DEGREE PROJECT IN INFORMATION AND COMMUNICATION TECHNOLOGY,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2021

Designing a text-based AI scheduling assistant chatbot for a business environment.

A case study of a mobile-based AI scheduling assistant app.

HAU-BEN BENJAMIN SHIH

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Examiner
Joakim Gustafsson
KTH Royal Institute of Technology

Academic adviser
Pavel Karpashevich
KTH Royal Institute of Technology

Jarmo Laaksolahti
KTH Royal Institute of Technology

Industrial adviser
Ather Gattami
The Research Institute of Sweden

Abstract

Scheduling a time to meet can be time-consuming, especially when coordinating with email. It could be challenging for business people when each participant is required to email back and forth to propose their availability, matching each he ime availability, and finding a suitable location to meet. It is even worse when participants must reschedule the entire meeting.

This thesis aims to design and develop an artificial intelligence (AI) scheduling assistant chatbot mobile app that could assist people in scheduling meetings efficiently in the business environment. The research process involves two rounds of design iterations. In the first design iteration, the goal was to explore and test the possible ways to design the chatbot. In the second design iteration, the goal was to learn from the first iteration and improve the design to fulfil the user needs. The results implied five options for designers to consider when designing an AI assistant chatbot for the business environment. The considerations include the (1) maturity of natural language processing, (2) instructions to new users, (3) feedback provided by the AI assistant, (4) effort of typing messages, and (5) personality of the AI assistant.

Keywords: Conversational agents, conversational user interface, AI assistant, natural language processing, chatbot

Sammanfattning

Det kan vara tidskrävande att planera en tid att träffas, särskilt när man samordnar med e-post. Det kan vara utmanande för affärsmän när varje deltagare måste skicka e-post fram och tillbaka för att föreslå deras tillgänglighet, matcha varandras tillgänglighet och hitta en lämplig plats att möta. Det är ännu värre när deltagarna måste planera om hela mötet.

Denna avhandling syftar till att utforma och utveckla en artificiell intelligens (AI) schemaläggningsassistent chatbot mobilapp som kan hjälpa människor att schemalägga möten effektivt i affärsmiljön. Forskningsprocessen innefattar två omgångar med design-iterationer. I den första designversionen var målet att utforska och testa möjliga sätt att utforma chatboten. I den andra designiterationen var målet att lära av den första iteration och förbättra designen för att uppfylla användarnas behov. Resultaten innebar fem alternativ för designers att överväga när de designade en AI-assistent-chatbot för affärsmiljön. Övervägandena inkluderar (1) mognad för naturlig språkbehandling, (2) instruktioner till nya användare, (3) feedback från AI-assistenten, (4) ansträngning att skriva meddelanden och (5) AI-assistentens personlighet.

Nyckelord: Konversationsagenter, konversationsanvändargränssnitt, AI-assistent, naturlig språkbehandling, chatbot

Acknowledgements

First, I would like to thank Pavel Karpashevich and Jarmo Laaksolahti, my thesis supervisors from the Division of Media Technology and Interaction Design at KTH Royal Institute of Technology. Pavel has provided me with helpful tips and guided me in the right direction throughout the research process. Jarmo has provided the final feedback for the thesis and guided me to finish up this thesis.

I would like to thank Dr. Joakim Gustafsson, my current examiner, for kindly agreeing to be my examiner and giving me the space to the direction of my research.

I would like to thank Dr. Mariët Theune, my observer at University of Twente, for providing her expertise and experience with conversational agents to assist me with my research process.

I would like to thank Dr. Ather Gattami, my industrial adviser, to meet with me every week to ensure that I am developing the project in the right track.

I would also like to thank the participants of the validation survey and testing for this research project. Without their participation, the research could not have been conducted.

Finally, I must express my very profound gratitude to my beloved TJ for providing me with unfailing support throughout the process of researching and writing this thesis.

Thank you.
Taipei, June 2021
Hau-Ben Benjamin Shih

Table of Contents

Table of Figures

Table of Tables

Chapter 1. Introduction

Companies and academic institutions worldwide have studied artificial intelligence (AI) for decades [1]. Artificial intelligence refers to the ability of computers or computer-controlled robots to perform tasks that are usually done by humans [2]. In short, AI learns from humans and takes actions similar to those that a human would [3]. According to Fortune Business Insights [4], a market research consulting firm, the AI market is expected to grow from USD 27 billion in 2019 to USD 266 billion in 2027.

According to Mordor Intelligence [5], a technology that is now trending due to AI technology growth is AI assistants (intelligent virtual assistants), such as chatbots or voice agent. Mordor Intelligence further predicted that the market for AI assistants, valued at USD 2.5 billion in 2019, is expected to reach USD 6.3 billion by 2025 [5]. AI assistants could help users deal with basic enquiries, allowing customer-facing industries, such as retail, healthcare, and e-commerce, to provide more reliable and immediate services to their customers [6]. Today, 1.4 billion people use AI assistants to access services and perform tasks, such as checking banking payments, setting daily reminders, finding flights and hotels, and so on [7]. Experts have estimated that, in 2025, most customer service will be replaced by AI assistants and that AI assistants will handle 95% of customer interactions [8].

Companies are now building AI assistants that work more closely with humans [9]. In other words, humans now have their own personal AI assistants to perform basic tasks [10]. A typical example of a personalised AI assistant is Apple s Siri, a virtual assistant introduced as part of the Apple iPhone in 2011, which helps its users perform personal tasks, such as setting alarms, navigating maps, or messaging others through a voice interface [11]. With mobile application (app) growth, companies have also started integrating chatbot as part of their mobile apps to replace customer service and enable customers to perform specific tasks with the chatbot [12]. A few success stories of implementing chatbots in their app include Duolingo, an app that helps users practice and learn new foreign languages by chatting with an AI assistant; Erica, an AI assistant from Bank of America that helps its customers perform basic banking services; and Emirates Vocation, an AI assistant that helps people organise trips. These examples are listed in Figure 1.

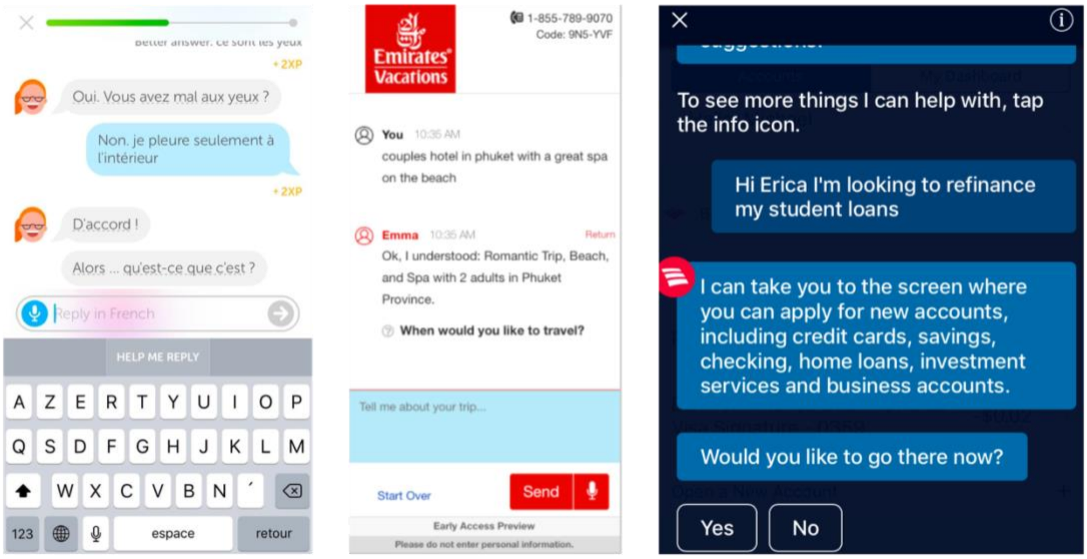


Figure 1. Example of application-based chatbot apps.

Left to right: Duolingo, Emirates Vocations and Bank of America.

This master thesis was conducted at the Research Institute of Sweden (RISE) in Stockholm at the department of Intelligent Dynamical Systems. The thesis aims to introduce an AI assistant chatbot mobile app called O.AI ha hel business people book meetings with colleagues. Today, scheduling a meeting often requires business people to send the email back and forth to the participants to find a time to meet. In a study conducted by Doodle [13], an employee spends an average of five hours each week scheduling meetings, and half of these meetings are rescheduled. An internal survey that the development team of O.AI conducted with 120 people from the business environment indicated that 89% of the participants had to send the email back and forth to schedule a single meeting, and 79% were willing to try any application that could solve this problem (Figure 2).

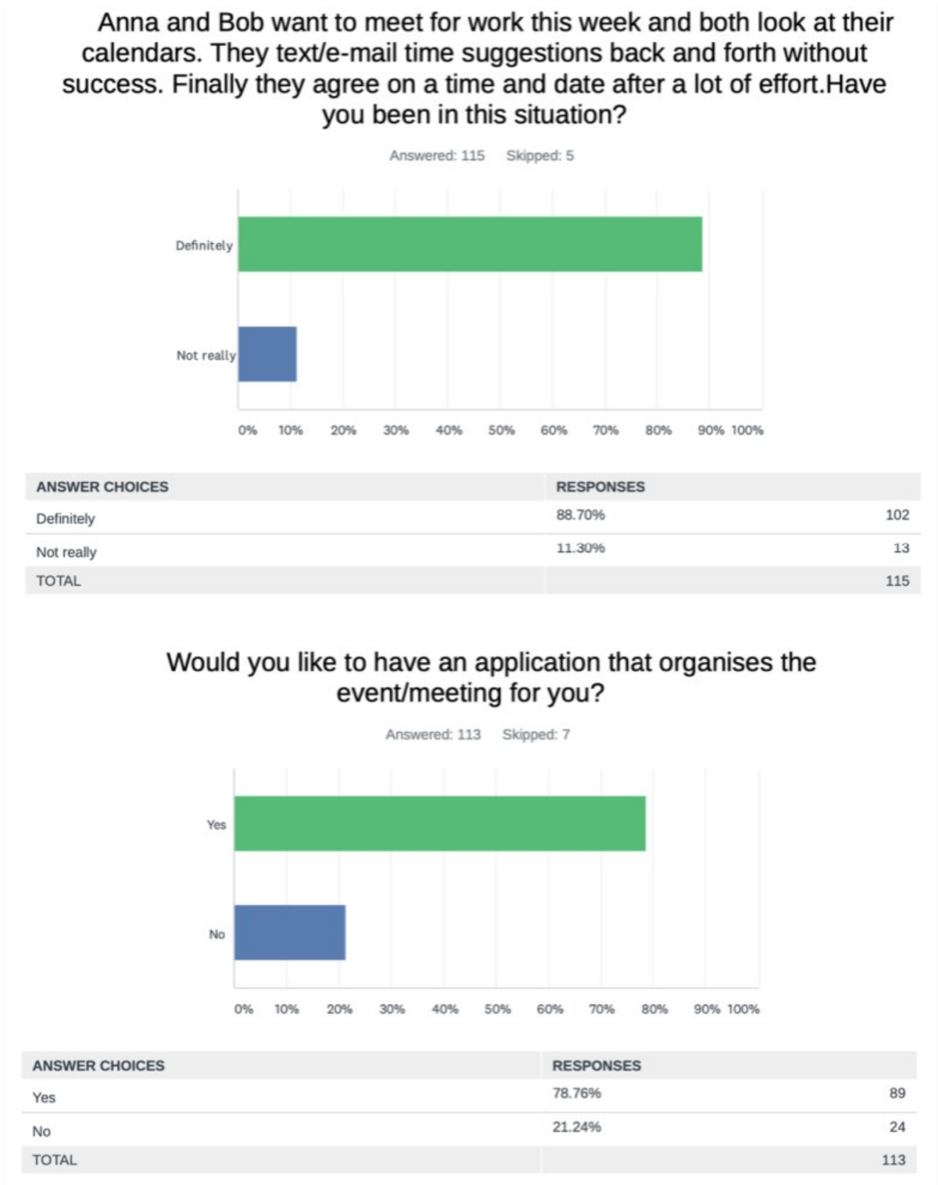


Figure 2. Internal survey results in May 2020.

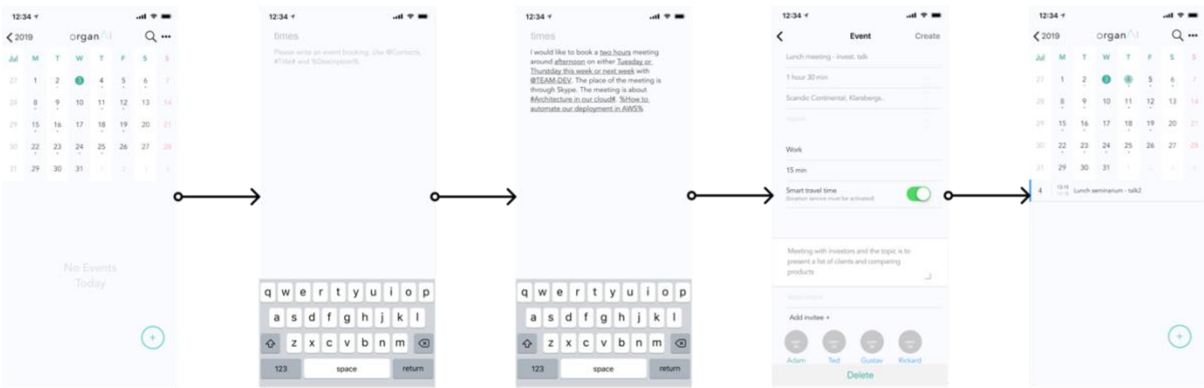
Due to a lack of design guidelines that detail how a chatbot interface can be designed [14], the thesis focuses on exploring different redesign methods for an existing AI scheduling assistant chatbot app and improving the usability through two rounds of design iterations. The thesis results deliver design implications to consider when designing an AI scheduling assistant chatbot that provides business people with a solution to schedule meetings with colleagues efficiently.

1.1 Project Status

Before I joined the company, an initial natural language processing (NLP) algorithm and an initial design prototype has already been developed and designed by previous colleagues.

The company developed the NLP on an NLP development platform, RASA, and can handle basic meeting information, such as how long the meeting will be, where is the meeting located, who the participants are, and the date and time of the meeting. It allows the users to send basic enquiries, such as Book a meeting with Laura at 4 p.m. tomorrow and handles the request to process it into a meeting request.

The initial design prototype of O.AI, designed by a previous designer of the company, was inspired by how business people send an email and offer a non-conversational interface type. The design requires the user to type all the required information and send it to the AI assistant at once. The AI assistant then processes the data using NLP and presents a confirmation page to ask the user to confirm the scheduled details. The design is presented in Figure 3. The detailed workflow of the original design is further explained in Chapter 4.



A) Calendar Homepage | B) Add New Event by Typing | C) Types meeting request | D) Confirm scheduling details | E) Added Event to Calendar

Figure 3. Initial design of the original design prototype.

1.2 Project Goal

The Stakeholders of O.AI are not happy with the initial design as the current design offers a rigid process. The user will have to type in all the required information at once. The app does not provide much instruction and feedback about the request. In order to replicate the experience of having an assistant, the main goal for this project is to convert the initial design of O.AI, which offers a non-conversational interface, into a conversational interface (e.g., chatbot) that offers users in the business environment a truly conversational AI assistant to help them schedule meetings. To achieve the goal, I am responsible for exploring the possible design opportunities to design the conversational-based AI scheduling assistant app and validating the design ideas by conducting user research.

1.3 Research Question

The goals of exploring the possibility of converting the current prototype into a chatbot-like interface led to the following research question:

Q: What type of chatbot is most suitable as an AI scheduling assistant in a business environment?

More specifically, because chatbots are roughly classified into text-based chatbots and button-based chatbots in the market, the team of O.AI wanted to validate the following two hypotheses:

H1: Users prefer a text-based chatbot because this interaction is more natural and easier to understand.

The hypothesis can be measured by evaluating the metrics generated from the System Usability Scale (SUS) questionnaire and interviewing the users to gather feedback on the design of a text-based chatbot compared to other kinds of chatbots.

H2: A button-based chatbot allows users to complete tasks more efficiently than a text-based chatbot.

The hypothesis is measured by evaluating how much time the participants spent completing tasks in a button-based chatbot and how much time it took to complete the same tasks in a text-based chatbot. In addition, interviewing the users could also help evaluate whether they experience more efficiency using the button-based chatbot than the text-based chatbot.

1.4 Approach

The approach taken to answer the research question involves two design iterations. In each of the iterations, new text and button-based chatbot designs are proposed based the research results. The first iteration focused on exploring different text and button-based chatbot design opportunities based on literature studies. The proposed designs in the first iteration are then tested with users to identify a preferred design approach The second iteration focused on redesigning the top-rated design preferred by the users in the first iteration to improve the usability. The proposed designs in the second iteration are also tested with users to validate its ease of use.

1.5 Thesis Overview

Chapter 2 discusses the history and background studies of chatbots and AI scheduling assistants. Chapter 3 presents the methods used to conduct in the two iterations. Chapter 4 presents the results of the two iterations. Lastly, Chapter 5 concludes this thesis by presenting the answer to the research question and hypothesis, a discussion of the research process, and the research limitations.

Chapter 2. Background

This chapter discusses the theoretical background of chatbots and AI scheduling assistants. In addition, because the thesis research was conducted at a company, it included a competitor analysis to understand the market and provide a better business strategy for the team responsible for O.AI.

2.1 History of Chatbots

Neff and Nagy [15] described chatbots (also known as chatter robots) as a type of program that engages users in a conversation. A user sends a message to the chatbot, and the chatbot responds to the message by choosing an appropriate expression from a preprogrammed set of data or through machine learning (ML) models [16].

The history of chatbots began in 1950 when Alan Turing [17] published an article in the journal *Mind* about whether machines could think. To answer the question, Turing proposed a simple method to test whether a machine could demonstrate human intelligence. The test had one human ask questions, and another human and a computer respond. The questioner continued to ask questions within a specific domain area, using a specific format and context, and the human and computer responded to each question. After a limited time and number of questions, researchers asked the questioner to determine which respondent was a human and which was a computer. The test is known as the Turing test [17] [19].

