Aalto University

School of Science

Master's Programme in ICT Innovation: Human-Computer Interaction & Design

Yuchen (Charles) Hu

Do People Want to Message Chatbots? Developing and Comparing the Usability of a Conversational vs. Menu-based Chatbot in Context of New Hire Onboarding

Master's Thesis

Espoo, September 30, 2019

Supervisor: D.Sc. (Tech.) Marko Nieminen, Aalto University

Aalto University
School of Science

AALTO UNIVERSITY
School of Science   ABSTRACT OF
Master's Programme in ICT Innovation   MASTER´S THESIS

Author: Yuchen (Charles) Hu

Title of the thesis: Do People Want to Message Chatbots? Developing and Comparing the Usability of a Conversational vs. Menu-based Chatbot in Context of New Hire Onboarding

Number of pages: 76	**Date:** Sept 30, 2019

Major or minor: ICT Innovation: Human-Computer Interaction & Design (SCI3020)

Supervisor: Marko Nieminen, Aalto University

Thesis advisors: N/A

How should people interact with chatbots? This question has become relevant as chatbots grow in recognition within the field of human-computer interaction. Should chatbots strive to have intelligent and realistic conversations with their users? Or, does a simplified, menu-based approach provide the better experience?

To answer these questions, a human-centred design process was used to design, develop, and evaluate the usability of two chatbots in context of new hire onboarding. A conversational chatbot with natural language processing was built using Google Dialogflow, while a technology-limited, menu-based chatbot was built with Landbot. 17 participants were split into three groups to perform a qualitative user test, where group 1 tested both bots, group 2 tested only the conversational bot, and group 3 tested only the menu-based bot. Afterwards, all participants were given a quantitative, Likert-scale survey to measure the usability, intelligence, and satisfaction of the chatbots.

The results indicate that users preferred a menu-based over a conversational chatbot experience due to its greater ease of use, less likelihood for errors, convenience of graphical user interface elements, and suitability for scenarios where information needs to be provided rather than requested. Conversational chatbot experiences were found to be more convenient when users had direct questions, although they are more complex to implement when compared to menu-based chatbots.

To my family, friends, coworkers, and professors, a sincere thank you as this thesis would not have been possible without your support.

Contents

1.   Introduction

*"Pretty much everybody today who's building applications ...
will build chatbots as the new interface ... they fundamentally
revolutionise how computing is experienced by everybody."*

— Satya Nadella (Johnson, 2016)

Microsoft CEO Satya Nadella captured the technological zeitgeist of the times in one succinct quote while on stage at the 2016 World Partner Conference in Toronto. Breathless hype around chatbots was at its zenith and they seemed poised to be the next great digital revolution after mobile computing.

After all, chatbots promised us the ability to obtain services and interact with businesses like how we do with our friends—by talking to them. Backed by powerful artificial intelligence, machine learning, and natural language processing capabilities, we could simply tell a restaurant bot to book a dinner reservation for two people at 7 pm on a Tuesday, instead of having to download and use a new mobile application. Bots pledged a more personal, intimate, and convenient experience for consumers, while businesses achieve cost savings through automation (Dale, 2016). At Facebook's F8 developer conference in 2016, CEO Mark Zuckerberg proclaimed "We think that you should be able to message a business in the same way you message a friend ... it shouldn't take your full attention like a phone call would. And you shouldn't have to install a new app (Heath, 2016)."

Fast forward three years later to 2019, and it appears that the revolutionary chatbot paradigm shift has yet to arrive. Several technology commentators have noted that user adoption of chatbots has lagged significantly in contrast to the initial optimism (Simonite, 2017). Despite recent progress, AI[1] and NLP[2] technology have yet to reach the capability required to reliably understand the complexity of human speech, leading to error and frustration. Indeed, Brandtzaeg and Følstad (2017) express that most available chatbots fail due to insufficient usability or nonsensical responses. Facebook admitted in 2017 that M, their AI-powered bot on its Messenger platform, could only

---

[1] Artificial intelligence

[2] Natural language processing

handle 30% of requests without human intervention and that they would scale back their scenarios to prevent further user disappointment.

The challenges facing intelligent bots today have led to the concurrent rise of so-called "dumb" chatbots, which utilise conventional GUI[3] elements with simple decision tree logic to minimise the potential for human-bot misunderstandings. Many successful bots in use today, such as Google Assistant, scale back on the amount of user typing required and instead introduce more graphical elements such as buttons (Piccolo et al., 2019). For example, the popular messaging app, Telegram, began removing friction by reducing the amount of typing required to interact with the bot only a few months after their chatbot platform went live (Lomas, 2016).

While several studies have been performed on the application of chatbots in various contexts, there is a lack of research comparing the usability of different chatbot implementation methods. Thus, this thesis extends existing chatbot research by developing and comparing the usability of a conversational "smart" chatbot with NLP, versus a menu-based "dumb" chatbot without any intelligent capabilities. The two chatbots are applied towards the scenario of new employee onboarding at Apegroup, a Stockholm-based design agency. New employee onboarding was the selected scenario because it is currently a widespread problem for many companies (Harpelund, 2019), while a technology-supported platform for obtaining HR[4] information has been identified as a key solution (Robinson et al., 2019).

It is challenging to develop a good chatbot, let alone two. However, given that the global chatbot market is expected to grow to USD $1.34 billion at a rate of 31% per year by 2024 (Bhutani & Wadhwani, 2018) and the technology still has major potential, it is a challenge worth tackling.

1.1. Partner company

Apegroup is the corporate collaboration partner for this thesis. Apegroup is an award-winning digital design consultancy based in Stockholm, Sweden. Composed of around 50 employees, it is dedicated to creating value through the design of meaningful digital products and experiences. Founded in 2001, it was also the first design studio in Sweden to create mobile apps.

The company's motto is to bring tomorrow's technology to today, making them an ideal partner for exploring the application of chatbots. Apegroup has also recently experienced a hiring spree, making new hire onboarding particularly relevant for them.

---

[3] Graphical user interface

[4] Human resources

1.2.   Structure of the thesis

The structure of this thesis is outlined below:

Chapter 1 includes the introduction of the thesis, partner company background, and structure of the thesis.

Chapter 2 presents the goals and research questions of the thesis.

Chapter 3 provides the academic and industry background of chatbots and new hire onboarding based on extensive literature review.

Chapter 4 presents the research, design, and evaluation methods of the study.

Chapter 5 includes the implementation details of the chatbots.

Chapter 6 presents the results of the chatbots evaluation.

Chapter 7 provides an overall discussion and analysis of the results.

Chapter 8 presents the final conclusions of the study, including limitations and future work opportunities.

2. Goals and research questions

As mentioned in the introduction, the overall goal of this thesis is to develop and compare the usability of two chatbots—a conversational chatbot with NLP capabilities that a user primarily communicates with by typing freely, versus an intelligence limited menu-based chatbot in which the user interacts by using a menu system with buttons.

The corresponding research questions are:

RQ1: Do users prefer a conversational or menu-based chatbot experience?

There is currently much debate regarding the design of chatbots, specifically whether they should continue to strive for realistic messaging conversations, or follow a simplified approach using buttons. To investigate this discussion, the goal of RQ1 is to conduct a structured study to determine which chatbot style is preferred by users using measures of usability and satisfaction.

RQ2: To what extent does a realistic conversational experience influence usability?

The current assumption is that providing an intelligent, human-like conversational experience in chatbots is considered better than an experience that does not. The aim of RQ2 is to investigate this assumption. Does the perception of intelligence, enabled by free text interaction, make a bot more usable? Or can a menu-based bot that does not try to be smart become the better experience?

3.   Background

This section includes the literature review on chatbots and new employee onboarding. Chatbot history, definition, design techniques, technology, and an overview of onboarding are provided.

3.1.   Chatbot history

3.1.1.   Chatbot origins

Chatbots have actually existed for quite some time, when the first bot named ELIZA was built by Joseph Weizenbaum in 1966 to simulate psychotherapy (Io & Lee, 2017). In the next few decades, bots were designed using simple pattern matching and template-based response mechanisms, with the primary goal of passing the Turing test, a measure of how human-like a machine could be. Indeed, new bots including ALICE, Mitsuku, and Rose were developed for the Loebner prize competition, which since 1991 has become the de-facto contest for implementing the Turing test. The chatbot fantasy of the time was to fool people into believing they were talking to real humans, rather than accomplishing any specific business or user goals (Dale, 2016).

3.1.2.   Growth of chatbots

With the rise of the internet and mobile computing, HCI[5] researchers in 2008 predicted a growing presence of "increasingly clever computers" in people's lives and a shift from simulating life-like behaviour to providing utility (Harper et al., 2008). The popular voice-based digital assistants including Amazon's Alexa and Apple 's Siri are now enjoyed by people throughout their homes to help check the weather, order items and play music. The growth of messaging platform based chatbots has also soared, with Facebook Messenger alone hosting more than 300,000 chatbots since 2016. Combined with Slack and Skype, the three platforms currently host over a million chatbots (Piccolo et al., 2019). These chatbots provide utility and accomplish a variety of use

---

[5] Human-computer interaction

cases, from ordering pizza (Domino's) to flight booking (Kayak) to shopping (Burberry) (Jain et al., 2018).

Recent studies have shown that the chatbot growth is expected to continue to increase. According to Gartner, chatbots will power 85% of all customer interactions by the year 2020, and that people will have more conversations with chatbots than with their spouse (Gartner, 2016). UK-based Juniper Research predicts that by 2022, bots will save businesses $8 billion per year (Dhanda, 2017).

3.1.3. Trends influencing chatbot growth

The growth of chatbots can be attributed to the following trends:

A. Advances in artificial intelligence, machine learning, natural language processing, and related technologies

The aforementioned technologies promise significant improvements in the ability for chatbots to understand human speech, including enhanced natural language interpretation and prediction capabilities. Additionally, progress in conversational modelling is enabling neural network based chatbots to surpass traditional pattern-based matching chatbots in effectiveness by a large margin (Brandtzaeg & Følstad, 2017). In short, the advances in technology are helping chatbots become smarter and more effective in understanding and responding to human input.

B. Adoption of messaging platforms

The way people communicate has dramatically shifted since the arrival of the Apple iPhone and mobile computing in general. 4.5 billion people worldwide use mobile messaging apps such as Facebook Messenger, WhatsApp, and WeChat as their preferred means of communication (Statista, 2019).
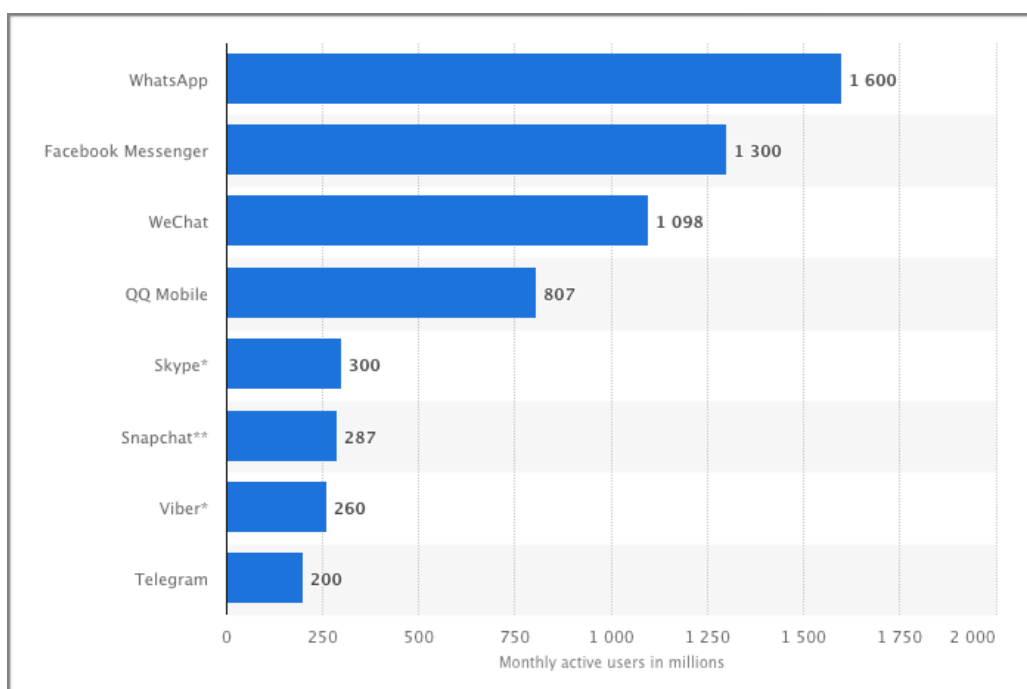

Figure 1. Most popular global mobile messenger apps as of April 2019 (Statista, 2019)

According to Dale (2016), messaging apps are a perfect environment for chatbots due to the ubiquity of the messaging interface effectively being a frictionless interface. Instead of requiring to download, install and open a new program just to book a restaurant, users could do so within the existing messaging apps that they already use daily.

C. Application fatigue

Not only has the popularity of messaging apps skyrocketed, but consumers are increasingly unwilling to download, install, and use new apps. According to Khorozov (2017), a typical user has an average of 30 apps on their mobile device, but uses less than five on a regular basis. Additionally, a 2017 report by Comscore found that 51% of US smartphone users don't download any apps in a single month, 75% downloaded two or fewer, and a third abandon an app after using it only once (Comscore, 2017). The increasing app fatigue among consumers correlates well with the bold but trendy belief that conversational interfaces could replace apps entirely. Instead of needing to download a new app to obtain a service, consumers could simply communicate with a chatbot inside their favorite messaging platform instead. Prominent tech investor Chris Messina penned an article in 2016, claiming "you and I will be talking to brands and companies over Facebook Messenger, WhatsApp, Slack and everywhere else before year's end, and will find it normal" (Messina, 2016).

D. Chatbot frameworks and lower development costs

Developing chatbots has become cheaper and more accessible than before. Over the past few years, the major technology giants including Microsoft, Facebook, and Google have all created frameworks to help speed up development while enabling engineers to utilise the latest in artificial intelligence and related technologies. For example, after launching the Bot Framework in 2016, Microsoft stated that they already had over 30,000 developers building bots for them on the Skype platform (Pall, 2016). When compared to traditional mobile apps, chatbots are typically cheaper to build and maintain due to them being essentially server-side applications with a simple user interface. They can also typically be deployed once without worrying about its adaptation to various screen sizes or operating systems.

3.2.   Chatbot challenges

The previous section highlighted the impressive growth of chatbots over the past decade. However, it must be noted that growth does not equal adoption. While there have been a remarkable number of chatbots built in the past few years, the adoption of those bots is still lacking. Per Jain et al. (2018), a majority of users are first time chatbot users, and 84% of internet users have not used a chatbot yet. The growth of chatbot development has also been

scrutinised, with researchers expressing that their spread is often due to marketing pressure, rather than solving real customer needs (Piccolo et al., 2019). So why haven't chatbots taken off with the consumer public as predicted? The reasons are expressed in the following section.

A. Technical difficulty

Despite advances in artificial intelligence and natural language processing, the technology is still far from the maturity level required to effectively converse with users. David Feldman, a former VP of conversational design at Google, expresses that even the best NLP today is limited compared to even a half-asleep human. He states that Siri and Alexa only understand words but not meaning, indicating a failure of NLP (Feldman, 2018). Furthermore, AI is still largely inaccessible and difficult to maintain because they require high attention to input, output, entity phrases and sentiment analysis in order to process the complexity of human language (Rahman et al., 2017). Jiaqi Pan, founder of chatbot startup Landbot, claims that the previous AI engine used to power their chatbot was growing increasingly chaotic, and that their own engineers could not understand why their bot was saying certain things (Lomas, 2018).

