practical task, as the apprentices in the Rebo Junior study, this misconception probably would not have occurred.

Finally, only two apprentices gave a completely negative feedback and these two stated two things. First, that it felt strange to talk to a computer and second, that sometimes very similar questions were presented. Of course it can feel strange to have a conversation with a computer, but such conversations are held daily for instance with chatbots in customer support. In a few years, when everyone is accustomed to chatbots, it will be part of our daily lives and probably feel less strange to people, if not completely normal. As for the second point that was criticized, it can be said that yes it is true that there exist similar questions in the different steps of the dialogue. However, these similar questions focus on different parts of the reflection levels. It seems that some of these questions are too similar to one another and therefore the apprentices felt like they had to answer the same question twice. To prevent this, the question pools need to be reworked in a way, so the phrases in two consecutive questions are not too similar to one another. Overall, it can be said that most of the apprentices had no negative attitude towards the chatbot and think that Rebo will help other apprentices in the future.

In the second flash light round the apprentices were asked if they gained anything from the interaction with the chatbot or if they think they could benefit from it in the future. Since it was assumed that the conversations would show a high degree of reflection, it was also assumed that the apprentices would notice they could benefit from them. It was unanticipated that only one apprentice stated that reflecting with a chatbot could be of value to them. All the other statements were either negative or neutral, and even some of the neutral comments were on the verge of being classified as negative. The negative feedback was mostly that the apprentices did not think they could benefit from interactions with the chatbot. The reasoning for that was often that the questions were superficial and not tailored to the topic they had been asked to study in the workshop. As mentioned above, it seems that the apprentices did not grasp the fact, that the chatbot is supposed help them reflect upon the task done and not help them study. From this it follows that this should be explained in more depth with the apprentices in the future, which should help them to better differentiate

between those two aspects. There is also the possibility that a more in depth explanation is not needed if they only have to reflect on practical tasks in the future. Therefore it can be said that the majority of the apprentices did not feel that reflecting with Rebo could be beneficial to them, which disproves the hypothesis that was set at the beginning of the study .

In the last round of gathering feedback we wanted to find out if the apprentices think that reflecting with Rebo is more suited for practical than theoretical tasks. This question arose during the second flash light round. Based on the feedback of the apprentices, the assumption was that reflecting with Rebo is not suited for theoretical task such as the one the apprentices had to fulfill in the workshop. The results for this question were unanimous. All the apprentices stated that they think it makes more sense to reflect on a practical task than on a theoretical one. This means the assumption, that reflecting with Rebo is not suited for theoretical task, could be confirmed with the feedback gathered and suggests that Rebo should be used for reflections only after practical tasks. Hence, in the future more communication between us and the instructors of the workshop has to take place. Especially for studies it should be clear to all parties how the workshop on that day should look like and whether reflection with Rebo makes sense on the planned tasks of the day.

Since Rebo adapts the conversation to the statements given by the user and guides the user through the reflection process, it was assumed that the conversations would show a high degree of reflection. This should happen automatically if the user answers the questions sincerely. Interestingly, only in two of the seven conversations reflection seems to have taken place. Four apprentices did not engage in reflection, even though their conversations with Rebo were coherent. In this study the apprentices had a theoretical task to do before using Rebo for reflection and it seems that reflecting upon a theoretical task was not that easy or even useful for them. As described above, all the apprentices said they felt that using Rebo would make more sense after fulfilling practical tasks. Since the conditions in this study were not the same as in Rebo Junior and since we do not know what the outcome would have been if the apprentices had been asked to reflect on a practical task rather than a theoretical task, it is not clear whether the assumption that an adaptive system helps achieve a higher degree of reflection was correct or incorrect.

## 7.2. Technical Assumptions

In this section the information regarding the technical assumptions is discussed.

By implementing the automatic restart script and the automatic database backup, it was assumed that human maintenance would rarely be required. At the time this thesis was written, the chatbot had already been running for a month without any human interference. The script restarted the server every week and additionally produced the required database backup files. At the beginning of the study regular restarts were required to keep the website and chatbot running. These two changes proved quite handy, since with these two features these things did not need to be carried out manually anymore. This shows that the required human maintenance was reduced, which confirms our assumption.

The goal of changing the visuals of the website was to have a pleasing experience when interacting with Rebo. This should be especially valuable for long term interaction with Rebo. During the workshop only one interaction per apprentice was done, therefore no long term interaction has taken place so far. Through the feedback it can only be said that the focus of the apprentices was not on the visuals and more on the conversation itself. In our opinion this means that the website visuals were designed in a way that it did not affect the users negatively during their interactions with Rebo. If the visuals were unpleasant, the apprentices would probably have stated that during the first feedback round. For the time being the assumption cannot be completely confirmed and therefore has to be checked after a more extensive field study.

The three different modes implemented have different behaviours for the website and different visual representations. Their objective is to add more options for future studies. It is assumed that the different modes will be utilized in future studies to measure how the reflective behaviour of apprentices has changed after interacting with Rebo for a certain amount of time. In the workshop of this study only the default mode with the chatbot was used and therefore no results regarding the usefulness of the other modes could be gathered. Irmtraud Wolfbauer is already planning to use the essay mode in future studies. Therefore, this assumption can only be

discussed further at the end of the next study when different modes of the website have successfully been integrated in the workshops.

To help researchers who want to analyse the collected data in the future, additional URLs for specific actions were implemented. With these URLs everyone can easily access the data without restrictions, which means that no developer has to extract the data manually if someone needs it. It is assumed that researchers will use these links to export the collected data themselves. In the documentation of the server, which can be found at [https://github.com/Tot333/WebServer\\_Bazaar](https://github.com/Tot333/WebServer_Bazaar), the available URLs are listed and explained in more detail. This should help to get these URLs known to whoever wants to access the conversation data. Since only a single workshop has been conducted so far not much data is available to this point of time. Other researchers will be more interested in the data after a bigger field study has been conducted and more data is available. The assumption that the URLs will be used to download the data may be confirmed in the future after more data has been acquired and more researchers get interested in the collected data.

The implemented automatic launch script for the chatbot, which was integrated in the server, was one of the most important additions for the Bazaar framework. By creating this script, it was assumed that Rebo can be automatically started at any point of time to reflect on something. This feature was also interesting for the head developer of the Bazaar framework, who stated that he would like to integrate it into their system. For the sake of contributing back to the developers of the Bazaar framework a merge request was created to integrate the changes into their repository. The website has been continuously reachable for a month and Rebo automatically starts up when connecting to a new chatroom. This means anyone who wants to use Rebo can interact with the chatbot at any point of time. This confirms the set assumption, even though so far only people connected to the developer team have used it. This is due to the fact that Rebo has only been used in one workshop and only a few people have interacted with Rebo. The stability and reachability of the system will be more important in the bigger field study by Irmtraud Wolfbauer.

