UPPSALA
UNIVERSITET

UPTEC STS 19022
Examensarbete 30 hp
Juni 2019

Identifying & Evaluating System Components for Cognitive Trust in AI-Automated Service Encounters

Trusting a Study- & Vocational Chatbot

Joakim Eklund
Fred Isaksson

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Identifying & Evaluating System Components for Cognitive Trust in AI-Automated Service Encounters

Joakim Eklund & Fred Isaksson

The intensifying idea that AI soon will be a part of our everyday life allows for dreams about the complex relationship we one day could have with non-biological social intelligence. However, establishing societal and individual acceptance of AIpowered autonomy in disciplines built upon to the reliance to human competence raises a number of pressing challenges. One of them being, what system components will engender respectively counteract cognitive trust in socially oriented AI-automated processes?

This masters thesis tackles the seemingly ambiguous concept of trust in automation by identifying and evaluating system components that affect trust in a confined and contextualised setting. Practically, we design, construct and test an AI-powered chatbot, Ava, that contains socially oriented questions and feedback about study- and vocational guidance. Through a comparative study of different system versions, including both quantitative and qualitative data, we contribute to the framework for identifying and evaluating human trust in AI-Automated service encounters. We show how targeted alterations to design choices constituting the system components transparency, unbiasses and system performance, identified to affect trust, has consequences on the perception of the cognitive trust concepts integrity, benevolence and ability. Our results display a way of conduct for practitioners looking to prioritise and develop trustworthy autonomy. More specifically, we account for how cognitive trust is decreased when system opacity is increased. Moreover, we display even more concerning effects on trust due to micking contextual bias in the conversation agent

Handledare: Maria Mattsson Mähl
Ämnesgranskare: Kristiaan Pelckmans
Examinator: Elísabet Andrésdóttir
ISSN: 1650-8319, UPTEC STS 19022

Sammanfattning

Automatisering - att utföra arbete utan mänsklig inblandning, kan vara en viktig del i att effektivisera olika processer. Automatisering, som vi känner till det idag, påbörjade sin utveckling mellan den första och tredje industriella revolutionen och har sedan dess helt förändrat spelplanen för alla produktproducerande industrier. Då digitaliseringen mer eller mindre snart underligger alla samhällsprocesser, drar även tjänsteleverantörer i större utsträckning nytta av automatiserade processer. Vanligtvis berör automatiseringen av tjänster möjligheter till att t.ex. erbjuda detaljhandel eller kundsupport utan någon mänsklig interaktion. För mer socialt orienterade tjänster där mänsklig interaktion och dialog anses vara en stor del av värdet (handledning, psykiatri, studie- och yrkesrådgivning etc.) har datorer ännu inte kunnat konkurrera med den tillit vi placerar hos en mänskligt övervakad process. Artificiell Intelligens (AI), skapandet av intelligenta och självtänkande maskiner, öppnar dock upp för nya möjligheter att efterlikna och potentiellt ersätta mänskliga beteenden och intelligens. AI används redan inom flera områden av vår vardag och dess tillämpning förväntas bara att intensifieras. Övergången från konversationer mellan ansikte och ansikte till ansikte och AI i sociala dialoger kräver kunskap om hur vi ska översätta mänsklig trovärdighet till binära siffror. Att etablera förtroende kommer att vara avgörande för att säkerställa den mänskliga acceptansen och utvecklingen av AI. Särskilt i tjänster där det mänskliga interaktionsvärdet traditionellt anses vara högt.

Detta examensarbete tacklar det till synes komplexa och flyktiga konceptet tillförlitlighet i automatiserade processer genom att identifiera och utvärdera systemkomponenter som påverkar förtroende i en begränsad och kontextualiserad miljö. I praktiken designas, konstrueras och testas en AI-driven chatbot, Ava, som för en socialt orienterade konversation om studie- och yrkesvägledning. Genom jämförande studier av olika systemversioner, innehållande både kvantitativa och kvalitativa data, bidrar studien till ramverket för att identifiera och utvärdera mänskligt förtroende för automatiserade tjänstebemötanden. Studien visar hur systematiska förändringar av designval som utgör systemkomponenterna transparens, kontextuell subjektivitet och systemprestation, som identifierats att påverka tillförlitlighet, har konsekvenser för uppfattningen av de kognitiva förtroendekoncepten integritet, välvilja och förmåga. Resultaten visar ett tillvägagångssätt för utövare som vill prioritera och utveckla tillförlitlig autonomi. Mer specifikt redogörs det i studien för hur kognitivt förtroende minskar när systemets transparens minskar. Vidare, exemplifierar vi ännu mer alarmerande effekter på förtroende genom att imitera kontextuell subjektivitet i konversationsagenten.

Acknowledgement

The master's project has been performed in collaboration with the company AlphaCE Coaching & Education located in Uppsala during the spring semester in 2019. Other companies or organisations providing help along the way and are worth mentioning is; Arbetsförmedlingen, JobTechdev, Lundellska Skolan and Rosendalsgymnasiet.

First, we would like to express our sincerest appreciation to Maria Mattsson Mähl, Senior Advisor, Partner and Board member at AlphaCE for being hugely inspirational whilst supervising this master's thesis. We would also like to thank AlphaCE and their fantastic staff for housing us and always being present to answer all of our work-market related questions. Furthermore, we would like to thank Kristiaan Pelckmans at the Department of Information Technologies, Uppsala University. Thank you for your assistance and advice throughout this thesis, steering the project in the right path. Finally, for a lot of help and inspiration along the way, we would like to thank Davide Vega D'aurelio at the Department of Information Technologies, Uppsala University.

Joakim Eklund & Fred Isaksson
Uppsala, June 2019.

Distribution of Work

This master's thesis is the last and final step in the master programme in Sociotechnical Systems Engineering at Uppsala University during the spring semester of 2019. The study has been conducted by Joakim Eklund and Fred Isaksson, who together have had an equal part in the development of this report as well as building and testing the studied product. To work as efficiently as possible, responsibilities have been divided between the authors. When developing the chatbot, Fred was responsible for the conversation flow and content whilst Joakim was responsible for the backend functionality and integration. Although these responsibilities were set, a lot of interdisciplinary work was conducted by both authors. Also, when writing the report, specific chapters were divided between the authors for an effective writing process. Altogether the whole report was continuously revised and changed by both authors to maintain an equal quality all the way through.

Glossary

The following glossary defines central concepts in this thesis. Explanations of the concepts account for their respective meaning in the presented case and NOT as a general empirically based definition. Most concepts are elaborated and explained further later on.

Ability ~ A cognitive concept of trust in AS. Ability is a human's perception of the AS capability in performing expected tasks. This research treats the concept as the perceived reliability, capability and predictability of Ava.

AS ~ An acronym for “Autonomous System”, and a reference to a system that is automated.

Automation ~ Enabling a process to run automatically without human intervention.

Ava ~ An acronym for “Artificial Vocational Advisor” and the provided name of the social chatbot service created and presented in this study.

Benevolence ~ A cognitive concept of trust in AS. Benevolence is a human's perception of the AS underlying positive intentions towards the human. This research treats the concept as the perceived prejudice, motives and beliefs of Ava.

Biases ~ The antonym to unbiasses.

Chatbot ~ A computer program which automatically conducts conversations via the means of different communication mediums.

Conversational Agent ~ A synonym for *Chatbot*. Occasionally shortened to *agent*.

HMI ~ An acronym for “Human-Machine-Interaction” and a reference to the interaction between a human and a machine.

Integrity ~ A cognitive concept of trust in AS. Integrity is a users perception of the AS loyalty to a set of principles that the human user has agreed upon. This research treats the concept as the perceived honesty, motives and character of Ava.

NLP ~ An acronym for “Natural Language Processing” and a reference to a script that enables the processing and interpretation of unsorted typed or spoken language, using machine learning.

Opaque/Opacity ~ The antonym to transparency. Opaque and Opacity are used in different grammatical situations.