B. Poor user experience

As a consequence of the technical challenges in chatbot development mentioned above, most bots today are built with decision tree logic and lack true linguistic or natural language learning capabilities. Therefore, chatbots lack contextual awareness and can't replicate the non-linearity of human conversation, where multiple topics weave around each other and discussions take unexpected left turns (Feldman, 2018).
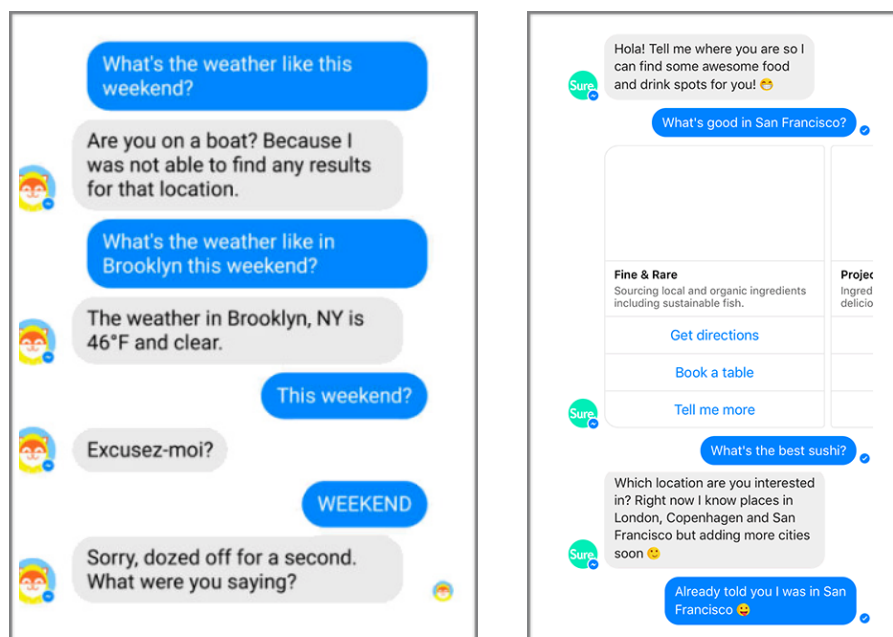
Figure 2. Examples of chatbot failures (Texeira, 2019)

Due to these limitations, users face a frustrating experience when interacting with chatbots. In a chatbot development study, Tavanapour and Bittner (2018) found that users were frustrated when bots consistently failed to understand them, and that users then began attempting to simplify their queries in an effort to predict what the bot would recognise. Pan also expressed that his users were becoming frustrated and attempting twenty different ways to indicate that they wanted to speak with a human (Lomas, 2018).

Furthermore, while chatbots promised a simpler experience by using existing messaging behaviour that was highly familiar, developers began trading app complexity for a different kind of complexity, as users were suddenly forced to type out lengthy instructions by hand instead of faster tap and swipe gestures (Feldman, 2018). The irony is clear when presented with the computing paradigm shift in the early 1980s, when text-based command lines were replaced with graphical user interfaces because the latter was easier and quicker to use.

C. Lack of focused use cases

When the chatbot fanfare was at its peak in 2016, technology experts were enraptured by its potential to make business services more personal, capable of small talk and quick messaging. Some technologists believed that chatbots could replace apps entirely (Dale, 2016). Sam Lessin, a respected Silicon Valley investor, expressed that "the 2016 bot paradigm shift is going to be far more disruptive and interesting than last decade's move from web to mobile apps" (Lessin, 2016).

The hyperbole has led to the development of bots that attempt poorly to accomplish many different things rather than succeed at a few things (Feldman, 2018). The messaging app Kik initially staked its company future on chatbot development, but scaled back their operations a few months later after finding that customers were frustrated by their bot's failure to accomplish several tasks (Griffith & Simonite, 2018). Feldman (2018) concludes that chatbots begin to fail when the language of *replace* is used instead of more nuanced concepts like *extend* or *augment*.

Furthermore, in a study investigating people's motivations for using chatbots, Brandtzaeg and Følstad (2017) found that the 68% majority of those surveyed said that productivity was their key reason for using them. While 20% of participants indicated that it was important for chatbots to be perceived as fun or entertaining, they still needed to help them accomplish a goal.

3.3. Chatbot definition

A quick internet search reveals many different chatbot definitions. The terminology varies based on the design techniques, interaction style, and technology used. For example, Ranoliya et al. (2017) define chatbots as "programs that mimic human conversation using artificial intelligence", implying that they must have AI to be considered a chatbot. Meanwhile, Rahman et al. (2017) highlight the interaction style, calling them "is [sic] a virtual person who can effectively talk to any human being using interactive textual skills".

For the purposes of this thesis, a broader chatbot definition will be used. Chatbots will later be segmented into different types based on the development methods mentioned above.

The definition of chatbots for this thesis is:

A chatbot is a digital service that can be interacted with through a chat interface in voice, written, or graphical input format.

This definition gives chatbots flexibility in their interaction modality, as users may use voice, written text, or traditional GUI elements to interact with it. Additionally, chatbots do not require artificial intelligence, as many employ simple rule-based logic and pattern matching to help users achieve their goals (McTear et al., 2016). Furthermore, the capacity to converse in a life-like manner is also not a prerequisite to being a chatbot. Given the current technological challenges in AI and NLP, several companies today have gradually used conventional tap and swipe gestures to replace the amount of chatting required in order to help users accomplish their goals faster (Brandtzaeg & Følstad, 2017). Therefore, the primary distinction between a chatbot and another service is that it must be contained within some form of chat interface.

According to Radziwill and Benton (2017), conversational agents are the overarching umbrella that encompass many types of conversation-based systems, including voice-based conversational agents, embodied conversational agents, and chatbots, the focus of this thesis. Voice-based conversational agents are probably the most well-known to the public, which are the voice-driven digital assistants from the major technology companies including Amazon's Alexa, Apple's Siri, and Google's Assistant (Dale, 2016). These digital assistants are considered distinct from chatbots because they primarily exist outside of a traditional messaging interface (Radziwill & Benton, 2017). Embodied conversational agents meanwhile are computer-generated animated characters that combine facial expression, physical body language and speech gestures to mimic humans. They are also distinct from chatbots because they are typically deployed in role-play, simulations, or immersive virtual environments (McTear et al., 2016). Figure 3 illustrates these categorisations.
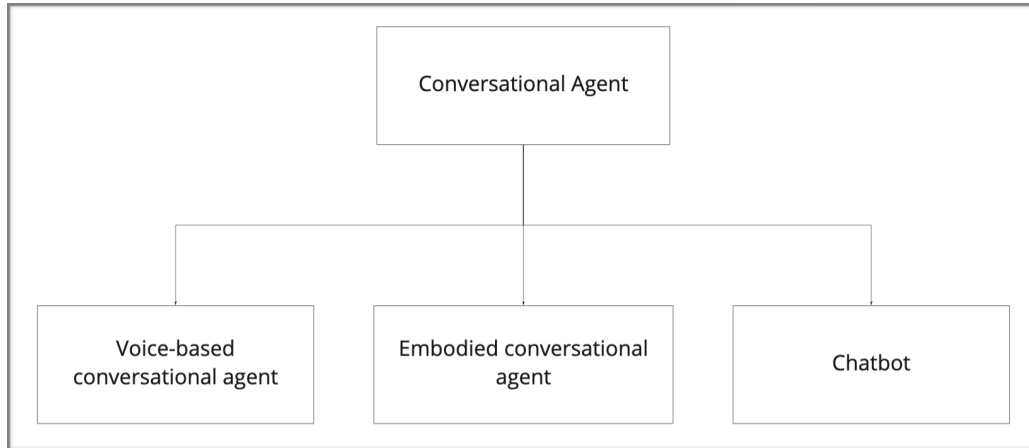
Figure 3. Conversational agent categorisation

3.4. Chatbot classification

With chatbots now categorised within the conversational agents group, a classification of chatbots will be provided, along with an overview of the associated design techniques and technologies.

3.4.1. Categories

Chatbots can be classified into different categories based on the user scenarios, related design techniques, and technologies. Ramesh et al. (2017) highlight three broad categories which are presented below.

A. Retrieval-based vs. Generative-based

Retrieval-based chatbots provide pre-defined responses based on the user's input and context. To determine the response, a variety of heuristics can be used, from simple rule-based expression matching, to a more complex combination using machine learning classifiers. Retrieval-based chatbots do not create new responses on their own—they choose from a pool of pre-defined responses that the developer creates. In contrast, generative chatbots are able to create new responses that are built from applying a set of machine translation and artificial intelligence techniques.

Both methods have distinct advantages and disadvantages. Retrieval-based chatbots are much easier to develop and make less grammatical errors because they pick from a pool of pre-defined responses. However, they are unable to handle unfamiliar inputs or understand context. Thus, information previously mentioned in a conversation like names or places can not be referred to again in this model. Generative models meanwhile are considered smarter because they can understand context and create their own responses. The challenge is that they are very difficult to build, prone to grammatical errors, and usually require huge amounts of training data.

B. Long vs. short conversations

Long versus short conversations relate to the complexity of the conversations that the chatbot need to handle. The longer the conversation, the harder it is to automate. Short conversations meanwhile typically require only one or two inputs and responses. An example of a long conversation scenario is customer care, where the customer's problem is unique to them and requires several rounds of questions and answers to complete. Information retrieval, such as finding a store's opening hours, is a typical example of short conversations.

C. Open vs. closed domain

Human conversations often weave from one topic to the next, without any particular reasoning. These conversations that can change domains are referred to as open domain conversations, where the chatbot models are not designed to serve a specific purpose. Social media conversations on Facebook and Twitter are examples of open domain conversations. They require considerable amounts of training and data to generate reasonable responses. Closed domain conversations meanwhile focus on a limited set of knowledge for a specific topic in order to provide an appropriate response. An example is a restaurant reservation system, where the system has a specific purpose (to book a table) and the conversation is based around that task. These systems generally use domain specific data for training.

3.4.2. Design approaches

Most chatbot systems today and historically employ pattern matching techniques, which utilise rules to generate appropriate responses. With the improvements in artificial intelligence and machine learning techniques, chatbot design is beginning to implement more sophisticated neural network modelling techniques. The two approaches are presented in broad terms in this section.

A. Pattern matching

Pattern matching is the most widely used methodology for designing chatbots, as some form of pattern matching algorithm is present in nearly every chatbot system. Pattern matching consists of identifying the structure of a sentence and providing a pre-defined response that can change according to the characteristic variables of the sentence (Ramesh et al., 2017). While pattern matching is relatively simple to implement, it suffers from responses that can become predictable and repetitive.

B. Neural network models

To overcome the limitations of pre-defined responses from pattern matching, modern approaches using neural networks have been utilised in chatbot design over the last few years. The general aim is to generate a target sequence by

examining the source sequence. In chatbots, the source sequence is the chat message from the user, while the target sequence is the machine's reply.

A basic sequence to sequence model uses two recurrent neural networks and an encoder-decoder which uses a sequence as input and generates another sequence as output. The encoder maps the incoming sequence to an encoded representation of that sequence, which is then used by the decoder to generate an output sequence. Using the neural network model technique allows chatbots to create their own responses based on past information, but these models are very hard to train, are prone to grammatical errors, and usually require huge amounts of training data (Ramesh et al., 2017).

3.4.3. Technologies

The technologies used to support the chatbot design approaches are presented broadly in this section.

A. AIML

AIML (Artificial Intelligence Markup Language) is an XML-based markup language used to design chatbot conversations and is fully based on pattern matching. It is made up of fundamental knowledge units called categories, which contain pattern tags that define what the user is saying, and template tags that define the computer's response. The three time Loebner prize winner in 2000, 2001 and 2004, ALICE, was built using AIML (Wallace, 2009).

B. ChatScript

While AIML is uncomplicated to learn, it has relatively weak pattern matching and is difficult to maintain. ChatScript is the successor to AIML and contains rules that are associated with topics. The best topic is associated to the user's input, which then fires the linked rule. ChatScript was released in 2011 by Bruce Wilcox for implementation in the Suzette chatbot, winning a Loebner prize (Bradeško & Mladenić, 2012).

C. Natural Language Processing and Natural Language Understanding

Natural language processing and natural language understanding are more flexible approaches to designing chatbots, as they parse conversations into intents, entities, and contexts. Intents are used to classify the goal of what the user is saying and the appropriate response, while entities extract specific variables from the intent, such as dates and times. Contexts are strings that store the context of the object that the user is talking about (Rahman et al., 2017).

3.5. Overview of onboarding

3.5.1. The business case for onboarding

New hire onboarding—who cares right? After all, from the company's perspective, the new employee has already joined and is eager to make a good first impression. In addition to meeting new colleagues, the new employee typically uses the first few weeks to get acclimated to their new workplace, including learning the company's policies, culture, and how to perform administrative tasks like time reporting. These activities have traditionally been considered to be inconsequential and improving the onboarding process is near the bottom of a company's priority list (Brown, 2007).

Critically, recent research has shown that onboarding plays a much bigger role in employee satisfaction and company success than previously thought, such as being positively related to job satisfaction, engagement, organisational commitment, performance, and inversely related to turnover (Meyer & Bartels, 2017). Harpelund (2019) states that in the US, 25% of new hires leave their company within the first 12 months, the cost of losing a new employee within the first 12 months equals to roughly 200–300% of their yearly salary, and that companies with good onboarding practices experience 54% higher productivity from their newly employed and twice the level of engagement. Despite these statistics, a Gallup poll found that only 12% of US employees strongly agreed that their organisation does a great job of onboarding new employees (Gallup, 2012).

Snell (2016) summarises the benefits of improved employee onboarding as the following:

- Reduced time and effort for HR, hiring managers, and others involved in onboarding.

- Improved speed and accuracy of data collection and transfer between payroll and HRIS systems.

- Ability to track new metrics for greater process efficiency.

- Better overall new hire experience including a single, self-service source of information during the crucial first days on the job.

- More effective employee–manager communication.

The business case for improving employee onboarding is becoming clearer and one of the keys to successful onboarding is a technology-supported platform that can handle data gathering, input, and display of HR information (Snell, 2006). Chatbots can assume the role of this technology-supported platform, as they open new possibilities for engaging new employees and customers alike (Robinson et al., 2019).

3.5.2. Scope of onboarding

The definition of onboarding can be very broad, encompassing a wide range of human resources related activities. Therefore, onboarding must be classified within the scope of this thesis to understand which scenarios the created chatbots will and will not solve.

Onboarding is the process by which newly hired employees learn about the business including daily functions, job responsibilities, culture, and values (Pike, 2014). The goal of successful onboarding is to bring new employees from being organisational outsiders to becoming organizational insiders quickly so they can acclimate and contribute (Bauer & Erdogan, 2011). Onboarding can be broken into two categories— person–organisation fit, where the employee learns about the organisation's values, people, benefits, and daily routines. This contrasts with person-job fit, where the employee is focused on learning about their job responsibilities and expectations (Pike, 2014). Furthermore, onboarding is distinct from orientation, which is often a discrete, stand-alone event conducted by an HR representative that focuses on transactional tasks like filling out benefits forms and paperwork (Graybill et al., 2013).

The chatbots developed in this thesis focus on the category of person-organisation fit within onboarding.

4.   Methods

This section provides an overview of the methods that were used to design, develop, and evaluate the conversational versus menu-based chatbot. The project was based on methods from the classic ISO 9241 standard for human-centred design. The standard provides a framework for human-centred design, is intended to be complementary to existing design methodologies, and can be integrated into different design and development processes in a way that is appropriate to the particular context (ISO 9241-210, 2010). In general, ISO 9241 identifies four primary human-centred design activities:

1. Understanding and specifying the context of use (Research)

In this step, users are studied to understand their context of use, which includes their characteristics, environment, problems, and needs.

2. Specifying the user requirements (Analysis)

Step 2 specifies the general features or specific requirements that the system needs to accomplish.

3. Producing design solutions (Design)

Step 3 represents the actual design and development of the system.

4. Evaluating the design (Evaluate)

The last step includes testing the completed product with the users of the system.

It should be noted that the ISO standard describes the design process as iterative in nature, and that designs should be updated based on feedback. However, iterations were not applied in this project, given that the two chatbots were being developed and evaluated under a tight timeframe.