It was necessary to develop an adaptive dialogue structure, which the chatbot can follow according to the answers of the user. It was designed in

such a way that reflection can take place easily and as the dialogue adapts to the answers given, it is assumed that the new dialogue structure makes it easier for the user to reflect on the task and that a high degree of reflection will take place. The second assumption regarding the dialogue structure is that since the user always experiences different paths and questions, the user does not get bored when interacting with Rebo regularly. The results of this study do not show the expected high level of reflection, but still some of the apprentices were able to reflect on the task on a low level. This may be connected to the fact discovered through the third flash light round and already discussed above, which is that reflecting with Rebo is more suited to a practical task than a theoretical one. Therefore, it can be seen as a positive outcome that still 28,5% of the apprentices in the study managed to achieve at least a low level of reflection during their conversations. In our opinion this shows that the dialogue structure did a good job guiding the apprentices through the reflection process. However, it needs to be used in a practical workshop to confirm its effects on the reflection depth. At this point of time the second assumption that the user does not get bored in the long term cannot be confirmed and has to be investigated further at after a field study of bigger scope.

The implementation for the introductory message was changed so different introduction phrases can be used in combination with the username. It is assumed that through the use of different introductions, the user does not get bored when interacting regularly with the chatbot. Additionally it is assumed, that with the use of the username the chat feels more personal, which should achieve a positive atmosphere during the conversation. In the interactions of the conducted workshop it can be seen that only one apprentice insulted Rebo throughout the conversation and one other apprentice tricked the system. All other five interactions appear to show a positive atmosphere during the conversation. After the last question of the dialogue was answered Rebo thanked them for participating and the interactions were finished, but all apprentices decided to respond with an additional closing message. This might also show that the apprentices had the feeling they were talking to a real person and not a chatbot, as they seemed to try and be polite by greeting the chatbot at the end of the conversation. In fact, one of the apprentices even stated in the feedback "that chatting to the chat-

bot had felt like talking to a real person<sup>10</sup>". As for our second assumption, in order to verify that the apprentices do not get bored that easily through the use of different introduction phrases another field study for a longer period of time has to be carried out. Since no comparable data is available, it cannot be confirmed that through the usage of the username in combination with these phrases the atmosphere was more positive than it would have been without it. Even though it cannot be confirmed, we still think that these introductory messages have a great impact on the perception of the conversation and we think that also the use of the apprentices' name has also contributed to this.

To get rid of the problem of matching partial words, the regex and implementation used for pattern matching was slightly changed. When such mismatches occurred, the chatbot often followed the wrong paths in the dialogue structure. Since it should not happen that the chatbot follows the wrong path, it was assumed that this change should eradicate this problem, which means that no sub words are accidentally matched anymore. After analysing the conversations of the workshop, not a single mismatch could be found. Even though only a small amount of conversations was held it seems that the problem was completely resolved. Therefore it can be said that the assumption is confirmed by the present data. Nevertheless, it would be good to investigate this point further in future studies when more data is available.

One of the extensions to the framework was to enable the user to use multiple messages to respond to questions. The assumption was that people often tend to answer messages with multiple small messages when using messenger apps. We assumed that adding three dots at the end of the message would be a simple and understandable approach to indicate multiple messages. The results show that not a single user has sent a multi message response to Rebo. There are several possible causes for this. The first one is that even though it was mentioned to the apprentices how this can be done, there is no reminder or info box on the website to show this. Second, the three dots need to be added at the end of the message and maybe this is not as straight forward as expected. Further research would have to be

---

<sup>10</sup>Verbatim Quote: "fühlt es sich an als ob man mit einem anderen Richtigen Menschen kommuniziert"

done to find out how to effectively trigger this feature. Thirdly, the reason for not using multiple messages might be a lack of motivation, which could be due to the fact the apprentices had to reflect on a theoretical task and that they did not really know how to answer the chatbot's questions in the first place. They might have had more to say if questions had been asked on a practical task they carried out beforehand, as the apprentices themselves felt this might have been more useful to them. Finally, another reason the feature was not used could be that it is just a behaviour of the author of this thesis and most people do not typically use multiple messages to respond. Therefore, it is possible that the assumption this change is based on is simply wrong. On the other hand there is the possibility that the assumption is right, but we did not implement enough hints that such a feature exists or did not implement the feature in a way that was simple enough, which lead to none of the apprentices using it. For the second assumption we still think that the three dots are a simple way to indicate that the response is splitted into multiple messages, but this needs to be looked at in more detail after the first assumption has been confirmed, which could be done by researching studies on chatting behaviour of teenagers that focus on multiple messages for responding. In addition, in future studies, a flash light round could be used to gather feedback on this feature by asking the apprentices specifically why they did not use the feature. It might not have been a conscious decision not to use the feature, however, probably some of the apprentices did have specific reasons they are aware of and they could tell us about, so we get more insights into this question.

The next feature implemented was a reminder message, which gets automatically sent to the user when it is suspected he or she is trying to trick the system. This message had the goal to remind the user to take the conversation seriously. The assumption is that whenever this reminder message gets sent during a conversation, the user will start writing longer sentences and stop tricking the system, since it has been detected. The results of the conversations show that this message was sent to three participants. Two of them were not taking the conversation seriously, but were not bothered by this message at all. In the third case the message was triggered, because the apprentice answered the questions really shortly. Even though this apprentice did not try to trick the system, after the message was received he started to increase the number of words per answer. The assumption that



the user would start to write longer sentences after the reminder message was true in one out of three cases, but interestingly this was the case with the person who was wrongly suspected of tricking the system. On the other hand those users who really tricked the system simply continued their behaviour in the same way. It seems that the reminder message is not a big enough stimulus to lead the user back to writing serious messages. In our cases, if the user started to trick the system from the beginning, they did not change their behaviour after being confronted with such a reminder message. Therefore, the assumption that this reminder message can help to stop tricking the system during a conversation has not been confirmed. However, since only two cases could be investigated, more research needs to be done on the matter to see whether the reminder message can at least be successful in some cases or whether it is completely useless. In addition, even if the implementation is not successful, it could still be used to directly flag conversations, when such behaviour is suspected. In the case of the apprentices, this information could be forwarded to their instructors, who could try and change their behaviour, e.g. by offering help in case they are frustrated because they have not understood the questions or by explaining to them again how they might benefit from reflecting with Rebo.