Service Encounter ~ The moment in which a human for the first time interacts directly with the frontline of a service. In some disciplines referred to as the "moment of truth”.

Social Chatbot ~ A computer program which automatically conducts socially oriented conversations via the means of different communication mediums.

System Component ~ A non-functional design aspect of the AS that is constituted by several functional design choices.

System Performance ~ In this research context, system performance is used as a general reference to a systems ability of performing expected tasks without faults or erroneous behaviour. Suggested to be one contributing system component, consisting of several individual design choices, to the perceived ability of Ava.

TIA ~ An acronym for "Trust In Automation" and a reference to the defined meaning of trust as a concept in the provided context of automation.

Transparency ~ In this research context, transparency is used as a general reference to the users overall possibility to view and understand the design, principles, functionality and limitations of a system. Suggested to be one contributing system component, consisting of several individual design choices, to the perceived integrity of Ava.

Trust ~ The overall psychological attitude achieved from beliefs and expectations about the AS trustworthiness. In this case, derived from the perceived integrity, benevolence and ability of a service encounter with a social chatbot, involving uncertainty and risk.

Unbiasses ~ In this research context, unbiasses is used as a general reference to objective and impartial system design choices. Optimising for mitigating the risk for including prejudisms and partisan beliefs in a certain context. Suggested to be one contributing system component, consisting of several individual design choices, to the perceived benevolence of Ava.

Table of Content

1. Introduction

In the following section, the framework for this thesis is defined. A brief problem statement acts as an introduction to the topic at hand, followed by explicit statements of the overall purpose and ambition of the research. Purpose, limitations and areas of focus are explained and motivated in the background section.

1.1 Problem Statement

Automation - doing things automatically without human intervention, is of crucial consideration for streamlining any process. Automation, as we know it today, evolved between the first and third industrial revolution and has completely changed the playing field for all product-producing industries. As digitalisation soon underlies all process of society, service-providing companies also benefit from automated processes to a growing extent, offering grocery shopping and travelling without the need for any human interaction. However, for socially oriented services where the human- factor and interaction is in itself considered to be of value (tutoring, medical care, psychiatry etc.), computers have not yet been able to compete with the trustworthiness associated with a humanly monitored process (Fearon & Maglio, 2018). Artificial Intelligence (AI), the science of making intelligent machines, is, however, opening up new possibilities for mimicking, and eventually possibly superseding human behaviours and intelligence. AI is already a part of our everyday life, and its encroachment is expected to intensify. This visionary dreaming that we one day could have artificially made social intelligence does, however, raise a number of pressing questions. One of them being, what system components will engender respectively counteract human trust in socially oriented AI-automated processes? What aspects of a human-machine-interaction (HMI) generates the trustworthiness needed for us to reveal personal information and to be influenced to the point where our image of AI is transformed from tools to teammates in all aspects of life. For practitioners, establishing trust will be vital in ensuring the acceptance and continuing progress and development of AI, especially in services where the social interaction value is considered to be high. Beyond the ability of machine learning algorithms, service developers will have to consider perceived integrity and benevolence as critical concepts to engender consumers trust (Hemment, 2018).

Addressing this comprehensive and interdisciplinary challenge demands research with regards to translating human trustworthiness in HMI into binary numbers. As such, this thesis aims to address a segment of the topic by identifying and evaluating significant components that seem to affect trust in a social dialogue with an AI-powered chatbot. By stationing the research at a vocational guidance company that provides labour-matchmaking services, this thesis utilises a self-evaluative conversation about one's professional career, as a suitable scenario to investigate the subject at hand.

1.2 Research Purpose

Trust in automation (TIA) consists of multiple cognitive concepts, each affected by several practical system components, individually constructed by numerous design choices. The main purpose of this research is not to isolate and quantify a singular design choice of a component targeted towards a concept of TIA. Rather, a broader and more commercially valuable approach is assumed. Firstly, the purpose of this research is to identify and confine significant system components that affect trust in the case context. Secondly, the purpose is to provide quantitative and qualitative data that displays that alterations to several design choices that construct the chosen components has effect on the cognitive concepts of TIA. In practice, this is done by reviewing existing literature, conducting market analysis as well as constructing and testing different versions of an AI-automated chatbot.

1.2.1 Research Question

Provided the purpose this thesis aims to answer the following research questions:

- What are significant system components that affect trust in a service encounter with an AI-powered study- and vocational chatbot?

- How do alterations to design choices related to the identified system components affect the targeted cognitive concepts of trust in automation?

1.2.2 Research Objective

Provided the research question, this thesis assumes the following objectives:

- Review existing literature to map previous findings, conduct market analysis to understand key context conditions and iteratively prototype and test an AI-powered chatbot. This is done to the extent where significant system components assumed to affect trust, that can be investigated in a quantitative, viable and credible way, are identified.

- Conduct a quantitative study on identified system components, by testing and evaluating deliberately altered design choices in different conversation designs.

1.3 Contribution to Science & Research Ambitions

In alignment with previous researchers this thesis aims to contribute to the framework for identifying, defining, measuring and evaluating human trust in AI-Automated service encounters. Suggesting that the research way of conduct may in itself act as a contribution on how to dissect and address the intricate subject at hands. Furthermore, the quantitative research results aim to contribute to the knowledge about how and to what extent the identified system components affect trust in a socially oriented HMI.

1.3.1 Focus & Delimitations

As stated in the purpose, TIA is an ambiguous and intricate topic consisting of a dynamic and multidimensional framework depending on contexts, definitions and perspectives. Therefore, an effort to conduct a generic and one-dimensional quantitative study on a singular design choice was neither academically or commercially motivated. Rather, the choice of research focus and method is based on a broader and more commercially valuable perspective. Focus is placed on contextualising existing findings in the chosen case and providing interesting results useful in further development. Consequently, we account for the validity, reliability and credibility of our results in the light of the chosen method. Provided the complex theoretical nature of evaluating TIA this study delimits related topics. Such as trust in performance-based chatbots, systematic empirical comparison between HMI and Human-Human-Interaction and trust associated to communication mediums other than text.

1.3.2 Master Student Ambitions

Using knowledge from relevant courses such as Interface Programming with a User Perspective, Science and Technology Studies, Artificial Intelligence, Data Mining, Software Engineering and Project Management to mention a few, we aim to develop our skills not only as researchers but as practitioners. Combining exercise within researching, prototyping, coding and project managing we aspire to gain interdisciplinary insights that test, confirm and summarize our learnings as masters in sociotechnical systems engineering.

2. Background

In the following section, a general project background is provided, containing a motivation to why the chosen case is justified with regards to the purpose of the thesis. Furthermore, by assuming the case perspective, previous research is contextualised through explanations of assumptions and chosen fields of research. Limiting and advocating the theoretical framework explained in detail in the next section. Finally, a brief introduction of relevant but delimited research fields is provided.

2.1 Case Significance on a Practional Level

The majority of existing research within human reasoning and decision making in professional pursuits takes entry point in assuming that individuals actively and consciously plan and perform steps in their professional career. This suggests that a person knowingly picks an occupation it seemingly will perform well in and become content with. However, more recent research suggests that self-perception is rarely one of the major drivers in occupational decision-making (Lundgren, 2012). For an optimal career, in terms of satisfaction and performance, this presumes individuals to be fully aware of their characteristics as well as their professional interests. However, individually and unbiasedly evaluating personal traits and comprehensively matching these towards the immense number of possible professional paths is an exception rather than the norm. Instead, parameters such as external social factors, availability and motivation are more likely to influence a person's professional route (Lundgren, 2012).