One of the first publicly known chatbots was named Eliza and was created by Joseph Weizenbaum at the Massachusetts Institute of Technology in 1966 [20]. Eliza was a chatbot that attempted to communicate with humans through pattern matching and substitution. Figure 4 reveals how Eliza communicated with a human. Eliza did not understand the conversation. However, by matching simple patterns and substituting a few words in each sentence, Eliza gave the illusion of understanding what the person was saying [18] [21].

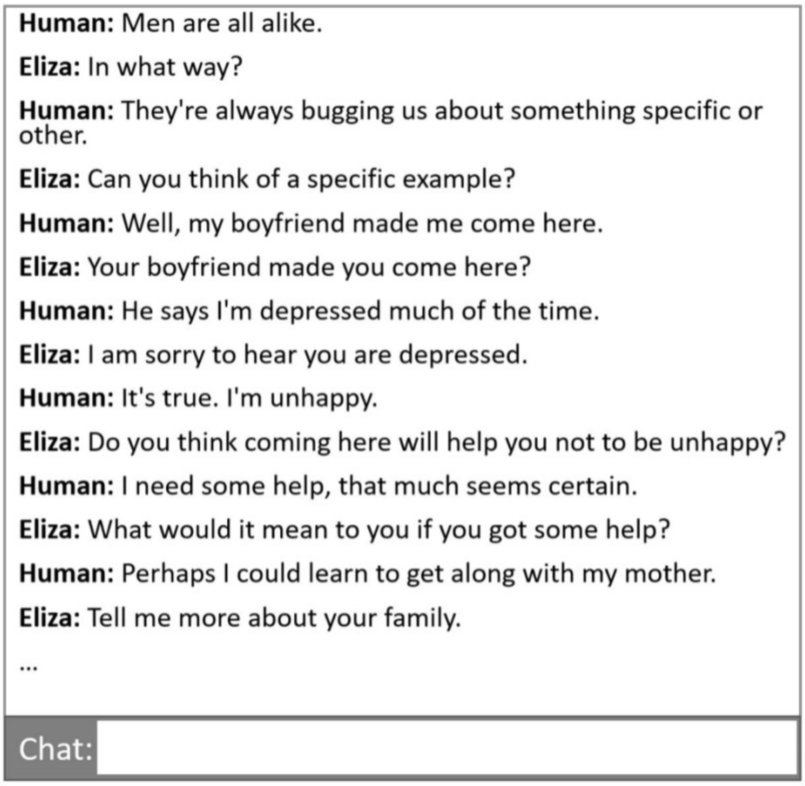


Figure 4. How Eliza communicated with a Human [20].

An enhanced version of Eliza, named Parry, introduced by Stanford computer scientist Kenneth Mark Colby, was developed in 1972 [21]. Parry was the first chatbot to pass a version of the Turing test [21]. However, critics insisted that both Eliza and Parry were cheating the Turing test because neither Eliza nor Parry could understand human language. Both programs created the illusion of understanding by looking for specific patterns and replying with preprogramed language [19], [22].

Although Eliza and Parry had a limited ability to communicate, they inspired programmers to improve chatbot technology [16]. In 1995, Alice, a chatbot inspired by Eliza, was the first chatbot to speak with correct inflection and grammar. In 2001, SmartChild became available through a messaging platform. In 2007, Apple Siri, Microsoft Cortana, and Amazon Alexa improved chatbot technology by using NLP, pattern recognition, and contextual analysis to make more intuitive leaps, perceptions, and judgements for better language recognition [16]. According to Scopus, chatbots have rapidly grown since 2016, and chatbot technology continues to grow [23]. Figure 5 presents a chart of the growth of research on chatbots.

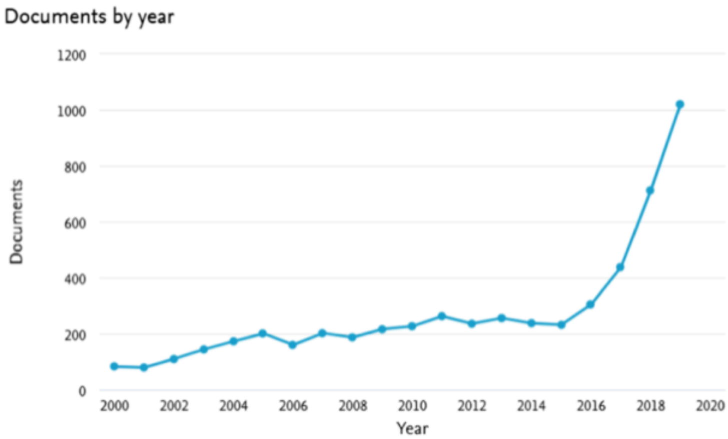


Figure 5. Number of documents on chatbot in Scopus from 2000 to 2020 [23].

2.2 Growth of Chatbots

The launch of chatbots in messaging platforms of the major technology firms Facebook and Microsoft in spring 2016 may have contributed to the rapid growth of chatbots because Facebook began encouraging companies to create chatbots for their fan pages to engage their fans and customers more closely [24]. In just two years, Forbes [25] reported that more than 300,000 active chatbots were created on Facebook Messenger.

According to Brandtzaeg and Følstad [24], The rapid acceptance of chatbots is due to their ease of use, speed, and convenience. Brandtzaeg and Følstad [24] further mentioned that companies have adopted chatbots to replace their customer service staff to reduce costs. They provided the example of KLM airlines, where KLM started to use a chatbot for basic services, such as check-ins, announcing delays, or issuing boarding passes. The chatbot managed to respond to 10% of the questions without a h ma age i e e i , which speeded up the response time by 20% [26] . Similar success stories of implementing chatbots have appeared with greater frequency. In 2020, even the World Health Organization (WHO) [27] launched a chatbot on WhatsApp, a popular messaging platform managed by Facebook, to respond to questions regarding the coronavirus disease 2019 (COVID-19). According to Brandtzaeg and Følstad [24] and Grand View Research [26], a marketing research firm, in addition to the speed of access and convenience, one of the major industry trends affecting why companies adopt chatbots is the popularity of AI and NLP in the industry in recent years, which is further explained below.

2.3 Rise of Artificial Intelligence: Natural Language Processing and Machine Learning

Natural language processing (NLP) i a ke ech l g ha e able he ba ic cha b , ch as Eliza and Parry, to make more intuitive leaps, perceptions, and judgements for better language recognition [16]. Specifically, NLP is a technology that explores how computers can understand human language and process information to perform tasks [29]. Using NLP, chatbots can process and understand the human language and determine how best to respond [30]. Machine Learning (ML) is another a key technology that enables chatbots to work smarter [30], with ML, chatbots can learn from human behaviour and execute actions similar to what humans would do, making the chatbot service more personalised for each user [30].

Both NLP and ML are subsets of AI, a technology that allows computers to mimic how the human mind perceives, learns, solves problems, and makes decisions [31]. Moreover, AI technologies can enhance chatbot features and enable them to be smarter [28], which can spur growth of the chatbot market from USD 17.17 billion in 2019 to a predicted USD 102.29 billion by 2025 [5].

2.4 Challenges of Designing a Chatbot

Despite the rapid advances in chatbot technology and supporting technologies, such as AI and NLP, which have led businesses to invest in them, chatbots have failed to meet user expectations [14]. Gartner [32] reported that 40% of the chatbots launched in 2018 on such platforms as Messenger and WhatsApp were abandoned by 2020. This might be due to the lack of design guidelines for the user interface and user experience [14]. Typical examples that Brandtzaeg and Følstad [14] described are whether a chatbot should have a text-based interface or a button-based interface and how best to motivate users to use chatbots, which remain unclear.

In addition to a lack of design guidelines on designing chatbots, researchers also found that users easily become frustrated with chatbots when they fail de a d he e language [33]. It is found that since most chatbots today rely on predefined questions and responses to interact with humans, the user experience of chatbots cannot be fully improved, as chatbots often provide unexpected responses when dealing with questions that are not predefined [34].

Lastly, Feldman [34] believes that the high expectation of chatbots from companies leads to a greater challenge for chatbots. Companies expect chatbots to perform any task rather than being limited to a particular task. Thus, companies quickly fail their chatbot development because chatbots cannot help customers perform all the tasks accurately.

2.5 Types of Chatbot Technology

Most chatbots follow the approach of simulating human language by following scripted rules and responding to users based on predefined responses [35], [36]. This approach is similar to what Eliza offered in 1966 and is known as the retrieval-based model.

Because of how it is designed, a retrieval-based chatbot relies heavily on message-response matching algorithms to respond to users based on the predefined set of rules [37], [38]. In other words, retrieval-based chatbots cannot handle cases where the message is unexpected or unstructured. However, because humans establish the predefined rules, retrieval-based chatbots can often accurately answer questions asked by humans [36], [39].

An alternative approach to the retrieval-based model is the generative-based model, which learns new sets of rules from the users and responds to the users input [39]. Generative-based chatbots are often smarter than retrieval-based chatbots, as AI and NLP aid the chatbot in understanding unstructured language [38]. They generate responses based on a wide range of training data from humans and simulate language similar to that a human would use in response. The main disadvantage is that it is difficult to train a generative-based chatbot because it normally requires a large set of training data [36], [39].

2.6 Types of Chatbot Interface

Chatbots can be categorised into three interface types: menu/button-based, keyword recognition-based, or text-based (contextual) interfaces [40]. Each of the types offers a different user experience and requires a different level of technology to support (See figure 6).

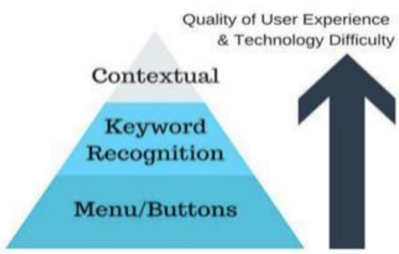


Figure 6. Chatbots can be categorised into contextual,
keyword recognition and menu/button [40].

2.6.1 Menu/Button-based Chatbots

Menu/button-based chatbots are the most popular type in the market today [40], [41]. For example, the chatbot service that Landbot [42] provides is a typical menu/button-based chatbot (Figure 7). Gupta [40] reported that a menu/button-based chatbot is often a retrieval-based model that responds to specific questions only. Instead of allowing users to type in text freely, a menu/button-based chatbot offers a limited number of menu or button options from which users can choose a list of designated messages.

The advantages of using a menu/button-based chatbot include that it can be built quickly and can handle simple questions efficiently [41]. Because all messages, including the questions users can ask, are prepared, it also guarantees that the user never becomes lost during the conversation. However, Wouters [41] also mentioned that the disadvantage of a menu/button-based chatbot is that users cannot ask complicated or advanced questions .

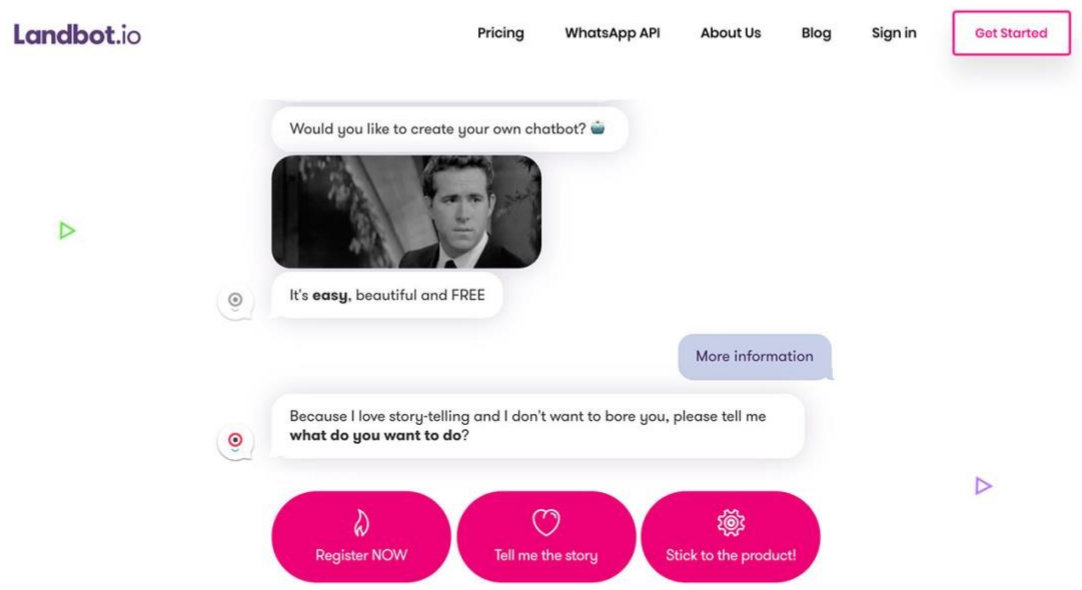


Figure 7. Landbot.io, a chatbot building platform [42].

2.6.2 Keyword Recognition-based Chatbots

Keyword recognition-based chatbots are an advanced version of the menu/button-based chatbots [40]. The chatbot that Facebook Messenger provides is a typical keyword recognition-based chatbot (Figure 8). These chatbots work with a limited AI to identify specific keywords. Based on the detected keywords, the chatbot determines which predefined message to present to respond to the user [40]. For example, if a user asks, What kind of food do you have in your restaurant? , the chatbot would likely recognise the keywords food a d restaurant a d provide a menu to the user.

Wouters [41] reported that the advantage of a keyword recognition-based chatbot is that users can type in messages and ask more advanced questions. However, because the chatbot is not connected to an AI or NLP system, the primary disadvantage arises from the primary advantage: because the user can type in messages freely, the chatbot may not be able to respond to the questions that the user asks, which could lead to user dissatisfaction [41].

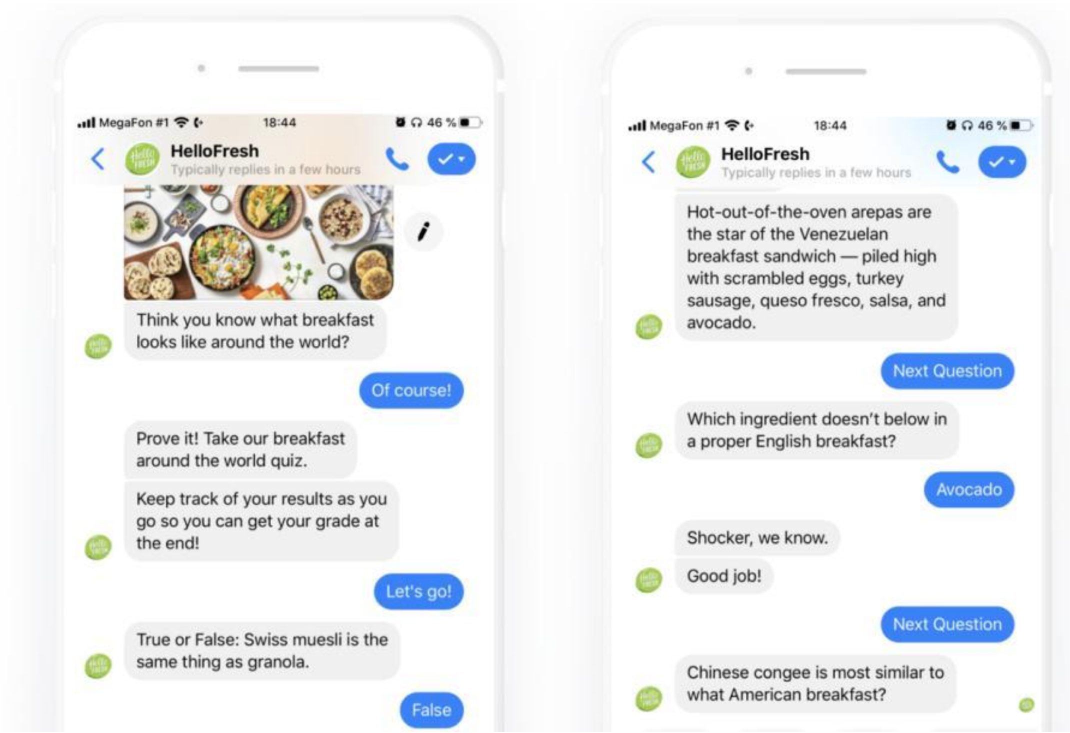


Figure 8. HelloFresh Chatbot on Facebook Messenger [43].

2.6.3 Text-based (Contextual-level) Chatbots

A text-based chatbot, or contextual-level chatbot, is the most advanced type of chatbot, which features a text-based interface that often works in conjunction with a pattern recognition system, such as NLP and text mining [44]. As Gupta [40] mentioned, text-based chatbots understand the user s language directly and remember conversations with a specific user to improve the experience. Figure 9 presents an example of a text-based chatbot from Capital Bank, enabling users to interact with it in natural language.

The advantage of a text-based chatbot is that users can communicate with the chatbot more naturally, similar to communicating with human agents [41]. Users are also able to ask advanced questions. However, building a text-based chatbot is more costly because of the need to integrate AI and ML [41]. Specifically, it is often difficult to build a text-based chatbot because it requires a large volume of user data to train the chatbot to handle user requests effectively.