### 7.3. Future Research

In this section different ideas for future research, which came up during the study, will be presented.

Firstly, since the results of the flash light rounds indicate that the current chatbot is not suited to reflect on theoretical tasks, it is important to find out whether it is really more suited for practical tasks. All apprentices stated during the third flash light round that they think it has more sense to reflect on a practical task. It would be a good idea to check this assumption in future research. If the assumption can be confirmed, it would be important to make sure Rebo is only used for reflecting on practical tasks in the future, since one of the main problems in this study appears to have been that the apprentices found it very hard to reflect on a theoretical task.

The next idea for future research would be to increase the intelligence of

the chatbot. Currently the chatbot uses pattern matching in combination with dictionaries, which is one of simplest approaches to analyse messages. The chatbot uses the matched words to adapt the conversation and lead the user onto different paths in the dialogue structure. Even though in the field study all messages were correctly matched, this type of pattern matching has its limitations. These are mostly related to the fact that it can only react to information stored in the dictionaries and cannot respond to new data. Therefore, it would be a good step to increase the intelligence of the chatbot, by using machine learning. This would ensure that the chatbot would not be restricted by the dictionaries. The apprentices who participated in this study are from a more rural area and therefore use a dialect in their messages. The chatbot only can react to known words in the dictionary, when using the pattern matching technique. With the use of a machine learning model it would be possible to react more reliable on unknown words, like when a user writes in dialect. The framework already supports LightSide machine learning models, but these can only be generated with enough data. This means more data has to be collected before this plan can be executed.

While analysing the results of the conversations it could be seen that the assumption that users typically use multiple messages to respond to questions was probably wrong. Therefore it would be an idea to find out if this assumption is indeed completely wrong and this is only a habit of the author of this thesis. Subsequently, if the assumption turns out to have been right it would be interesting to investigate further why the apprentices did not use this feature.

The results show that the reminder message was successfully triggered when a user was suspected of tricking the system. Even though one false positive was detected, the system also detected two other apprentices who really tricked the system. It should be possible to decrease the amount of false positives by waiting for at least two answered questions before calculating the average word count. The reminder message was not successful at persuading the apprentices who were tricking the system to write serious sentences. Therefore, it would be an idea for future research to focus on the topic of tricking the system and how to lead the user back to normal behaviour. No research done on this topic could be found, so there appears to be a gap in the research in this field which could provide an interesting topic for future studies.

The final idea for future research is to tackle the biggest limitation of this thesis, which is that only seven apprentices interacted with the chatbot one time. This amount of interactions is quite low for a study and therefore bigger field studies should be conducted in the future. Such studies could focus on the long term engagement with Rebo, which automatically leads to a lot of interactions between users and Rebo. The assumptions of this thesis could be analysed in more detail, especially the ones for which very little data was available. The assumptions which should be looked at after a bigger field study are: that the user does not get bored that easily of the website through the updated visuals, that the additional modes are useful and add more options for future studies, that the additional URLs are useful to researchers who want to access our data, that the different paths of the dialogue reduce the probability of a user getting bored after interacting with Rebo regularly and that people tend to answer messages with multiple small messages. A long term study is currently being planned by Irmtraud Wolfbauer and will again be conducted in cooperation with apprentices.



## 8. Conclusion

The goal of this study was to use a chatbot to lead apprentices through a reflective dialogue. For this, the chatbot Rebo was created, by using the Bazaar framework as a starting point. Further, it was important to study the acceptance of the chatbot by apprentices. The apprentices interacted with Rebo during their workshop in the apprenticeship training. This workshop was also used to collect feedback from the apprentices, which was a first step for a full study.

In this master thesis we had to fix known bugs of the Bazaar framework, which caused the created chatbot to be stuck during the conversation. To prevent mismatches during analysing the messages of the user, we improved the used pattern matching technique. To extend the functionality of the chatbot, it was implemented that a user can respond to a question with multiple small messages and a reminder message was implemented, which gets only sent when Rebo suspects that the user is tricking the system. The adaptive dialogue structure, created by Irmtraud Wolfbauer, was used for the chatbot. This dialogue structure consists of different paths, where different messages are available for each step. This should ensure that the user does not get bored that easily after interacting with Rebo frequently. Additionally, a webserver was set up to host the chatrooms for the conversations. The website provided by the developers of the Bazaar framework was redesigned to achieve a more pleasing experience when chatting with Rebo. On the server also two scripts were implemented, which should help to keep the maintenance cost of the server low.

The procedure for the evaluation was taken over from the Rebo Junior study and adapted during the workshop. The chatbot was designed for rather practical tasks, but due to too little communication with the instructor of the workshop we had no choice but to use Rebo for reflection on a theoretical task. This was not ideal for the study and in future studies a tighter collaboration with the instructors should be maintained, in order to prevent such a problem from occurring again. Nevertheless, some apprentices were able to achieve a low level of reflection by conversing with Rebo. We have shown in this study that a chatbot can lead a user through a reflective

dialogue and that such a chatbot can be created successfully with the Bazaar framework.

The acquired results have laid the foundation for a bigger future study with Rebo and until the next study Rebo can be improved further according to the results of this study. Through the feedback of the apprentices we know that the reflection should take place after practical tasks and this information should help to design the next study. The next study should probably aim at the long term engagement of users with such a reflective chatbot and at how their reflective behaviour changes through the use of such a system.