Individuals are most likely to feel satisfaction in their occupation when their personality aligns with their delegated tasks as well as their working environment (Allport, 1935). Basing career decisions on something else than personal traits and interests then becomes contradictory to the theory about occupation-personality-compatibility. Well-renowned studies conclude that individuals that feel content with their circumstances and surroundings, will perform better (Holland, 1997). This will either occur if the individual actively seeks personal compatibility in their occupational pursuits or if they adjust and adapt traits in a given working scenario. Assuming there is a common conception of not having actively planned one’s professional life, but rather "ending up" somewhere, due to a set of seemingly more random variables then seems conflicting. Either way, self-perception seems to play a crucial role in establishing efficiency and performance through proper work-role-matching. Furthermore, recent and extensive results display the impact of job productivity depending on the personality type (Najam-us-Sahar, 2015). For practitioners, at a strategic level, this motivates the acknowledgement that the personality of an employee could have a profound effect on the productivity of an employee. However, rather than solely focusing on the demand side of the equation, that is finding ways for organisational entities to sort and recruit individuals based on suitable characteristics. It seems beneficial to assume the individual’s perspective and provide tools that allow people to more actively and efficiently establish self-perception and awareness in their professional ambitions.

AlphaCE - Coaching and Education is a study- and vocational guidance company that supports people in their professional careers by offering support, advice and labour-matchmaking services. The described fundamental value of their service is prioritizing the individual's interests, needs, ambitions, possibilities and then matching them to the current market situation, rather than primarily focusing on labour market demands. This means that beyond tangible services like matchmaking and network, the value of their guidance is based around increasing self-perception and self-development by asking the "right" evaluative questions. As face-to-face advisory sessions are time-consuming and difficult to scale, even just partial automation would allow for larger scale possibilities in terms of distribution and service value (AlphaCE, 2019). As to such, the shared vision between AlphaCE and the authors of this thesis is to, by the means of an AI-powered chatbot, research possibilities for automating sections of the "self-evaluative-dialogue". The ambition is not, even in the case of successful and sophisticated chatbot, to replace sections of a human-to-human dialogue, but rather to explore it as an easily accessible tool for participants looking to quickly better their self-perception. As advisory sessions are service encounters - when a customer interacts directly on the frontline of the organization for the first time, it quickly became clear that creating trust in such a service would be of crucial consideration, even beyond technical implementation. Furthermore, given the nature of the aspired conversation, where the goal is to influence individuals to evaluate personal traits and interests, the very value of such a service relies on establishing a trustworthy relationship. Meaning that for such an interaction to withhold its value and quality, consumers would have to perceive similar trustworthiness in the chatbot as they place in advice from a human counsellor. For this reason, AlphaCE together with the authors coherently defined the scope of this project to be targeted towards researching trust in the given scenario. Furthermore, envisions of that encounters with a machine rather than a human, it the context of discussing personal traits, could have natural benefits such as talking to something neutral and judgement free, took place, motivating the chosen field of research even further (AlphaCE, 2019).

2.2 Research Contextualisation

As technical developments and implementations of AI and intelligent machines proceed, so does the envisions of what complex processes robots will be able to become active participants within. As future scenarios with intricate software are being drawn, our expectations of machines aiding us in everything from work assistance to crisis response increase. This visionary process converges our image from AI acting as tools to instead assuming the role of teammates in all aspects of life (Lewis et al., 2018). Although it is tempting to solely focus on the complex relationships we one day could have with artificially made intelligence, these scenarios raise a number of pressing questions. One of them being, what factors will engender respectively counteract human trust in AI-automated processes?

In order to address this comprehensive topic in a viable way, the scope needed to be delimited beyond general trust in chatbots. Although there are numerous tasks, situations and environments where HMI can be envisioned, there is a general distinguishment between two types. It can either be seen as performance-based, where the main objective is for the human to influence and control the machine in such a way that it performs wanted and useful tasks. Secondly, there is social-based interaction, which focuses on how the behaviour of the machine influence the human's beliefs and behaviour. In either case, the human is always the trustor and the machine the trustee (Devitt, 2018). Provided the case of a study- and vocational chatbot, a performance-based version would have a particular task with a clear performance goal. With reference to the explained service value, this interaction type does not seem very fitting. This is due to that the primary ambition is not to provide a recommendation system, that takes input from the user, in terms of personal traits and interests, finally returning an output within the scope of a job suggestions. Instead, the proposed chatbot focuses on influencing, by asking questions and providing feedback, the interacting person in such a way that the overall self-perception is increased, leading to the person coming to conclusions of its own. As to which, the case falls into the realm of a social interaction, where the goal is not as crisply defined. Ideas for performance measurements could for example be targeted towards influencing a human to reveal private knowledge, or investigating how well the chatbot can influence a human to do useful mind-exercises to increase self-perception.

2.3 Research Background

The explosive establishment of AI-powered processes in all disciplines has fuelled a mass of literature that addresses the impact on business models and implications for firm scalability and growth. However, less focus is placed on a customer-centric viewpoint and end-user impacts. For frontline interactions, establishing knowledge and making predictions about what applications will be accepted and which will not and why will be imperative (Juma, 2016; Leung et al., 2018). Anchoring AI as a part of our society will require insight beyond technical advancement, considering and understanding components for the human acceptance of such services. As to why this thesis aims to contribute with results about user-trust of AI in the context of frontline service encounters. More specifically focus is placed on identifying and evaluating system components that engender trust and its effect on human acceptance of AI-powered services. Trust is what bridges the gap between a humans perception of characteristics and abilities of automation and the individual's intentions to use and rely on a service (Lee & See, 2004). Furthermore, trust is particularly critical in the early stages of a relationship, which fits the case of researching at the time of first interaction.

However, trust is an interdisciplinary and multidimensional concept where investigations have been made in a broad domain of disciplines. Only with regards to understanding the relationship between human and machine, trust has been studied in perspectives ranging from social psychology to industrial organisation. As a result, the definitions and theories regarding trust in the context of HMI are various and many. Studies suggest that, depending on the setting, trust can be seen as a behaviour, intention or attitude (Madsen & Gregor, 2000; Mayer et al., 1995; Moray et al., 2000). Consequently, both within the general literature and the contextual results from researching TIA, there lacks a generally agreed upon definition. Upon review of existing literature, there seems to be a general conception that the absence of a universal definition of trust, even in specified settings, is a result of the complexity and multidimensionality of the concept rather than a symptom of contradictory research results. As to such, there seems to be a controversial aspect of studying trust in technology (Fryer & Carpenter, 2006). However, there is, in contrast, a rapidly increasing body of research addressing the notion. Furthermore, as trust is a function of personal definition, there is a likely possibility that participants in different studies will evaluate trust differently. For this reason, a case-to-case contextualisation seems to be the general approach.

The existing literature provides clues to what cognitive concepts, system components and design choices seems to influence trust in chatbots. However, due to the set of highly contextual characteristics, the need of exploring trust in continuous regards to the specific case context is emphasised. In this study, this is done by market research, company interviews and prototyping. The theoretical section explains previous findings of the identified and evaluated components, their connection to the cognitive concepts of trust and how they their constituting design choices can be altered. However, how and why these components were identified to be of interest and value and how we practically altered design choices is saved for the method and results section. For now, just note that the identification of the general framework for evaluating trust in AS provided the basis upon identified research components could be justified.

2.4 Related Research Fields

A consequence of trust being such an interdisciplinary concept is the difficulty to confine its study to a certain area. In this study, we do our best to stay within the defined scope but might at times unconsciously touch related research fields. Therefore, we below provide a brief introduction to closely associated subjects.

2.4.1 Automation VS Human Trustee

Although there was no explicit ambition of envisioning a chatbot to relieve even just sections of a dialogue made with a career coach at AlphaCE, the idea of to some extent compare trust between an equivalent human-human-interaction and a HMI was repeatedly considered. However, due to difficulties in establishing equivalent dialogues as well as practical challenges, it was determined that this method would stay outside the scope of the project. Nonetheless, some previous research has explored the concept of trust by comparing automation and human trustee (Lewandowsky et al., 2000).