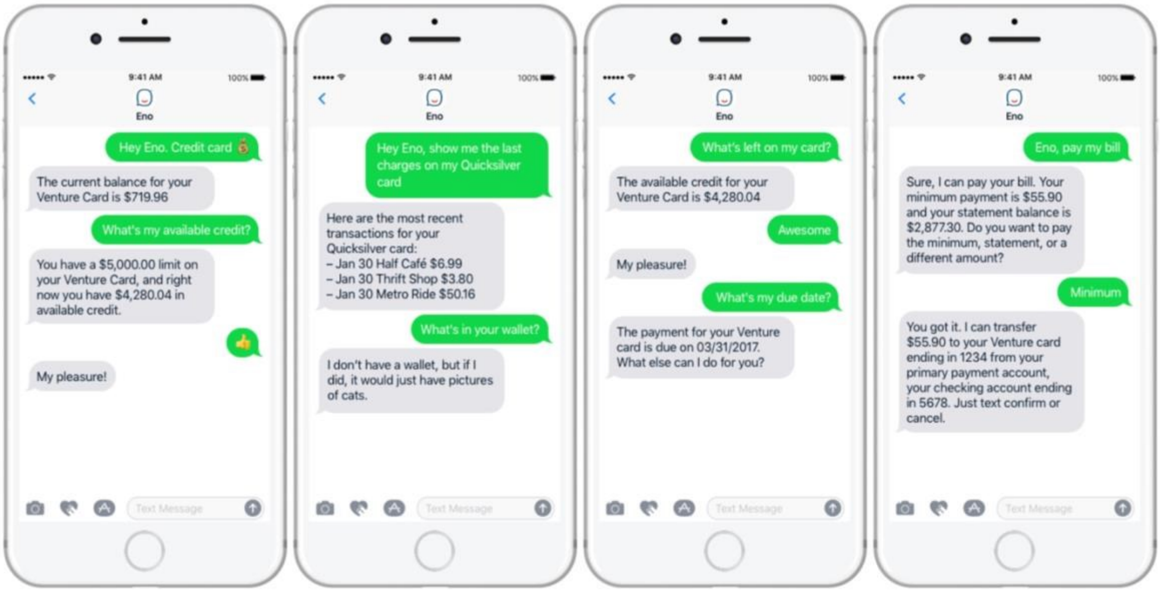


Figure 9. Cha b E f m Ca i al Ba k [45].

2.7 Overview of Scheduling Assistants

Scheduling a time to meet can often be time-consuming and requires much effort, especially when using email to arrange the meeting [46]. This section presents some background studies and existing competitors on the subject.

2.7.1 Artificial Intelligence Scheduling Assistant

Doodle [13], a meeting scheduling platform, conducted research in 2019 that included interviews of more than 6,000 employees in the US, Germany, Switzerland, and the UK. The research found that an employee spends two hours per week on average attending pointless meetings and five hours on average scheduling meetings [13]. An AI-based scheduling software, X.ai [47] also found that it takes an average of eight emails to schedule a single meeting. It causes professionals who attend many meetings, such as information technology workers, recruiters, and salespeople, to spend 25 hours per month scheduling meetings [47]. Even in Sweden, where this project is conducted, employees spend more than 13 hours on average waiting for a response from others to schedule a time to meet [13]. A similar study by Microsoft [48] found that scheduling meetings is time-consuming for business people.

2.7.2 Existing Solutions

A few AI-based scheduling assistants are already available on the market. One of the first companies that provided this service was X.ai. In 2014, it launched its AI assistant, Amy, to help its users arrange meetings with other people easily [47]. X.ai works by having the user send a meeting request email to all participants and a carbon copy to the AI assistant, Amy. Amy then sends an email to all participants to ask for available times and sends out the final meeting invitation to all participants after the AI finds a time that suits all participants.

An alternative service to X.ai, Julie Desk [49], provides a similar service and is known for having a strong AI behind the AI scheduling assistant, which allows the AI to learn the behaviour of the users, such as remembering their meeting time preferences, habits, and so on. It offers a similar interaction as X.ai, in which users send a carbon copy email to the AI assistant while scheduling a meeting with the participants, and the AI assistant sorts out the meeting request automatically.

Microsoft [46] launched a research project in 2015 to deal with scheduling meetings using its email program Outlook and virtual assistant Cortana. In 2017, Microsoft launched the product Calendar.Help that integrated scheduling services with its voice assistant Cortana to provide users with a way to arrange meetings and schedule tasks [46]. Calendar.Help offers the same interaction as X.ai and Julie Desk, which still requires the user to send a carbon copy email to Cortana. Cortana accesses the calendars of all parties to find an available time to meet. The difference is primarily that, because Cortana is also a voice agent, users can ask Cortana to schedule a meeting directly using voice [50].

In addition to AI scheduling assistants, platforms like Calendly [51] and Doodle [52] provide a more traditional service. Participants manually pick the time that best suits them on these platforms, and the meeting host uses the platform to visualise a ici a availability and select a time to hold the meeting. Moreover, calendar software, such as Google Calendar [53] and Microsoft Outlook Calendar [54], allows internal workers to view colleagues working calendars within the organisation and helps users check whether the colleague is available at a certain timeslot.

2.7.3 Challenges of Existing Solutions

Although these competitors provide a reliable scheduling method, many drawbacks exist to their service approach. The primary problems are their reliance on other platforms (i.e., email providers), the assistants are too human-like (i.e., people not knowing they are dealing with an AI), and users must manually pick available times.

Existing competitors, such as X.ai, Julie Desk, and Calendar.Help, all depend on email services, which means they cannot completely control the user experience because part of that experience depends on which email service the user has. Moreover, when a user switches to another email service, the user experience could change.

Business software review websites, G2.com [55] and Trustradius.com [56], indicates that X.ai virtual agent Amy is too human-like, which makes users believe the virtual agent is a live person. Feedback has included the following:

> People sometimes don t realise Amy isn t real [57].
>
> It would be useful for Amy to identify herself as a bot to avoid confusion. I had one client who thought I was in LA when I was really in Nashville but kept insisting to Amy (without cc ing me) on an in-person meeting [58].

Relying on email services compounds the problem, as people are not used to receiving emails from AI virtual agents, leading to confusion on certain occasions.

In all currently available solutions, meeting participants must still manually select times that suit them. The existing solutions provide the most benefits to the meeting host because the host does not have to check with all participants to find a common meeting time. However, participants must still manually check their calendars and select a suitable meeting time.

2.8 Summary and Design Opportunity

In this chapter, the background literature on chatbots and AI scheduling are discussed. The identified challenges indicate that designing a chatbot with a good user experience can be difficult because not much information or guidelines exist on designing a chatbot interface. Furthermore, based on exploring the literature on AI assistants and reviewing the relevant competitors in the market, most other AI scheduling assistants offer services through email only. Due to the fact that there are no existing solutions that offer an AI scheduling assistant as a mobile app, this thesis aims to fill this gap. In the remaining part of this thesis, different ideas on how an AI scheduling assistant chatbot app for business people could be designed are explored to help future researchers clarify design decisions when designing a chatbot mobile app for a business environment.

Chapter 3 describes the methods that were used for the two design iterations. Figure 10 provides a visual roadmap of the steps involved in both iterations. The first iteration focused on exploring different design opportunities that an AI scheduling assistant chatbot could present. The second iteration focused on redesigning the top-rated design preferred by the users in the first iteration to improve the usability.

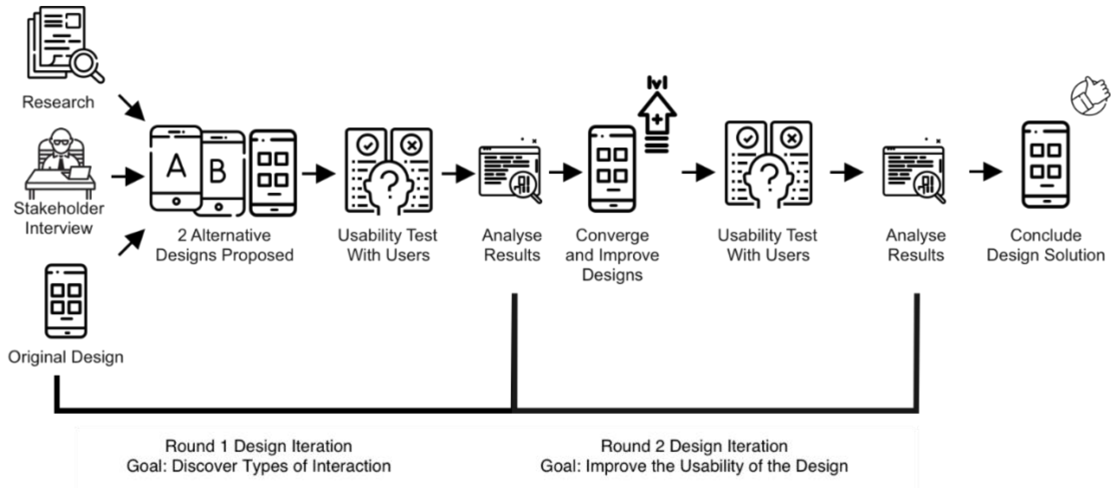


Figure 10. Roadmap of the design process.

3.1 First Design Iteration: Explore Design Opportunities

The methods used in this iteration include analysing how the current prototype is designed and what the stakeholder expects, studying background literature, exploring and implementing different design ideas, and evaluating the designs with the users. The evaluation process includes both the current prototype design and the newly proposed designs so that the results can be compared against each other.

3.1.1 Analysing the Current Prototype

The original prototype that the previous designer of the company designed provides hints about what functionalities and features the AI scheduling assistant chatbot should include. It also reveals how the chatbot could handle and process meeting requests. Thus, analysing how the current prototype works serves as the basis of the research and ideation process. The proposed new designs in later stages are also based on the requirements and user flow that the current prototype illustrates.

3.1.2 Exploring Stakeholder Expectations

As this is a company-based project, a meeting with the stakeholders (i.e., the CEO and CTO of the company) was important to understand their goals and expectations from this project. A report by the Interaction Design Foundation [59] made it clear that it is important to involve the stakeholders in the research process as soon as possible to ensure that the outcome of the design matches the akeh lde de i ed direction.

3.1.3 Background Literature and Competitor Review

The research results presented in Chapter 2 also serve as an important resource for the design process in later stages. The proposed designs strongly rely on the research to ensure the designs are relevant to what the users might experience in other kinds of mobile apps or chatbots.

3.1.4 Design Exploration and Implementation

The design exploration began by exploring the possible opportunities to design the AI scheduling assistant chatbot. The designs were inspired by analysing how the current prototype works, knowing the stakeholders expectations, and evaluating the studied background literature.

In this iteration, multiple design ideas were proposed and prototyped simultaneously to compare how different methods of designing the AI scheduling assistant chatbot app would affect how the users perceive and experience the meeting scheduling process. In addition to discovering which type of interaction and interface the users prefer, possible feature requests and user needs were expected to be revealed from the users after they tested different designs.

The design ideas were implemented directly on a smartphone to enable users to experience the proposed designs in the most realistic setting. This is achieved by using XCode 12 and Swift 5.2, a development environment and programming language for iOS applications. The implemented designs were also connected to RASA, the NLP recognition system that O.AI uses so that users can experience how the AI scheduling assistant would communicate with them when they send a request. Part of the implementation was done using Landbot.io, a tool for creating online chatbots, to accelerate the implementation process.

3.1.5 User Testing and Evaluation

The final step of the iteration was to test and evaluate the original prototype and proposed prototypes that represented different presentations of the AI scheduling assistant. The testing session included a usability testing session, an interview, and a System Usability Scale (SUS) questionnaire. The process of the testing sessions is further explained below.

3.1.5.1 Qualitative Measurement: Usability Testing and Interview

A summative in-lab moderated usability testing method was selected to understand the participants experience with the designs. The participants were asked to complete three assigned tasks related to scheduling meetings with each of the designs. The tasks are presented in Appendix A. An interview was also conducted along with the usability testing to seek for feedback and comments. The interview questions are presented in Appendix B.

The qualitative data analysis from the testing sessions was organised using an affinity diagram approach. An affinity diagram is a popular method for organising a large set of ideas into groups. In human-computer interaction, the affinity diagram has been used widely as a tool to categorise research results [60]. The data are grouped into cluster groups, and their importance is prioritised based on the frequency of occurrence for each participant.

3.1.5.2 Quantitative Measurement: Questionnaire and Task Time Completion Rate

The designs were also measured by having the participants in the testing sessions complete a System Usability Scale (SUS) questionnaire to learn how people perceive and experience each prototype. The SUS questionnaire is a simple way to assess usability using a 10-item scale [61]. It contains a mixture of extreme questions that prompt participants to choose from options on a scale from *disagree* to *agree* with different statements about design usability. The questions in the SUS questionnaire are shown in Appendix C.

The SUS questionnaire results can be measured using a few different measurement methods to compare each design prototypes. The results were measured using a Net Promoter Score (NPS) approach, acceptable approach, and adjective approach. Figure 11 summarises how these approaches measure the results from the SUS questionnaire. In addition to the SUS questionnaire, the time that the participant spent to complete each task in the usability testing sessions was also recorded to compare how each design affects the efficiency of completing the tasks.

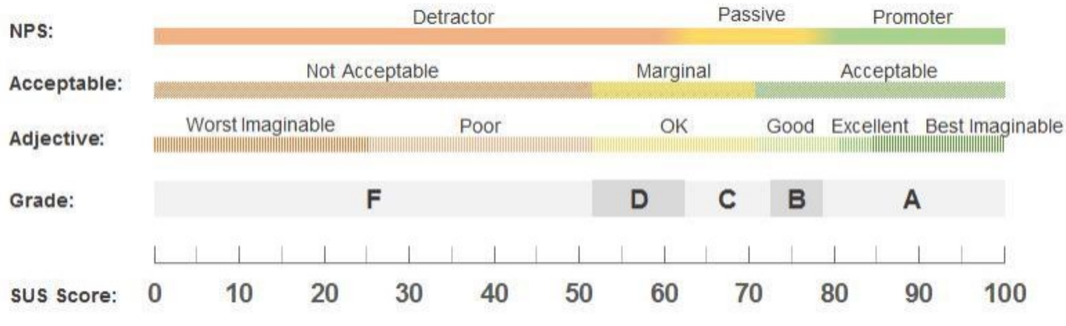


Figure 11. Different evaluation methods for system usability scale surveys [62].

3.1.5.3 Participants

Due to the COVID-19 pandemic, which happened in the middle of this thesis project, I returned to Taiwan while working on this thesis project. Thus, all participants in the study were Taiwanese and non-native English speakers. The participants were recruited through a Facebook group of professionals, such as designers, product managers, and engineers, in Taiwan. All people who were interested in participating in the study were asked to complete a screening survey (Appendix D), and the participants were invited based on the number of meetings they were involved in per week, how much time they spent on meetings, and how many of the meetings they initiated.

3.1.5.4 Procedure

After the participants were selected from the screening survey, the participants were invited to participate in an in-person testing session. Before each testing session, the participants were asked to complete and sign a consent form granting permission to record the interview process, take notes about them and include their photos in the thesis. They were also informed that the collected data would only be used for research purposes. The consent forms are presented in Appendix E.

At the beginning of each testing session, the participant was provided with a list of tasks to complete using the design prototypes, and the testing order for the different design prototypes was randomised. The participants were asked to complete the SUS questionnaire immediately after completing the task for each design prototype. Finally, after the participants completed all tasks for all the design prototypes and completed the SUS questionnaire, the participants were interviewed to share feedback regarding their experience using each of the design prototypes.

To support the research analysis, the voices, performed actions, and testing process were recorded during each testing session using a smartphone screen recorder (a built-in function) and by positioning a Go-Pro camera above the phone. In addition, a webcam was placed in front of the participants to record the testing process. Figures 12 and 13 provide details on the setting and the photographs from some sessions.

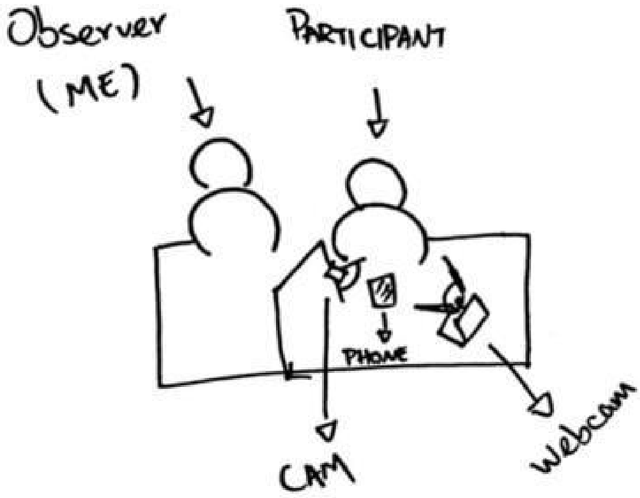


Figure 12. Sketch of the test setting.

Figure 13. Photographs from the test sessions.

3.2 Second Design Iteration: Improve the Top-rated Design

The method used in this iteration included redesigning the top-rated design in the first iteration by prioritising the feedback collected from the results in the first iteration and conducting another round of user testing to evaluate the newly proposed designs in this iteration.

3.2.1 Redefining Designs Based on First Iteration Feedback

The top-rated design prototype in the first iteration was selected as the main design to assess in the second iteration. The advantages and disadvantages identified from testing different designs in the first iteration were also gathered and prioritised to facilitate the process of improving the usability of the top-rated design prototype.

3.2.2 Design and Implementation

Several new design ideas were proposed in this iteration based on the organised feedback. Some further review of the literature and related products was also explored to improve the top-rated design. The implementation method remained similar to the first iteration, all the implementations were completed using XCode 12 and Swift 5 and were connected to RASA so that the participants in the testing sessions could experience how the app works in a realistic setting.

3.2.3 User Testing and Evaluation

The same approach used in the first iteration was repeated to test and evaluate the improved design. The participants were selected from the same Facebook group and prescreening survey, the testing process involved a usability testing session, interview, and SUS questionnaire. The tasks and settings in the testing sessions were also the same as those in the first iteration to make the results comparable with those of the first iteration. Finally, similar to the first iteration, the interview results were organised using an affinity diagram.

This chapter presents the result of the two design iterations. The first section presents the results of different designs explored and tested in the first iteration. The second section presents the improved design of the top-rated design in the first iteration and the evaluation result of it.

4.1 First Design Iteration Results

The first design iteration explored different ways of designing the AI scheduling assistant chatbot app. The exploration started by analysing the existing design, talking with the stakeholder, and analysing the literature. The different design ideas were then evaluated by conducting testing sessions with users to identify each newly explored design's advantages and disadvantages.

4.1.1 Analysing the Original Design Prototype

The current prototype provides users with an email-inspired interface in which users can type in a meeting request that contains all related information about the meeting, including the location, participants, and possible dates and times. Then, the AI assistant handles the request and searches for availability among all participants within the possible dates and times and presents the users with a confirmation containing all meeting details. Figure 14 illustrates the original prototype.