## References

- AbuShawar, Bayan and Eric Atwell (Dec. 2015). "ALICE chatbot: Trials and outputs." In: *Computación y Sistemas* 19. DOI: [10.13053/cys-19-4-2326](https://doi.org/10.13053/cys-19-4-2326) (cit. on p. 21).
- Breines, Juliana G. and Serena Chen (2013). "Activating the inner caregiver: The role of support-giving schemas in increasing state self-compassion." In: *Journal of Experimental Social Psychology* 49.1, pp. 58–64. ISSN: 0022-1031. DOI: <https://doi.org/10.1016/j.jesp.2012.07.015>. URL: <http://www.sciencedirect.com/science/article/pii/S0022103112001503> (cit. on p. 25).
- Chaudhuri, Sourish et al. (Jan. 2009). "Engaging Collaborative Learners with Helping Agents." In: vol. 200, pp. 365–372. DOI: [10.3233/978-1-60750-028-5-365](https://doi.org/10.3233/978-1-60750-028-5-365) (cit. on p. 21).
- Ciechanowski, Leon et al. (2019). "In the shades of the uncanny valley: An experimental study of human–chatbot interaction." In: *Future Generation Computer Systems* 92, pp. 539–548. ISSN: 0167-739X. DOI: <https://doi.org/10.1016/j.future.2018.01.055>. URL: <http://www.sciencedirect.com/science/article/pii/S0167739X17312268> (cit. on p. 22).
- Crutzen, Rik et al. (May 2011). "An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study." In: *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 48, pp. 514–9. DOI: [10.1016/j.jadohealth.2010.09.002](https://doi.org/10.1016/j.jadohealth.2010.09.002) (cit. on p. 20).
- Cui, Lei et al. (July 2017). "SuperAgent: A Customer Service Chatbot for E-commerce Websites." In: *Proceedings of ACL 2017, System Demonstrations*. Vancouver, Canada: Association for Computational Linguistics, pp. 97–102. URL: <https://www.aclweb.org/anthology/P17-4017> (cit. on p. 19).
- Di Stefano, Giada et al. (Jan. 2014). "Learning by Thinking: How Reflection Aids Performance." In: *SSRN Electronic Journal* 2015. DOI: [10.2139/ssrn.2414478](https://doi.org/10.2139/ssrn.2414478) (cit. on p. 19).
- Fessl, Angela et al. (2017). "In-App Reflection Guidance: Lessons Learned Across Four Field Trials at the Workplace." In: *IEEE Transactions on Learning Technologies* 10.4, pp. 488–501. DOI: [10.1109/TLT.2017.2708097](https://doi.org/10.1109/TLT.2017.2708097) (cit. on pp. 17, 19).

- Fitzpatrick, Kathleen Kara, Alison Darcy, and Molly Vierhile (June 2017). "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial." In: *JMIR Ment Health* 4.2, e19. ISSN: 2368-7959. DOI: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785). URL: <https://doi.org/10.2196/mental.7785> (cit. on p. 24).
- Graesser, Arthur C., Danielle S. McNamara, and Kurt VanLehn (2005). "Scaffolding Deep Comprehension Strategies Through Point&Query, AutoTutor, and iSTART." In: *Educational Psychologist* 40.4, pp. 225–234. DOI: [10.1207/s15326985ep4004\\_4](https://doi.org/10.1207/s15326985ep4004_4). eprint: [https://doi.org/10.1207/s15326985ep4004\\_4](https://doi.org/10.1207/s15326985ep4004_4). URL: [https://doi.org/10.1207/s15326985ep4004\\_4](https://doi.org/10.1207/s15326985ep4004_4) (cit. on p. 20).
- Graesser, Arthur C., Kurt VanLehn, et al. (Dec. 2001). "Intelligent Tutoring Systems with Conversational Dialogue." In: *AI Magazine* 22.4, p. 39. DOI: [10.1609/aimag.v22i4.1591](https://www.aaai.org/ojs/index.php/aimagazine/article/view/1591). URL: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1591> (cit. on p. 20).
- Hayashi, Yugo (2013). "Pedagogical Conversational Agents for Supporting Collaborative Learning: Effects of Communication Channels." In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '13. Paris, France: Association for Computing Machinery, pp. 655–660. ISBN: 9781450319522. DOI: [10.1145/2468356.2468472](https://doi.org/10.1145/2468356.2468472). URL: <https://doi.org/10.1145/2468356.2468472> (cit. on pp. 20, 21).
- Hettige, Budditha and Asoka Karunananda (Nov. 2015). "Octopus: A Multi Agent Chatbot." In: Proceedings of 8th International Research Conference, KDU (cit. on p. 21).
- Inkster, Becky, Shubhankar Sarada, and Vinod Subramanian (Nov. 2018). "An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study." In: *JMIR Mhealth Uhealth* 6.11, e12106. ISSN: 2291-5222. DOI: [10.2196/12106](http://www.ncbi.nlm.nih.gov/pubmed/30470676). URL: <http://www.ncbi.nlm.nih.gov/pubmed/30470676> (cit. on p. 25).
- Jain, Mohit et al. (Dec. 2018). "FarmChat: A Conversational Agent to Answer Farmer Queries." In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.4. DOI: [10.1145/3287048](https://doi.org/10.1145/3287048). URL: <https://doi.org/10.1145/3287048> (cit. on p. 24).
- Kaleem, Mohammed, James O'Shea, and Keeley Crockett (July 2014). "Development of UMAIR the Urdu Conversational Agent for Customer



- Service." In: *Lecture Notes in Engineering and Computer Science* 1, pp. 86–91 (cit. on p. 23).
- Klenner, Manfred, Angela Fahrni, and Stefanos Petrakis (May 2009). "Po1Art: A Robust Tool for Sentiment Analysis." In: *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)*. Odense, Denmark: Northern European Association for Language Technology (NEALT), pp. 235–238. URL: <https://www.aclweb.org/anthology/W09-4635> (cit. on p. 44).
- Kocielnik, Rafal, Daniel Avrahami, et al. (2018). "Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent." In: *Proceedings of the 2018 Designing Interactive Systems Conference*. DIS '18. Hong Kong, China: Association for Computing Machinery, pp. 881–894. ISBN: 9781450351980. DOI: 10.1145/3196709.3196784. URL: <https://doi.org/10.1145/3196709.3196784> (cit. on pp. 19, 26, 27).
- Kocielnik, Rafal, Gary Hsieh, and Daniel Avrahami (2018). "Helping Users Reflect on Their Own Health-Related Behaviors." In: *Studies in Conversational UX Design*. Ed. by Robert J. Moore et al. Cham: Springer International Publishing, pp. 85–115. ISBN: 978-3-319-95579-7. DOI: 10.1007/978-3-319-95579-7\_5. URL: [https://doi.org/10.1007/978-3-319-95579-7\\_5](https://doi.org/10.1007/978-3-319-95579-7_5) (cit. on p. 26).
- Kocielnik, Rafal, Lillian Xiao, et al. (July 2018). "Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity." In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.2. DOI: 10.1145/3214273. URL: <https://doi.org/10.1145/3214273> (cit. on pp. 19, 25, 26).
- Kopp, Stefan et al. (2005). "A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application." In: *Intelligent Virtual Agents*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 329–343. ISBN: 978-3-540-28739-1 (cit. on pp. 22, 24).
- Kovanović, Vitomir et al. (2018). "Understand Students' Self-Reflections through Learning Analytics." In: *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. LAK '18. Sydney, New South Wales, Australia: Association for Computing Machinery, pp. 389–398. ISBN: 9781450364003. DOI: 10.1145/3170358.3170374. URL: <https://doi.org/10.1145/3170358.3170374> (cit. on p. 19).
- Lee, Minha et al. (2019). "Caring for Vincent: A Chatbot for Self-Compassion." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing*

- Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, pp. 1–13. ISBN: 9781450359702. DOI: 10.1145/3290605.3300932. URL: <https://doi.org/10.1145/3290605.3300932> (cit. on p. 25).
- Leung, Cane and Stephen Chan (Jan. 2008). "Sentiment Analysis of Product Reviews." In: (cit. on p. 35).
- Li, Ian, Jodi Forlizzi, and Anind Dey (2010). "Know Thyself: Monitoring and Reflecting on Facets of One's Life." In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '10. Atlanta, Georgia, USA: Association for Computing Machinery, pp. 4489–4492. ISBN: 9781605589305. DOI: 10.1145/1753846.1754181. URL: <https://doi.org/10.1145/1753846.1754181> (cit. on p. 25).
- Liao, Q. Vera et al. (2018). "All Work and No Play?" In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, pp. 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3173577. URL: <https://doi.org/10.1145/3173574.3173577> (cit. on p. 24).
- Lucas, Gale M. et al. (2014). "It's only a computer: Virtual humans increase willingness to disclose." In: *Computers in Human Behavior* 37, pp. 94–100. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2014.04.043>. URL: <http://www.sciencedirect.com/science/article/pii/S0747563214002647> (cit. on p. 19).
- McNamara, Danielle, Irwin Levinstein, and Chutima Boonthum (June 2004). "iSTART: Interactive strategy training for active reading and thinking." In: *Behavior research methods, instruments, and computers : a journal of the Psychonomic Society, Inc* 36, pp. 222–33. DOI: 10.3758/BF03195567 (cit. on p. 20).
- Mulyana, E., R. Hakimi, and Hendrawan (July 2018). "Bringing Automation to the Classroom: A ChatOps-Based Approach." In: *2018 4th International Conference on Wireless and Telematics (ICWT)*, pp. 1–6. DOI: 10.1109/ICWT.2018.8527810 (cit. on p. 23).
- Pammer, Viktoria, Birgit Krogstie, and Michael Prilla (Jan. 2017). "Let's talk about reflection at work." In: *International Journal of Technology Enhanced Learning* 9, p. 151. DOI: 10.1504/IJTEL.2017.10005175 (cit. on p. 19).
- Renner, Bettina et al. (May 2016). "Effects of Prompting in Reflective Learning Tools: Findings from Experimental Field, Lab, and Online Studies." In: *Frontiers in Psychology* 7, p. 820. DOI: 10.3389/fpsyg.2016.00820. URL:

- <https://www.frontiersin.org/article/10.3389/fpsyg.2016.00820> (cit. on p. 19).
- Ruan, Sherry et al. (2019). "QuizBot: A Dialogue-Based Adaptive Learning System for Factual Knowledge." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland UK: Association for Computing Machinery, pp. 1–13. ISBN: 9781450359702. DOI: [10.1145/3290605.3300587](https://doi.org/10.1145/3290605.3300587). URL: <https://doi.org/10.1145/3290605.3300587> (cit. on p. 20).
- Schön, D (1983). *The reflective practitioner. How professionals think in action*. London: Temple Smith (cit. on p. 19).
- Silvervarg, Annika and Arne Jönsson (May 2012). "Subjective and Objective Evaluation of Conversational Agents in Learning Environments for Young Teenagers." In: (cit. on p. 22).
- Suppala, Kavya and Narasinga Rao (June 2019). "Sentiment Analysis Using Naïve Bayes Classifier." In: *International Journal of Innovative Technology and Exploring Engineering*. Blue Eyes Intelligence Engineering and Sciences Publication, pp. 265–269 (cit. on p. 34).
- Tanaka, Fumihide and Shizuko Matsuzoe (July 2012). "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning." In: *J. Hum.-Robot Interact.* 1.1, pp. 78–95. DOI: [10.5898/JHRI.1.1.Tanaka](https://doi.org/10.5898/JHRI.1.1.Tanaka). URL: <https://doi.org/10.5898/JHRI.1.1.Tanaka> (cit. on p. 22).
- Wallace, Richard (2004). *The elements of AIML style*. ALICE AI Foundation. (Cit. on pp. 17, 21).
- Wolfbauer, Irmtraud, Viktoria Pammer-Schindler, and Carolyn Rose (2020). "Rebo Junior: Analysis of Dialogue Structure Quality for a Reflection Guidance Chatbot." In: *EC-TEL Impact Paper Proceedings 2020: 15th European Conference on Technology Enhanced Learning, CEUR Workshop Proceedings xxxx, 2020* (cit. on pp. 17, 18, 65).
- Xu, Anbang et al. (2017). "A New Chatbot for Customer Service on Social Media." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, pp. 3506–3510. ISBN: 9781450346559. DOI: [10.1145/3025453.3025496](https://doi.org/10.1145/3025453.3025496). URL: <https://doi.org/10.1145/3025453.3025496> (cit. on p. 23).
- Xue, Zijun et al. (2018). "Isa: Intuit Smart Agent, A Neural-Based Agent-Assist Chatbot." In: *2018 IEEE International Conference on Data Mining*

- Workshops, ICDM Workshops, Singapore, Singapore, November 17-20, 2018.* Ed. by Hanghang Tong et al. IEEE, pp. 1423–1428. DOI: [10.1109/ICDMW.2018.00202](https://doi.org/10.1109/ICDMW.2018.00202). URL: <https://doi.org/10.1109/ICDMW.2018.00202> (cit. on p. 23).
- Zivin, Kara et al. (2009). “Persistence of mental health problems and needs in a college student population.” In: *Journal of Affective Disorders* 117.3, pp. 180–185. ISSN: 0165-0327. DOI: <https://doi.org/10.1016/j.jad.2009.01.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0165032709000093> (cit. on p. 24).
- Al-Zubaide, Hadeel and Ayman A Issa (2011). “Ontbot: Ontology based chatbot.” In: *International Symposium on Innovations in Information and Communications Technology*. IEEE, pp. 7–12. DOI: [10.1109/ISIICT.2011.6149594](https://doi.org/10.1109/ISIICT.2011.6149594) (cit. on p. 21).