An illustrative overview of the ISO 9241 design process is provided in Figure 4.
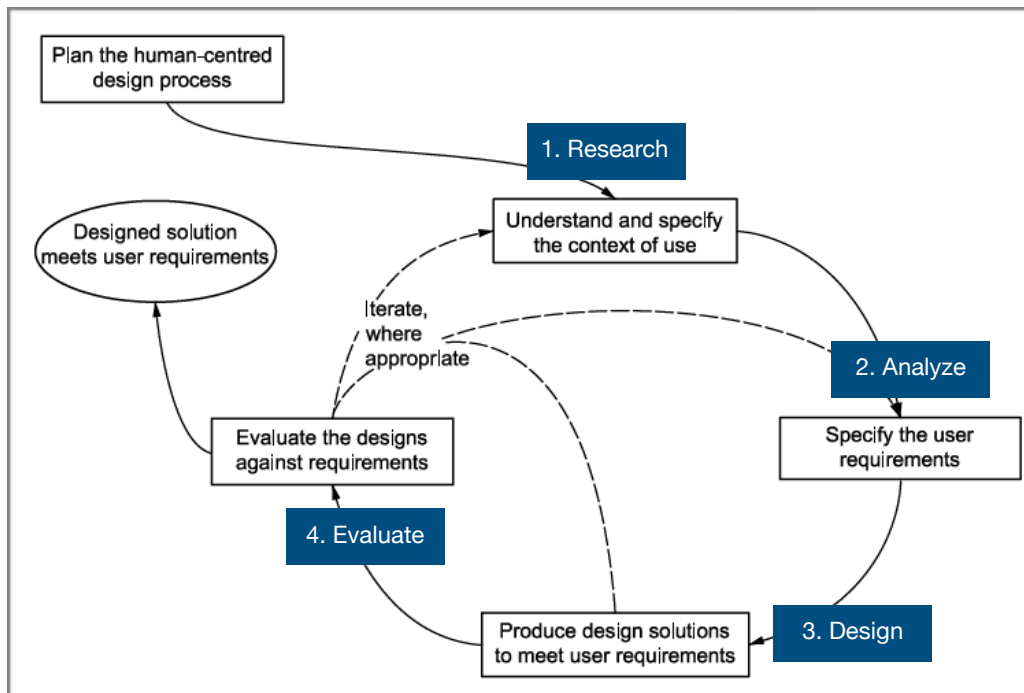
Figure 4. ISO 9241 Human-centred design process

4.1. Research

In this phase, the users of the system and relevant stakeholders are identified, along with their context of use, which includes the user characteristics, goals, and system environment.

4.1.1. Users and stakeholders

The primary users of the chatbots were identified as new employees at Apegroup, defined as those who have been with the company for three months or less. However, given that the primary scope of employee onboarding for this project is related to obtaining relevant company information, all employees can still benefit from the chatbot. Therefore, all employees, representing those that have been with the company for longer than three months, were identified as secondary users. The stakeholders are the human resources employees at the company who will provide insights on their current process and how employees obtain company information today.

4.1.2. Context of use

To understand the context of use, individual semi-structured interviews lasting 30 minutes each were conducted with four new employees, two existing employees, and two HR professionals at Apegroup. Semi-structured interviews were used because they provide in-depth experiences of individuals by enabling a predetermined set of open-ended questions while allowing for more

probing when needed (DiCicco-Bloom & Crabtree, 2006). The interviews were recorded with permission from the interviewees. The interview questions were designed to first break the ice with the interviewees and build rapport rapidly, an essential component of semi-structured interviews (DiCicco-Bloom & Crabtree, 2006). The rest of the questions were designed to identify their needs regarding onboarding, and finally their experience with chatbots.

The interview templates for regular and HR employees are shown in Figures 5 and 6.

Name of interviewee
Date

1. What's your role at Apegroup?
2. How long have you been with the company?
3. What was your onboarding experience like at Apegroup?
   a. Why do you feel that way?
4. What type of HR information would you be looking for?
5. What were some questions you had when you first joined the company?
6. Do you have any HR-related questions?
7. Do you know what Apegroup benefits are?
8. Do you know how to file for maternity/paternity leave, work from home, or vacation policy?
9. Have you used chatbots before? How do you feel about them?
10. If there was a chatbot for HR, what type of things do you wish it could do?

Figure 5. Interview questions for new and existing employees

Name of interviewee
Date

1. What's your role at Apegroup?
2. How long have you been with the company?
3. How does the onboarding process work at Apegroup?
4. What would successful onboarding look like to you?
5. How do you wish the onboarding process worked?
6. What do new employees typically ask about?
7. What would make your life easier?
8. Have you used chatbots before? How do you feel about them?
9. If there was a chatbot for HR, what type of things do you wish it could do?

Figure 6. Interview questions for HR employees

The relevant characteristics for the employees were that they were educated and technically savvy, which was expected as they were all technology professionals. Their experience with chatbots however was consistently

limited, with one new employee expressing that she has never used one before. Most of them have used chatbots only once or twice, primarily for online customer service. Six out of eight interviewees expressed moderate attitudes towards chatbots, saying that they liked them only if they were useful and not too complex. Two out of eight interviewees had negative attitudes, saying that their experiences with them had many problems due to the bot being very limited and was unable to understand what they were saying.

Regarding problems and needs, a major pattern emerged amongst the six non-HR employees— the new hire onboarding process at Apegroup was poor in general. Five of the six new and existing employees expressed negative attitudes towards the process. Three of them indicated that it felt rushed, disorganised, and lacked clear information regarding benefits and policies, while two of them said they didn't have onboarding at all because they were needed on an urgent project. One of the six employees indicated that her onboarding experience was fine because the HR representative had extra time to answer her questions, and she was diligent in writing down the answers. Another interesting insight was that two of the employees said that not having good onboarding was expected—they had bad experiences at their previous companies as well. On a positive note, all six of the new and existing employees said that the technology setup portion of the onboarding process was very good.

From the HR representative side, they acknowledged that the onboarding process needed improvement because they had a severe lack of time and resources. There were only two HR employees at the company—one of them handled technology setup while working full-time as a developer, while the other was solely responsible for all HR management, including recruiting. They expressed that a chatbot that could answer common questions about HR would be useful. Cultural fit and values were also critical to hiring at Apegroup, so they said that a chatbot that could communicate the company's values would be useful.

More details regarding the needs and goals of the users will be provided in section 4.2 Analyze method.

4.1.3. Environment

It was also observed and recognised across Apegroup that all of the communication occurs in Slack, a popular workplace communication platform. Therefore, it was important that the chatbots are deployable to Slack.

4.2.　Analyze

To determine what requirements or features should be in the chatbots, an affinity diagramming exercise was conducted. Affinity diagraming is a popular industry exercise for identifying design ideas from large sets of research findings (Pernice, 2018). Affinity diagramming consists of writing down all the observations, findings, or ideas from research sessions onto post-it notes, and then clustering similar post-it notes together. The clusters represent insights, which can later be prioritised.

All eight recorded interviews were reviewed, and findings from each of them were placed onto sticky notes. Similar sticky notes were then clustered together, which revealed the needs that the chatbots should fulfil. The affinity diagram was created using Miro[6], an online SaaS tool for brainstorming. The affinity diagram can be found in Appendix A. The needs from the affinity diagram were translated into high level requirements, as described below:

- **Ability to show key people at the company:** A pattern emerged in that getting to know the people in the company was very important to new hires. In particular, they wanted to know key stakeholders such as managers and HR representatives so they could ask questions. Knowing their team members was also important, but this was excluded from the MVP for the chatbots due to time constraints.

- **Ability to show popular lunch spots and explain lunch habits:** New hires, especially on their first day, wonder about how lunch works at the company, particularly if it's common for people to eat out together or bring their own food. They also want recommendations on popular lunch spots.

- **Ability to show key employee benefits:** A pain point for new hires, the interviewees said they lacked clear information regarding their benefits, such as the fitness and educational budget. The HR employees also said that emphasising these benefits would be a good way for new employees to feel happy and engaged with the company.

- **Ability to show company values:** Although the newly hired employees did not indicate this was a need, the interviewed HR staff said this was very important for the company. Therefore, this was prioritised into the MVP.

- **Ability to show key HR policies:** Knowing the policies around common HR activities such as work from home were important. Interestingly, the policies were still very unfamiliar to even the existing employees.

---

[6] https://miro.com/

- **Ability to show unspoken office culture:** New hires also expressed interest in knowing more about the office culture. These are things that they may feel uncomfortable asking someone in person, such as when do people typically leave from work, or whether it's appropriate to drink alcohol in the office.

4.3.   Design

Prior to implementation, the usability, personality traits, and conversation flow of the chatbots must be designed. Those elements are described in this section.

4.3.1. Usability and User Experience

There is currently a lack of academic literature regarding user experience and usability best practices for developing chatbots. This is due to chatbot design being a relatively new field when compared with traditional web design, which has numerous established principles such as Nielsen's ten usability heuristics. Abdul-Kader & Woods (2015) conclude that researchers are reluctant to divulge improvement techniques, and that chatbot design practices are still a matter of debate with no common approach yet identified.

Nonetheless, industry experts have created their own set of best practices. Design principles from Shevat (2017), Hanson (2019), and Facebook (2019) were reviewed, and the relevant strategies are described below:

4.3.1.1. Set expectations: Begin by clearly explaining what the chatbot is for and what it can do. List out everything it can do if possible.

4.3.1.2. Be brief: People have limited attention. Messages should be brief, clear, and straight to the point.

4.3.1.3. Provide context: Ensure users understand where they are, what is being asked, and what will happen next.

4.3.1.4. Fail gracefully: If the bot does not understand a request, clearly state that it doesn't understand and reiterate the possible options.

4.3.1.5. Provide flexible navigation: Allow users the flexibility to go back to their previous step or escape the path altogether.

4.3.1.6. Use contractions: Use shortened versions of words or groups of words to feel less robotic and more casual.

4.3.1.7. Be conversational: Avoid plain, impersonal statements which sound robotic, such as "The email you entered is invalid."

4.3.1.8. ***Allow multiple expressions:*** Humans express the same thing in multiple ways ("Hi", "Hello", "Hey there"). Chatbots should allow diverse inputs and also have diverse responses

4.3.1.9. ***Use emoji and rich content such as images sensibly:*** Emoji and images can give chatbots more personality, but avoid excessive use.

*4.3.1.10. **Don't leave dead ends:*** Ensure every path has an appropriate ending, or offer a return to the primary path.

4.3.2.   Personality and tone of voice

Given that chatbots are conversational in nature, they have personality and can humanise a product (Radziwill & Benton, 2017). Norman (2004) concludes similar sentiments in his book, Emotional Design, stating that "People can more easily relate to a product, a service, a system, or an experience when they are able to connect with it at a personal level."

Given that the corporate identity of Apegroup is very lighthearted and aloof, the chatbots should also feel casual, like talking to a friend. Personality in chatbots is manifested in the tone of voice through the copy. The copy will therefore be casual and informal while respecting the design principles in section 4.3.1, which highlights that conversations should be brief and concise without feeling robotic. The personality will also be reflected in bright colours and humour where appropriate.

4.3.3.   Conversational flow diagram

A conversational flow diagram was also created. They are critical because they serve as the blueprint of the bot and provide an overview to ensure that users are led throughout the entire journey without friction (Jassova, 2019). The diagram is depicted in Figure 7. The white boxes represent bot actions, while the black boxes represent user input.
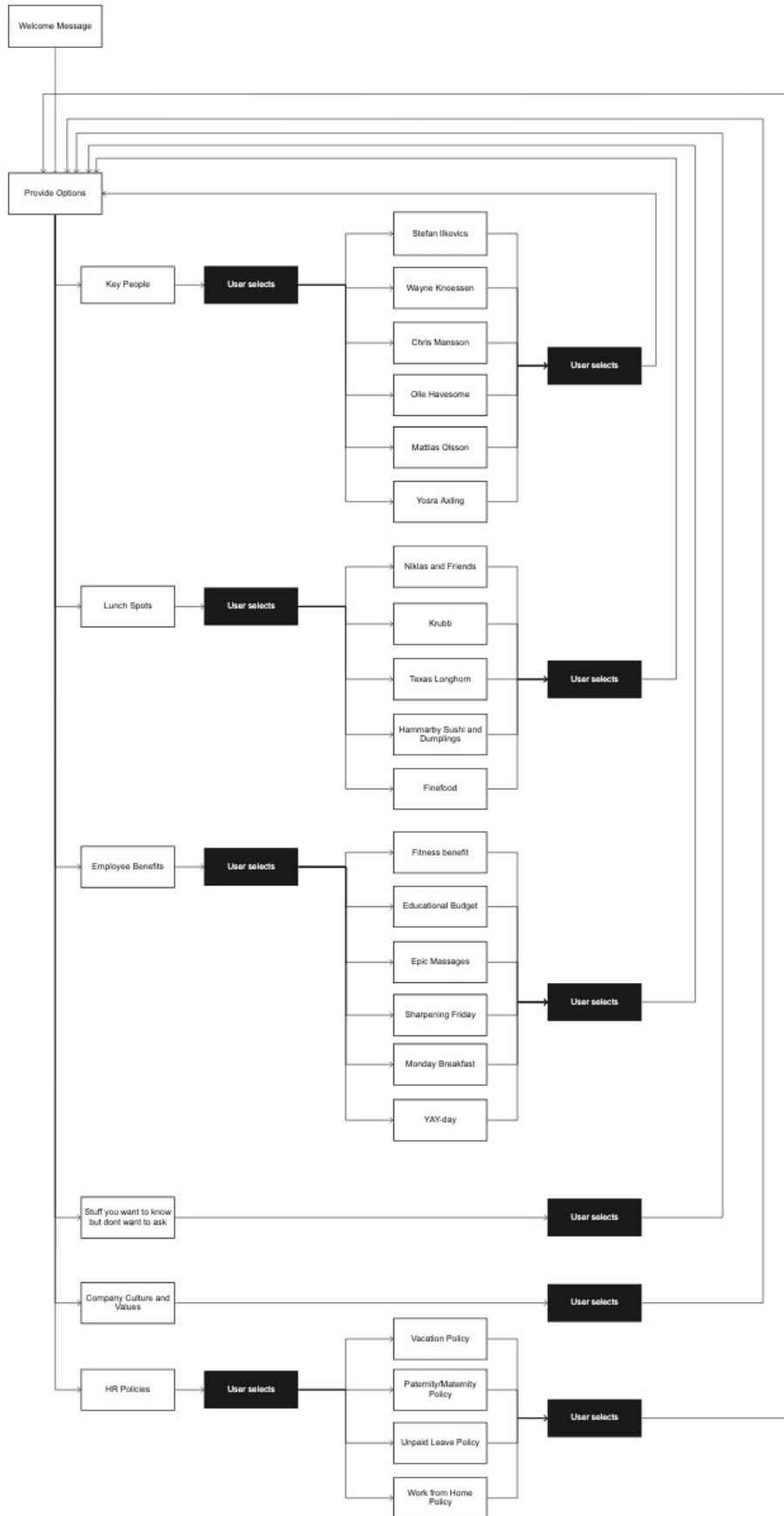
Figure 7. Conversational Flow Diagram

After the flow diagram was created, basic conversational scripts were written based on the design principles mentioned in section 4.3.1 Usability and User Experience. The contents of the script and details regarding navigation patterns will be further elaborated on in section 5 Implementation.

4.4.   Evaluate

This section describes the participants and evaluation methods for the chatbots. 17 participants tested the chatbots and were evaluated with qualitative and quantitative methods.

4.4.1. Participants

A total of 17 participants were involved with testing and evaluating the two chatbots. The 17 participants were separated into three groups for qualitative evaluation. Group 1 consisted of seven people who tested both chatbots, one after the other. The sequence in which they tested the two chatbots was randomised in order to minimise any unexpected dependencies. Group 2 consisted of five people who tested only the conversational chatbot and group 3 consisted of five people who tested only the menu-based bot. After the tests were completed, all the participants were given a survey for further quantitative evaluation. The participants were chosen by the researcher to include as many new hires as possible from the company, and include employees from as many different teams as possible to maximise the sample diversity. In total, the participants consisted of seven new hires (those employed three months or less) and ten regular employees (employed over three months), from three different teams (business, design, engineering).

4.4.2. Qualitative evaluation

Qualitative evaluation was performed using the simplified cognitive walkthrough method, in which users are given a high level scenario with a goal and instructed to use the service (chatbots in this case) to achieve the goal (Rieman et al., 1995) without explicit guidance. The walkthrough was supplemented with specific tasks when needed to guide the users through all functions of the chatbot. The simplified cognitive walkthrough is suitable because the chatbots' primary use case, obtaining HR information, is exploratory instead of task-based in nature. Employees at Apegroup choose information that is relevant to them at their own leisure.

Test participants were encouraged to think aloud as they explored the interface, because it provides a quick and informal way for the researcher to understand the participants' thoughts and ask follow-up questions when needed (Van Someren et al., 2015). The cognitive walkthrough tests were conducted in the participant's real world context, the Apegroup office. Each

test consisted of one participant and the researcher. The researcher took notes, indicating positives and negatives about the experience.