Interestingly, the results primarily show that the main components of trust are similar in the two cases, where fault decrease trust in either case. Secondly, it was shown that the sole distinguishment of expected reliance on automated interaction vs human was the distinction between trust and self-confidence. Leading to the third and final conclusion which is that participants in a human-human-interaction were more likely inclined to increase control over a task if their own self-perceived trustworthiness in the given scenario was high. However, in a HMI, the self-aware trustworthiness had no effect. More fittingly for this case, studies have been done with regards to comparing human trust in identical advice given by either a machine or a computer (Lerch et al., 1997). Yet, the results of the research are partially contradictory. One study indicated that participants were biasedly trusting of the human's advice even though the agreement of the content of the advice was equivalent. Another study within the same research instead implied that participants tended to agree with the advice from the machine more but had overall lower confidence in the system than in the human (Lewis et al., 2018). Similar conclusions have been made by showing that fault made by a machine did not disturb the level of influence a piece of given advice had on the human (Salem et al., 2015). However, faults did affect the subjective evaluation of the reliability of the machine and therefore overall trustworthiness (Robinette et al., 2016). Even though this study delimits a comparative study between human-human-interaction and HMI, the above-mentioned results were acknowledged as guidelines on as how to select interesting areas of focus as well as recommendations on how to conduct the study. Furthermore, other findings showed the importance of how reliance and trustworthiness is a function of the situations perceived risk. Meaning that depending on the seriousness of the topic in the interaction and the consequences it may have, leaning towards human or machine trustworthiness may differ (Lyons & Stokes, 2011).

2.4.2 Different Communication Mediums

Commonly concluded within the literature is that humans are likely to judge chatbot characteristics based on their anthropomorphic properties, subconsciously making comparisons to a human-human-interaction. In order to do so, effective social chatbots are not uncommonly able to interact with humans in several communication modalities, such as text, speech and vision. In addition, a correlation between trust and pedagogical and expressive interfaces have been suggested (Lester et al., 1997). There are studies that elaborate on the impact on trust from anthropomorphic properties by means of expressive graphical interfaces and advanced communication modalities (Ciechanowski et al., 2018). This study considers anthropomorphic features such as how empathic language engenders trust (Tapus et al., 2007) and how the level of expressiveness affects the likeliness of disclosing personal information. However due to time and technical limitations the communication modality is limited to text. Furthermore, personally tailored responses through identity profiling had to be delimited (Heung-Yeung et al., 2018).

3. Theory

The following section highlights and explains previous findings in the area of research as well as various key concepts that are of importance to the investigation. The presented references constitute the theoretical framework which this thesis results, analysis and conclusions lay upon. The headlines and paragraphs of this section follow a logical order, where new information is layered upon previous. Finally, a summary is provided, followed by the analysis model.

3.1 Researching Trust in Automation: Contexts & Concepts

One of the major challenges within automation is to establish appropriate reliance. Because humans tend to approach technology in a social manner, trust lays the basis for the level of reliance in an automated process. More specifically, trust guides reliance. For this reason, the philosophical literature on trust differentiates reliability and trust. Trust occurs between seemingly conscious entities while reliability is a property that an inanimate machine has. For example, we don't trust a shelf to hold our books, but we rely on it to do so (Hawley, 2012). Trust involves psychological relationship components like the capability to apologise if a mistake occurs. As to why trust can be treated like a psychological attitude in which a human allows themselves to be vulnerable since they are confident that the autonomous system (AS) will not exploit them (Nave & Camerer, 2015). Further, a state of trust is a social feeling of mutual confidence that comes from truth-telling, loyalty and empathy in an interaction (Arrow, 1974). However, a shelf has no attitude towards its purpose. Similarly, automated processes are designed to complete a set of tasks within a domain where there is no self-driven desire to maintain their reputation. To tackle the realisation that TIA is cognitively based rather than just a summary of behaviour, assessing trust should be done with regard to the context in which the automation is performed. By limiting the context and acknowledging that the situation will influence perceived performance, a goal-oriented perspective can be obtained (Lewis et al., 2018).

3.1.1 Contextualising Trust in a Chatbot Service Encounter

Service encounters are a central feature within service management and have a strong connection to customer satisfaction and loyalty (Bitner & Wang, 2014; Gupta & Zeithaml, 2006). Before AI and even automation, a service encounter was defined as "*the dyadic interaction between a customer and a service provider*" (Surprenant & Solomon, 1987, p. 87). Back then, service providers where human and represented the "face" of an organisation. Nowadays the "face" has in most cases faded into the vast online and our phones, more recently in the shape of AI-powered applications. A contemporary definition of a service encounter has therefore been suggested as "*any customer-company interaction that results from a service system that is comprised of interrelated technologies (either company- or customer-owned), human actors* *(employees and customer), physical/digital environments, and company/customer processes.*" (Larivière et al., 2017, p. 239).

In this study, this definition is contextualised and simplified as; "*The moment in which a human for the first time interacts directly with the frontline of an automated service*".

A chatbot is a computer program which automatically conducts conversations via the means of different communication mediums (Følstad & Brandtzæg, 2017a). Chatbot technology has existed for decades but proceedings in AI and machine learning has spiked the interest for conversational agents in customer service (Vinyals & Le, 2015). Furthermore, the establishment of messaging platforms has contributed to the uprise of conversational agents in a broad domain of industries (Følstad & Brandtzæg, 2017b). Although novel technology allows for more sophisticated natural language processing, customer service typically requires highly personalized customer interaction, involving skilled customer service personnel (Dixon et al., 2010). However, developments in sentiment analysis and contextual understanding act as tools for a more accessible and efficient intelligent automation (Xu et al., 2017). Allowing for AI-powered chatbots to be used in more high-risk socially oriented service encounters. Such areas could be education (Friedman et al., 2007) and therapy (Fitzpatrick et al., 2017).

For service providers in socially-oriented areas, the quality of automated customer service is going to be even more crucial than customer services regarding less complex information retrieval. Drawing upon the distinguishment between socially-oriented chatbots and performance-based chatbots there are several differences to be considered with regard to the overall purpose and its measurements. Social-based interaction targets how the behaviour of the machine influences the human's beliefs and behaviour (Lewis et al., 2018). This implies dealing with a closer resemblance to how humans interact with each other. Previous studies of this kind highlight the level of influence a machine has over a human's perception of parameters like trustworthiness, companionship, comfortability. Although performance-based interactions are seemingly more quantifiable, there is a considerable amount of results that display how chatbot design principles affect ratings of trust, without explicitly defining a social performance goal (Følstad et al., 2018). However, the current body of contextualised research on how trust affects perceived quality and experience of social chatbot interaction is sparse. Nonetheless, findings from the generic literature on TIA can be used to create contextualized frameworks to evaluate trust in particular services (Lewis et al., 2018). Furthermore, based on proven cognitive concepts of trust a systematic operationalisation of engendering trust through different system components can be achieved. These components can then be isolated and their constituting design choices varied to make analysis of their impact on the cognitive concepts of TIA (Følstad et al., 2018; Hieronymi, 2008; Keren, 2014; Simpson, 2013).

3.1.2 Cognitive Concepts of Trust in a Social Chatbot

Based on that system components have an impact on user trust, comes the process of distinguishing their significant constituting design choices in a certain context. Dissecting TIA into fundamental system components needs careful consideration of the goal of the interaction. Accepting that there are numerous design choices that can be made with regard to establishing trust in dialogue with a chatbot, a goal-centred viewpoint helps in prioritising relevant aspects (Devitt, 2018). Drawing upon the objectives of a social chatbot raises the importance of considering psychological concepts of trust. Although the past literature (Muir, 1994; Barber, 1983) differs in the cognitive division of trust, three general concepts are distinguished. Users will consider the *integrity, benevolence and ability* of a service in relation to their individual expectations and experiences. *Integrity* is the AS loyalty to a set of principles that the human has agreed upon, *benevolence* is the AS underlying positive intentions towards the human and *ability* is the AS capability in performing expected tasks. Each one of these concepts that constitute trust in AS can be affected and targeted by practical system components (Lewis et al., 1997).

Integrity consists of perceived honesty, motives and character. Humans tend to trust an AS that is trying its best and that takes responsibility for its actions. Achieved through translucent actions and empathic characterisation. Suggesting that design choices such as a reference to the hosting brand's legitimacy and the chatbots self-presentation will affect trust (Følstad et al., 2018). In this research, the cognitive concept of *integrity* is targeted by linking alterations to design choices related to system *transparency* as a contributing component to the trust of an AS (Devitt, 2018).