## A. Transcript of the Feedback Rounds

Patrick

Irmtraud Wolfbauer

Apprentices

Wir machen jetzt zwei schnelle Runden wo jeder einfach ein kurzes Statement jeweils dazu abgibt. Wir gehen dann einfach reihum. Die erste Frage wäre: Wie hat euch die Interaktion mit Rebo gefallen und was sind eure Gedanken dazu?

Einer nach dem anderen. Fängst du an?

Die Interaktion mit dem war ganz cool muss ich sagen. Es war, also wenn man es nicht weiß dann fühlt es sich an als ob man mit einem anderen richtigen Menschen kommuniziert. Sicher es ist noch in der Entwicklungsphase und alles aber ich glaube, dass es in ein paar Monate beziehungsweise in 1 bis 2 Jahre sicher eine ausgereifte Geschichte sein wird und sicher eine große Hilfe für andere Lehrlinge sein wird.

Dankeschön.

Ja für mich wars halt nicht hilfreich... ich für andere wäre es es sich nichts Schlechtes und es ist eh in der Entwicklungsphase und wenn er halt besser entwickelt ist dann wird er sehr groß werden wahrscheinlich.

Und wie wars jetzt? Hast du dich gern mit ihm auseinander gesetzt einfach? War das lustig, hat er dir getaugt?

Ja halbert halbert. Er hat ab und zu fast eine gleiche Frage gestellt und also schlecht wars net aber ab und zu fast die gleiche Frage gestellt wo ich halt nur die gleiche Antwort geben könnte.

Cool danke.

Komisch eindeutig ich was net in was für einer Hinsicht wie weit des helfen soll, weil ich es nicht fachspezifisch auf Lehrling und wie des fachspezifisch.

Na die Frage war jetzt einfach, hat es dir Spaß gemacht mit ihm zu interagieren mit Rebo.

Ja. Anders. Komisch. Nicht wirklich wie ein Mensch oda ka Ahnung was das wirklich is.

Danke. Wie wars für dich?

Ja es is interessant weil manche Leute sozial schwach sand und keinen haben zum Reden und so haben sie wen zum Reden. Du kannst da manche Sachen überlegen, der red mit dir ja gscheid. Aber i sag halt des is daweil no net so ausgereift, i mein das Programm. Es gibt ganz sicher schon genug apps wo es sowas scho gibt. Die eigene Gschichten erzählen die i irgendwo eini geschrieben hab.

Siri

Ja sowas in der Art. Es gibt ja scho vü solcher Sachen. Aber speziell aufs lernen bezogen. . . in Moment versteh ichs no net.

(Alle reden zeitgleich - unverständlich)

Ahm Sachen wie mir lernen gefallen hat und wie halt das Ganze für mich halt war und eigentlich regt das mi zum Überlegen an. Wie war das für mi und so. Des hilft halt net viel fürs Lernen. I man vielleicht ok du denkst nach über den Gedanken. Aber wenn es halt spezifische Fragen wären oder so und mit ihm interagieren kannst. Waast? Hätt i vielleicht Antworten geben dürfen und des würd sehr hilfreich sein.

Ok danke.

Darf i no was sagen?

Ja klar

Vielleicht is es eigentlich mehr für Leut die nicht mit Leut reden können.

Mhm.

Kann ja sein, dass du schüchtern bist oda so. Da kann i ma das dann scho gut vorstellen.

Geh ma dann mal weiter. Wie hat es sich angefühlt, war es lustig?

Das is fürs Lernen das is, das is ganz normal Kontaktaufbau so wie jetzt mitn Jessy red is es halt nur übers Internet. Wal fürs Lernen wennst lieber

a Frag stellst dann kommt er halt wieder mit irgendeiner Gegenfrage her und des bringt ma halt net viel dass i da halt irgendwas lernen kann.

Und wie hat sich die Interaktion für dich angefühlt? War es cool?

Ja, komisch.

Komisch, ok.

Ja wallst halt mit am Computer schreibst. Ja es wär halt besser wenn er schreibt wie war die Aufgabe für dich und wenn du da halt hinschreibst ich habe sie nicht ganz verstanden, dass er dann halt Tipps gibt wie du es machen kannst. Zum Beispiel vor dich hinsprechen, net nur lesen weil es gibt ja auditive äh kinästhetische und was ahm und diverse Lerntypen.

Visuelle. Und wie war jetzt die Interaktion mit Rebo für dich? War es lustig? Hast du ihn cool gefunden? War es schräg?

Na es war schon ok. Aber es war halt das Problem dass wenn du zum Beispiel was zurück geschrieben hast ahm. Es war sehr interessant für mi und dann hat dann halt die Frage fast genauso wieder gestellt also was hat dir am besten gefallen und vorher hat er mich gefragt gehabt hat mir die Aufgabe gefallen. Er kann halt ausm Antwortsatz kann er einfach 2 Wörter nehmen, a Nomen oder was und dann sagt zum Beispiel warum war Induktion so spannend für dich. Sodass er halt das Nomen reflektiert weil das regt dich dann halt nochmal dazu an, dass du spezifisch mehr darüber nachdenkst.

Ok. Mhm das ist gut.

Man muss halt a no sagen, er wirkt etwas leblos mit den Fragen, die vorprogrammiert sind. Es ist irgendwie vorhersehbar. Also da merkt ma das die vorprogrammiert san, des war vorhersehbar. Des heißt er hat leblos gewirkt, anfoch weil, technisch und deshalb wirkst wahrscheinlich auch komisch. Weil immer so ähnliche Fragen kommen. Da fühlt man sich halt echt so, keine Ahnung. Was tut der oda so.

I glaub es is deswegen gut, weil man dann nochmal darüber nachdenkt was weiß i oder was weiß i net. Oder durch die Fragen zum Beispiel was i anders mach wenn ichs am nächsten Tag nochmal mach. Und dann denkst mal selber drüber nach morgen würd ichs vielleicht schneller machen. Und

dann kommst vielleicht selber drauf je öfter i des durchlies desto besser kann ich es.

Du beantwortest grad die Frage die als zweites Stellen wollt. Könnt ihr das Handy kurz weggeben es wär cool wen ma kurz bei der Sache bleiben. Die erste Frage wär eigentlich gewesen Wie war die Interaktion für dich und is es cool für dich mit Ihm zu chatten. Wie fühlt sich das an.