Prior to beginning the simplified cognitive walkthrough test, each participant was asked three questions to put them at ease and assess their experience with chatbots.

The three questions are below:

1. What is your role?

2. How long have you been working at Apegroup?

3. Have you used chatbots before?

Next, the participants were presented with the following scenario and encouraged to explore the chatbots on their own:

You are a new employee at Apegroup and this is your first week. Your goal is to learn HR-related information about the company, and this chatbot has been provided for you to explore.

To ensure that the full spectrum of functions was experienced by the participant during the test, a list of six tasks was available for the researcher to ask the participant to perform if they did not do so on their own. Each task was marked with an X to indicate whether it was completed successfully or needed help.

Once the test was finished, the participant was asked what they thought about the chatbot experience. Participants that tested both chatbots were also asked which experience they believed was better. Figure 8 depicts the qualitative user test template used.

Name, Date, Type of bot	What is your role?	How long have you been working at Apegroup?	Have you used chatbots before?
Scenario			
You are a new employee at Apegroup and this is your first week. Your goal is to learn HR-related information about the company, and this chatbot has been provided for you to explore.			
Tasks	**Successfully done**	**Needed Help**	**Additional Comments**
Find out about a key person			
Find a lunch spot			
Find out about company culture and values			
Learn about something you wanted to know but didnt want to ask			
Learn about a benefit			
Learn about a policy			
Researcher observations and notes			
Additional questions			
What do you think about this experience?			
Which of the two chatbots did you prefer? (two chatbot test only)			

Figure 8. User test template

4.4.3. Quantitative evaluation

After the qualitative chatbot test was completed, each participant was given a quantitative survey created on Typeform[7] with ten questions to complete. The questions were intended to assess how the chatbots successfully fulfil the ISO definition of usability—effectiveness, efficiency, and satisfaction, and also measure whether they were perceived as intelligent. Group 1, which tested both chatbots, was given two surveys, one for each bot, after they finished testing both bots.

The survey questions were based on a 7-point Likert scale, as it is an extremely popular and fundamental psychometric tool used in user experience and social sciences research. A 7-point scale may also perform better than a 5-point scale as it provides more variety of options which in turn increase the probability of meeting the objective reality of people (Joshi et al., 2015). Numerous usability metric evaluation systems were assessed, including the Usability Metrics for User Experience (UMUX), Technology Acceptance Model (TAM), and System Usability Scale (SUS). The UMUX questionnaire was ultimately adopted because it consists of only four questions but is very correlated (0.8) to the SUS method, which consists of ten questions (Finstad, 2010). In a separate study, Lewis et al. (2013) also confirms that UMUX is a viable alternative to the SUS if the researcher requires a shorter questionnaire. Furthermore, Finstad (2010) asserts that the UMUX questionnaire is representative of the ISO definition of usability. The TAM was rejected due to question overlap with the UMUX. Therefore, the four questions from UMUX were included, along with six additional questions designed to measure the user's perception of intelligence in the chatbot and overall satisfaction. The popular net promoter score question was one of the six included, as it is an effective indicator for user satisfaction (Grisaffe, 2007).

The ten questions and their categorisation is listed in Figure 9.

Category	Question
Usability	The possibility of this chatbot meet my requirements.
Usability	Using this chatbot is a frustrating experience.
Usability	This chatbot is easy to use.
Usability	I waste too much time on correcting things in this chatbot.
Intelligence	To what extent did this chatbot seem personal?
Intelligence	To what extent did this chatbot seem intelligent?
Intelligence	To what extent did this chatbot seem human-like?

[7] https://www.typeform.com/

Category	Question
Satisfaction	How likely are you to use this chatbot again if it were implemented for real?
Satisfaction	Overall, how satisfied are you with the system?
Satisfaction	How likely are you to recommend this chatbot to a friend or colleague?

Figure 9. Quantitative survey questions

5.   Implementation

This section describes the implementation details for the two chatbots.

5.1.   Platform selection

Before developing the chatbots, their level of intelligence must be defined in order to select the most appropriate development platform. The intelligence definition is based on the classification mentioned above in section 2.4, where chatbots are classified into categories (retrieval vs. generative, long vs. short conversation, open vs. closed domain), design approaches (pattern matching vs. neural network models), and technologies (AIML, Chatscript, NLP/NLU).

The intelligence definition and selected chatbot development framework for each bot is defined below.

Chatbot 1: Conversational bot with natural language processing

This chatbot interacts with the user primarily through text messaging, therefore it is critical that some level of natural language processing is included. AIML and ChatScript languages are excluded because the bot will be built with an existing development platform. Given that the use case of new hire onboarding has a specific and limited knowledge base, the bot will be retrieval-based, utilising short conversations in a closed domain. Doing so also limits development complexity.

Several popular platforms were considered for the development of this chatbot, including Google Dialogflow[8], IBM Watson[9], Wit.ai[10], Chatfuel[11], and

---

[8] https://dialogflow.com/

[9] https://www.ibm.com/watson

[10] https://wit.ai/

[11] https://chatfuel.com/

Manychat[12]. Ultimately, Dialogflow was chosen because it was the easiest to use while maintaining AI/NLP capabilities, had the most documentation, and could be deployed to Slack, the primary communication channel for all Apegroup employees. Watson and Wit.ai were eliminated because they required more technical knowledge, while Chatfuel and Manychat were eliminated because they could only be deployed to Facebook Messenger. Dialogflow had also been experimented with on a previous project at Apegroup and was recommended by the company.

Chatbot 2: Menu-based bot

The menu-based bot is very simple and does not use traditional chatbot elements like intelligence or messaging. Instead, this chatbot wraps conventional, menu-based interaction into the form of a bot. Therefore, it does not use any of the previous technologies mentioned, nor pattern matching or neural networking models. Because this bot serves the same use case of employee onboarding, it is also retrieval-based with short conversations in a closed domain.

Landbot[13] was the platform selected, as it was the only platform identified that offers an easy to use, drag and drop interface for creating purely menu-based bots. A weakness of Landbot is that it does not have native Slack integration, but a workaround was reached by creating a dedicated Slack channel with a link to the chatbot as a default message.

5.2.   Conversational bot implementation

Link to conversational chatbot: N/A, Dialogflow does not allow public access to bots that are integrated with Slack.

Link to conversational chatbot demo: https://youtu.be/GXHMDZfV9Us

The conversational flow diagram was implemented into the structure provided by Dialogflow. Dialogflow contains four primary building blocks used to create functioning chatbots.

- **Training phrases:** Training phrases are the messages that the user sends to the chatbot during interaction.

- **Intents:** Intents are configured by the chatbot developer and are used to interpret the user's goal, based on the training phrase. For example, an intent called GetLunchSpots could be mapped to a training phrase that asks "where can I go get some food?".

---

[12] https://manychat.com/

[13] https://landbot.io/

- **Entities:** Entities are used to pick out specific pieces of information that users mention in their inputs—anything from product names, to dates, or amounts with units.

- **Responses:** Responses are how the chatbot responds to the user, after it has analysed the related intents and entities.

The relationship between the building blocks within Dialogflow is depicted in Figure 10. The list of intents, entities, and responses can be found in the conversational bot architecture in Appendix B.
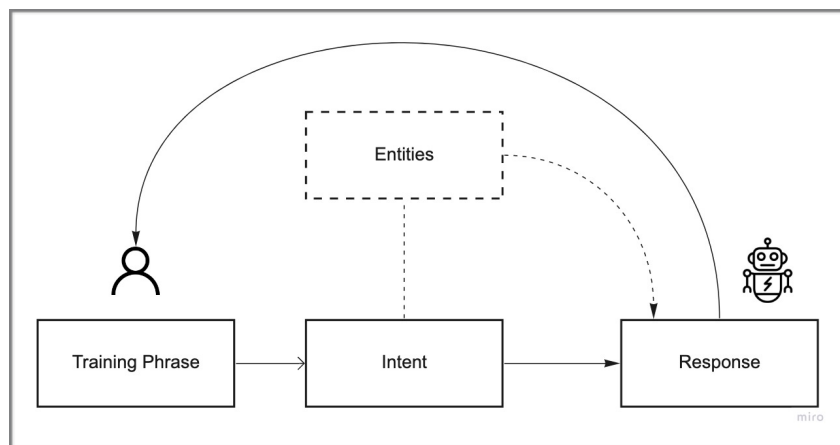
Figure 10. Dialogflow bot structure

To increase the conversational and personal feel for the intelligent bot, the conversation was designed so that the bot begins the chat by first asking the user for their name and team at Apegroup. After the initial introduction, the chatbot communicates its purpose and shows its available options in accordance to design principle *4.3.1.1 Set expectations.* This is done so that the user immediately knows what the chatbot is capable of and can ask questions accordingly. This example is shown in Figure 11.
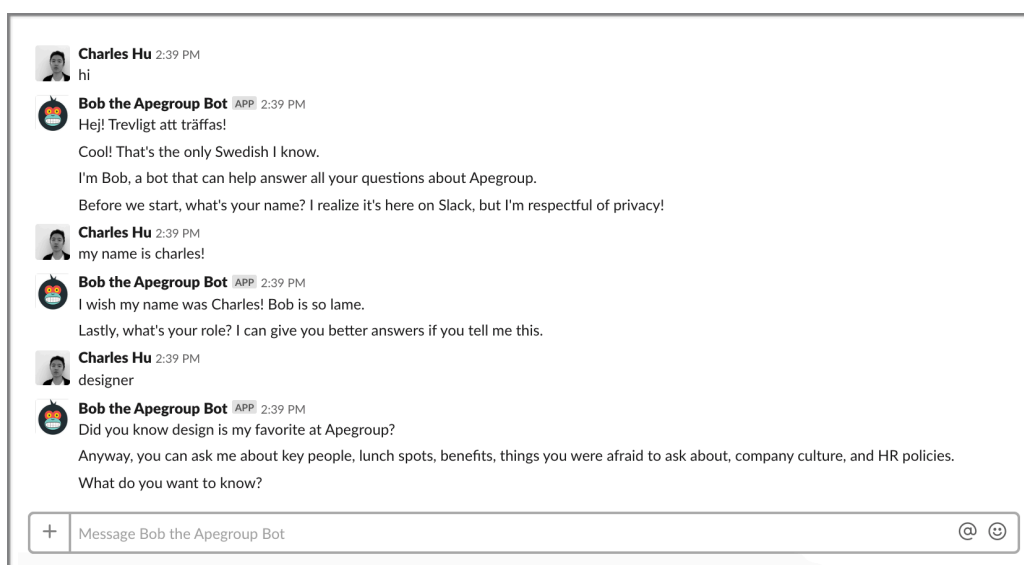
Figure 11. Introduction of conversational chatbot

In accordance to design principle *4.3.1.5 Provide flexible navigation*, the user always has the ability to escape the current path and view all available options that the chatbot provides before choosing another topic. This is done to provide user flexibility and create a more natural, life-like conversation, where topics can shift quickly. Figure 12 depicts this example.
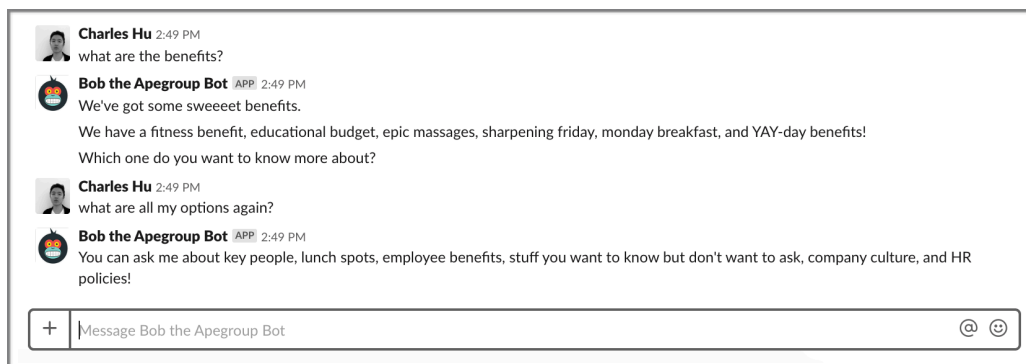
Figure 12. Flexible navigation in conversational bot

Given that the intelligent bot is conversational, it inevitably fails to understand certain user utterances given the diversity of human language and limitations of NLP. When this happens, the chatbot follows design principle *4.3.1.4 Fail gracefully*. The chatbot apologises and rather than ending the conversation, it reiterates the available options. Additionally, the chatbot follows design principle *4.3.1.8 Allow multiple expressions*. It can respond and apologise in multiple ways ("Sorry could you say that again?", "I missed that, say that again?"). This makes the chatbot feel more natural and less robotic. These examples are shown in Figure 13.
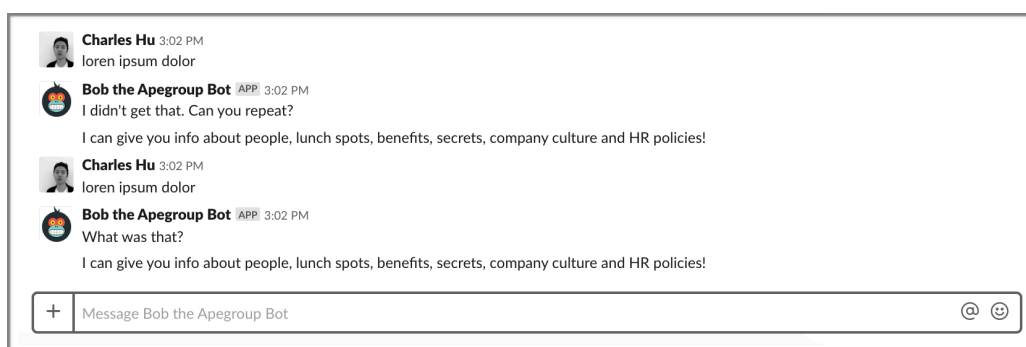
Figure 13. Providing error fallback with multiple expressions in conversational bot

The chatbot also follows design principles *4.3.1.2 Be brief*, *4.3.1.6 Use contractions*, *4.3.1.7 Be conversational*, and *4.3.1.9. Use emoji and rich content such as images sensibly* through out the experience. These principles are designed to make the chatbot feel more like a real conversation, and reflect

the casual personality defined in section 4.3.2 Personality and Tone of Voice. Examples of these principles can be found in Figure 14.
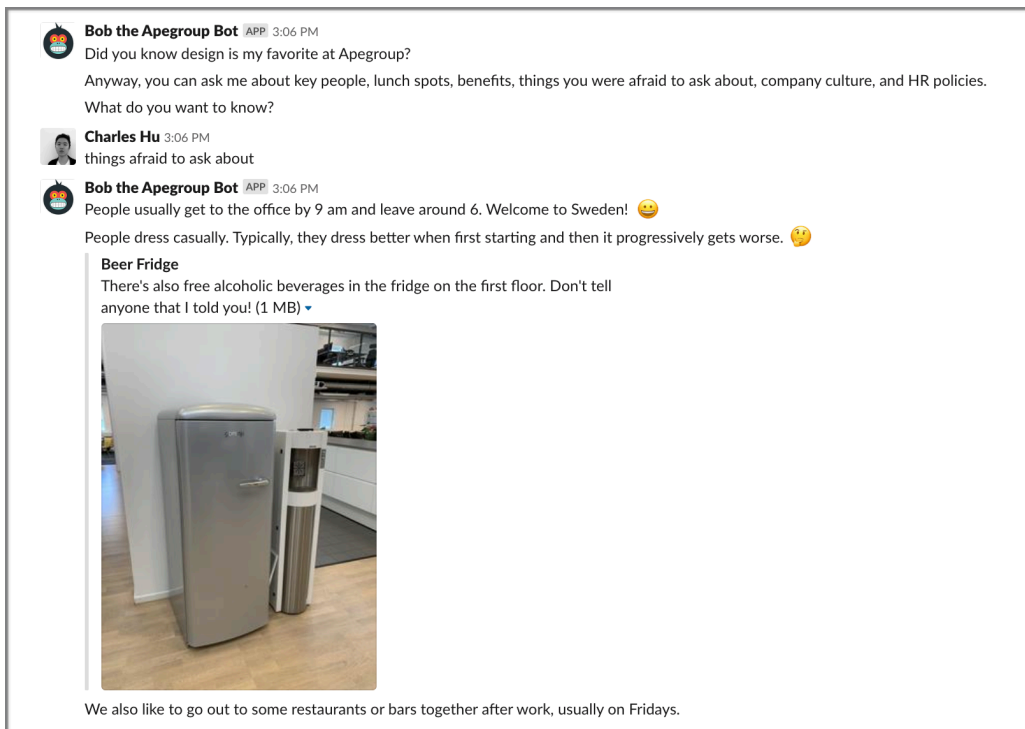
Figure 14. Personality and brief messages in conversational chatbot