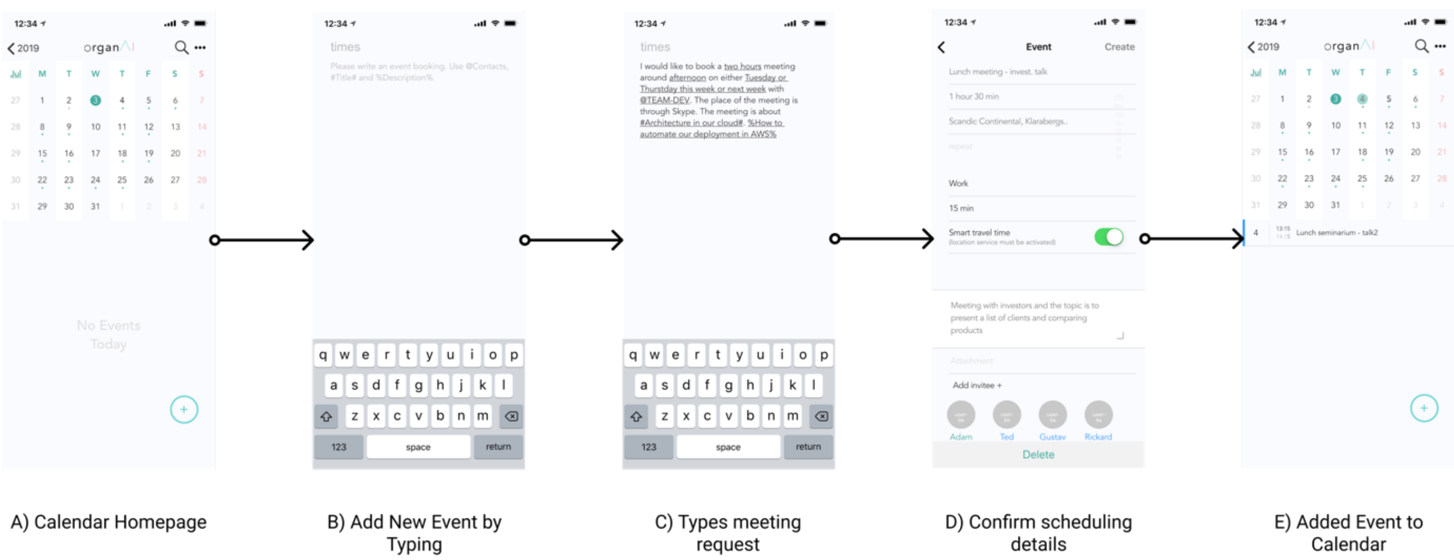


Figure 14. Original prototype.

The working flow is as follows. The first view the user saw when the app launched was a calendar screen (Figure 14.A), where the user finds all events on a particular date. An Add icon in the bottom-right corner enabled the user to trigger the AI assistant and add new meeting e e . O ce he Add ic i a ed, an email-like interface appears (Figure 14.B), which allows the user to type in messages. After the message request is typed (Figure 14.C), the user can send the request. The AI assistant then triggers NLP to understand what the user typed, processes the request, and shows the confirmation information for the meeting (Figure 14.D). The meeting is added he e cale da he he e confirms the details (Figure 14.E). The process is simplified in the user flow chart in Figure 15.

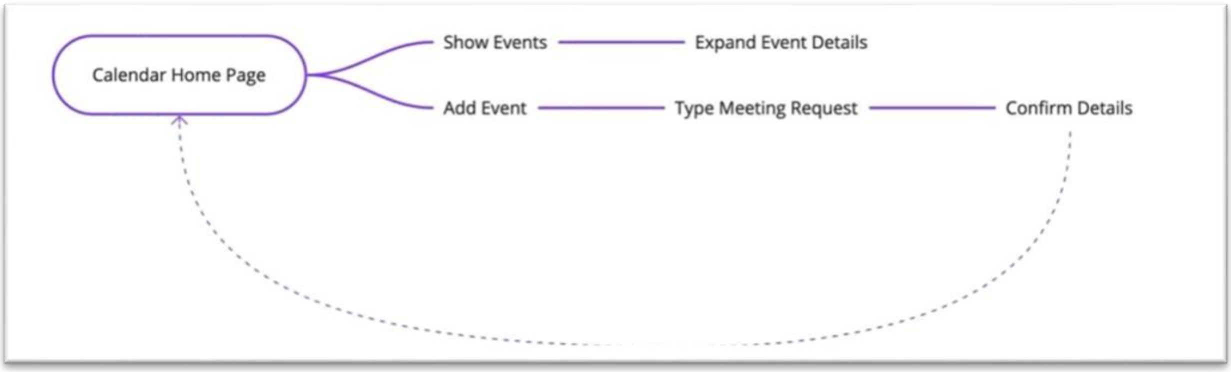


Figure 15. Simplified user flow chart.

4.1.2 Exploring Stakeholder Expectations

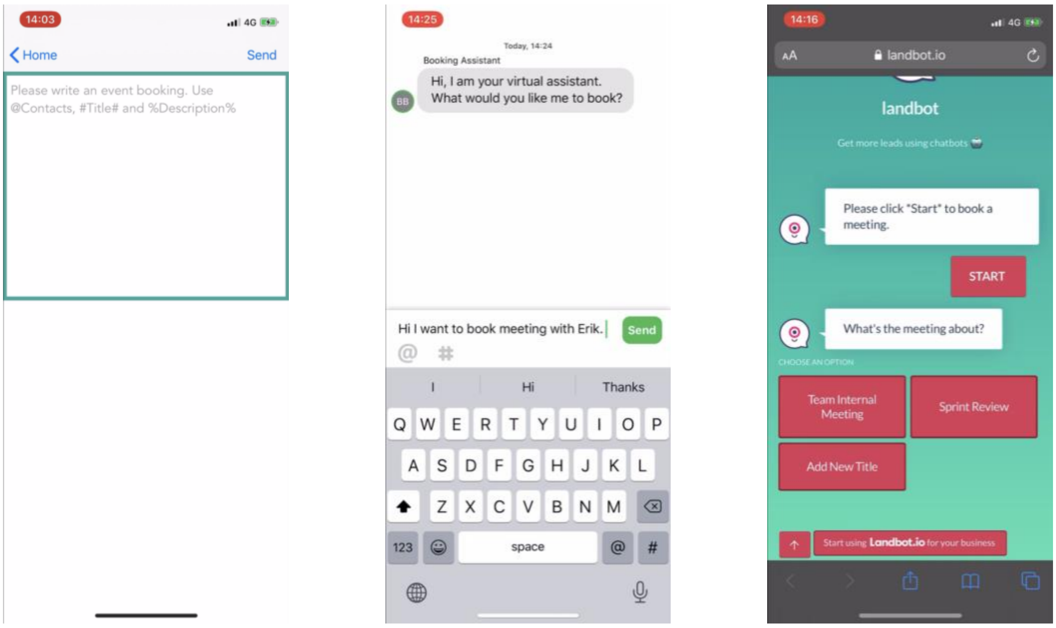
At the beginning of the project, an initial meeting with the stakeholders provided an overview of their expectations on the outcome of the project. That meeting developed the following goals:

1. The final deliverable should be a well-implemented app running on a smartphone and i eg a ed i h O.AI NLP on RASA.
2. The current design (existing design from previous designer) should be tested and evaluated to learn its advantages and disadvantages.
3. As the current design offers a non-conversational interface, the stakeholder is willing to turn the design into a conversational interface so that the users can communicate with the AI assistant and be more engaged.
4. While evaluating the existing design and newly proposed designs, it is possible to implement new features or remove existing features as long as the usability of the final design can be improved.

4.1.3 Design Exploration and Implementation Results

The proposed design prototypes consisted of three different designs representing different ways to interact with the AI assistant. Besides the original design prototype, a button-based and text-based chatbots inspired by the theoretical background research were designed to evaluate different interaction methods with the chatbot. Both methods of designing the chatbot enable users to complete the scheduling process while communicating with the AI assistant, which aligns with the stakeholders expectations.

Figure 16 presents the main screen of each of the three designs. Each design contains a user flow that enables users to experience the flow of triggering the AI assistant to book meetings to adding a meeting to the calendar to make the designs comparable with the original design. The detailed designs and interactions are presented separately below.



Original Design Text-Based Design Button-Based Design

Figure 16. Main screen of the prototypes.

4.1.3.1 Original Design Prototype

Figure 17 presents the results after the original prototype was implemented on an actual smartphone. The result appears slightly different from the design in Figure 12 as the c fi ma i age i e laced i h iPh e defa l cale da c fi ma i age h e he implementation process, but the entire flow and logic is the same as what the original design shows. This prototype is a non-conversational interface but contains a large rectangular box where users can type in any request and send it at once. Depending on how the NLP processes the request, it automatically creates an event creation page that contains the information the user typed. The user can then enter the missing information or correct any misunderstood information on this screen and add the event to their calendar.

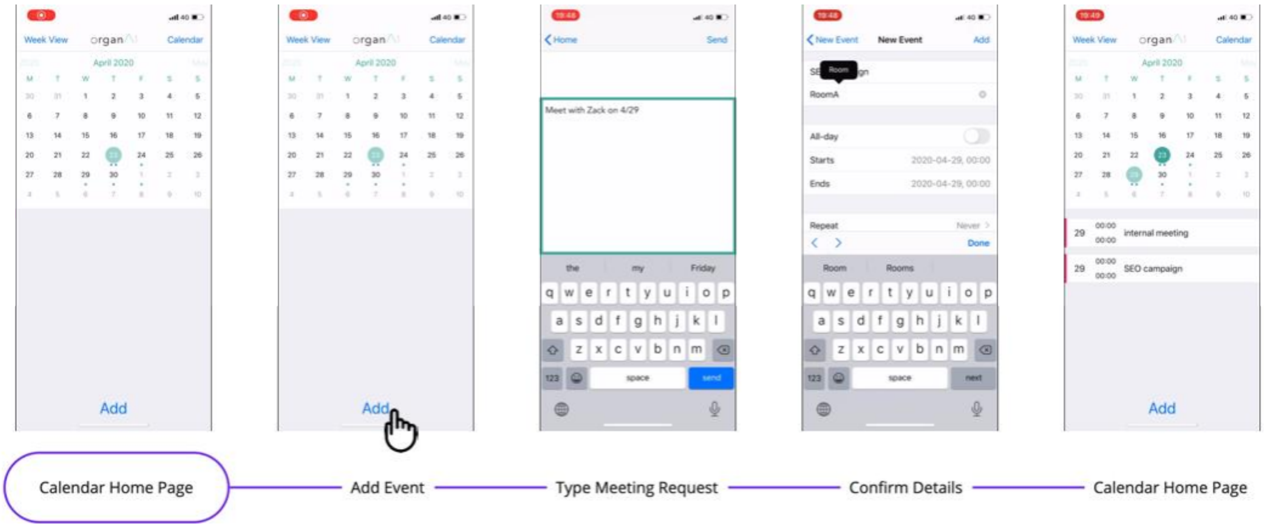


Figure 17. Original prototype.

4.1.3.2 Text-based Design Prototype

The first alternative to the original prototype, a text-based prototype, was proposed to offer a different type of interface and interaction. With the text-based prototype, the user could communicate with the AI assistant through a chatting interface. Figure 18 illustrates the design.

The AI assistant starts with a prompt in the conversation, asking users to type in a meeting request. The e me age m i cl de he title, time/date, location, and participants to handle a request. If the user misses one of these inputs in the message, the AI assistant will continue to ask the user to type in the missing information so that all the required information can be handled. In practice, if the user types a full message that contains all required information, such as Book a meeting with Zack on SEO meeting at 4 p.m. tomorrow at the office , the request is handled directly. However, if the user types in a partial message, such as Book a meeting with Zack , the AI assistant asks, Wha he mee i g ab ? . After the user answers, the AI assistant continues to ask for time, date, and location to handle the required information.

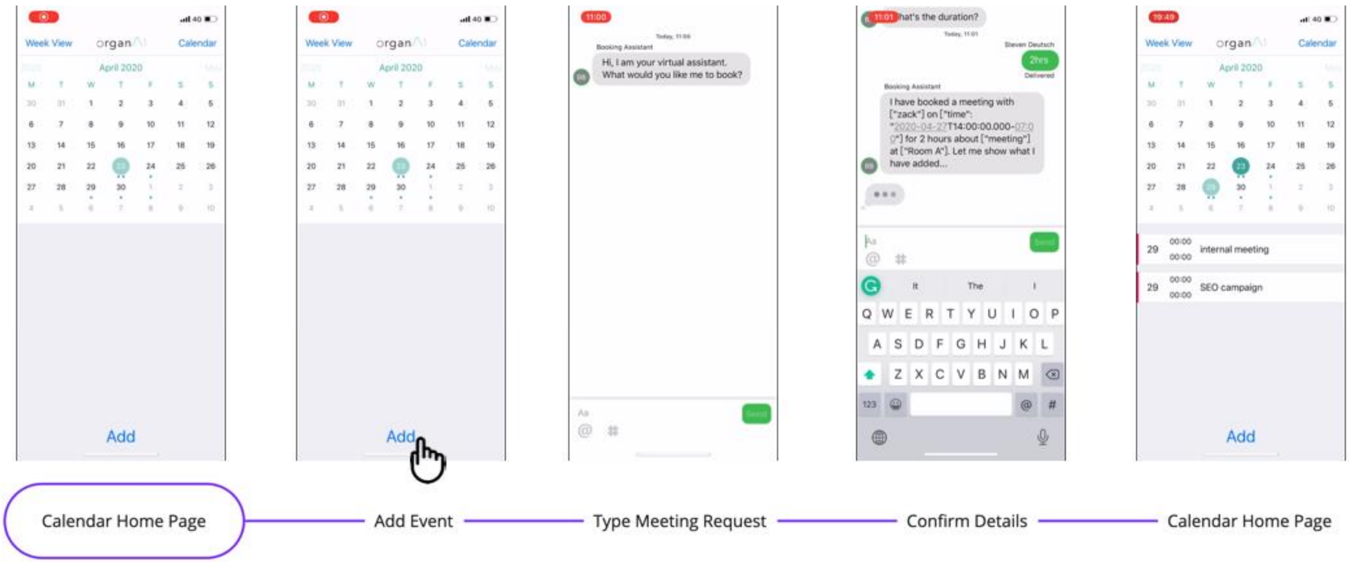


Figure 18. Text-based prototype.

4.1.3.3 Button-based Design Prototype

The second alternative to the original design was the button-based design. In this design, the AI assistant asks question about the meeting and the users were asked to select the appropriate buttons from a list of options (or add a new option manually). The AI assistant continues to ask questions until all the required information about the meeting is received. The questions and available action options are presented in Figure 19.

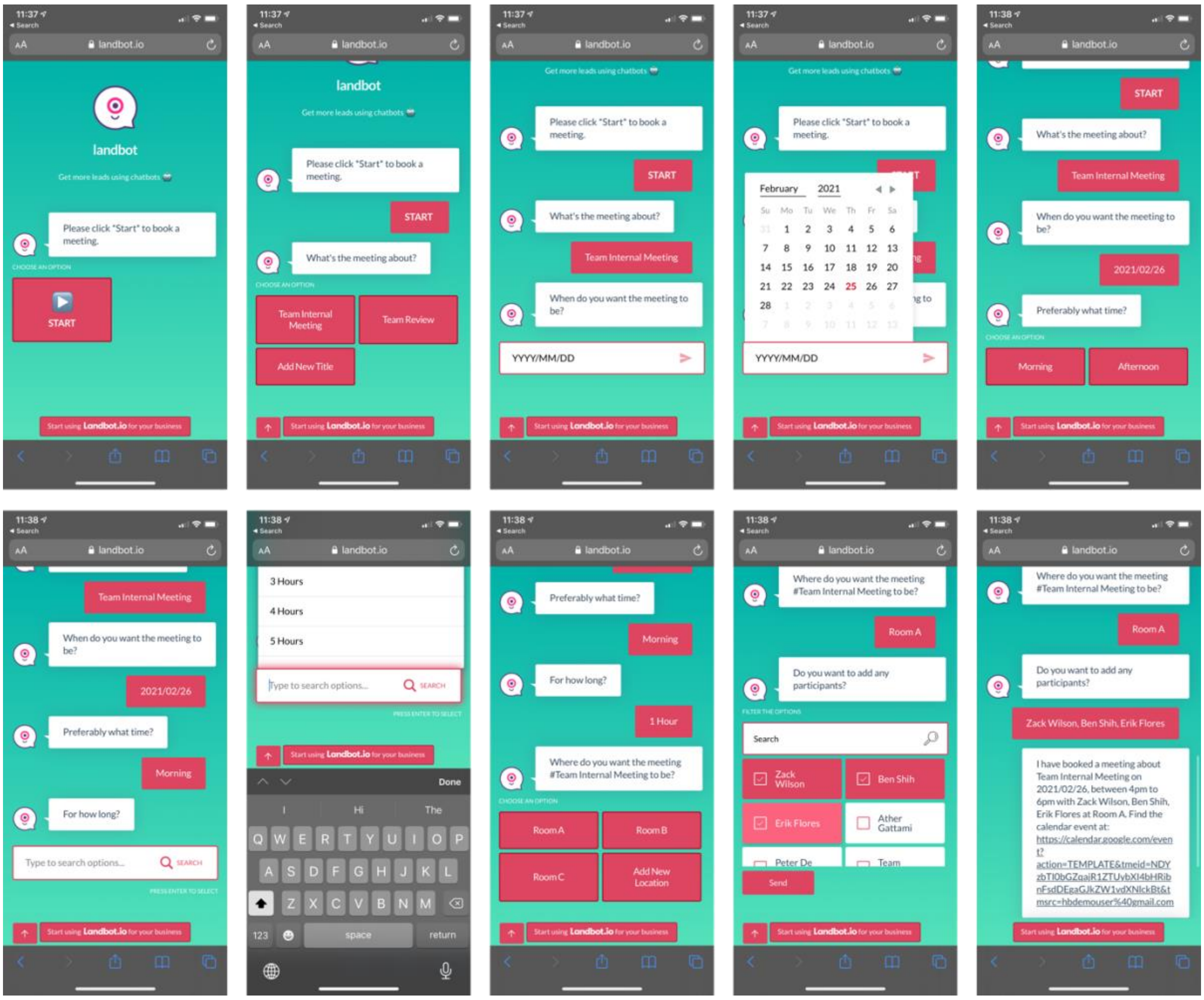


Figure 19. Complete flow of the button-based prototype.