*Benevolenc*e consists of perceived prejudice, motives and beliefs. Trust is influenced by individual faith in AS motives. Individuals base trust in their preconceived notions about what they think the AS want's to achieve. Suggesting that design choices such as the professional appearance and perceived altruistic character of the conversational agent will affect trust from the user (Følstad et al., 2018). In this research, the cognitive concept of *benevolence* is targeted by linking alterations to design choices related to contextual *unbiasses* as a contributing component to the trust of an AS (Robinette et al., 2015).

Ability consists of perceived system reliability, skills and accuracy. An AS can have good capabilities, yet not have the right skills for a certain task, or occasionally fail to perform a task expected to be in their domain. Furthermore, unexplained erroneous behavior and lack of interaction accuracy will affect trust. (Følstad et al., 2018). In this research, the cognitive concept of *ability* is targeted by acknowledging and linking differences in design choices related to *system performance* as a contributing component to the trust of an AS (Hoffman et al., 2013).

Although further distinguishments can be made to the complex framework that constitutes trust in AS, these divisions allow for a structure to ground viable investigations (Connelly et al., 2015; Connelly et al., 2012; Kim et al., 2004; Luo, 2002). Furthermore, it is important to note that these concepts are closely connected and interlaced. Although this study assumes particular targeting between certain system components and cognitive concepts of trust, their merge is what will collectively affect trust in an AI-powered service encounter. Moreover, the volatility of established trust will depend on the stability of each system component and to what extent the cognitive concepts are perceived to be satisfied (McKnight et al. 1998). As such, this study is based on the previous agreed upon conceptualisation of trust as:

"*A multidimensional psychological attitude involving beliefs and expectations about the trustee's trustworthiness derived from experience and interactions with the trustee in situations involving uncertainty and risk*." - (Lewis et al., 2018, p.137).

However, provided above referenced previous findings about the need for contextualisation. This investigation specifies and contextualises the definition of trust in the certain case as; "*An overall psychological attitude achieved from beliefs and expectations about the AS trustworthiness. Derived from the perceived integrity, benevolence and ability of a service encounter with a social chatbot, involving uncertainty and risk".*

3.2 Targeting Integrity with System Transparency

There are several system components that contribute to what previous research refers to as the *integrity* of an AS. Within the realm of chatbots, integrity can be treated as the result of overall perceived honesty and character of the dialogue (Følstad et al., 2018). Furthermore, there are several design choices that contribute to how a human will experience honesty and character of a chatbot interaction. In this research, chosen design choices are gathered under the system component transparency. Moreover, two sources of trust affected by transparency are distinguished. Transparency in the provider and transparency in the transaction medium (Følstad et al., 2018). Previous research suggests that humans tend to trust an AS that is explicit and transparent about its nature, functionality, limitations and behaviours (Devitt, 2018; Brandtzaeg et al., 2019; Mone, 2016; Luger & Sellen, 2016; Castelvecchi, 2016). Furthermore, results regarding the brand legitimacy, referenced content and promised privacy and security are linked to trust affected by the transparency and integrity of the service provider (Kretzschmar et al., 2019; Følstad et al., 2018).

3.2.1 Transparency in the Conversational Agent

Within the distinguishment, that trust in social chatbots has a closer resemblance to interpersonal relationships comes the sought for imitating anthropomorphic features of human conversational agents. In some contexts, it might even be tempting not to be transparent regarding the conversational agent's nature as a bot or human. However, the lack of transparency about the systems true nature might cause uneasy feelings in the user. Admitting that interpersonal interaction will differ from HMI due that human and chatbot capabilities differ, transparency of a conversational agent and its limitations becomes significantly important. (Brandtzaeg et al., 2019). By designing chatbots to be upfront about their machine status, negative implications of users perceiving the agent as a human can be avoided (Mone, 2016). Furthermore, this decreases the probability of mismatches in user expectations and system capabilities which can have a serious impact on the overall system experience. Arguing for chatbots to be open about their limitations (Luger & Sellen, 2016). In practice, the explicitness of conversational agents abilities and limitations is done through proper self-presentation. Studies show that by communicating what the system is able to do and to what extent it can provide assistance, user expectations can be managed, increasing trust as a result. More specifically, self-presentation is argued to be done at the beginning of interaction and should include a declaration of system nature, abilities and limitations (Kretzschmar et al., 2019).

Beyond transparency in limitations and system abilities, there are findings that relate trust to the perceived dialogue character. Suggesting that depending on the degree to which the conversation has been thoughtfully developed will affect user attitude. More specifically, this concerns using suitable, adequate and correct language in a certain context (Brandtzaeg et al., 2019). Another aspect of system character in a provided context and a common barrier to the adoption of AI is algorithmic transparency. There are numerous studies that address the vulnerabilities of presenting AI-powered applications as "black-box" systems. In a similar fashion, conversations agents can suffer from complex and intricate reasons to why the conversation is behaving the way it is (Castelvecchi, 2016). Previous research implies that trust is engendered from that users are provided motivations from how responses are chosen and interpretations of system algorithmics that are interpretable and their technical level. This means providing visibility to why the system is doing as it is and providing ways for the user to understand the influence and justifications of machine reasoning (Hepenstal et al., 2019). Furthermore, it has been considered of importance to provide motivations to certain areas of topics mapped against the goal of the interaction. By providing the goal and the constraints of dialogue, key elements of the context are visualised, motivating certain priorities (Kretzschmar et al., 2019).

3.2.2 Transparency in the Service Provider

Beyond trust affected by the conversational agent's character and self-presentation, there are aspects of trust engendered by transparency of the service provider (Følstad et al., 2018). More specifically this partly relates to trust in the brand hosting the chatbot. Suggesting that the branding of the chatbot and where and how the conversation is accessed is of importance. Findings imply that reference to a legitimate brand increases the trust in content provided by the agent representing the brand. This is due that the service provider needs to take responsibility in the case of that the chatbot fails or provides misleading information. Furthermore, provided that the conversational agent provides sensitive information, the user should be informed about the extent to which the service is backed up by research and evidence (Kretzschmar et al., 2019).

Previous research shows that trust in a chatbot is subject to contextual issues, such as concerns for security and privacy. Service providers need to create the perception that the automated chatbot service is just as secure as interpersonal interaction. Chatbots should early in the conversation convince the user that the service provider guarantees a sufficient level of security for a certain context. With an emphasis on trust engendered from integrity, findings further suggest that chatbots should be transparent about to what extent the service provider stores personal data from the interaction. More specifically, the results imply that chatbot preferably should store as little personal data as possible (Følstad et al., 2018).

3.3 Targeting Benevolence with System Unbiasses

Trust related to the concept of *benevolence* of an AS has been treated differently in previous research. With regards to chatbots, the concept can be conceptualised as the perceived altruistic character of a conversational agent, consisting of its prejudice, motives and personal beliefs (Følstad et al., 2018; Robinette et al., 2015). This research targets trust affected by the perception of benevolence through alterations of design choices constituting the practical system component of unbiasses. More specifically, how user notions about the agent's personal opinions and how differences in semantics affect user trust are considered. Previous results suggest that trust engendered from the natural characteristics of AS due to their unbiased nature might be harmed from the development of prejudiced opinions and partisan semantics (Caliskan et al., 2017; Følstad & Brandtzaeg, 2017b; Fuchs, 2018; Zheng & Jarvenpaa, 2019). Although a closer resemblance to human dialogue might increase trust (Tapus et al., 2007; Følstad & Brandtzaeg, 2017b),

3.3.1 Unprejudiced Opinions

Constructing technology by learning and mimicking human behaviours and expectations is the very basis for AI and consequently its resemblance to human intelligence. In order for such behaviour to be trustworthy, the design and characteristics of AS must be overseen by socially accepted ethical principles. However, there is a rising concern that such technology could imitate, with or without intention, the prejudices, failings and unfairness that characterise many of society's institutions. Whether it be automated service encounters that practices gender and racial biases or a chatbot with an NLP human-like semantic bias (Caliskan et al., 2017). Moreover, the discussion about biased within AS is closely related to the one about transparency and integrity. Ambiguous evidence backing content or suggestions prompted by the conversational agent might lead to opacity when transparency is needed. Furthermore, unreasonable evidence might lead to biased responses when the functionality of a system reflects the designer's values (Zheng & Jarvenpaa, 2019).