After the user is finished obtaining information from one path, the chatbot follows design principle *4.3.1.10 Don't leave dead ends* by presenting all of its different options again. This ensures that a user who wants more information is not stuck after completing one flow. Figure 15 depicts this example.
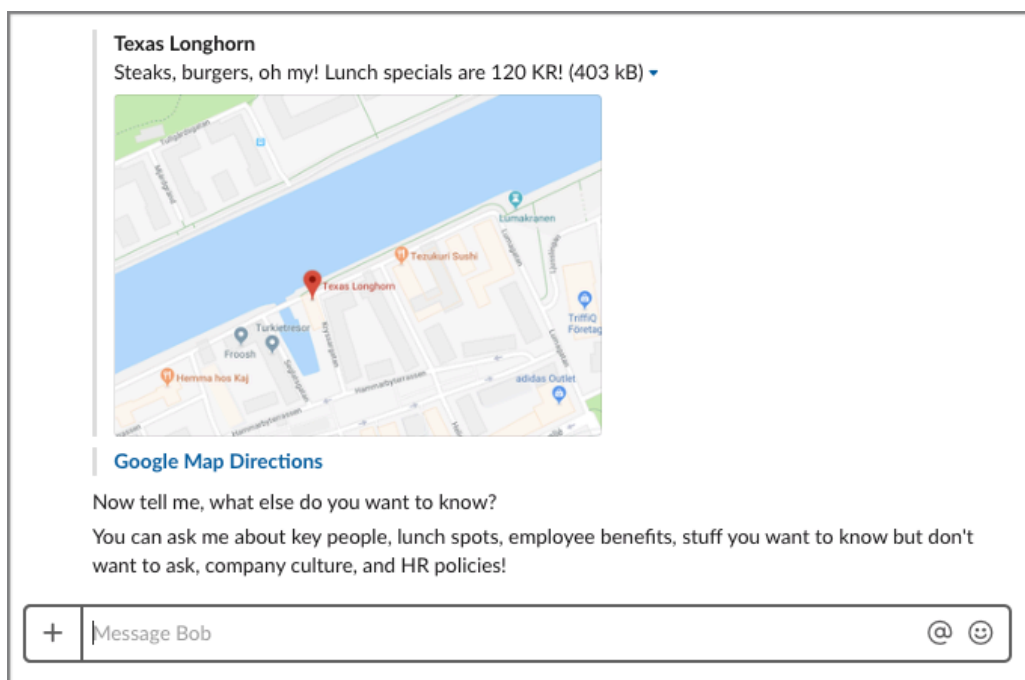
Figure 15. Removing dead ends in conversational chatbot

5.3.   Menu-based bot implementation

Link to menu-based chatbot: https://landbot.io/u/H-162644-XUKSM0HRZF86UAAQ/index.html

Link to menu-based chatbot demo: https://youtu.be/sYl0Q5BGIbw

The conversational flow diagram was mapped into the structure provided by Landbot. Landbot does not allow any text input from the user—rather, all user interaction comes from buttons. Meanwhile, bot messages can appear as text, menu systems with buttons, or images. Using the drag and drop interface, the conversational user flows were applied to the Landbot structure. The menu-based bot architecture built in Landbot can be found in Appendix C.

In accordance to design principles *4.3.1.1 Set expectations* and *4.3.1.5. Provide flexible navigation*, the chatbot was designed so that after a friendly introduction, it immediately explains its purpose and what services it can perform. In addition, the user always has the option to go backwards in the flow by hitting the back button. This example is shown in the Welcome flow in Figure 16.
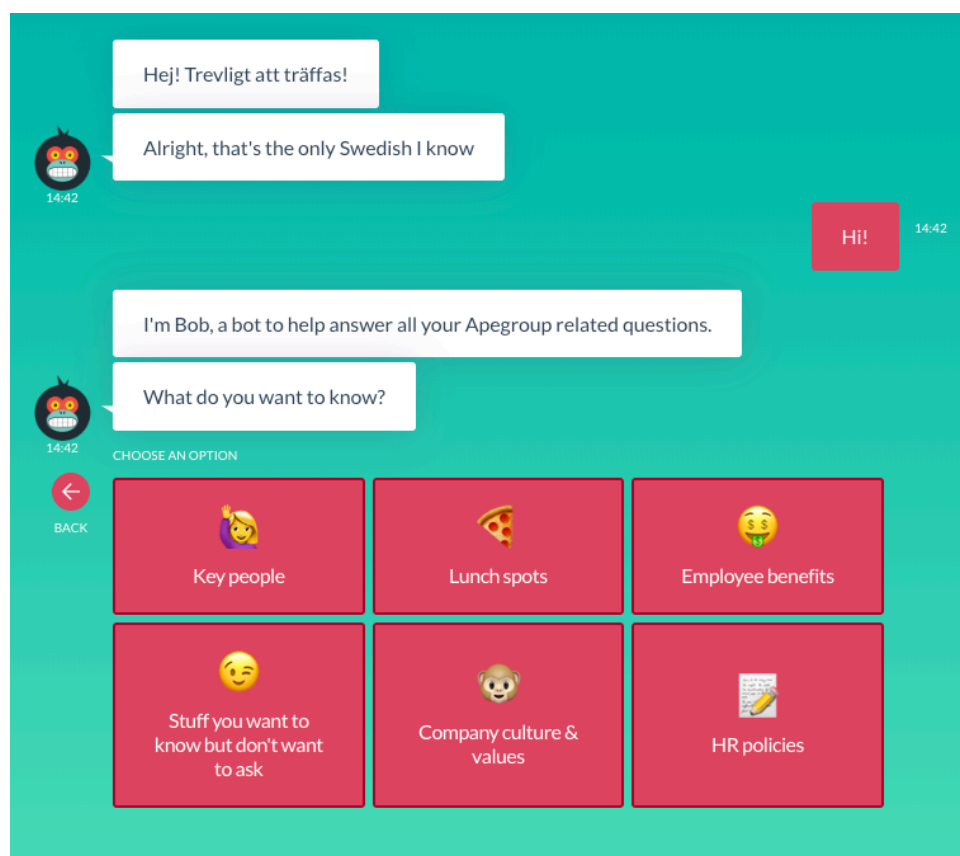
Figure 16. Introduction of menu-based bot, including flexible navigation

It should be noted that the menu-based bot does not strive to create a realistic and personal experience like the conversational bot, therefore it does not ask the user for their name or team.

After the user selects the type of information they want, the chatbot responds according to design principles *4.3.1.2 Be brief*, *4.3.1.6 Use contractions*, *4.3.1.7 Be conversational*, and *4.3.1.9. Use emoji and rich content such as images sensibly*. The chatbot keeps a casual, playful personality as determined in section 4.3.2 Personality and Tone of Voice through use of emoji and humour. This example is reflected in Figure 17.
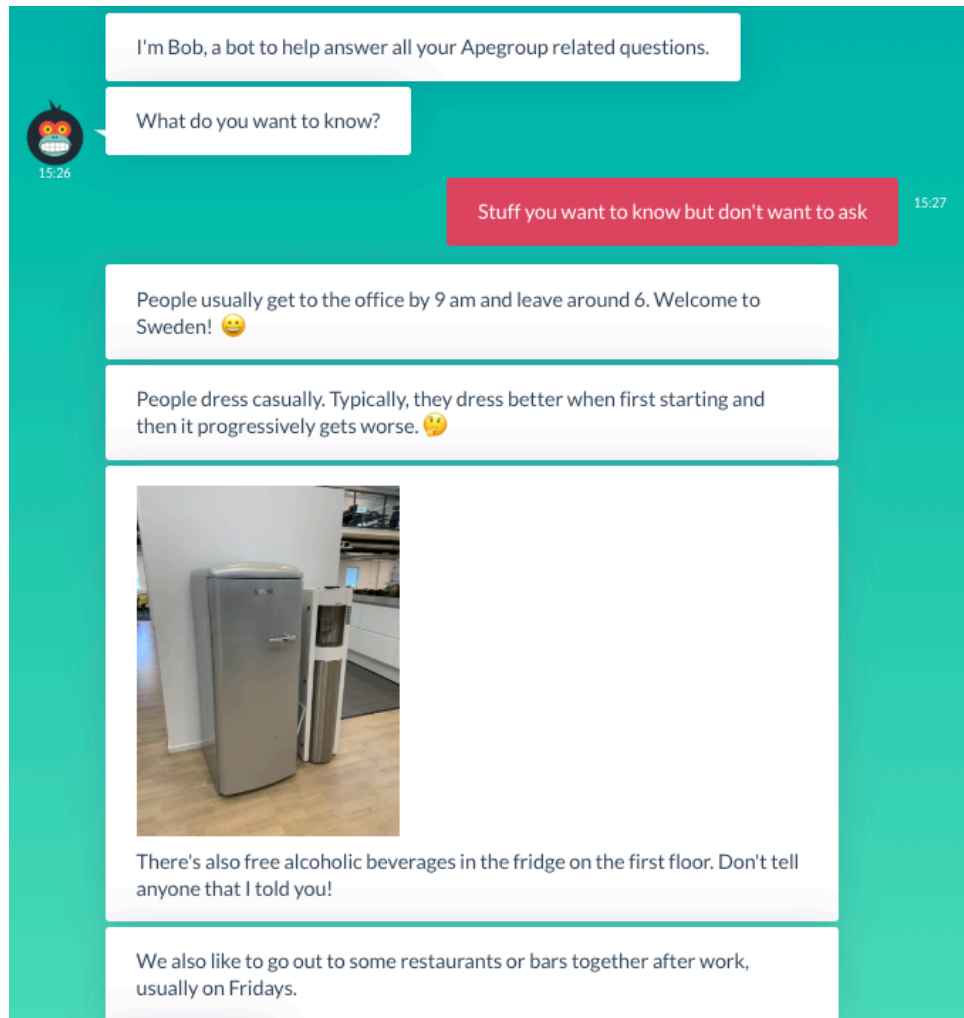
Figure 17. Personality and messaging style of menu-based bot

After the user has received the information they requested, the chatbot follows design principle *4.3.1.10 Don't leave dead ends* by responding with all available options again so that the user does not get lost and can continue obtaining more information. This example is shown in Figure 18, after the user has obtained information about the Monday breakfast benefit.
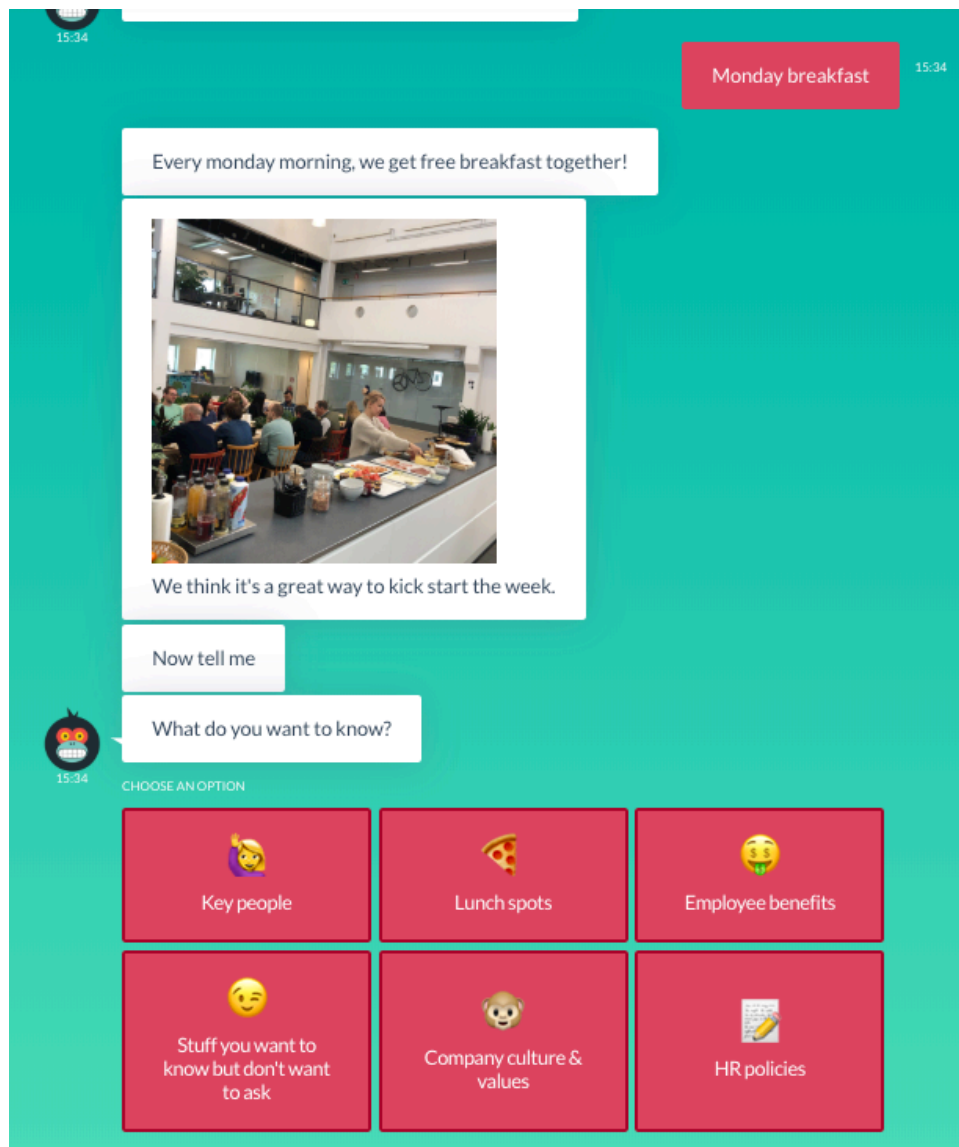
Figure 18. Removing dead ends in menu-based bot

5.4.   Comparison of chatbot implementations

The conversational chatbot built with Dialogflow was much more difficult and time-consuming to implement when compared to the menu-based chatbot built with Landbot. Whereas the menu-based bot took two days to complete, the conversational bot took over three weeks to complete and fine-tune. This is due to the complexity of natural language processing that is required in a

conversational chatbot. Mapping training phrases to logical intents, entities, and responses is extremely difficult and often leads to unpredictable responses.

6.   Results

This section provides the results of the qualitative and quantitative evaluation methods mentioned in section 4.4 Evaluation.

6.1.   Qualitative evaluation results

The user tests showed that in general, of the participants that tested both chatbots, a majority preferred the menu-based bot. When asked "Which of the two chatbots was the better experience?" following the test, roughly 70% of participants (five out of seven) preferred the menu-based over the conversational experience. Of the two participants that preferred the conversational experience, one acknowledged its usability challenges, but indicated that he still preferred it because of its greater potential. The other participant had an error-free experience and liked that he could directly ask a question and receive the answer.

The qualitative findings from group 1, who tested both chatbots, are elaborated as follows:

- **Menu-based bot has fewer errors:** Six of seven participants said that the menu-based chatbot was easier to use because it was much more difficult to make errors. This is due to the fact that the menu-based bot does not allow for real conversational messaging—the user simply selects pre-defined menu options that are wrapped up into the look and feel of a conversation. Meanwhile, the conversational chatbot sometimes provided unpredictable responses.

- **Menu-based bot interaction is more convenient:** All seven participants indicated that while sending real messages was nice, pressing a menu button was more convenient because it was quicker. Once participant noted "I don't want to send full messages to a bot, it's not fooling anyone into thinking it's human."

- **Menu-based bot is more visually pleasing:** From a visual design perspective, all seven participants preferred the menu-based bot. This can largely be attributed to the aesthetic of the menu-based platform, Landbot, which offers colourful themes, large emojis, and an attractive design. Generally speaking, menu-based bots have more graphical user interface elements that can be attractively designed, whereas conversational chatbots are limited in terms of visual design opportunities because because they primarily rely on text to communicate.