The button-based prototype was the only prototype implemented with the online building software, Landbot.io, as it was much faster to build it using Landbot.io than using other development tools. However, due to the software limitations, the chatbot is not connected to the NLP service. Thus, all the responses and messages that the AI assistant sends are based on pre-set phrases. The prototype was exported as a mobile app that allows users to interact with it using smartphones to simulate an experience similar to other designs. The prototype pre-assumed that all the shown dates in the calendar are available to book. Thus, once the person booked the timeslot, the meeting is confirmed immediately. Figure 20 illustrates the flow of the button-based prototype.

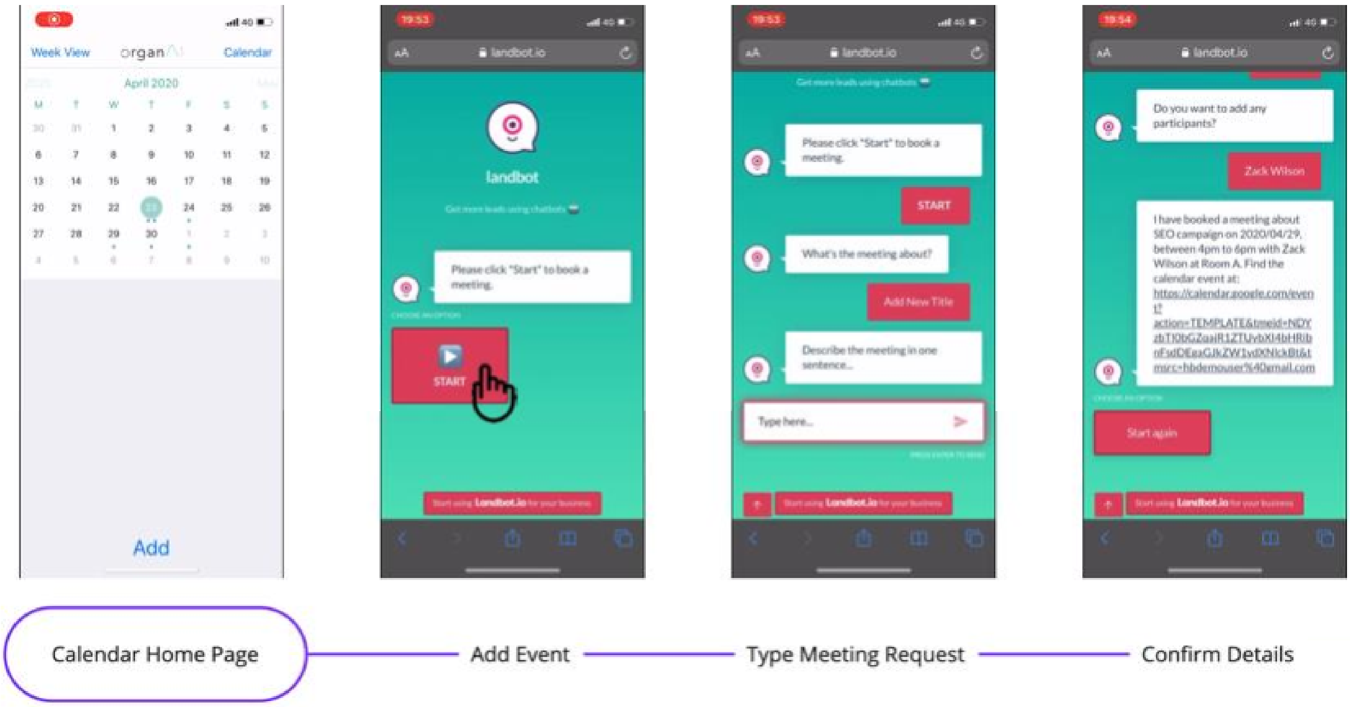


Figure 20. Button-based prototype.

4.1.4 User Testing and Evaluation Results

Seven participants evaluated the prototypes. The participants ages were between 20 and 30 years old. The participants were involved in 3 to 10 meetings per week, initiated around 3 to 5 meetings per week and spent around 3 to 5 hours in meetings per week. The order that each of the participants tested the three prototypes was randomised using a random number generator.

4.1.4.1 Quantitative Data Results: Questionnaire and Task Time Completion Rate

The SUS questionnaire results show that the text-based prototype scored the highest, with a SUS score of 80.0 and a learnability score of 1.4 out of 5.0 (the closer to 1.0, the better). The text-based prototype had a SUS grade of A . The button-based prototype was the second most favoured design, with a SUS score of 68.2, a learnability score of 2.0, and a SUS g ade f C . The least favoured design was the original prototype, with a SUS score of 57.8, a learnability score of 2.1, a d a SUS g ade f D . Table 1 presents an overview of the scores. The results are also presented in the spectrum diagram in Figure 21, which presents a different interpretation of the SUS questionnaire. The individual SUS results of the designs are presented in Appendix E.

Table 1. Results from the first-round survey

Prototype	Mean Score	Median Score	Standard Variation	Learnability
Original Prototype	57.8	50	16.5	2.1
Text-Based Prototype	80	80	5.4	1.4
Button-Based Prototype	68	82.5	29.0	2.0

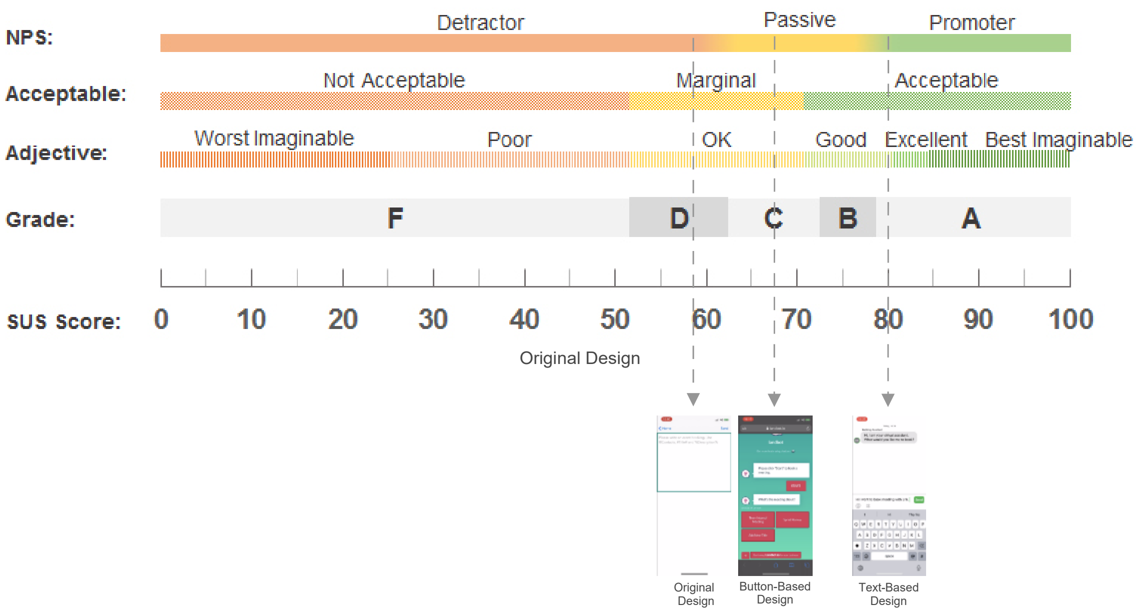


Figure 21. Spectrum diagram of the survey results.

In addition to the SUS survey results, the time that the participants spent on completing each task using different prototypes was also recorded. The timer starts when the user initiated the AI assistant and stops when the AI assistant shows the confirmed meeting details. Table 2 shows the average completion time for each task. The result shows that the text-based design allows the user to complete all three tasks most efficiently, with the button-based ranking as the second.

Table 2. Average completion time for each task with different prototypes (in seconds)

	Original	Text-Based	Button-Based
Task 1	152.3 s	93.3 s	106.3 s
Task 2	162.0 s	81.2 s	98.0 s
Task 3	106.0 s	105.4 s	107.9 s

4.1.4.2 Qualitative Data Results: Usability Testing and Interview

The following sections present the organised results of the affinity diagram for each prototype. The full affinity diagrams of the results from each prototype are in Appendix F.

Results from Testing the Original Prototype

(A) Lack of Instructions

When the users started to use the app, they could not understand what kind of message they should send to the AI assistant because the instructions were unclear. The current instructions are presented in Figure 22. In addition, from observing how the users interacted with the prototype, they could not recheck the instructions after they started because the instructions disappeared when the user started typing. As a result, first-time users spent quite a bit of time thinking about and typing messages they thought were suitable. For example, instead of typing Book a meeting with Erik , they typed, Mee E ik A   il 28, 2 m .

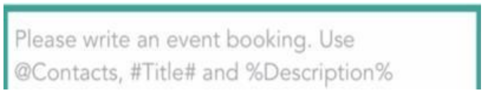


Figure 22. Current instructions.

(B) Natural Language Processing Detection Issues

In addition to lacking instructions, the current NLP algorithm also lacked training data, causing the text recognition system (i.e., NLP) to identify text incorrectly. In the case of the message Hi, please schedule a meeting for Erik and me at his convenience, thanks , the NLP might only recognise Erik as the person s name and omit all other information, causing the user confusion and requiring the user to retype the information in the confirmation page.

(C) Lack of Feedback from the Artificial Intelligence Assistant

Currently, the AI assistant does not provide any feedback when the NLP successfully identifies a keyword. The only feedback the user receives is when they send the request and the app asks them to confirm the meeting details, with the identified keywords automatically filled in for the meeting details. However, this lack of feedback confuses users because when the NLP is not working perfectly, the confirmation page often causes confusion. For example, when a user types, I d like to have a chat with Erik on the SEO campaign , the NLP might recognise chat as the meeting title instead of SEO campaign (Figure 23).

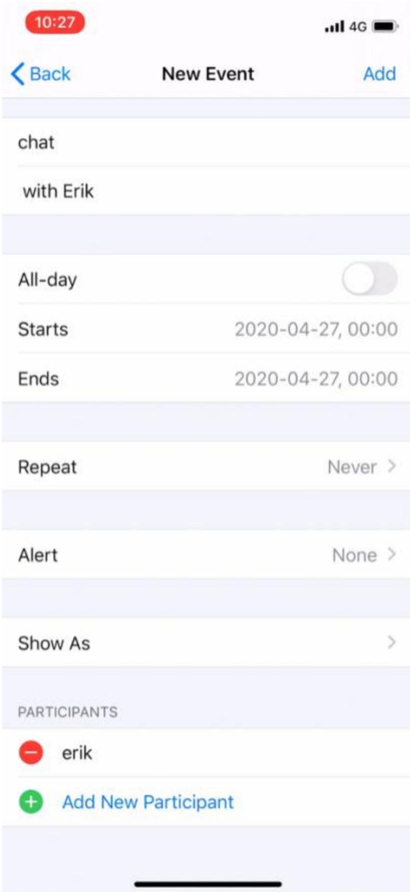


Figure 23. Example of misunderstanding.

(D) Interaction Similar to Sending Email

Users commented that the process was similar to sending an email to a secretary to schedule meetings. Because of this similarity, although the instructions were not given clearly initially, the users were able to learn quickly from past experiences on how to send the meeting request to the AI assistant.

Results from Testing the Text-Based Prototype

(A) NLP Detection Issues

From our observations, participants tended to use abbreviations when typing, such as Wanna f a , I ll f I ill , a d hr f h . Simila the limitations of the original prototype, the NLP algorithm cannot always handle these abbreviations. This limitation created some dissatisfaction among the participants because they had to type the message again using a longer sentence.

(B) Typing Process Takes Too Much Time

Following the lack of instructions previously mentioned, the users spent considerable time typing and tended to send one piece of the message when interacting with the AI assistant, although no instructions guided them to type it this way. For example, instead of typing Book a meeting with Erik at 4 p.m. tomorrow , participants typed Book a meeting , waited for the AI assistant to ask for more details, and answered the questions one by one. The users mentioned that this type of interaction requires too much typing, and they might not always have that much time to type when working on other things.

(C) Calendar Is Not Visible

As the conversation between the user and AI assistant happens in the chatting interface, three users complained that the calendar was not visible while interacting with the AI assistant. Thus, when asked to provide the meeting date or date range, they had no idea how to respond without seeing the calendar. The users also mentioned that they would like to check their calendar to confirm the available date visually.

(D) User Flow Is Intuitive

Four users responded that they preferred this prototype because the interaction with the AI assistant was much more intuitive and human-like than the other prototypes. They also liked that the AI assistant asked questions only when they did not or forgot to enter certain information. As the AI assistant asks for information piece by piece, the users also did not have to think about what type of information they should type in at once and could follow the AI a i a i c i to complete the book meeting request.

(E) AI Assistant Needs More Personality

Two of the users preferred to have an AI assistant with more personality because they claimed that having a personality would make her feel much more like she was interacting with a human assistant and would allow her to trust the AI assistant more.

Results from Testing the Button-based Prototype

(A) Process Is Too Rigid, Takes Time to Complete

Because the users had to follow a specific flow to confirm the meeting details, it was difficult for them to change previously selected options. If the users selected the wrong option, they had to restart the entire flow to correct it. In addition, the users had to follow the same flow every time they wanted to schedule an appointment, and they complained that the process was too rigid.

(B) Difficult to Add New Items

Another issue was that the button options were displayed as a default option for the user; thus, extra steps were required if they wanted to create new options. For example, if they wanted to add more participants who were not in the default options, they either had to scroll through all the contacts or enter a new participant name and details manually.

(C) Less Effort on Typing

Six of the users commented that the button-based design allowed them to type less, which allowed them to complete the tasks faster because they did not have to type anything. They liked that the design provided them with default options and that they could choose from the list directly.

4.2 Second Design Iteration Result

In the second design iteration, the top-rated design (i.e., the text-based design) was selected to improve usability. The improvement process was conducted by exploring new design ideas that solve the identified problem in the first iteration.

4.2.1 Redefining Designs Based on First Iteration Feedback

The improvements were categorised into the chatting experience and learning experience of the chatbot. The classification was based on the results identified in the first iteration.

4.2.1.1 Chatting experience

The chatting experience refers to the entire experience of the user with the chatbot, including how and what they type and how the AI responds to the user. In the first iteration, I learned from the original design and the text-based design that the users do not like to type too much. One of the users also mentioned in the first iteration that they did not have time to type that many messages when they were working and hoped that there could be some default messages or options that they could tap to save time. From interviewing the users, I also learned that the users perceived the AI assistant as a robot and did not expect the AI assistant to understand their language initially. A participant mentioned that the AI a i a lack of personality made her feel insecure and did not make her feel engaged.

4.2.1.2 Learning experience

From the first iteration, many participants expressed a desire to have more instructions to guide them on interacting with the AI assistant. The users also had problems when they used the app for the first time because they were unsure of what they could type and how much the AI assistant understood human language.

4.2.2 Design and Implementation Results

The design aims to fix the problems identified in the first iteration. Figure 24 illustrates the new design. The details of the new design are explained below.

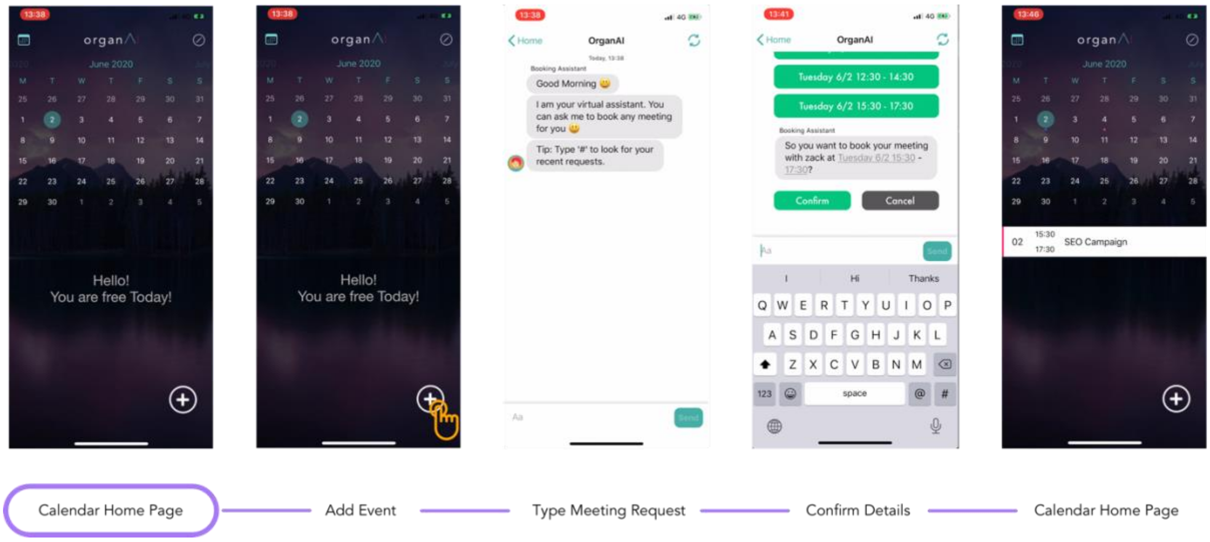


Figure 24. New design.