Als ob i mit irgendwem schreib, wie a email zB. Und er stellt halt durchgehend Fragen

Weil die Frage die wir als zweites Stellen war, was du jetzt als erstes beantwortet hat. Was hat es dir gebracht und was glaubst du das es dir bringen kann.

Also das ich nochmal über das ganze Thema drüber nochmal nachdenk.

Ja aber generell es tun sich viele leichter beim Lernen wenn i es irgendwem anders erklären kann, also wenn i irgendwem Nachhilfe gib in dem Thema. Dann merk i Sachen leichter und meistens vü bessa als wenn i es nur so hinschreib und dann wars des für mi. Dann geht es nur in Kurzzeitspeicher und spätestens 2 Tag später hab i es eh wieder vergessen.

Mach ma jetzt kurz die zweite Runde auch. Du bist jetzt auch scho halbert dort. Was glaubst du was dir ein Chatbot bringen kann. Hat er dir was gebracht oder was glaubst du kann er in der Zukunft bringe. Hat es dir was gebracht, einfach das reflektieren mit dem Rebo.

Ja es war halt eine Konversation und geht halt nicht ins fachliche. Er fragt halt nur ganz oberflächlich. Wie war die Aufgabe für di, hats da was bracht und i man dahin lügen „ja sia“ kann man schnell amal. Aber wenn die jetzt ausdrucken würd und dann spezifisch was fragen würd, was du geschrieben hast, dann muast du drüber nachdenken.

Cool danke.

Und das er deine Fehler dann wennst was falsch schreibst dann korrigiert.

Na des kummt arrogant



Na i man wenn du jetzt irgendwas Fachliches falsch schreibst und der merkt ah ha, da is jetzt a Fehler drinne das er dir dann sagt des gehört net hin und aber des gehört dazu.

Also du fändest es gut wenn er dich korrigieren würd. Ahm und so wie er jetzt is findest du es bringt was? Oder kann es den nachfolgenden Lehrlingen was bringen so zu reflektieren. Was manst du?

I sog i bin no imma da Meinung wenn wer glabt das er lernt dann lernt er und wenn er sagt na, dann glaub i net des der do wenn er sagt er will net. Kann da kummen was will. Bringt sich des nix.

Und wenn jemand motiviert darauf zu geht bringt die app was, bringt der chatbot was?

Ja wenn du motiviert darauf zugehst dann scho.

Ok Dankeschön. Was meinst du? Was bringts?

Die Lehrlinge im ersten Lehrjahr haben scho damit gearbeitet oda?

Ja. Net mit dem chatbot sondern einer einfacheren Version.

Da haben sie auch geschattet oda was

Da sind sie zum Reflektieren angeleitet worden aber er hat nicht auf die antworten reagieren können. Das war so die Vorstufe.

Also im Moment weiß i net was ma des beim Lernen helfen würd. Er fragt mi ja nur wie die Aufgaben glaufen san, also es geht eigentlich genau um gar nix. Es einfach nur normal small talk. Wie wenn ich daham mit meiner Mutter red die sich fachlich net auskennt und die mi fragt wie is es ma dabei gangen. Wie wenn i von da schul ham kum

Ok und das bringt dir persönlich jetzt weniger

Speziell net, außer i würds eini schreiben . . . Einfach dast wen zum Zuhören hast . . .

Ok. Kannst du noch ein Statement abgeben, was bringt da Rebo?

Mir persönlich bringt der sicher nix weil ich sowieso ein leichter Lerner bin. Also brauch i des net. I weiß net wie. . . es gibt ja verscheiden Methoden, . . . wie ausgereift die Methode ist. I glaub i stells ma. . . Wennst nochmal

darüber nachdenkst dann vielleicht wennst es dann wieder erwähnst aba sonst schwierig.

Ok danke

Im Grunde genommen hat jeder Satz a Abfolge und wenn ma irgendwie übers Skriptum drüber laufen und der nach gewissen Kriterien a Frage stellt oda gegenfrag oder die Antwort

Du hast gemeint du hättest gern da Rebo Hintergrundinformationen über das Fach hat.

Ja das wäre a Möglichkeit. Aba das macht nur das Programm an sich groß oder. Wenn du jetzt Fach Information eini pflanzt. Es wär leichter wenn am ausm Satz raus reflektiert weil jeder Satz a geregelte Abfolge hat und durch die Abfolge kann man ganz schnell bestimmen an welcher Stelle das Nomen steht oda wie es außer gefiltert werden kann, weil basiert ja alles auf Grammatik also

Also einfach wenn du das Winkelschleifen erwähnst das Rebo das dann zurück schmeist.

Nur das Problem ist glaub ich der Satz selber. Denn wer kann heutzutage no an ganzen Satz bilden. Das wird imma a Problem sein.

Das is a Challenge da muss ma. Cool magst du jetzt noch a Statement geben. Was glaubst du das er bringt, der Rebo? Oder bringt er dir was?

Ja momentan net so guat, weil i halt anders lern wie die andern und des was anders is. (alle reden zeitgleich)

Glaubst du das er was bringt? Bringt er dir was.

Also persönlich im Moment in dem wie es jetzt is wenig bis gar net. Aber was wie gesagt aber cool wär. . . interessant wär, wär zum Beispiel wenn i a Antwort schreibt wenn er di fragt mit was für nem Thema hast du dich zuletzt beschäftigt oda zuletzt gearbeitet und du schmeist ihm nen Begriff ummi. Das er dir dann a paar Grundfragen sand zB wenn ma zum Beispiel an Begriff gsagt hat ok i hab mi zuletzt mit dem Fehlerstromschutzschalter beschäftigt, dass er dir dann a allgemeine Frage zu dem Fach stellt.

Also wenn er dich ausfragen würde, net nur dich dazu anleitet selber nach zu denken. Sondern wenn er

Ja grad a paar Fragen oder was kannst du ma alles über des Thema erzählen.

Es is wirklich a unterschied weil ihr grad mitten im Lernen drinnen seid. Weil wir haben ihn bisher nur zum Reflektieren nach praktischer Arbeit verwendet. Das is a anderer Punkt

Da kann ichs ma vorstellen, weil praktisch

Dürf ma noch a Runde machen? Habts noch Energie? I glaub das wär jetzt interessant. Geh ma jetzt noch einmal im Kreis. Wenn ihr euch vorstellts ihr habts grad a praktische Aufgabe gemacht. Ihr habts grad irgendwas herstellen müssen und nachher würd er euch so durchleiten. Zum Reflektieren. Würd das was ändern, wär das vernünftig für euch.