Bias within AI and machine learning in its purest form is merely a reference to prior information, a necessary prerequisite. Social chatbots are generally constructed to converse like a human, offering personal perspectives and prompting new topics to maintain the dialogue. However, biases could become harmful when the data is generated from subjective precedents. Such biases require deliberate action based on awareness of ethical challenges that the certain context has (Caliskan et al., 2017). Society has historically had shortcomings with regards to biases and inequality in numerous institutions, the vocational and educational scene being one of them (Byrnes & Kiger, 1992). As such, it is reasonable to assume (Mittelstadt & Floridi, 2016) that there is a risk that an AS could make similar contextually biased suggestions with regard to education and labour. For example, promoting choices that society historically has perceived as better or of higher class.

Previous research within the field of trust in social chatbots shows that interaction with a conversational agent could have beneficial characteristics over interpersonal interaction due to its unbiased nature. More specifically, qualitative interviews with users of social chatbots suggest that the threshold for answering truthfully might be lower, knowing that the agent will not judge or value the answer. Participants in tests imply that the fear of thinking or saying stupid or silly things is lower, knowing that the receiver does not have an opinion (Følstad & Brandtzaeg, 2017b). Assuming that one of the key elements of engendering trust in social chatbots is their very unbiased nature, allowing and designing for biased responses could have a significant impact on user trust and experience (Fuchs, 2018).

3.3.2 3.3.2 Nonpartisan Semantics

Furthermore, findings have shown that empathic language engenders trust (Tapus et al., 2007). It has been found that the level of expressiveness in the communication provided by the bot in dialogue with a human affects the likeliness of disclosing personal information. Furthermore, politeness and positive attitude from the agent has been reported to benefit communication with chatbots (Følstad & Brandtzaeg, 2017b). However, semantics - the meaning of words, just as opinions, might also reflect prejudiced regularities latent in our culture. Results imply that widely used NLP tools are prone to similar biases that humans have in psychological studies (Caliskan et al., 2017). Since these tools are based on data from the ordinary web, they are exposed to the same language biases that a human would. For example, previous literature has empirically proposed the existence of gendered wording in job recruitment materials. More specifically, the results suggest that job advertisements not uncommonly maintain masculine wording, words associated with stereotypes, for traditionally male-dominated occupations (Gaucher et al., 2011).

Although human-like semantics might to some extent increase trust in social chatbots, there is a risk that partisan semantics might affect the users perceived benevolence of the conversational agent. Suggesting that the use of loaded terms in a certain context might create uneasy feelings in the user (Følstad & Brandtzaeg, 2017b; Singh, 1999).

3.4 Targeting Ability with System Performance

Similar to previously described trust concepts, ability can be conceptualised in different fashions with regards to an AS. In the case of a chatbot, ability refers to the perception of system reliability, capability and predictability. Suggesting that trust will be affected by the system skill, fault and erroneous behaviours and the acknowledgment of system vulnerability and mistakes (Følstad et al., 2018; Lewis et al., 2018). This investigation targets trust affected by perceived ability by acknowledging design choices related to the practical component of system performance. More specifically, this study examines how faults, lack of accuracy in responses and accountability for erroneous behaviour affect trust in a chatbot (Hoffman et al., 2013). Previous findings suggest that there is a correlation between trust in AS and differences in performance (Salem et al., 2015; Moray et al., 2000; Lewis et al., 2018; Følstad et al., 2018; Lewandowsky, 2000: Hancock et al., 2011).

3.4.1 System Reliability

There is an extensive body of literature that confirms the relationship between system faults and trust (Lewis et al., 2018). Previous findings suggest that mistakes performed by an AS significantly affect the subjective assessment of an agent in terms of trustworthiness (Salem et al., 2015). Furthermore, findings suggest that declining system reliability can lead to the decline of system trust and expectations. Moreover, it has been shown that the magnitude of erroneous behaviour has differential effects on trust, smaller mistakes have a smaller effect while large faults have a more significant impact (Moray et al., 2000). There is an interesting difference between interpersonal trust and HMI trust when it comes to tolerance for faults. Humans tend to build trust inductively between each other, allowing for mistakes to be made. However, in the case of HMI, the trust relationship can be broken by a single instance of error (Lewis et al., 2018). Moreover, there have been suggestions that erroneous behaviours influences user willingness to comply with further suggestions and questions. These ideas yet lack empirical grounding but it has been shown that user likeliness to reveal information decreases as erroneous and ambiguous behaviour increases (Salem et al., 2015).

3.4.2 System Capability & Predictability

Although less explicit than obvious system failures it has similarly been shown that a chatbots ability to interpret and accurately reply to specific input affects user trust. Moreover, the chatbots ability to provide helpful and precise responses has been considered as a key component for engendering trust. In other terms, its efficiency in achieving and performing wanted tasks. (Følstad et al., 2018). The aspect of accuracy in responses is closely related to the one about the perceived capabilities of an AS. Previous research has suggested that trust and the skill of an AS forms a quadratic relation (See figure 1). Trust is increased the more competent an agent is perceived. However, as the AS competence exceeds human levels beyond user understanding, then trust is decreased (Lewis et al., 2018).

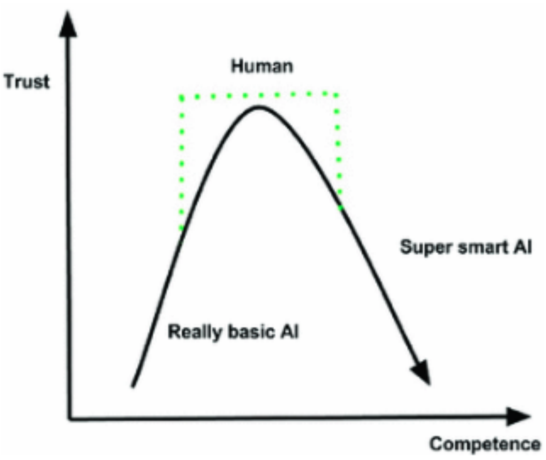


Figure 1. Quadratic relationship between trust and agent competence (Lewis et al., 2018).

From assuming that system accuracy and the perception of chatbot competence plays a role in establishing trust, comes the aspect of system predictability. Findings suggest that although mistakes and erroneous behaviour affects trust in the AS, this is partly due to that the user has little knowledge of why the failure is occurring (Lewandowsky, 2000). Studies suggest that when users have previous experience or are informed about system vulnerability, faults may have a smaller impact on the trust relationship (Riley, 1994). An explanation for this is suggested to be that when the user knows that the agent might fail, then user uncertainty and confusion is decreased in the case of a failure. Beyond perceptions of integrity this connects effects of system *transparency* to experiences of ability. Implying that predictability might be equally important to reliability (Lewis et al., 2018).

Beyond admitting and explaining system shortcomings that might occur, it has been shown that a limiting factor for establishing trust is the dissonance between expected skills and actual ability. Participants in tests have reported disappointment in agents capability of answering questions thought to be viable to ask (Følstad et al., 2018). Findings suggest that current chatbots are unable to answer more complex questions or questions more specific to their certain conversation. The problem is partly due to technical limitations but also failings in managing expectations through transparent limitations of skills. Furthermore, this relates to agents ability to mimic human intelligence. Studies have highlighted anthropomorphic characteristics to have a strong impact on trust engendered from technical capabilities (Hancock et al., 2011). Findings imply that ratings within the scope of technical performance, such as reliability and skills, are affected by perceived human-like properties (Bainbridge, 2008).