4.2.2.1 Improve Chatting Experience

(A) Integrated Button-Based Interaction into Text-Based Prototype

From the testing sessions in the first iteration, I learned that users preferred the text-based prototype but found that the button-based prototype saved time because they could simply select different options and tap buttons to schedule a meeting. Moreover, as the button-based prototype allowed them to choose from a list of default options, it also decreased the chance the user would type a wrong request (e.g., requests that the NLP could not recognise) and increased its usability. Hu [37] found that users perceived menu-based chatbots to be easier to use because they do not have to spend time thinking about what to type. As a result, the second design included more buttons and menu options wherever possible for interaction with the AI assistant. Figure 25 presents the new design.

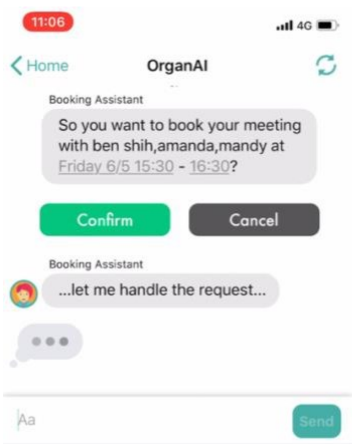


Figure 25. Adding button interactions to the text-based prototype,

(B) Added Auto-complete Feature to Minimise Typing Effort

In addition to implementing buttons and menu options in the text-based flow, I have also tried to minimise the effort required from users by implementing a feature called auto-complete, depicted in Figure 26. The auto-complete feature was implemented using a programming library called Me ageI Ba , to fasten the development process. The auto-complete feature is triggered every time the users type in a keyword that matches the phrases saved in the app. The phrases were collected from the words and sentences the participants typed during the testing sessions in the first iteration.

In the literature studies, I have found that using auto-complete (i.e., auto-suggestion) could reduce the time required to complete tasks by approximately 10% [63]. A few years ago, G gle Gmail also implemented a similar function called smart-compose , illustrated in Figure 27. Google found that this function allowed a user to type and respond to email more efficiently [64]. Implementing this feature should reduce the time users spend typing while maintaining the flexibility of typing freely.

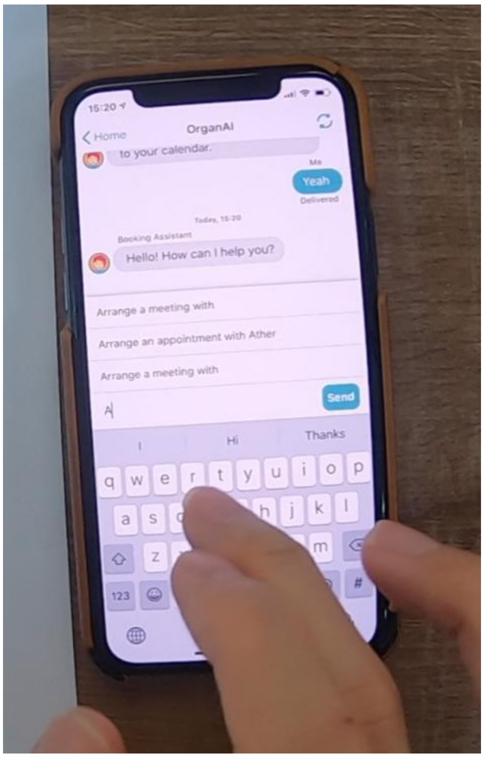


Figure 26. Auto-complete feature.

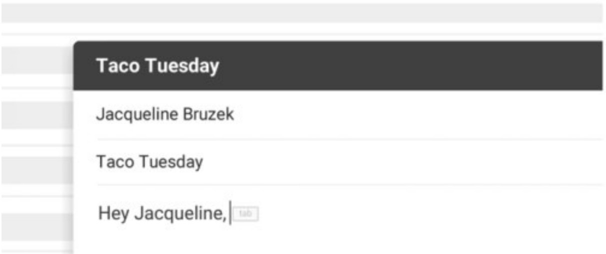


Figure 27. Gmail mart-compose function (Google, 2018).

(C) Added Personality Traits and Profile Image to the Avatar to Seem More Human

During the first iteration, one user reported that they would trust and prefer an AI assistant that was more human-like. As a result, this version added some personality traits, a profile image, and a name to the AI assistant and included small delays with typing indicators to give users the impression that the AI assistant was thinking about its response. The changes are presented in Figure 28.

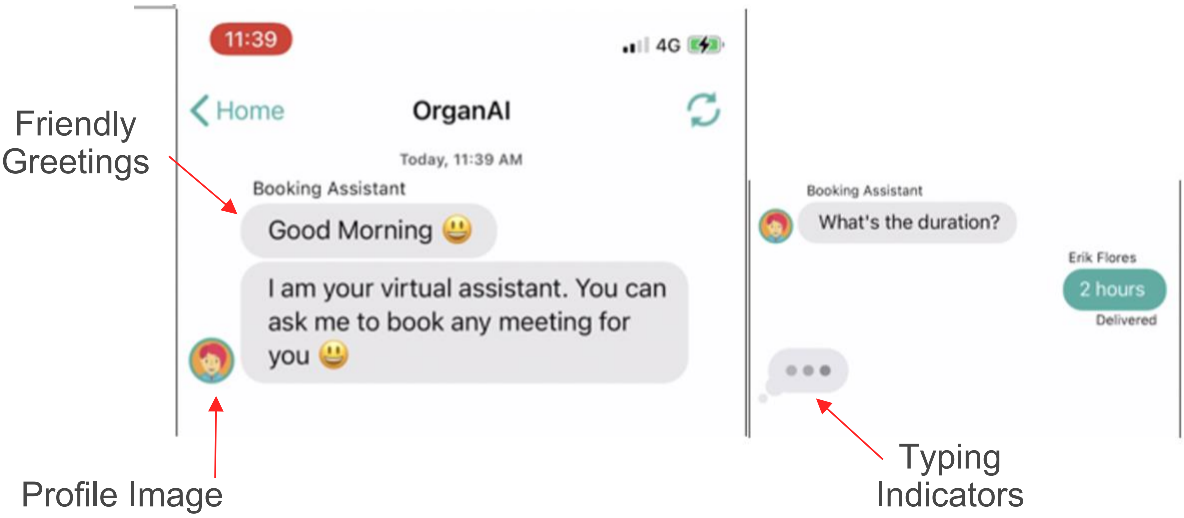


Figure 28. Improved artificial intelligence appearance.

4.2.2.2 Improve Learning Experience: More Instructions and Guidance Added

From the first iteration, I found that users were having problems interacting with the AI assistant because only a few instructions were provided. In this design version, more instructions were added to the beginning of the conversation. The instructions might not be perfect for guiding them on communicating with the AI assistant, but they provide users with some hints to start and understand what the AI assistant can achieve for them. Figure 29 presents an example of these changes.

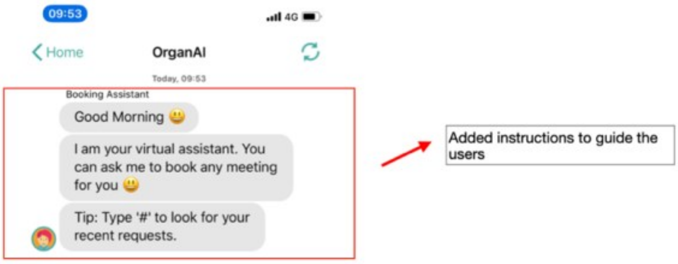


Figure 29. Added some guidance and hints to the instructions.

4.2.3 Test and Evaluation

Nine participants evaluated the prototype. The participants ages were between 20 and 40 years old. The participants were involved in around 5 to 10 meetings per week, initiated around 3 to 10 meetings per week and spent around 3 to 10 hours in meetings per week.

The results confirmed that the auto-complete feature helped users type requests faster, and the button options also helped reduce the typing effort. More results from the testing sessions are presented below.

4.2.3.1 Quantitative Results

All nine participants completed a SUS questionnaire immediately after they completed the tasks during the usability testing session. Table 3 reports the results. The mean of the SUS score was 75.0 (Grade: B), with a standard deviation of 20.5. The spectrum diagram in Figure 30 indicates that the SUS score was slightly lower than the text-based prototype tested in the first iteration. However, the median total of the SUS score was 82.5 (Grade: A), which is higher than all the prototypes in the first iteration. The full results of the SUS questionnaire are in Appendix H.

Table 3. System usability scale survey results for the new prototype

Mean Score	Median Score	Standard Variation	Learnability
75.0	82.5	20.5	1.7

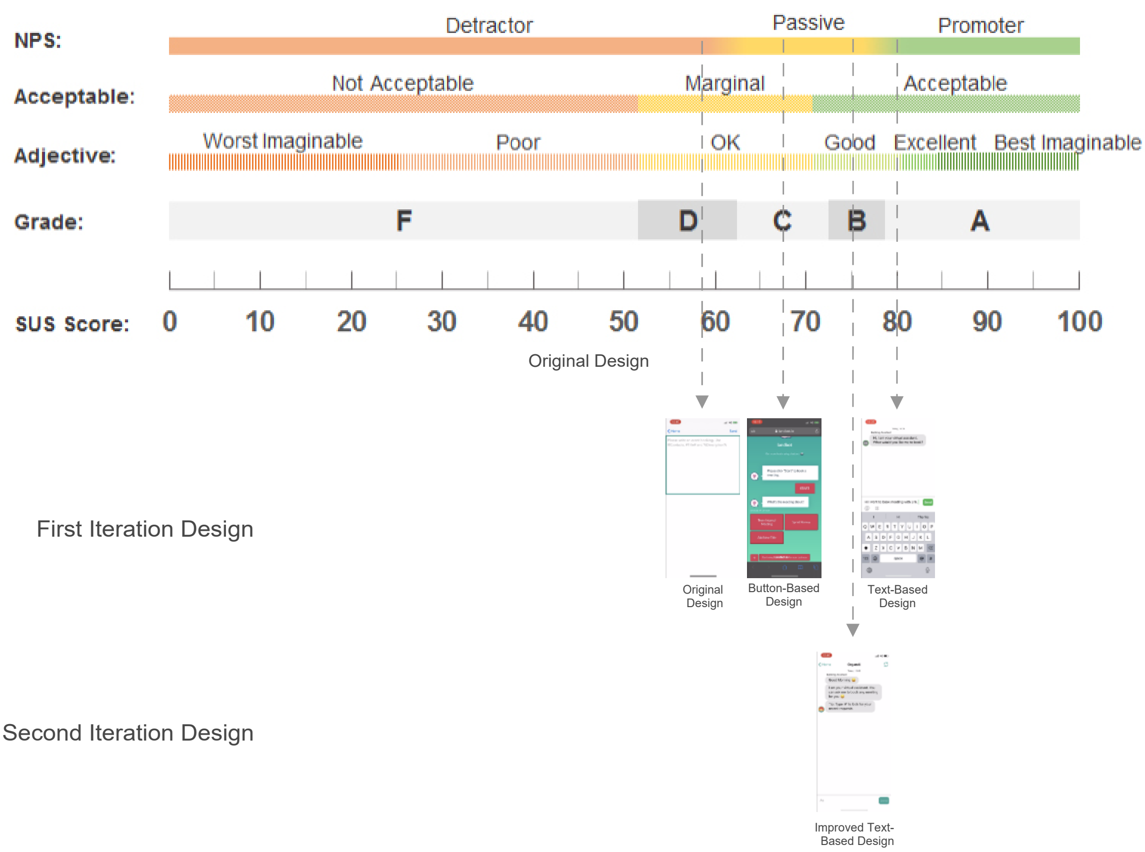


Figure 30. Spectrum diagram of the results.

4.2.3.2 Qualitative Data

Generally, the feedback for this version was much better than for the previous testing round. The auto-complete feature was the most notable function that the participants liked. The user feedback i cl ded The auto-complete feature gave me a hint on how to use the chatbot correctly , I will always tap on the suggested words when possible , The auto-complete suggestion saves a lot of time , and similar feedback from eight of the nine participants. Users also mentioned that the new design not only kept the openness of allowing them to type any request they wanted but also had the auto-complete feature to help them understand how to use the chatbot and send requests. H e e , he de ig f i g a ha h ag # a a igge l k up recent messages from the user caused confusion and should be further evaluated. In addition, adding a few instructions at the beginning of the conversation helped the users know what kind of messages they should send to the AI assistant. Less confusion and misunderstanding occurred in this iteration of the study than in the last iteration.

4.2.3.3 Main Findings from the Affinity Diagram

Through organising the qualitative data with the affinity diagram approach, six findings have been discovered. The full affinity diagram is presented in Appendix I.

(A) Auto-complete can reduce the effort of typing.

Seven of the participants believed that the auto-complete feature helped them send out requests faster. It also helped the participants learn how to use the app faster because the auto-complete feature suggested what to type when the user started typing.

(B) More button/menu options can be added.

Six of the nine users believed that some questions could be answered by simply tapping. For example, when asked for the duration or location of a meeting, the AI assistant could provide them with a list of actions so that they could choose their intended option from a list, rather than using the auto-complete feature to type in the whole word. They believed that, in some cases, the options in the auto-complete lists should be shown directly.

(C) The interaction with the AI assistant was intuitive.

Five of the participants found that the interaction with the AI assistant was smooth and intuitive. The participants were able to interact naturally with the AI assistant and perform the intended actions intuitively.

(D) The use of the hashtag to trigger auto-complete caused problems.

Part of the instructions, in the beginning, indicated that the users should type a hashtag in the textbox to search for recent requests. Six users wondered why they needed to type a hashtag, as it was not very intuitive. Three users even thought they had to type a hashtag every time they entered a new query.

(E) The personality that the AI assistant has made the user feels more engaged.

Two participants commented that they liked the AI a i a e ali (profile image, name, and tone). They also recognised a response delay from the AI assistant when the participant sent a message because the AI assistant displayed a typing indicator during these delays. The participants believed that these features made them feel more trusting and engaged when interacting with the AI assistant, as they felt more like they were interacting with a human assistant.

(F) Extra functionalities are needed.

All nine users asked for more functions to fit their work styles. They asked for functions, such as the ability to add notes, an agenda, and extra information. Additionally, six of the users me i ed ha he c e c fi ma i me age Y mee i g ha bee b ked did provide them with enough assurance. They wanted to know whether the meeting invitation was sent correctly and received by all participants. Two of the users thought that the AI assistant would be more useful if it could understand voice commands. The participants mentioned that typing messages takes considerable time, and they would prefer an interactive voice bot to use while working on other tasks.

4.3 Summary

In the second design iteration, a new design was proposed based on the findings from the first design iteration. The testing sessions indicated the design direction for an AI scheduling assistant chatbot can be a text-based interface but should include some button interaction or auto-complete (auto-suggestion) features to minimise the typing time. In Chapter 5, the findings from the two iterations are presented and discussed.

Chapter 5. Discussion

This chapter discusses the research question and presents the conclusions drawn from the research process. Because the project is not a typical academic research project, this chapter also discusses some limitations that affected the research process. Finally, future directions for researchers designing and developing an AI assistant are discussed at the end of the thesis.

5.1 Answering the Research Question

This thesis focused on investigating the research question 'What type of chatbot is most suitable as an AI scheduling assistant in a business environment? . The stakeholders formed two hypotheses before conducting the research. H1: Users prefer a text-based chatbot because this interaction is more natural and easier to understand. H2: A button-based interface allows users to learn more efficiently than a text-based interface.

Based on the results from the two iterations, H1 is true. From qualitative (usability testing and interviews) and quantitative (questionnaires and task time completion rate) analyses, the text-based prototype was the preferred method of interaction because it often supports more flexible interaction, allowing the users to type in messages in any format they preferred. However, the drawbacks are that users might have to spend time typing messages and that the NLP must be well-trained to interpret user messages correctly.

Regarding the second hypothesis, based on the SUS survey results and evaluating the time spent to complete the usability tasks, no clear evidence indicates that the button-based interface is more efficient than the text-based interface. The learnability score for the button-based prototype (2.0) was much higher than the score for the text-based one (1.4),[1] and the users also spent less time completing the tasks using the text-based interface. However, this might also be due to the limitation and the bad design of the button-based interface, as the user has to follow a rigid process every time to book a meeting. Direct observation suggests that users tended to learn how to use the button-based prototype more quickly, as they could simply tap buttons to test how to use the AI assistant. In contrast, the text-based prototype relied more

[1]The lower the learnability score, the easier it is to learn how to use the software.

on providing instructions to the users to understand what message to type upon first interacting with the AI assistant.

Given the answers to these two hypotheses, the answer to the research question is that the interface for an AI scheduling assistant for the business environment should combine features of text-based and button-based interactions so that users can minimise their efforts when interacting with the AI assistant, as business people are often busy with other tasks while interacting with the AI assistant.

The solution that this thesis proposed is to integrate the two interaction types by implementing an auto-complete feature where the app recommends phrases or keywords while the user types and by adding button options whenever possible to save time for the users. For example, rather than typing, Yes, I can confirm the meeting details , the interface presents button options f Confirm or Cancel , allowing users to tap on the options to confirm or cancel the meeting request. In addition, while seeking to answer the research question, the research also derived several findings, presented in Section 5.2.

5.2 Design Implications

By running two rounds of design iterations, several design details that designers should consider were revealed. These findings are presented below.

5.2.1 Maturity of Natural Language Processing

Because the NLP running on RASA was lacking training data, I found that the NLP could not fully recognise all messages sent by users, which affected the usability of the app and caused frustration for the users. In one case, a participant did not place a space between the words with Zack , so the NLP could not recognise the name Zack when the participant sent the request. Other problems occurred when the users typed in abbreviations, such as hr for hour or wanna for want to . As a result, the NLP might not handle the request correctly because it did not understand the abbreviations. These minor recognition problems caused the participants to trust the AI assistant less in terms of usability and recognisability.

5.2.2 Instructions for New Users

Ami She a book *Designing Bots* [65] describes onboarding as the following:

> The first interaction users see from the bot it could be a message that the bot sends to the installing user or a general message to a team. It sets the first impression and tackles a set of tasks that can best be accomplished at the start of the conversation [61, pp.80].

In the case of good onboarding, the bot declares its purpose in the context of the conversation, making it transparent to the user or the team. The bot should be very clear about what it does and how it can help the user .