- **Conversational bot did not appear more personal or intelligent than the menu-based bot:** Arguably the most surprising result was that the conversational chatbot did not appear to be more personal than the menu-based bot despite the ability to have natural conversations. Four testers explicitly stated that the use of real messaging did not make the chatbot feel more human, as they still knew they were interacting with a bot. Meanwhile, they also mentioned that the aesthetics and presentation of the menu-based bot gave it a more personal feel. One user also commented that the conversational bot felt automated because it did not have any delay between messages, whereas the menu-based bot included an artificial delay with typing indicator between messages, making it feel more human. It should be noted that the Dialogflow platform did not support artificial delay with typing indicators, or else it would have been implemented.

- **Conversational bot is more convenient if the user has specific questions:** Four of the seven participants noted that while the menu-based bot is easier to use, the conversational bot is more useful if the user has a specific question. This is because natural messaging allows the conversational bot to jump to particular answers, whereas the menu-based bot forces the user into a menu flow. For example, if the user wants to know how the vacation policy works, they can type a relevant message and the conversational bot can activate the GetVacationPolicy intent to provide the response. In the menu-bot, the user would be forced to first navigate to all benefits before being able to select the vacation policy.

- **Users don't like to read or follow conversational bot instructions:** Despite relatively concise messages, three of the seven participants did not read and ignored the conversational bot's suggestions on what information it could provide. Instead, they jumped immediately into asking questions, which led to the bot not understanding, or providing unpredictable responses.

The findings from the ten remaining participants in groups 2 and 3, whom each tested either the conversational or menu-based bot, were similar to those of group 1. Of the five who tested the menu-based bot, nearly all perfectly executed each task. There were only minor issues in the experience, as one of the testers did not identify the back button initially, while another did not immediately realise that the menu options were clickable buttons. In general,

the menu-based bot testers remarked that it was easy to use because the buttons were convenient for navigation, and that it was a much better experience than reading the existing internal Wikipedia page because the bot made the information more digestible and memorable. Another advantage was that due to the nature of the new hire onboarding scenario, where a new employee is unaware of what they should know, the menu system was able to convey all of the available information that is relevant to them so they could browse accordingly.

The feedback was mixed for the group that tested only the conversational bot. Four of five had errors when using the bot, where the bot couldn't understand their input or provided an incorrect response. Two people commented that the bot forced them to guess what they could type in order for the bot to understand. However, all five of them still acknowledged that the bot was more convenient than reading through the internal Wikipedia page.

6.2.   Quantitative evaluation results

The summary of the survey results from Typeform can be found in Appendix D.

6.2.1.   Results for group 1

The quantitative survey results for group 1, consisting of seven participants who tested both bots, are presented in Figure 19.

Category	Question	Group 1, Conversational		Group 1, Menu-Based		Mean difference
		Mean	SD	Mean	SD	
Usability	The possibility of this chatbot meet my requirements	3.20	5.00	5.40	0.89	**-2.20**
Usability	Using this chatbot is a frustrating experience.	5.00	1.58	2.80	2.39	**2.20**
Usability	This chatbot is easy to use.	3.40	2.51	6.00	1.00	**-2.60**
Usability	I waste too much time on correcting things in this chatbot.	5.60	1.14	1.60	0.89	**4.00**
Intelligence	To what extent did this chatbot seem personal?	3.20	1.64	3.00	1.58	**0.20**
Intelligence	To what extent did this chatbot seem intelligent?	2.80	1.79	2.80	0.45	**0.00**

Category	Question	Group 1, Conversational		Group 1, Menu-Based		Mean difference
Intelligence	To what extent did this chatbot seem human-like?	2.80	1.48	3.40	1.14	**-0.60**
Satisfaction	How likely are you to use this chatbot again if it were implemented for real?	3.40	2.30	5.40	1.52	**-2.00**
Satisfaction	Overall, how satisfied are you with the system?	3.40	1.52	4.80	0.45	**-1.40**
Satisfaction	How likely are you to recommend this chatbot to a friend or colleague?	2.80	3.42	8.00	1.22	**-5.20**

Figure 19. Group 1 survey results

The results show that overall, the menu-based bot consistently outperforms the conversational bot on measures of usability and satisfaction, as calculated by the difference in mean score.

For the first four UMUX questions measuring usability, the mean differences of -2.20, 2.20, -2.60, and 4.00, are quite large when considering that the highest point value possible is 7.00, based on the 7-point Likert scale. This result is not surprising, as the menu-based bot made it nearly impossible to make errors, while the conversational bot could be unpredictable due to it allowing the user to say anything.

The last three questions, measuring satisfaction, were also in favour of the menu-based bot. The mean difference of the NPS score, -5.20, was particularly large. However, this can be somewhat mitigated by the high standard deviation for the conversational bot, 3.42. The deviation is explained by the fact that some users had very smooth experiences because they followed the suggestions provided by the bot, while others ignored the suggestions and tried to explore it independently.

Interestingly, the mean difference of the three questions measuring perceived intelligence, 0.2, 0.0, and -0.6, reveal that the conversational chatbot was not seen as smarter or more human-like than the menu-based bot. A potential explanation for this is that based on the comments from the qualitative evaluation, most of the participants were skeptical of chatbots in general due to previous poor experiences. Chatbot errors during the evaluation of the conversational chatbot could have also contributed to its low intelligence score.

6.2.2. Results for groups 2 and 3

The results of the quantitative survey for groups 2 and 3, consisting of five participants each who only tested one of the chatbots respectively, are presented in Figure 20.

Category	Question	Group 2, Conversational		Group 3, Menu-Based		Mean difference
		Mean	SD	Mean	SD	
Usability	The possibility of this chatbot meet my requirements	4.29	1.60	5.14	1.95	**-0.85**
Usability	Using this chatbot is a frustrating experience.	3.71	1.50	2.00	1.15	**1.71**
Usability	This chatbot is easy to use.	4.86	2.04	6.14	1.46	**-1.28**
Usability	I waste too much time on correcting things in this chatbot.	4.14	1.35	1.57	1.13	**2.57**
Intelligence	To what extent did this chatbot seem personal?	2.71	1.80	2.71	1.89	**0.00**
Intelligence	To what extent did this chatbot seem intelligent?	3.71	0.95	2.43	1.62	**1.28**
Intelligence	To what extent did this chatbot seem human-like?	2.86	1.07	2.43	2.07	**0.43**
Satisfaction	How likely are you to use this chatbot again if it were implemented for real?	4.57	1.72	5.43	1.40	**-0.86**
Satisfaction	Overall, how satisfied are you with the system?	4.29	1.50	5.57	1.40	**-1.28**
Satisfaction	How likely are you to recommend this chatbot to a friend or colleague?	5.86	2.73	7.43	2.82	**-1.57**

Figure 20. Groups 2 and 3 survey results

The results for the two groups that tested each chatbot showed similar results as group 1, with the menu-based outperforming the conversational bot overall on measures of usability and satisfaction.

Interestingly however, while still relatively small, the conversational bot was considered more intelligent than the menu-based bot, with a mean difference more in favour of the conversational bot when compared to group 1 where the participants tested both bots. An explanation could be that without a

comparison to the menu-bot, the conversational bot seemed more intelligent as users could freely talk to it.

Another insight is that when comparing the average ratings of group 1 with groups 2 and 3, the conversational bot consistently had better scores across questions measuring usability and satisfaction when it was tested independently vs. tested together with the menu-based bot. This could also be due to the reasoning that without a direct comparison to a simpler, menu-based bot, the conversational bot is actually perceived as quite usable. The mean ratings for group 1 compared with mean ratings for group 2 and 3 side by side are presented in Figure 21.

Question	Group 1, Conversational	Group 2, Conversational	Group 1, Menu-based	Group 3, Menu-based
	Mean	Mean	Mean	Mean
The possibility of this chatbot meet my requirements	3.20	4.29	5.40	5.14
Using this chatbot is a frustrating experience.	5.00	3.71	2.80	2.00
This chatbot is easy to use.	3.40	4.86	6.00	6.14
I waste too much time on correcting things in this chatbot.	5.60	4.14	1.60	1.57
To what extent did this chatbot seem personal?	3.20	2.71	3.00	2.71
To what extent did this chatbot seem intelligent?	2.80	3.71	2.80	2.43
To what extent did this chatbot seem human-like?	2.80	2.86	3.40	2.43
How likely are you to use this chatbot again if it were implemented for real?	3.40	4.57	5.40	5.43
Overall, how satisfied are you with the system?	3.40	4.29	4.80	5.57

Question	Group 1, Conversational	Group 2, Conversational	Group 1, Menu-based	Group 3, Menu-based
How likely are you to recommend this chatbot to a friend or colleague?	2.80	5.86	8.00	7.43

Figure 21. Groups 1, 2, and 3 survey results side-by-side comparison

	Group 1, Conversational	Group 2, Conversational	Group 1, Menu-based	Group 3, Menu-based
How likely are you to recommend this chatbot to a friend or colleague?	2.80	5.86	8.00	7.43

7.   Discussion

In this study, we found that a technologically-limited, menu-based chatbot experience had better usability and was therefore preferred over a conversational chatbot experience powered by elements of artificial intelligence and natural language processing. The usability of the two chatbots was tested and compared under the context of new hire onboarding and obtaining general human resources information from a company. From this study, the following insights are obtained.

- **Menu-based chatbots are perceived as easier to use due to lack of errors:** Due to the controlled nature of menu-based chatbots where inputs are restricted to buttons, there is no guesswork around what the bot will understand. This dramatically limits potential for errors, contributing to menu-based bots to be perceived as easier to use. Furthermore, the use of buttons and menus is more convenient than typing a message.

- **Conversational chatbots come with higher expectations:** The user interviews during the research and evaluation phases of this study revealed that people have higher expectations towards the capabilities of a conversational chatbot when compared to a menu-based chatbot, and need a high degree of accuracy in order to meet those expectations. When users can freely message a chatbot, they expect it to understand what they say to a reasonable degree. As soon the chatbot makes an error, their perceived usability and intelligence falls greatly.

- **The scenario for which the chatbot is deployed is very important:** The user tests showed that for onboarding employees into a company, it is important that they are presented with information because they do not yet know what they can ask for. A menu-based chatbot is typically more visual because it uses primarily GUI elements, making it more suited for presenting content in an aesthetic and memorable way. However, for scenarios where the user has a specific question or request, a conversational chatbot could be more convenient as it allows for an immediate response without navigating through a tree structure.

- **Conversational chatbots are not considered more intelligent or personal by default:** The user tests indicated that a conversational chatbot experience does not inherently make it feel more intelligent or personal than a menu-based bot. Rather, users are influenced more by the ease of use. Additionally, the user's perception of chatbot intelligence and personality is also influenced by the visuals and interaction from the platform used. For example, the visual design, large emojis, and artificial message delay of the Landbot platform made it feel equally as human as the conversational bot built in Dialogflow.

- **Conversational chatbots are much more time consuming and difficult to implement than menu-based chatbots:** Despite platforms like Google Dialogflow with natural language processing built in, it is still very difficult to classify different user utterances into the appropriate intents, as well as respond to sudden conversational shifts. Therefore, the user scenario must be carefully considered before determining whether a conversational or menu-based chatbot should be developer.

8.   Conclusion

8.1.   Summary

The aim of this thesis was to determine whether users preferred a conversational or menu-based chatbot experience based on usability and satisfaction in context of new hire onboarding. Another goal was to investigate the assumption that an intelligent and realistic conversational experience is an inherently better user experience.

The results presented are based off an extensive literature review and human-centred design process. A conversational chatbot with natural language processing was built with Google Dialogflow, while an intelligence-limited, menu-based chatbot was developed with Landbot. Both bots were evaluated with a qualitative cognitive walkthrough and quantitative, Likert-scale survey. The bots were tested by employees at the partner company, Apegroup, for new hire onboarding and obtaining general human resources related information.

Two research questions were presented in section 2.

To answer *RQ1: Do users prefer a conversational or menu-based chatbot experience?*, the conclusion reached is that for new hire onboarding, a menu-based chatbot experience is more optimal. This is due to a menu-based experience being easier to use, more convenient, and better suited for a scenario where a user needs to be told what they do not know.

For *RQ2: To what extent does a realistic conversational experience influence usability?*, the conclusion reached is that a realistic conversational experience does not inherently provide better usability or satisfaction. A conversational experience has advantages in that it can answer questions directly without a user having to navigate through branches of a menu tree. It also has the potential to show more intelligence, empathy, and keep a conversation going. To enable these advantages however is very difficult, as conversational chatbots are very complex and time-consuming to implement, and much more

prone to error. The platform used for a menu-based chatbot also influences how realistic or personal it is perceived.

8.2.   Limitations

The limitations of this study are discussed in this section.

8.2.1. Dialogflow implementation difficulty

This study aimed to provide a fair evaluation of a conversational vs. menu-based chatbot experience. Due to the technical complexity of creating a conversational chatbot in Dialogflow and limited time, it is debatable whether the conversational chatbot met the level of usability required for a fair comparison to a menu-based chatbot. This may have affected the test results of the conversational chatbot. However, this risk is mitigated by the academic and industry literature review, which acknowledge that even the biggest technology companies routinely struggle with developing good conversational experiences.

8.2.2. User sample

This study primarily focused on evaluating the two chatbots in a new hire onboarding scenario. However, given that the partner company is quite small with about fifty employees total, of which only eight can be classified as a new hire (employed less than three months), the 17 users who evaluated the chatbots were a mix of new and existing employees. This could have impacted the results. This risk however is mitigated by the fact that a large component of new hire onboarding is obtaining basic human resources information, which was still unclear to many of the existing employees.

8.3.   Future work

To further this study, more time and quality assurance could be dedicated to improving the conversational chatbot in order to provide a fairer evaluation. Additionally, the conversational chatbot could be enhanced with advanced capabilities that a menu-based chatbot can not provide, such as understanding context and have better small-talk.

Additionally, it would be very interesting to compare a conversational versus menu-based chatbot experience in other scenarios, where more empathy and understanding is required, such as healthcare. Perhaps a conversational experience in this scenario would be evaluated with better results, as a conversational chatbot has the means to show more emotion and sympathy.

9.   References

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, *6*(7).

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003, April). Consolidating the ISO usability models. In *Proceedings of 11th international software quality management conference* (Vol. 2003, pp. 23-25).

Anatoly Khorozov. (2017, March). Trends Driving the Chatbot Growth. Retrieved May 18, 2019, from Chatbots Magazine website: https://chatbotsmagazine.com/trends-driving-the-chatbot-growth-77b78145bac

Bauer, T. N., & Erdogan, B. (2011). Organizational socialization: The effective onboarding of new employees. *APA handbook of industrial and organizational psychology*, *3*, 51-64.

Bhutani, A., & Wadhwani, P. (2018, June 11). Global Market Insights. Retrieved May 12, 2019, from Gminsights.com website: https://www.gminsights.com/pressrelease/chatbot-market

Bradeško, L., & Mladenić, D. (2012, October). A survey of chatbot systems through a loebner prize competition. In *Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies* (pp. 34-37).

Brandtzaeg, P. B., & Følstad, A. (2017, November). Why people use chatbots. In *International Conference on Internet Science* (pp. 377-392). Springer, Cham.

Brown, J. (2007). Employee orientation: Keeping new employees on board. *humanresources. about. com/library/weekly/nosearch/nuc042102a. htm*, *20*(02), 2014.

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811-817.

Dhanda, S. (2017). Chatbot Conversations to deliver $8 billion in Cost savings by 2022. Retrieved May 18, 2019, from Juniperresearch.com website: https://

www.juniperresearch.com/analystxpress/july-2017/chatbot-conversations-to-deliver-8bn-cost-saving

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical education*, *40*(4), 314-321.

Facebook. (2019). General Best Practices - Messenger Platform - Documentation - Facebook for Developers. Retrieved from Facebook.com website: https://developers.facebook.com/docs/messenger-platform/introduction/general-best-practices/

Feldman, D. (2018, April 10). Chatbots: What Happened? Retrieved May 18, 2019, from Chatbots Life website: https://chatbotslife.com/chatbots-what-happened-dcc3f91a512c

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323-327.

Gallup. (2012). State of the American Workplace. Retrieved April 13, 2019, from Gallup.com website: https://www.gallup.com/workplace/238085/state-american-workplace-report-2017.aspx#

Graybill, J. O., Taesil Hudson Carpenter, M., Offord Jr, J., Piorun, M., & Shaffer, G. (2013). Employee onboarding: Identification of best practices in ACRL libraries. *Library Management*, *34*(3), 200-218.