3.5 Surrounding Influences on Trust

The main focus of this investigation is to evaluate system components, expected to affect trust with regards to different cognitive concepts. However, due to treating trust as a cognitive attitude achieved through individual experience, individual perception needs to be considered. Although individual expectations, prerequisites and contextual risk lie beyond the scope of the quantitative part of this study, such influences on trust are acknowledged in accordance with previous scholars (Hubal et al., 2008; Louwerse et al., 2015; Nass & Brave, 2007; Van Mulken et al., 1999).

3.5.1 Expectations & Experience

The discussion about predicting and overseeing expectations is closely related to the establishment of trust in AS. This is due to that design choices relate to changes in perception and human analogies by engendering stereotypes and presumptions, suggesting that expectations need to be treated as variables that indirectly affect trust (Louwerse et al., 2015; Nass & Brave, 2007). User perception of a chatbot as a high-quality artefact has been found to engender trust, further emphasising the question about what user expectations are to be considered (Van Mulken et al., 1999). As to such, TIA needs to be put in the perspective of how well users expectations are met. In the case of a performance-based chatbot, where it per definition has a clear performance objective in mind, user expectations are relatively easily contained towards efficiently achieving a particular goal. Take a chatbot for forecasting the weather for example, in this scenario the user has a clear expectation on what the chatbot will deliver, namely the weather on a given time. Allowing for a more explicit process of predicting and meeting expectations as a dimension of a trustworthy interaction. However, as social chatbots are designed to influence humans in different regards without a quantifiable performance metric, encompassing expectations becomes more intricate (Lewis et al., 2018). As to such, beyond the perspective of designing and evaluating a social conversational agent on trust engendering system components, a user-centred viewpoint for acknowledging the basis for the perception and expectations of chatbots is further needed (Van Mulken et al., 1999). However, there is a numerous amount of variables that shape a person's expectations for any given chatbot. Accounting for how demographics, previous experiences and personality affects expectations lies outside the scope of this study, even though such influences are acknowledged (Hubal et al., 2008). Rather, the focus is placed on assuming that considering expectations play an important role to legitimately evaluate trust in human-chatbot-interaction. The topic of recognising and managing expectations raise in importance for practitioners who are looking to design and evaluate chatbots for a diverse user base, succeeding in the general audience.

3.5.2 Contextual Level of Risk

Although there are several definitions of trust, there seems to be a general agreement that trust is particularly relevant in situations that are characterised by risk, where the user depends on the actions of the AS (Lewis et al, 2018). For example, there is a significant difference between the threshold for establishing trust in a chatbot that delivers opening hours and one that prescribes medicine. Depending on the consequences of a certain context, the required level of trustworthiness in the agent is increased. Previous findings suggest that the impact of decisions made by influences from the agent increases the difficulty in establishing trust (Devitt, 2018). Life-threatening contexts, such as the choice of medical care have a higher threshold for establishing trust than an agent for grocery shopping (See figure 2). Although contextual differences are delimited in this investigation, the provided case context is considered to affect the threshold of establishing trust. Since AS is applicable in a huge variety of context, such considerations are a vital part of understanding user trust (Devitt, 2018). Furthermore, certain contexts and conversational topics may affect the willingness of sharing sensitive information, and contribute to privacy concerns. Moreover, biases might be affected by the nature of the context as certain conversations may include a stronger fear of being judged (Følstad et al., 2018).

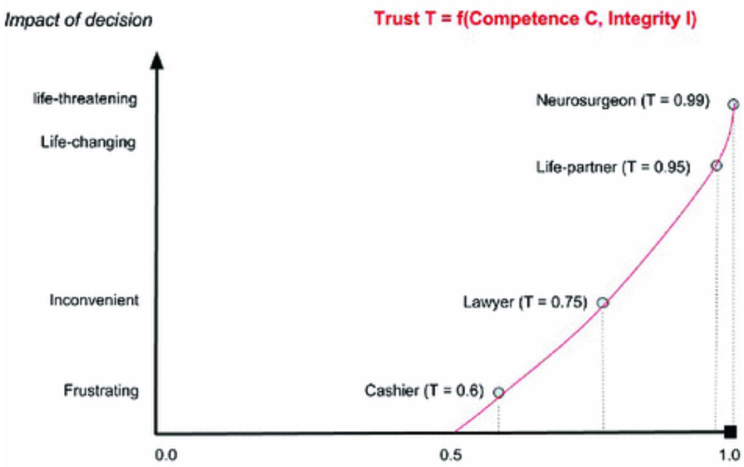


Figure 2. Threshold for establishing trust in certain contexts (Devitt, 2018).

3.6 Evaluating & Quantifying Trust

Drawing upon the acknowledgement that socially-oriented chatbots are prone to be designed and evaluated by anthropomorphic features, TIA can be understood through its analogy to interpersonal trust. Consequently, most refined scales of TIA are based on correlations to the dimensions of interpersonal relations, such as ability, integrity and benevolence (Lewis et al., 2018). Generally speaking, there are three well-known measurements of TIA, SHAPE Automation Trust Index (Jian et al., 2000), Human-Computer Trust (HTC) (Madsen & Gregor, 2000) and Empirically Derived (ED) (Goillau et al., 2003). These metrics have gone through systematic development and validation, however their viability of proper application differ in different contexts. For example, ED measures TIA without a clear reference to a specific case, but rather in an abstract environment, resulting in a metric for the propensity of trust and not trust in a particular system. There was a recent effort (Chien et al., 2014; Chien et al., 2015) to develop a general measure of TIA which provides a scale for measuring the effects of manipulations of components expected to affect trust. However, due to the deceptive nature of defining and quantifying trust in HMI, measurements and metrics still remain elusive. Previous literature tends to have fairly implicitly designed frameworks applied to highly contextualised conceptualisations of trust (Abbass et al., 2018). Instead of a universal metric, these studies rely on situation-based worded questionnaires and interviews (Lewis et al., 2018). However, due to the numerous amount of reported components (biases, reliability, compliance etc.) that have been supported to play a role in TIA, a basis for criticism towards the topic has been provided (Dekker & Hollnagel, 2004). Striving to establish conclusions within the area has been commented to use fabricated assessment frameworks and that the modelling of this diversity is unfalsifiable and lacks empirical grounding (Dekker & Woods, 2002). In order to address this criticism, empirical research within the area needs to be conducted by isolating and diverging well-motivated system components, assumed to effect trust, within a single task, performed by a homogenous test group (Parasuraman et al., 2008).

3.7 Analysis Framework

This theoretical section has referenced and highlighted findings made by previous scholars in the field of research. Emphasis is put on those results that serve as the basis for later analysis and discussion. Following a short summary is a visual representation of the final framework.

3.7.1 Theoretical Summary

Initially, it is acknowledged that trust is seemingly ambiguous and difficult to evaluate for the general audience (Hawley, 2012; Nave et al., 2008). With reference to previous findings it is found that a goal-centred viewpoint allows TIA to be contextualised and dissected into smaller contributing cognitive concepts, practical system components and design choices (Lewis et al., 2018). Provided the case context studies with relevance to service encounters are accounted for (Bitner et al., 2014; Gupta et al., 2006). Furthermore, a modern viewpoint is provided (Larivière et al., 2017) implementation of AI in socially oriented service encounters is accounted for (Fitzpatrick et al., 2017). More specifically references to definitions of social chatbots and how their goal and different design choices relate to trust are presented (Lewis et al., 2018). After accounting for the context, a general summary of plausible system components expected to affect trust, targeted towards the cognitive concepts of TIA, are motivated and accounted for in accordance with previous scholars (Følstad et al., 2018). Furthermore, is is supported that these components can be evaluated by alterations to their constituting design choices to make an analysis of the cognitive concept of TIA (Følstad et al., 2018; Hieronymi, 2008; Keren, 2014; Simpson, 2013). More specifically, that integrity, benevolence and ability constitutes a legitimate basis for grounding and evaluating trust, although it is acknowledged that further distinguishment can be made (Muir, 1994; Barber, 1983). Furthermore, it is supported that this division allows for a structure to framework viable investigations (Connelly et al., 2015; Connelly et al., 2012; Kim et al., 2004; Luo, 2002). A more in-depth accounting for the theoretical meaning of integrity, benevolence and ability is then provided together with their practical linkage to transparency (Luger & Sellen, 2016), unbiasses (e.g Caliskan et al., 2017) and system performance (e.g Lewis et al., 2018) respectively. This is done with reference to previous findings that have found that design choices of respective system component have shown to affect trust. Finally, a brief acknowledgement of circumstantial considerations that can affect trust, such as individual expectations (e.g Louwerse et al., 2015; Nass & Brave, 2007) and contextual level of risk is presented (Lewis et al., 2018; Følstad et al., 2018). Lastly references to previous methods and metrics for measuring TIA to provide a theoretical basis for choices made in the methodology section are highlighted (Jian et al., 2000; Madsen & Gregor, 2000; Goillau et al., 2003).