The results from the testing sessions confirm She a a g me . Users relied heavily on the provided instructions and tips to learn how to communicate or interact with the AI assistant. Testing the first design iteration revealed that the original prototype provided too few instructions. As a result, the users were unable to send messages that the AI assistant could recognise.

When the new prototype provided a tip that users could type a hashtag (#) to search for recent requests (or frequent requests), users were confused. This case further confirmed that the instructions provided by the AI assistant must be accurate and clear because the users relied heavily upon them.

5.2.3 Feedback from the Artificial Intelligence Assistant

Providing feedback is one of the most important processes when an AI assistant is interacting with a human. In the first round of testing sessions, the importance of providing feedback when the NLP recognises a keyword was especially clear in the case of the original prototype. In which the AI assistant automatically filled in the blanks (the event title, location, participants, etc.) for some keywords but did not provide feedback to the users when it recognised a keyword, causing the users to retype the information.

Even during the second round of testing sessions, users wanted the extra certainty of knowing whether the meeting had been scheduled correctly. The simple message The meeting has been booked did not provide this assurance. Instead, the users wanted the AI assistant to display that it had sent the invitation to the correct invitees and had successfully reserved the meeting space. This confirmation was necessary because some meetings might be important for the users, who would otherwise need to check that the meetings were scheduled correctly.

5.2.4 Effort of 'Typing'

In both the original prototype and text-based prototype, users typed requests to the AI assistant. Many users preferred typing less and tapping buttons more, even though the process was more rigid. The second iteration used the text-based prototype as the basis to improve the usability and added an auto-complete feature and button options whenever possible to minimise typing to balance these demands. These functions minimised the effort required by the users, and they no longer complained about the effort of typing messages with the new prototype. Instead, the users began using the auto-complete feature when typing, and they commented that tapping recommended phrases allowed them to save time typing.

5.2.5 Personality of the Artificial Intelligence Assistant

The tone and appearance of the AI assistant also affected its usability. Users believed that the AI assistant was more professional when they could view its profile image. In addition, a conversational tone and typing indicators that simulated a delay to produce a human-like experience caused users to trust the AI assistant more. This perception is especially important for an AI scheduling assistant because business people require more assurance when scheduling important meetings and would prefer to communicate with an AI assistant they trust.

5.3 Limitations

As the thesis was conducted within a company, many limitations exist from the company that might affect the research process and direction. First, as the previous designer at the company had already designed the prototype, the stakeholders hoped that I could work on the new designs based on the original prototype. As a result, instead of conducting user interviews ahead of proposing new designs, I could only propose designs based on literature and the original prototype. In addition, as the NLP running behind the AI assistant chatbot still suffers from recognition issues, the users experienced frustration while interacting with the AI assistant chatbot because it sometimes did not understand the users. If the NLP worked better, the usability of the AI assistant chatbot would increase as well.

Furthermore, the COVID-19 pandemic in 2020 also affected the research process. Due to relocation, the participants in the testing sessions were all Taiwanese individuals who do not speak native English. This language barrier might have affected the research results because the user had to type in English to interact with the AI assistant.

Lastly, only seven individuals participated in the first iteration, and nine participated in the second iteration. The lack of participants in the SUS questionnaire might have affected the questionnaire results. Moreover, as the SUS questionnaire was designed in English, similar language issues that affected the usability testing process might have affected the SUS questionnaire results.

5.4 Conclusion and Future Research

This thesis evaluated the AI scheduling assistant, O.AI, a chatbot designed to provide business people with a means of efficiently scheduling meetings with peers. The project began with evaluating the existing prototype designed by a former colleague on the team and background studies. During the evaluation sessions, the users preferred to interact with the AI assistant through text-based interactions; thus, the project focused on designing a text-based AI assistant that minimised the user s efforts.

One notable feature that was proposed is the auto-complete feature, through which the AI assistant automatically recommended phrases to the users when they typed any keywords that matched the recommended phrases. For example, when a user typed the letter B , recommended phrases, such as Book a meeting and Book an event , were displayed, and the user could tap an option to complete the sentence automatically. Another notable feature was integrating button interactions wherever possible throughout the interaction to minimise the time required to complete tasks by allowing users to select from a list of options to perform the intended action. The project also revealed the importance of having a well-trained NLP to handle user requests and an AI assistant with a suitable personality to gain user trust.

Future researchers could analyse how the auto-complete feature can be designed systematically to provide users with an efficient way to type messages using the AI assistant. In addition, the best situation in which to combine button interaction during the e conversation with the AI assistant remains unknown. Lastly, as this thesis focus on designing the AI scheduling assistant on a mobile app, future researchers are encouraged to continue exploring different ways, such as voice-interaction or haptic-interaction, of allowing users to interact with AI assistants to provide people in the business environment with a well-considered, effective, and efficient AI assistant.

References

[1] F. We e heide, The A ificial I ellige ce I d a d Gl bal Challe ge , *Forbes*, Nov. 27, 2019. https://www.forbes.com/sites/cognitiveworld/2019/11/27/the-artificial-intelligence-industry-and-global-challenges/ (accessed Apr. 13, 2021).

[2] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*. Wiley-Blackwell, 1993.

[3] J. Shabbi a d T. A e , A ificial I ellige ce a d i R le i Nea F e , *ArXiv*, vol. abs/1804.01396, Apr. 2018.

[4] A ificial I ellige ce Ma ke Si e, G h, Sha e A al i [2020-2027] , *Fortune Business Insights*. https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114 (accessed Apr. 13, 2021).

[5] I ellige Vi al A i ant (IVA) Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026) , *Mordor Intelligence*. https://www.mordorintelligence.com/industry-reports/intelligent-virtual-assistant-market (accessed Apr. 13, 2021).

[6] M. Robinson, J. Gray, A. Cowley, a d R. Ta , Ad i g he e f c e a i al UX , *Deloitte*, 2019. https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-chatbots-adopting-the-power-of-conversational-ux.pdf (accessed Apr. 13, 2021).

[7] D. Jovic, The F e I N - 37 Fa ci a i g Cha b S a i ic , *smallbizgenius*, Nov. 20, 2020. https://www.smallbizgenius.net/by-the-numbers/chatbot-statistics/ (accessed Apr. 13, 2021).

[8] AI ill e 95% f c me i e ac i b 2025 , *Finance Digest*. https://www.financedigest.com/ai-will-power-95-of-customer-interactions-by-2025.html (accessed Apr. 13, 2021).

[9] H. J. Wil a d P. R. Da ghe , H H ma a d AI A e W ki g T ge he i 1,500 C m a ie , *Harvard Business Reviews*, Aug. 2018. https://webcache.googleusercontent.com/search?q=cache:qKR82p2QfHoJ:https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces+&cd=3&hl=zh-TW&ct=clnk&gl=tw (accessed Apr. 13, 2021).

[10] T. M. B ill, L. M , a d R. J. Mille , Si i, Ale a, and other digital assistants: a d f c me a i fac i i h a ificial i ellige ce a lica i , *J. Mark. Manag.*, vol. 35, no. 15 16, pp. 1401 1436, Oct. 2019, doi: 10.1080/0267257X.2019.1687571.

[11] Si i , *Apple*. https://www.apple.com/siri/ (accessed Apr. 14, 2021).
[12] A. Kh , T e d D i i g he Cha b G h , *Medium*, Sep. 21, 2017. https://chatbotsmagazine.com/trends-driving-the-chatbot-growth-77b78145bac (accessed Apr. 13, 2021).
[13] D dle The a e f mee i g e , *Doodle*, 2019. https://meeting-report.com/ (accessed Apr. 13, 2021).
[14] A. F l ad a d P. B. B a d aeg, U e e e ie ce i h cha b : fi di g f m a e i ai e d , *Qual. User Exp.*, vol. 5, no. 3, pp. 1 14, Apr. 2020, doi: 10.1007/s41233-020-00033-2.
[15] G. Neff, Talki g B : S mbi ic Age c a d he Ca e f Ta , *Int. J. Commun.*, vol. 10, no. 17, pp. 4915 4931, 2016.
[16] E. Adam l a d L. M iade , Cha b : Hi , ech l g , a d a lica i , *Mach. Learn. Appl.*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/j.mlwa.2020.100006.
[17] A. M. T i g, I. COMPUTING MACHINERY AND INTELLIGENCE , *Mind*, vol. LIX, no. 236, pp. 433 460, Oct. 1950, doi: 10.1093/mind/LIX.236.433.
[18] G. O a d D. D e, The T i g Te , i *The Stanford Encyclopedia of Philosophy*, Winter 2020., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2020. Accessed: Apr. 13, 2021. [Online]. Available: https://plato.stanford.edu/archives/win2020/entriesuring-test/
[19] T. Zem k, A B ief Hi f Cha b , *DEStech Trans. Comput. Sci. Eng.*, Oct. 2019, doi: 10.12783/dtcse/aicae2019/31439.
[20] J. Wei e ba m, ELIZA a computer program for the study of natural language c mm ica i be ee ma a d machi e , *Commun. ACM*, vol. 9, no. 1, pp. 36 45, Jan. 1966, doi: 10.1145/365153.365168.
[21] A Hi f Cha b , *ChatBot Pack*, Jun. 07, 2018. https://www.chatbotpack.com/a-history-of-chatbots/ (accessed Apr. 13, 2021).
[22] J. L. H che , H Pa he T i g Te b Chea i g , . 33, A . 1996.
[23] Sc - Seach Re l , *Scopus*. https://www.scopus.com/term/analyzer.uri?sid=a85c8c01a04ce3e0bbb3511868b9d58c&origin=resultslist&src=s&s=TITLE-ABS-KEY%28chatbot%29+AND+PUBYEAR+%3e+1999+AND+PUBYEAR+%3c+2021&sort=plf-

f&sdt=b&sot=b&sl=60&count=2091&analyzeResults=Analyze+results&txGid=7473e72fe5dbb5f9d7eaf842bd565bdc (accessed Apr. 13, 2021).

[24] P. B. B a d aeg a d A. F l ad, Wh Pe le U e Cha b , i *Internet Science*, vol. 10673, I. Kompatsiaris, J. Cave, A. Satsiou, G. Carle, A. Passani, E. Kontopoulos, S. Diplaris, and D. McMillan, Eds. Cham: Springer International Publishing, 2017, pp. 377 392. doi: 10.1007/978-3-319-70284-1_30.

[25] G. Neal , C cil P : U i g Faceb k Me e ge A d Cha b T G Y A die ce , *Forbes*, Jun. 04, 2018. https://www.forbes.com/sites/forbesagencycouncil/2018/06/04/using-facebook-messenger-and-chatbots-to-grow-your-audience/ (accessed Apr. 13, 2021).

[26] G. Caff , H KLM e a ificial i ellige ce i c me e ice , *Digiday*, Oct. 04, 2016. https://digiday.com/marketing/klm-uses-artificial-intelligences-customer-service/ (accessed Apr. 13, 2021).

[27] WHO la che a cha b Faceb k Me e ge c mba COVID-19 mi i f ma i , *World Health Organization*, Apr. 15, 2020. https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation (accessed Apr. 13, 2021).

[28] Cha b Ma ke W h $1.25 Billi B 2025 CAGR: 24.3% , *Grand View Research*, Aug. 2017. https://www.grandviewresearch.com/press-release/global-chatbot-market (accessed Apr. 13, 2021).

[29] G. G. Ch dh , Na al la g age ce i g , *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51 89, Jan. 2005, doi: 10.1002/aris.1440370103.

[30] S. Ayanouz, B. A. Abdelhakim, a d M. Be hmed, A Sma Cha b A chi ec e ba ed NLP a d Machi e Lea i g f Heal h Ca e A i a ce , i *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, New York, NY, USA, Mar. 2020, pp. 1 6. doi: 10.1145/3386723.3387897.

[31] Wha i A ificial I ellige ce (AI)? IBM , *IBM Cloud Education*, Jun. 03, 2020. https://www.ibm.com/cloud/learn/what-is-artificial-intelligence (accessed Apr. 13, 2021).

[32] K. Bl m, H Ma age C me Se ice Tech l g I a i , *Gartner*, Mar. 12, 2019. //www.gartner.com/smarterwithgartner/27297-2/ (accessed Apr. 13, 2021).

[33] N. Ta a a a d E. Bi e , A ma ed Facili a i f Idea Pla f m : De ig a d Evalua i f a Cha b P e , e e ed a he Thi Ni h I e a i al C fe e ce Information Systems (ICIS) 2018, San Francisco, Oct. 2018.

[34] D. Feldma , Cha b : Wha Ha e ed? , *Medium*, Apr. 16, 2018. https://chatbotslife.com/chatbots-what-happened-dcc3f91a512c (accessed Apr. 13, 2021).

[35] T. Kl e , F m Cha b Dial g e S em , i *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*, 2011, pp. 1 22. doi: 10.4018/978-1-60960-617-6.ch001.

[36] D. Jurafsky and J. H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, 2nd ed. Upper Saddle River, N.J: Pearson Prentice Hall, 2009.

[37] Y. H , D e le a me age Chatbots?: developing and comparing the usability of a conversational vs. menu-ba ed Cha b i c e f e hi e b a di g , Thesis, Aalto University, 2019. Accessed: Apr. 13, 2021. [Online]. Available: /paper/Do-people-want-to-message-Chatbots%3A-developing-and-Hu/8272287e6edf9b8bed01264abfb3127116642953

[38] D. B i , Dee Lea i g f Cha b , Pa 1 Introduction WildML , *Wild ML*, Apr. 06, 2016. http://www.wildml.com/2016/04/deep-learning-for-chatbots-part-1-introduction/ (accessed Apr. 13, 2021).

[39] J. Kim, H.-G. Lee, H. Kim, Y. Lee, and Y.-G. Kim, T -Step Training and Mixed Encoding-Dec di g f Im leme i g a Ge e a i e Cha b i h a Small Dial g e C , in *Proceedings of the Workshop on Intelligent Interactive Systems and Language Generation (2IS&NLG)*, Tilburg, the Netherlands, 2018, pp. 31 35. doi: 10.18653/v1/W18-6707.

[40] A. G a, I d c i AI Cha b , *Int. J. Eng. Res.*, vol. V9, no. 07, pp. 255 258, Jul. 2020, doi: 10.17577/IJERTV9IS070143.

[41] J. W e , 3 cha b e : Which i be f b i e ? , *Chatimize*, Sep. 16, 2019. https://chatimize.com/chatbot-types/ (accessed Apr. 13, 2021).

[42] I i i e C e a i al Cha b B ilde , *Landbot.io*. https://landbot.io (accessed Apr. 13, 2021).

[43] M. Rybak a, Hell F e h C ea e C me Se ice Cha b Sla h Re e Time b 76% . h ://cha f el.c m/bl g/ /hell f e h-reduces-support-wait-times-with-chatfuel-messenger-bot (accessed Apr. 13, 2021).

[44] J. He a d J. Lee, CiSA: A I cl i e Cha bot Service for International Students and Academic , i *HCI International 2019 – Late Breaking Papers*, Jul. 2019, pp. 153 167. doi: 10.1007/978-3-030-30033-3_12.

[45] K. D deli , Le Cha : E i Read Take Ba ki g C e a i he Ne Le el , *Capital One*. https://www.capitalone.com/learn-grow/tech-innovation/eno-chatbot-banking-conversations-next-level/ (accessed Apr. 13, 2021).
[46] J. Cranshaw *et al.*, Cale da .Hel : De ig i g a W kfl -Based Scheduling Agent i h H ma i he L , S cial Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2940295, Mar. 2017. doi: 10.2139/ssrn.2940295.
[47] I a Mee i g Sched li g , *x.ai*. https://x.ai/ (accessed Apr. 13, 2021).
[48] A. Monroy-He de a d J. C a ha , H We B il a Vi ual Scheduling A i a a Mic f , *Harvard Business Review*, Jul. 28, 2017. Accessed: Apr. 13, 2021. [Online]. Available: https://hbr.org/2017/07/how-we-built-a-virtual-scheduling-assistant-at-microsoft
[49] J lie De k - AI-based personal assistant wh ched le mee i g ia email . https://www.juliedesk.com/ (accessed Apr. 13, 2021).
[50] P. La e a d P. Wa a, A C e a i al U e I e face f S ck A al i , i *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 5298 5305. doi: 10.1109/BigData47090.2019.9005635.
[51] F ee O li e A i me Sched li g S f a e - Cale dl , *Calendly*. https://calendly.com/ (accessed Apr. 13, 2021).
[52] F ee li e mee i g ched li g l , *Doodle*. https://doodle.com/en/ (accessed Apr. 13, 2021).
[53] See me e cale da a ailabili - Android - Cale da Hel , *Google*. https://support.google.com/calendar/answer/6294878?co=GENIE.Platform%3DAndroid&hl=en (accessed Apr. 13, 2021).
[54] L k me e ched le i O l k f Mac , *Microsoft Outlook*. https://support.microsoft.com/en-us/office/look-up-someone-s-schedule-in-outlook-for-mac-468e6b1a-c5c2-4378-adc0-dd1e210669c4 (accessed Apr. 13, 2021).
[55] .ai Re ie 2021: De ail , P ici g, & Fea e G2 , *G2*, Dec. 20, 2018. https://www.g2.com/products/x-ai/reviews (accessed Apr. 13, 2021).
[56] .ai Re ie & Ra i g 2021 , *TrustRadius*. https://www.trustradius.com/products/x-ai/reviews (accessed Apr. 14, 2021).
[57] D. B a , .ai Re ie & Ra i g 2021 , *TrustRadius*. https://www.trustradius.com/reviews/x-ai-2018-12-05-10-00-32 (accessed Apr. 13, 2021).