Ja wie wie durchleiten. Also i hab jetzt was gemacht und jetzt geht's eigentlich nur darum er fragt mi wie is ma dabei gangen oda was.

Genau er versucht dich dann nochmal dazu zu bringen das ganze nochmal das ganze durchzudenken was du gelernt hast und was es bringt

Wenn i meine Arbeitsschritte nochmal überdenk und vielleicht kum i auf irgendeinen Fehler drauf den was i... also ob i irgendwas vergessen hab

Also glaubst du das des mehr bringen würde wenn das auf a praktische Arbeit folgt. Für dich? Nummer 1 hat gesagt ja das würde was bringen.

I würd gern no was sagen. Des reagiert ja auf gewisse Wörter. Je nachdem was für Wörter i schreib umso, dessen Satz reflektiert der eben.

Ja

Bei dem elektrischen Arbeitsbeispiel haben wir imma a feste Arbeitsabfolge, also wennst dann schreibst habe angeschlossen. De zwa Wörter, würd der bot dann nachfragen ob du die 5 Sicherheitsregeln angewendet hast

Nein Nein Nein. Der bot is eine Lerntechnologie keine Technologie die auf eure Fächer spezialisiert ist. Der leitet euch nur zum selber Nachdenken an.

Also generell für jedes Fach, für jede Branche

Es is daweil nicht fachspezifisch. Das heißt wenn du jetzt was Praktisches gemacht hättest, wär des für dich sinnvoller gewesen so angeleitet zu werden?

Ja

Ja

Wer hatten die Zeit das er dann nachschauen kann

Die Zeit nehmen

Nachdem du was Praktisches gemacht hast, dann nochmal darüber nachdenken würd dir das was bringen.

Ja aba da fahrt das ganze Gerät aba scho

Ja und

Genau

Aba es geht darum das du wiedergibst was du gemacht hast und vielleicht dir vorher dir merkst oda merkst ok bei dem Vorgang war dieses Mal des falsch dann hätt i eher was anders gemacht

Ja aba i wird jetzt. . .

Es geht jetzt net für di in da Praxis sondern vü welche die grad lernen oda wie oda was

In da Ausbildung

Aaaah

(Gelächter)

Du glaubst ja nit das sich a dreißig jähriger Chef sich sowas antuat, der muss ja nimma lernen. Was isn jetzt los. Denk amol noch

Ich hol euch jetzt nochmal ganz kurz zurück. Die ersten beiden haben gemeint es würde was bringen. Du bist da net sicher, was glaubst du in der Ausbildung würd des dir was bringen? Ja oder nein? Darf ich dein

Statement haben, auf eine praktische Arbeit so nachzudenken. Würd dir was bringen?

Wie i? Was?

Nach einer praktischen Aufgabe hier in der Ausbildung so theoretisch nochmal durch zudenken würd dir was bringen?

Ja

Glaubst schon

I glaub mehr als wenn i jetzt was lern

Mhm ja i glaub des könnt besser zammen passen. Was meinst du?

Eindeutig, weil i dann vielleicht die Schritte und vielleicht die Sicherheit dann nochmal durch geh und des fürs nächste Mal. Keine Ahnung vielleicht hab i an Fehler gemacht und gib des wieder. Dann weiß ich des beim nächsten Mal, i hab an Fehler gemacht und jetzt wird ich es anders machen. I glaub das da die Wiedergabe schon was bringt.

Bringt was. Was meinst du?

Ja, ja alles was da Jessy halt a grad gesagt hat. Es würd scho was bringen.

Cool cool. Was meinst du?

Also i denk scho das es was bringen würd, weil ma zum Nachdenken angeregt wird, weil ma stellt sich selber net so die Fragen. Aba so hat ma die Möglichkeit das ma so Fragen gestellt kriegt und darüber dann nachdenken kann und dann wie gesagt Fehler erkennen, noch andere Sachen die was Fehlen oder was als Bonus no angesehen werden noch erkennen und noch zusätzliche Sachen erkennen. Denk scho

Cool, Dankeschön. Also ihr habts uns jetzt auf was hingeführt was echt wirklich interessant für uns ist. Für praktische Arbeiten is das viel relevanter als für lernen. Das is eigentlich a interessantes Ergebnis für uns. Danke fürs ausprobieren und danke fürs Feedback.





## B. Python Dictionary Splitter

Listing 3: Self-written Python program to split the polart dictionary.

```
colordeepblueimport os
colordeepblueimport sys

colordeepblueif __name__ == "__main__":
    colordeepblueif colordeepbluelen(sys.argv) != 3:
        #Requires path to polart dictionary and its filename.
        colordeepblueprint('Invalid number of arguments.')

    path = sys.argv[1]

    positiveFilePath = path + "\\positive.txt"
    negativeFilePath = path + "\\negative.txt"
    neutralFilePath = path + "\\neutral.txt"

    colordeepblueif os.path.exists(positiveFilePath):
        os.remove(positiveFilePath)

    colordeepblueif os.path.exists(negativeFilePath):
        os.remove(negativeFilePath)

    colordeepblueif os.path.exists(neutralFilePath):
        os.remove(neutralFilePath)

    inputFile = colordeepblueopen(path + "\\" + sys.argv[2], "r", encoding='utf-8')
    posFile = colordeepblueopen(positiveFilePath, "a+")
    negFile = colordeepblueopen(negativeFilePath, "a+")
    neuFile = colordeepblueopen(neutralFilePath, "a+")

    positveWords = []
    negativeWords = []
    neutralWords = []

    colordeepbluefor word colordeepbluein inputFile:
        lineSplitted = word.split()
        extractedWord = lineSplitted[0].lower()

        colordeepblueif(lineSplitted[1][:3] == "POS"):
            positveWords.append(extractedWord.replace("_", " ") + "\n")
        colordeepblueelif(lineSplitted[1][:3] == "NEG"):
            negativeWords.append(extractedWord.replace("_", " ") + "\n")
        colordeepblueelif(lineSplitted[1][:3] == "NEU"):
            neutralWords.append(extractedWord.replace("_", " ") + "\n")

    positveWords.sort()
    posFile.writelines(positveWords)

    negativeWords.sort()
    negFile.writelines(negativeWords)

    neutralWords.sort()
    neuFile.writelines(neutralWords)

    neuFile.close()
    negFile.close()
    posFile.close()
    inputFile.close()
```