Griffith, E., & Simonite, T. (2018, January 9). Facebook's Virtual Assistant M Is Dead. So Are Chatbots. Retrieved May 18, 2019, from WIRED website: https://www.wired.com/story/facebooks-virtual-assistant-m-is-dead-so-are-chatbots/

Hanson, A. (2019, June). Principles of Bot Design: Successful Chatbot Design in 2019 [+Tips]. Retrieved June 8, 2019, from Intercom website: https://www.intercom.com/blog/principles-bot-design/

Harpelund, C. (2019). Onboarding: Getting New Hires off to a Flying Start. Emerald Publishing Limited.

Harper, E. R., Rodden, T., Rogers, Y., Sellen, A., & Human, B. (2008). Human-Computer Interaction in the year 2020.

Heath, A. (2016, April 21). Facebook's grand plan to simplify your life is off to a rough start. Retrieved June 27, 2019, from Business Insider website: https://www.businessinsider.com/the-present-and-future-of-facebook-messenger-bots-2016-4?r=US&IR=T

Io, H. N., & Lee, C. B. (2017, December). Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE International Conference on*

Industrial Engineering and Engineering Management (IEEM) (pp. 215-219). IEEE.

ISO 9241-210 (2010) SFS-EN ISO 9241-210: Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. European Committee for Standardization, Brussels, Belgium. 65 p.

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018, April). Convey: Exploring the use of a context view for chatbots. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 468). ACM.

Jassova, B. (2019, May 7). The Ultimate Guide to Conversational... Retrieved June 9, 2019, from Landbot.io website: https://landbot.io/blog/guide-to-conversational-design/

Johnson, K. (2016, July 12). Microsoft CEO: Chatbots will 'fundamentally revolutionize' computing. Retrieved June 7, 2019, from VentureBeat website: https://venturebeat.com/2016/07/11/microsoft-ceo-chatbots-will-fundamentally-revolutionize-computing/

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 396.

Lariviere, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., et al. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. Journal of Business Research, 79, 238e246. https:// doi.org/10.1016/j.jbusres.2017.03.008.

Lessin, S. (2016). On Bots, Conversational Apps and Fin. Retrieved May 18, 2019, from The Information website: https://www.theinformation.com/articles/on-bots-conversational-apps-and-fin

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013, April). UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2099-2102). ACM.

Lomas, N. (2016, May 5). A few words on chatbots. Retrieved May 12, 2019, from TechCrunch website: https://techcrunch.com/2016/05/05/a-few-words-on-chatbots/

McTear, M., Callejas, Z., & Griol, D. (2016). Conversational interfaces: Past and present. In *The Conversational Interface*(pp. 51-72). Springer, Cham.

Messina, C. (2016, January 19). 2016 will be the year of conversational commerce. Retrieved May 18, 2019, from Medium website: https://medium.com/chris-messina/2016-will-be-the-year-of-conversational-commerce-1586e85e3991

Meyer, A. M., & Bartels, L. K. (2017). The impact of onboarding levels on perceived utility, organizational commitment, organizational support, and job satisfaction. *Journal of Organizational Psychology, 17*(5), 10-27.

Most popular messaging apps 2019 | Statista. (2019). Retrieved from Statista website: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Civitas Books.

Pall, G. (2016, August 3). Progress in the shift to conversational computing - The Official Microsoft Blog. Retrieved May 18, 2019, from The Official Microsoft Blog website: https://blogs.microsoft.com/blog/2016/08/03/progress-in-the-shift-to-conversational-computing/

Peras, D. (2018). Chatbot evaluation metrics. Economic and Social Development: Book of Proceedings, 89-97.

Pernice, K. (2018, February 18). Affinity Diagramming: Collaboratively Sort UX Findings & Design Ideas. Retrieved June 8, 2019, from Nielsen Norman Group website: https://www.nngroup.com/articles/affinity-diagram/

Piccolo, L., Mensio, M., & Alani, H. (2019). Chasing the chatbots: Directions for interaction and design research.

Pike, K. L. (2014). New Employee Onboarding Programs and Person-Organization Fit: An Examination of Socialization Tactics.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579.*

Rahman, A. M., Al Mamun, A., & Islam, A. (2017, December). Programming challenges of chatbot: Current and future prospective. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 75-78). IEEE.

Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017, May). A survey of design techniques for conversational agents. In *International Conference on Information, Communication and Computing Technology* (pp. 336-350). Springer, Singapore.

Ranoliya, B., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related FAQs - IEEE Conference Publication. *Ieee.Org.* Retrieved from https://ieeexplore.ieee.org/abstract/document/8126057

Rieman, J., Franzke, M., & Redmiles, D. (1995, May). Usability evaluation with the cognitive walkthrough. In *Conference companion on Human factors in computing systems* (pp. 387-388). ACM.

Robinson, M., Gray, J., Cowley, A., & Tan, Ri. (2019). Adopting the power of conversational UX. In *Deloitte* (p. 2). Retrieved from Deloitte website: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-chatbots-adopting-the-power-of-conversational-ux.pdf

Shevat, A. (2017). Designing bots: Creating conversational experiences. " O'Reilly Media, Inc.".

Simonite, T. (2017, April 14). Facebook's Perfect, Impossible Chatbot. Retrieved May 12, 2019, from MIT Technology Review website: https://www.technologyreview.com/s/604117/facebooks-perfect-impossible-chatbot/

Snell, A. (2006). Researching onboarding best practice: Using research to connect onboarding processes with employee satisfaction. *Strategic HR Review*, *5*(6), 32–35. https://doi.org/10.1108/14754390680000925

Tavanapour, N., & Bittner, E. A. (2018). Automated facilitation for idea platforms: design and evaluation of a Chatbot prototype.

Teixeira, F. (2019). Why chatbots fail. Retrieved May 18, 2019, from Why chatbots fail website: https://chatbot.fail/

The 2017 U.S. Mobile App Report. (2017). Retrieved May 18, 2019, from Comscore, Inc. website: https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/The-2017-US-Mobile-App-Report?cs_edgescape_cc=US

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). The think aloud method: a practical approach to modelling cognitive. Academic Press, London.

Wallace, R. S. (2009). The anatomy of ALICE. In *Parsing the Turing Test* (pp. 181-210). Springer, Dordrecht.

Appendix A - Affinity diagram

People

- want to know team members
- key people at the company
- who are the managers?
- HR people to ask questions
- first thing is usually trying to know who is who

Lunch

- first day I usually want to know what to do for lunch
- do people eat lunch together?
- best places that people like to go to for food?
- popular lunch places nearby
- should I bring my own food or eat outside?

Benefits

- im unclear what the benefits are, I hear we have a gym benefit?
- One place to see all company benefits
- I see sharpening friday mentioned alot, want to understand what that is in more detail
- what are the benefits at the company?

Secrets

- what time should I arrive and leave the office?
- not sure when I can leave
- do people go out after work together?
- what's the dress code?
- do I have to dress in a certain way?

Culture

- From HR: important to include culture and values

HR Policies

- I'm still unsure when I can take vacation
- all the information about policies is scattered
- consolidate work from home and sick leave submission
- what can I do if I'm sick and can't come to work?

Discarded from MVP

- how to use the coffee machine?
- how do I setup printer?
- how to access office after hours?

miro

Appendix B - Conversational bot architecture

List of intents and responses

Intent	Response
Default Welcome Intent	• Hej! Trevligt att träffas! Hallå! Trevligt att träffas! • Alright, that's the only Swedish I know. • Cool! That's the only Swedish I know. • I'm Bob, a bot that can help answer all your questions about Apegroup. • First, can you tell me your name? I realize your name is here on Slack, but my creator is respectful of privacy and won't give it to me automatically. • Before we start, what's your name? I realize it's here on Slack, but I'm respectful of privacy!
GetName	• $given-name is a nice name! Much better than Bob. • I wish my name was $given-name! Bob is so lame. • $given-name is a better name than Bob for sure. • Last question, what team are you on? I can give you better answers if you tell me. • Lastly, which team are you on? I can give you better answers if you tell me. • Lastly, what's your role? I can give you better answers if you tell me this.
GetTeam	• Yay! $Team is my favorite team at Apegroup. • Did you know $Team is my favorite at Apegroup? • Anyway, I can tell you about the people, lunch spots, benefits, stuff you want to know but don't want to ask, company culture, and HR policies.

Intent	Response
GetPeople	• So you want to stalk someone. I like that. • Stefan Ilkovics (Slack card) • Wayne Knoessen (Slack card) • Chris Mansson (Slack card) • Mattias Olsson (Slack card) • Yosra Axling (Slack card) • Olle Havesome (Slack card) • Now tell me, what else do you want to know? • You can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture and HR policies!
GetLunchSpots	• Great! People generally eat out together at nearby places 2-3 times per week. They bring their own lunches otherwise. • People from different teams commonly eat together as well, so don't be shy 🙂! • Here are some of our employees' favorite lunch spots nearby. • Niklas and Friends (Slack card) • Finefood (Slack card) • Krubb (Slack card) • Hammarby Sushi & Dumplings (Slack card) • Texas Longhorn (Slack card) • Now tell me, what else do you want to know? • You can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetBenefits	• We've got some sweeeet benefits. • We have a fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day benefits! • Which one do you want to know more about?
GetBenefits - sharpening	• Every 2nd Friday, we get the whole day to work on our own passion project, or Apegroup invites amazing guest speakers! • Sharpening (Slack card) • Now tell me, what else do you want to know? • Our benefits include fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day! • Or, you can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetBenefits - massage	• Our boy, certified massage therapist *Daniel Yeoh* gives massages in our office every 2 weeks at a big discount! • Massage (Slack card) • Now tell me, what else do you want to know? • Our benefits include fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day! • Or, you can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!

Intent	Response
GetBenefits - education	• Every employee gets 3000 kr per year to spend on anything related to personal learning and training! • Just talk to Stefan, our head of design, to help set this up. • Stefan Ilkovics (Slack card) • Now tell me, what else do you want to know? • Our benefits include fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day! • Or, you can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetBenefits - yayday	• Is it your birthday? If so, you get a free paid day off 🥳! Just let an HR member know. • Now tell me, what else do you want to know? • Our benefits include fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day! • Or, you can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetBenefits - Monday breakfast	• Every Monday morning, we get free breakfast together! • Breakfast (Slack card) • Now tell me, what else do you want to know? • Our benefits include fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day! • Or, you can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetBenefits - Fitness	• Every employee gets a 3000 SEK per year health and fitness budget to spend on physical activities! • SATS gym location (Slack card) • Mattias Mangberg (Slack card) • Now tell me, what else do you want to know? • Our benefits include fitness benefit, educational budget, epic massages, sharpening friday, Monday breakfast, and YAY-day! • Or, you can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetStuffYouWanttoKnow	• People usually get to the office by 9 am and leave around 6. Welcome to Sweden! 😀 • People dress casually. Typically, they dress better when first starting and then it progressively gets worse. 🤔 • Beer fridge (Slack card) • We also like to go out to some restaurants or bars together after work, usually on Fridays. • You can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!

Intent	Response
GetCompanyCulture	• We love to collaborate and push the boundaries in everything we do 🤷‍♀️ . • We also like to keep things fun and casual, but honesty and transparency are also essential to the company 🕺 . • Company values (Slack card) • Company mission statement (Slack card) • Here are some of our favorite projects • McDonalds (link) • ICA (link) • Now tell me, what else do you want to know? • You can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetPolicies	• We have a vacation policy, maternity/paternity policy, unpaid leave policy, and work from home policy. • You can submit requests for each of them here as well! • Which do you want to know more about?
GetPolicy - work from home	• If you need to be at home for the day, just fill out the next form to inform HR and your manager! • Fill out the work from home request form here: • https://docs.google.com/forms/d/1Q4tOtBHcCxqvg4t8N0x1_MGRhj2zuTJmejJ_VdV7g-U/viewform?edit_requested=true • Now tell me, what do you want to know? • We have a vacation policy, maternity/paternity leave policy, unpaid leave policy, and work from policy. • Or, You can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetPolicy - vacation	• Every full time employee can accrue up to 25 days of paid holiday vacation every year! Thank you Sweden 😎 . • Fill out the vacation request form here: • https://docs.google.com/forms/d/1ZSVoJzaXbv5bsJkc_K8p5ABxzxSrPLdqECFaZVmRKPA/viewform?edit_requested=true • Now tell me, what else do you want to know? • We have a vacation policy, maternity/paternity policy, unpaid leave policy, and work from home policy. • Or, you can also ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!

Intent	Response
GetPolicy - unpaid leave	• If you need to be away from work for an extended period of time, fill out the form and a member of HR will get back to you. • Fill out unpaid leave request here: • https://docs.google.com/forms/d/1atSrBW7t95wJwF9pHqX3DWUAE9EmDzbvwlBew40fUe0/viewform?edit_requested=true • Now tell me, what else do you want to know? • We have a vacation policy, maternity/paternity policy, unpaid leave policy, and work from home policy. • Or, you can also ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
GetPolicy - paternity and maternity	• Congratulations! Expecting mothers and fathers get 9 months of paid leave each. Fill out the form, and a member of HR will get back to you. • Fill out paternity and maternity leave form here: • https://docs.google.com/forms/d/1nLJUwBwXZ3p0-Ji8RCZSr9DLAwt8F-1k0Z9TXrnTkag/viewform?edit_requested=true • Now tell me, what else do you want to know? • We have a vacation policy, maternity/paternity policy, unpaid leave policy, and work from home policy. • Or, you can also ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies!
Default Fallback intent	• I didn't get that. Can you say it again? • I missed what you said. What was that? • Sorry, could you say that again? • Sorry, can you say that again? • Can you say that again? • Sorry, I didn't get that. Can you rephrase? • Sorry, what was that? • One more time? • What was that? • Say that one more time? • I didn't get that. Can you repeat? • I missed that, say that again? • You can ask me about key people, lunch spots, employee benefits, stuff you want to know but don't want to ask, company culture, and HR policies! • I can give you info about people, lunch spots, benefits, secrets, company culture and HR policies!

List of entities

Entity	Keyword	Synonyms
@benefits	breakfast	breakfast, monday breakfast
@benefits	education	education, knowledge, learning, books
@benefits	massage	massage, masseuse, chiropractor
@benefits	sharpening	sharpening, sharpening Friday
@benefits	fitness	fitness, gym, health, workout, training
@benefits	yayday	yayday, birthday
@policy	paternity/maternity	baby, maternity, paternity, kid
@policy	unpaid	leave, sick, sick leave, unpaid, unpaid leave
@policy	vacation	holiday, holidays, vacation, vacations, PTO
@policy	work from home	work from home, wfh
@team	design	design, designer, designing
@team	engineering	developer, development, engineer, engineering, tech, QA, QA tester, tester
@team	producer	producer, product, product management
@team	management	CEO, CFO, management, manager
@sys.given-name	Given-name	Dialogflow default set of names

Appendix C - Menu bot architecture

Appendix D - Quantitative survey results

Conversational bot results
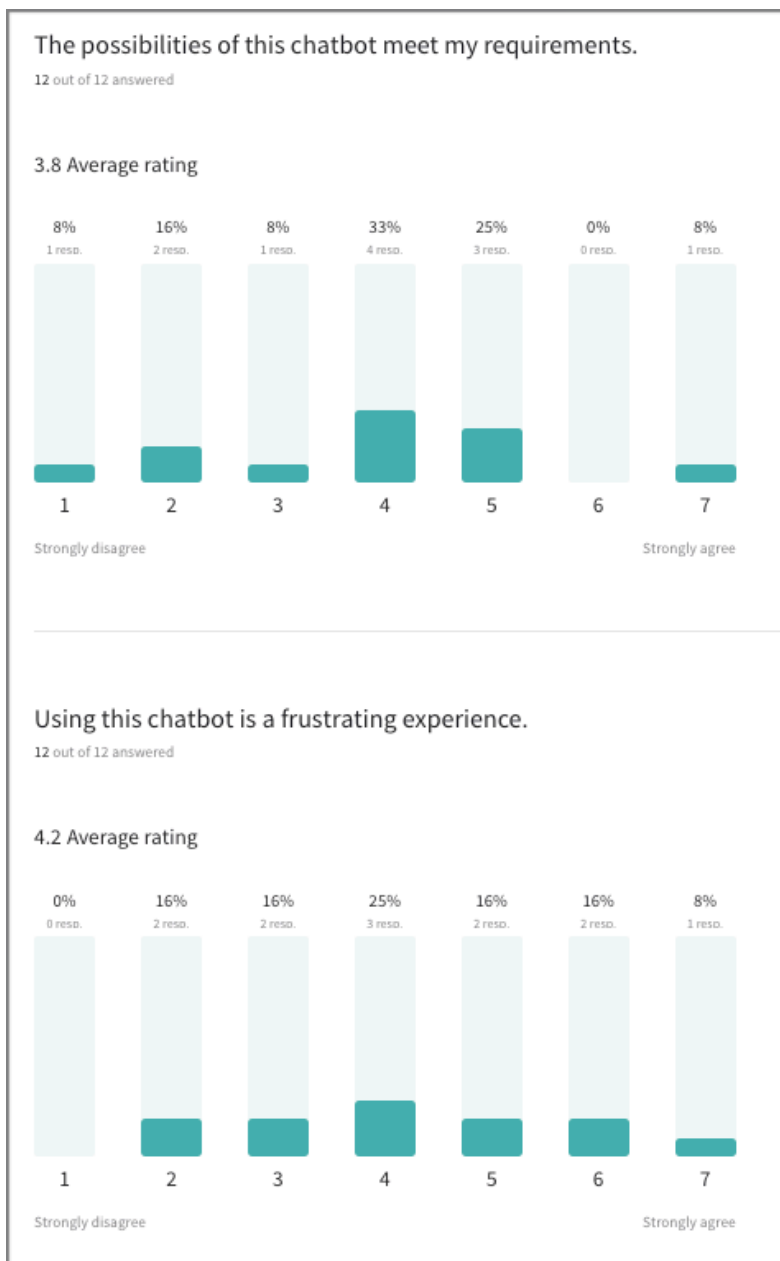
The possibilities of this chatbot meet my requirements.