[58] M. Mei el, .ai Re ie 2021: De ail , P ici g, & Fea e G2 , *G2*, Dec. 07, 2018. https://www.trustradius.com/reviews/x-ai-2018-12-05-13-07-45 (accessed Apr. 13, 2021).
[59] D. H. M e e , H I l e S akeh lde i Y U e Re ea ch , *The Interaction Design Foundation*. https://www.interaction-design.org/literature/article/how-to-involve-stakeholders-in-your-user-research (accessed Apr. 13, 2021).
[60] K. Pernice, Affi i Diag ammi g: C llab a i el S UX Fi di g & De ig Idea , *Nielsen Norman Group*. https://www.nngroup.com/articles/affinity-diagram/ (accessed Apr. 13, 2021).
[61] P. W. J da , B. Th ma , I. L. McClella d, a d B. Wee dmee e , Ed ., SUS: A Q ick a d Di U abili Scale , i *Usability Evaluation In Industry*, 0 ed., CRC Press, 1996, pp. 207 212. doi: 10.1201/9781498710411-35.
[62] J. Sa , Mea i gU: 5 Wa I e e a SUS Sc e , *Measuring U*, Sep. 19, 2018. https://measuringu.com/interpret-sus-score/ (accessed Apr. 13, 2021).
[63] T. Mandl, K. Furtner, and C. Womser-Hacke , Effec f A -Suggest on the U abili f Sea ch i eC mme ce , e e ed a he 14 h I e a i al S m i m Information Science (ISI 2015), Zadar, Croatia, May 2015. doi: 10.5281/zenodo.17948.
[64] P. Lambe , W i e email fa e i h Sma C m e i Gmail , *Google*, May 08, 2018. https://blog.google/products/gmail/subject-write-emails-faster-smart-compose-gmail/ (accessed Apr. 13, 2021).
[65] A. Shevat, *Designing bots: creating conversational experiences*, First edition. Beiji g ; B : O Reill , 2017.

Appendix A: Tasks in Usability Testing Session

Scenario 1:

You are currently a project manager at the marketing team of the Company LJJ. This morning, your manager has just let you know that the CEO decided to put more effort and resources on your team's SEO campaign. To make sure that the resources can be wisely distributed, your manager wanted you to meet Zack, a colleague from the engineering team to discuss the project.

Your Task:

Book a meeting with your virtual assistant with the following information:

- Topic: SEO Campaign
- Person to meet: Zack
- Time: Next Tuesday (28th of April), In the Afternoon
- Duration: Approximately two hours.
- Location: Room A

Scenario 2:

After meeting Zack, you have decided to arrange an internal team meeting to announce the news. Since everyone is busy at the moment, your team members suggest everyone meet online.

Your Task:

Book a meeting with your virtual assistant with the following information:

- Topic: Internal Meeting
- Person to meet: Ben, Ather, Peter (Marketing Team)
- Time: Between Wednesday (29th of Apr.) and Friday (1st of May)
- Duration: One hour.
- Location: Skype

Scenario 3:

Finally, you found that Erik, an external consultant of your company has a strong background in conducting successful SEO campaigns, and you are willing to invite him for a coffee chat at Starbucks next week.

Task:

Arrange a meeting with Erik.

Appendix B: Interview Questions

- Which interface do you like the best
- Do you think this chatbot can help you complete your tasks more efficiently?
- Overall, what's your experience been with the app?
- Would you consider using this app
- Do you like the interface? Is it easy to use?
- What do you most like about the chatbot?
- What do you least like about the chatbot?
- If you could change one thing about the app, what would it be? Why?
- What one thing are you most excited about with the app? Why?
- Why will you continue to use this app? What will stop you from using this app in the future?
- How likely are you to refer to this app? Why or why not?

Appendix C: SUS Questionnaire

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Appendix D: Screening Survey

1. How old are you?
2. What's your working industry?
3. What's your position?
4. Which calendar software do you use at work?
5. How often do you get involved in a meeting per week?
6. How many meetings do you schedule per week?
7. How much time do you spend on scheduling meetings with others?
8. How much time do you spend on meetings?
9. When was the last time you scheduled a meeting with others?

Appendix E: Consent Forms

Information Sheet, 8th April 2020

As a participant in the research thesis: ***The appropriate design techniques for an AI agent***, on the __________________, I agree to participate in the study conducted and recorded by the KTH Royal Institute of Technology and The Research Institute of Sweden (RISE).

I understand and consent to the use and release of the recording by KTH Royal Institute of Technology and The Research Institute of Sweden (RISE). I understand that the information and recording is for research purposes only and that my name and image will not be used for any other purpose. I relinquish any rights to the recording and understand the recording may be copied and used by KTH Royal Institute of Technology and The Research Institute of Sweden (RISE). without further permission. I understand that participation in this usability study is voluntary and I agree to immediately raise any concerns or areas of discomfort during the session with the study administrator.

Please sign below to indicate that you have read and you understand the information on this form and that any questions you might have about the session have been answered.

If I agree to participate, are there any risks for me?
We do not think there are any risks in taking part in this study. You will not be identified in our data analysis, or any public or internal presentations of the data about the workshop. If you consider that the data collection will or has had an adverse effect upon your participation then please contact researcher: Hau-Ben Benjamin Shih, hbbshih@kth.se. Supervisor: Pavel Karpashevich, pavelka@kth.se.

Thank you very much for reading this information sheet.

If you have any questions or please feel free to contact:

Hau-Ben Benjamin Shih
MID, EECS, KTH
hbbshih@kth.se

1 (1)

Consent Form for Participants

I give consent to participate in the study: ***The appropriate design techniques for an AI agent*** being carried out by Name of your team, KTH.

Please **initial** each box

- I have read and understood the information sheet about taking part. ☐
- A team member has answered any questions that I had / I have no further questions. ☐
- I understand that I will be interviewed following the study and that this interview may be audio recorded. ☐
- I understand that the data collected for this study will be stored in a secure location. ☐
- I understand that the data collected about me will be used only for research purposes. ☐
- I understand that I will not be mentioned by name on any documents or in any presentations about the research. ☐
- I understand that I can withdraw from the study at any time without needing to give a reason. ☐
- Withdrawing from the study will not affect any services that I am currently receiving now or might receive in the future. ☐

Signature of participant..……………………………………...............................

Name (in capitals) …………………………………….Date……………………..

Appendix F: SUS results of the first design iteration

Original Design Results:

Participant	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	SUS Score
p1	1	4	1	3	3	3	2	5	2	3	90
p2	2	2	3	2	2	4	4	3	4	2	55
p3	3	2	4	2	3	5	2	5	3	2	47.5
p4	3	3	2	2	2	4	2	4	5	1	50
p5	2	3	2	1	1	3	3	4	3	2	45
p6	4	4	3	3	2	3	2	3	3	2	47.5
p7	4	2	4	2	4	3	4	2	3	2	70
Mean	2.7	2.9	2.7	2.1	2.4	3.6	2.7	3.7	3.3	2.0	57.8571428
Median	3.0	3.0	3.0	2.0	2.0	3.0	2.0	4.0	3.0	2.0	50
STD	1.1	0.9	1.1	0.7	1.0	0.8	1.0	1.1	1.0	0.6	16.5

<= 60	5
60-70	1
70-80	0
80-90	0
>= 90	0
Learnability	2.1

Text-based Design Results:

Participant	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	SUS Score
p1	2	1	4	1	3	3	5	4	5	1	90
p2	3	2	4	2	4	2	4	2	4	2	72.5
p3	4	3	5	2	4	1	5	4	4	1	77.5
p4	4	2	5	1	2	2	4	2	5	1	80
p5	4	4	4	1	4	2	5	1	5	1	82.5
p6	4	1	4	2	4	2	4	2	4	1	80
p7	4	2	4	1	4	2	4	2	4	2	77.5
Mean	3.6	2.1	4.3	1.4	3.6	2.0	4.4	2.4	4.4	1.3	80
Median	4.0	2.0	4.0	1.0	4.0	2.0	4.0	2.0	4.0	1.0	80
STD	0.8	1.1	0.5	0.5	0.8	0.6	0.5	1.1	0.5	0.5	5.4

<= 60	0
60-70	0
70-80	6
80-90	1
>= 90	0
Learnability	1.4

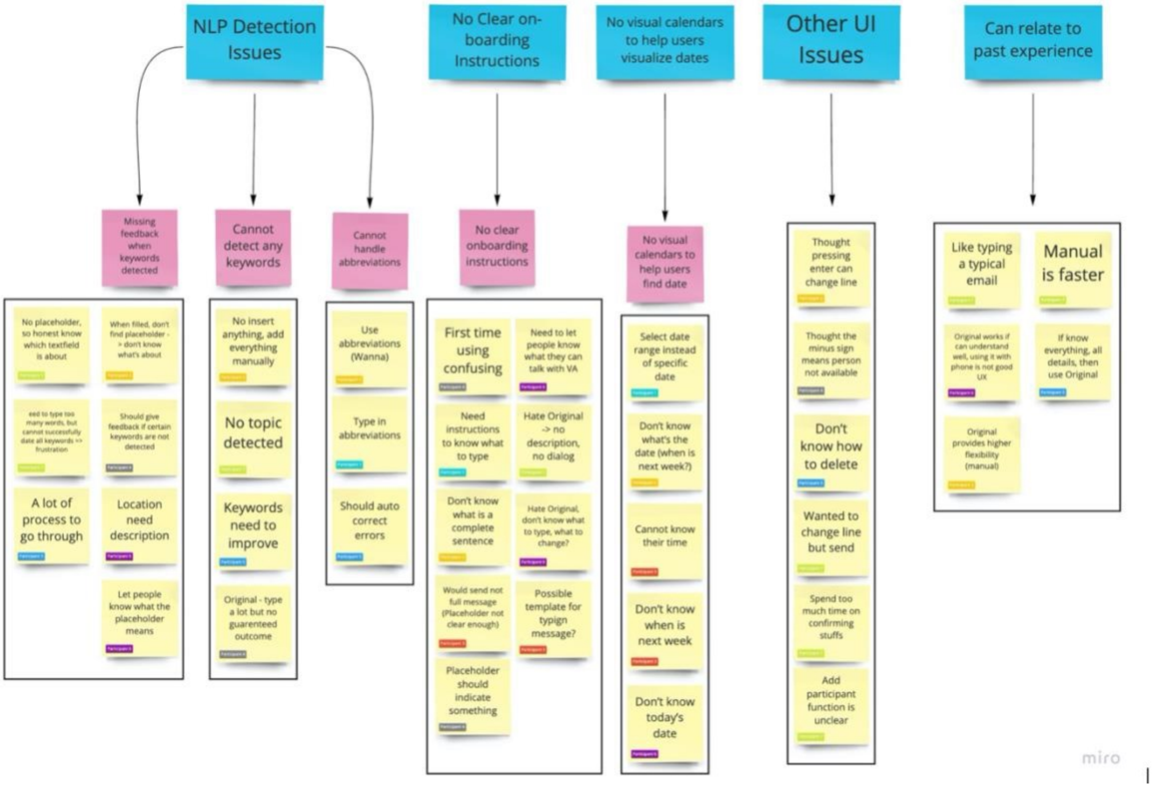
Button-based Design Results:

Participant	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	SUS Score
p1	5	1	5	1	4	2	5	1	5	1	90
p2	4	1	4	1	4	1	4	2	4	2	82.5
p3	2	3	1	5	1	4	1	5	2	5	12.5
p4	4	2	4	1	4	1	5	1	4	1	87.5
p5	5	1	5	1	3	2	5	1	5	1	92.5
p6	3	3	3	3	3	3	3	3	3	2	52.5
p7	3	3	3	2	3	2	4	3	3	2	60
Mean	3.7	2.0	3.6	2.0	3.1	2.1	3.9	2.3	3.7	2.0	68.2142857
Median	4.0	2.0	4.0	1.0	3.0	2.0	4.0	2.0	4.0	2.0	82.5
STD	1.1	1.0	1.4	1.5	1.1	1.1	1.5	1.5	1.1	1.4	29.0

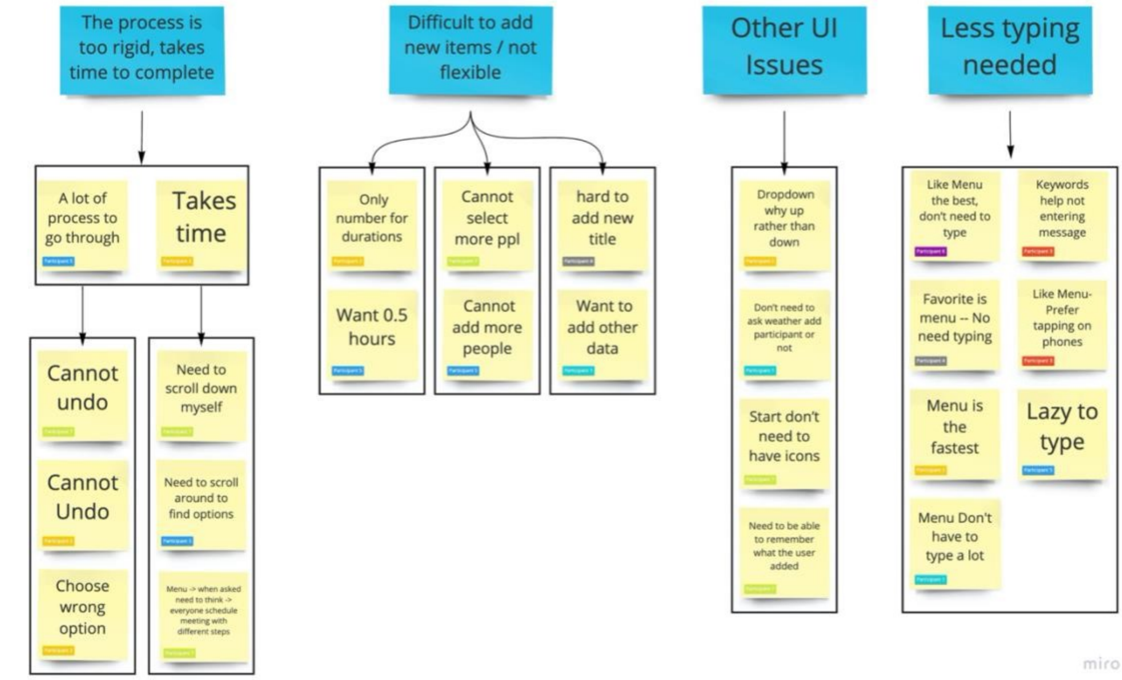
<= 60	2
60-70	0
70-80	0
80-90	2
>= 90	1
Learnability	2.0

Appendix G: Affinity diagram of first design iteration

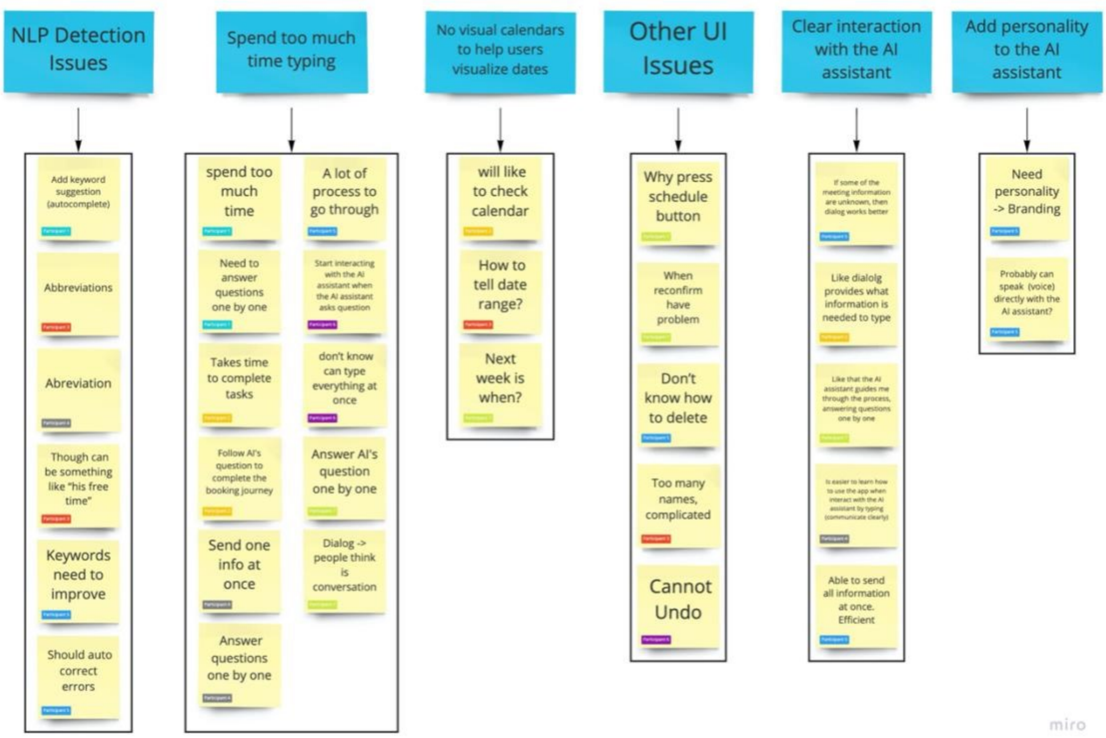
Affinity Diagram of Original Design



Affinity Diagram of Button-based Design



Affinity Diagram of Text-based Design



Appendix H: SUS results of the new design

Participant	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	SUS Score
p1	4	1	5	1	4	2	4	1	5	1	90.0
p2	3	4	2	1	3	2	2	3	2	1	52.5
p3	4	1	4	1	3	1	4	1	4	1	85.0
p4	4	1	5	2	3	3	5	1	5	2	82.5
p5	3	2	4	2	3	1	5	2	4	1	77.5
p6	5	2	5	2	5	1	5	1	5	1	95.0
p7	5	1	5	1	4	2	5	1	5	1	95.0
p8	3	2	4	2	3	2	4	3	4	2	67.5
p9	2	5	2	4	3	4	1	2	3	4	30.0
Mean	3.7	2.1	4.0	1.8	3.4	2.0	3.9	1.7	4.1	1.6	75.0
Median	4.0	2.0	4.0	2.0	3.0	2.0	4.0	1.0	4.0	1.0	82.5
STD	1.0	1.5	1.2	1.0	0.7	1.0	1.5	0.9	1.1	1.0	21.7

<= 60	2
60-70	1
70-80	1
80-90	2
>= 90	3
Learnability (Average of Q4 and Q10)	1.7

Appendix I: Affinity Diagram of the new design

TRITA -EECS-EX-2022:39