12 out of 12 answered

3.8 Average rating

8%	16%	8%	33%	25%	0%	8%
1 resp.	2 resp.	1 resp.	4 resp.	3 resp.	0 resp.	1 resp.
1	2	3	4	5	6	7

Strongly disagree — Strongly agree

Using this chatbot is a frustrating experience.

12 out of 12 answered

4.2 Average rating

0%	16%	16%	25%	16%	16%	8%
0 resp.	2 resp.	2 resp.	3 resp.	2 resp.	2 resp.	1 resp.
1	2	3	4	5	6	7

Strongly disagree — Strongly agree

This chatbot is easy to use.
12 out of 12 answered

4.2 Average rating

16%	16%	0%	8%	33%	0%	25%
2 resp.	2 resp.	0 resp.	1 resp.	4 resp.	0 resp.	3 resp.
1	2	3	4	5	6	7

Strongly disagree — Strongly agree
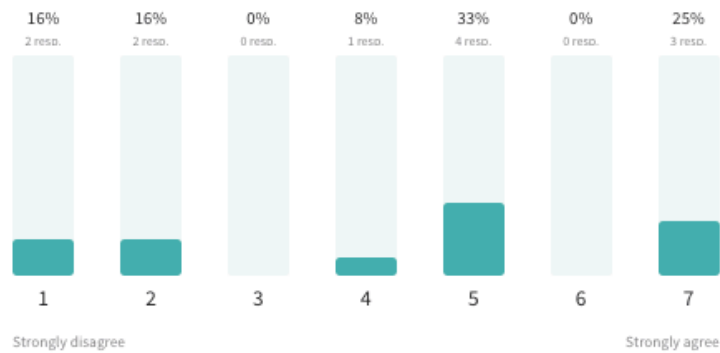
I waste too much time on correcting things in this chatbot.
12 out of 12 answered

4.8 Average rating

0%	8%	8%	25%	25%	25%	8%
0 resp.	1 resp.	1 resp.	3 resp.	3 resp.	3 resp.	1 resp.
1	2	3	4	5	6	7

Strongly disagree — Strongly agree
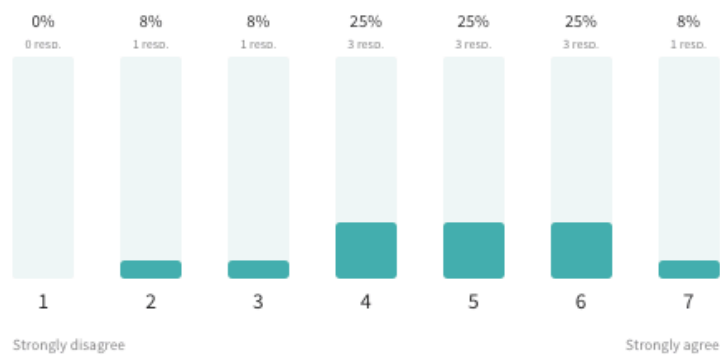
To what extent did this chatbot seem personal?

12 out of 12 answered

2.9 Average rating

25%	25%	8%	25%	8%	8%	0%
3 resp.	3 resp.	1 resp.	3 resp.	1 resp.	1 resp.	0 resp.
1	2	3	4	5	6	7

Very impersonal — Very personal
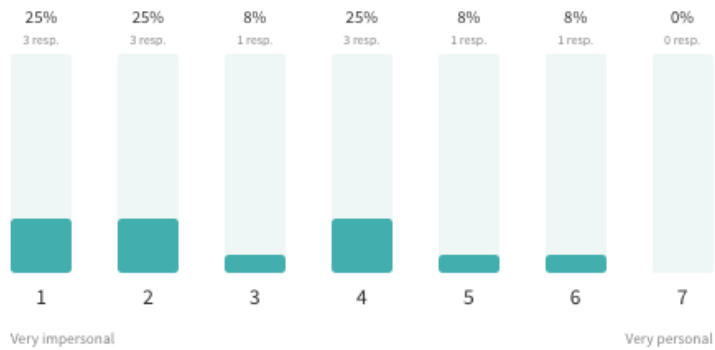
To what extent did this chatbot seem intelligent?

12 out of 12 answered

3.3 Average rating
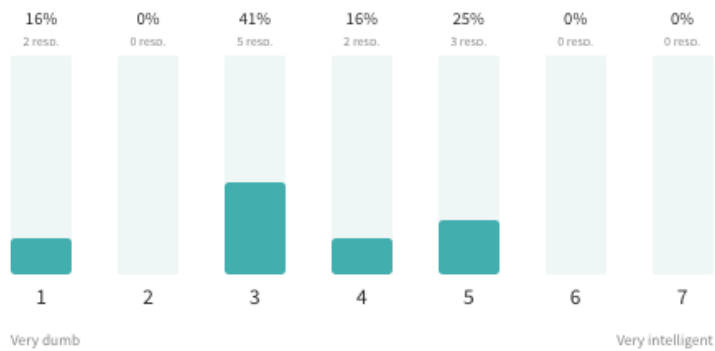
16%	0%	41%	16%	25%	0%	0%
2 resp.	0 resp.	5 resp.	2 resp.	3 resp.	0 resp.	0 resp.
1	2	3	4	5	6	7

Very dumb — Very intelligent

To what extent did this chatbot seem human-like?

12 out of 12 answered

2.8 Average rating

8%	33%	41%	0%	16%	0%	0%
1 resp.	4 resp.	5 resp.	0 resp.	2 resp.	0 resp.	0 resp.
1	2	3	4	5	6	7

Very un-human · Very human-like
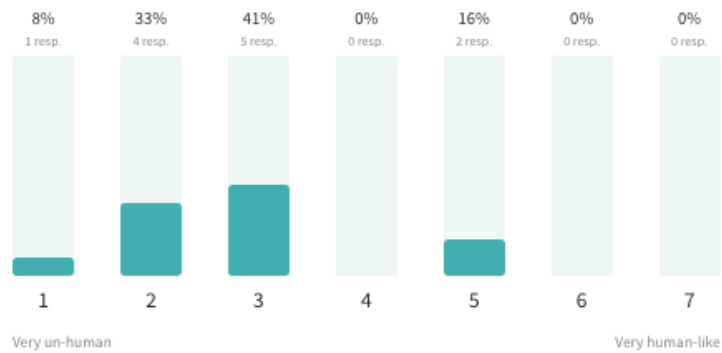
How likely are you to use this chatbot again if it were implemented for real?

12 out of 12 answered

4.1 Average rating

16%	8%	8%	16%	25%	16%	8%
2 resp.	1 resp.	1 resp.	2 resp.	3 resp.	2 resp.	1 resp.
1	2	3	4	5	6	7

Very unlikely · Very likely

Overall, how satisfied are you with the system?

12 out of 12 answered

3.9 Average rating

8%	8%	8%	50%	16%	0%	8%
1 resp.	1 resp.	1 resp.	6 resp.	2 resp.	0 resp.	1 resp.
1	2	3	4	5	6	7

Very dissatisfied — Very satisfied

How likely are you to recommend this chatbot to a friend or colleague?

12 out of 12 answered

4.6 Average rating

16%	8%	8%	8%	0%	0%	33%	8%	8%	0%	8%
2	1	1	1	0	0	4	1	1	0	1
0	1	2	3	4	5	6	7	8	9	10

Not at all likely — Extremely likely

Menu-based bot result

The possibilities of this chatbot meet my requirements.
12 out of 12 answered

5.2 Average rating

0%	8%	0%	25%	16%	25%	25%
0 resp.	1 resp.	0 resp.	3 resp.	2 resp.	3 resp.	3 resp.
1	2	3	4	5	6	7

Strongly disagree — Strongly agree

Using this chatbot is a frustrating experience.
12 out of 12 answered

2.3 Average rating

33%	41%	8%	8%	0%	0%	8%
4 resp.	5 resp.	1 resp.	1 resp.	0 resp.	0 resp.	1 resp.
1	2	3	4	5	6	7

Strongly disagree — Strongly agree

This chatbot is easy to use.
12 out of 12 answered

6.1 Average rating

0%	0%	8%	0%	16%	25%	50%
0 resp.	0 resp.	1 resp.	0 resp.	2 resp.	3 resp.	6 resp.
1	2	3	4	5	6	7

Strongly disagree   Strongly agree

I waste too much time on correcting things in this chatbot.
12 out of 12 answered

1.6 Average rating

66%	16%	8%	8%	0%	0%	0%
8 resp.	2 resp.	1 resp.	1 resp.	0 resp.	0 resp.	0 resp.
1	2	3	4	5	6	7

Strongly disagree   Strongly agree

To what extent did this chatbot seem personal?

12 out of 12 answered

2.8 Average rating

33%	8%	25%	16%	8%	8%	0%
4 resp.	1 resp.	3 resp.	2 resp.	1 resp.	1 resp.	0 resp.
1	2	3	4	5	6	7

Very impersonal · Very personal

To what extent did this chatbot seem intelligent?

12 out of 12 answered

2.6 Average rating

25%	16%	41%	8%	8%	0%	0%
3 resp.	2 resp.	5 resp.	1 resp.	1 resp.	0 resp.	0 resp.
1	2	3	4	5	6	7

Very dumb · Very intelligent

To what extent did this chatbot seem human-like?

12 out of 12 answered

2.8 Average rating

16%	41%	16%	8%	8%	0%	8%
2 resp.	5 resp.	2 resp.	1 resp.	1 resp.	0 resp.	1 resp.
1	2	3	4	5	6	7

Very un-human   Very human-like
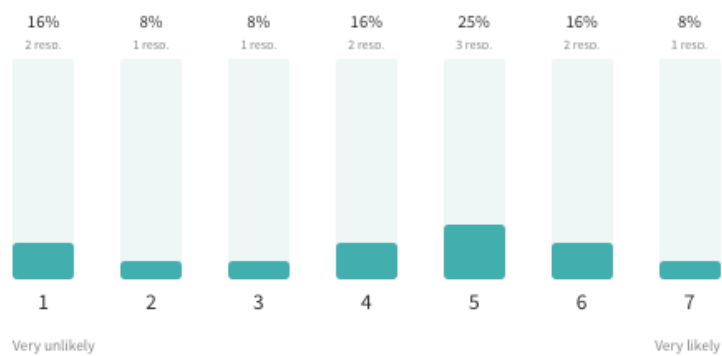
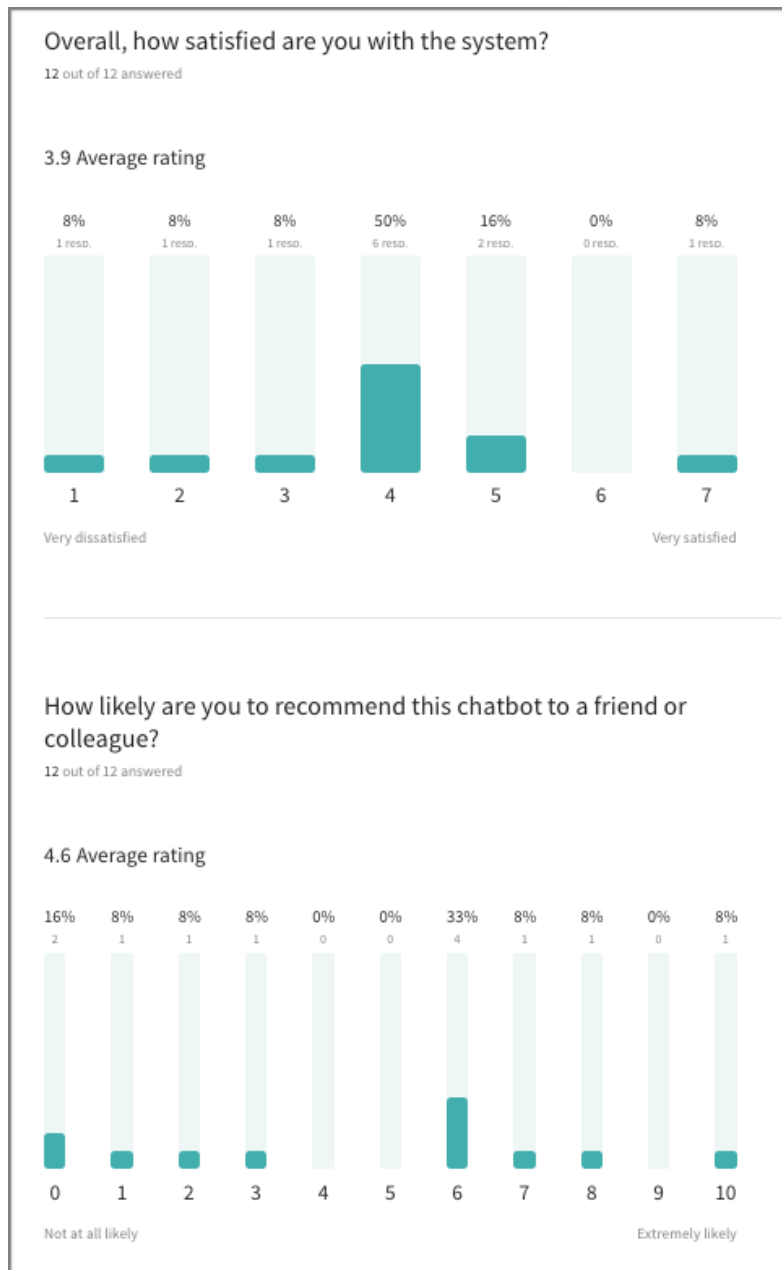## How likely are you to use this chatbot again if it were implemented for real?

12 out of 12 answered

5.4 Average rating

0%	0%	16%	0%	33%	25%	25%
0 resp.	0 resp.	2 resp.	0 resp.	4 resp.	3 resp.	3 resp.
1	2	3	4	5	6	7

Very unlikely   Very likely

Overall, how satisfied are you with the system?

12 out of 12 answered

5.2 Average rating

0%	0%	8%	8%	50%	16%	16%
0 resp.	0 resp.	1 resp.	1 resp.	6 resp.	2 resp.	2 resp.
1	2	3	4	5	6	7

Very dissatisfied   Very satisfied

How likely are you to recommend this chatbot to a friend or colleague?

12 out of 12 answered

7.7 Average rating

0%	0%	8%	0%	0%	0%	8%	25%	25%	8%	25%
0	0	1	0	0	0	1	3	3	1	3
0	1	2	3	4	5	6	7	8	9	10

Not at all likely   Extremely likely