3.7.2 Visual Representation of Theoretical Framework

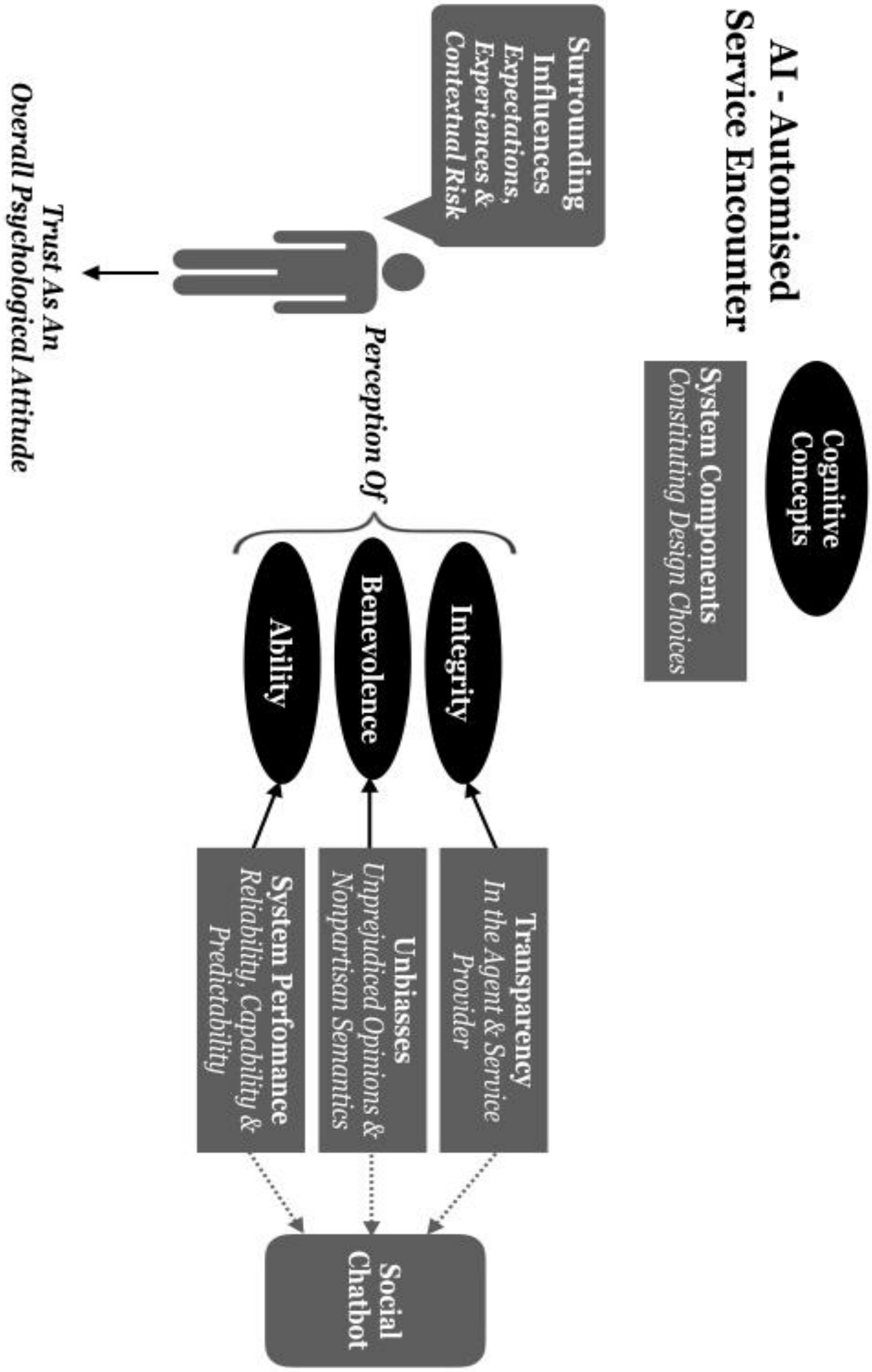

AI - Automised Service Encounter
Cognitive Concepts
System Components
Constituting Design Choices
Surrounding Influences
Expectations, Experiences & Contextual Risk
Trust As An Overall Psychological Attitude
Perception Of
Integrity
Benevolence
Ability
Transparency
In the Agent & Service Provider
Unbiasses
Unprejudiced Opinions & Nonpartisan Semantics
System Perfomance
Reliability, Capability & Predictability
Social Chatbot

Figure 3. Visual representation of the theoretical framework. In black are cognitive concepts of trust. In grey are contextualised system components suggested to target the cognitive trust concepts. In italic are the design choices deliberately altered to examine effects on trust (Eklund & Isaksson, 2019).

4. Methodology

In the following section, the methodology choices are explained and justified. The selection of research model, workflow and all other necessary decisions made in the study are explained and motivated, followed by a description of the course of action. This study is a quantitative study with qualitative elements and all decisions are made with regard to the purpose of the study, given limitations and assumptions.

4.1 Research Design

This project has been carried out in continuous compliance with our partner's interests. Practically this meant that we were stationed at AlphaCE's office in Uppsala where we had our own working space, ensuring smooth communication, supervising and arrangements of meetings with the company. To ensure a transparent process for all involved parties, a schema for consecutive and frequent meetings with both the company supervisor and the university reviewer was set at the beginning of the project.

By acknowledging the advantages stated by Race (2008), literature review has in this study been used as a mechanism to gain an increasing understatement about the investigated subject together with expanding the knowledge on how to test and evaluate the formed research hypothesis. The majority of the literature used within this study consists of research articles gathered from various academic databases such as Uppsala University library and Google Scholar. Literature from other more informal sources has also been used, primarily regarding the technical subjects and solutions. Because chatbots and the frameworks that constitutes their existence is relatively new technology and in constant change, sufficient relevant information can be found on different types of blogs and forums. Information that cannot be classified as academic but still holds great value to this study due to its ability to provide the latest within the field. According to Race (2008) and Adams et al. (2001), it is up to the authors of a study to themselves consider the trustworthiness and usability of the given source of information relative to the study (Race, 2008; Adams et al., 2001). In addition, reports from the Swedish Labour Agency, Arbetsförmedlingen, has been reviewed in order to gain an overall understanding of the current Swedish labour situation. This information has also been obtained through informal conversations with the people working at AlphaCE as well as with our supervisor Maria.

The purpose of this study is to evaluate different system components and how they contribute to the perceived trust in the studied chatbot, predominantly as quantitative investigation. Initial focus was on distinguishing relevant and significant components, done mainly from beta-testing and market research stated above. Moreover, the aim was to showcase the effect of alterations to trust engendering system components through large scale comparative tests resulting in generalisable data. Tests of a qualitative nature focus on objective measurements such as statistical or numerical data gathered through surveys in order to find correlations between different variables or tests and are highly dependant on the researcher's point of view (Muijs, 2010). In contrast to the quantitative research method, there is qualitative methods which focuses on the bigger picture, contextualising a phenomenon from a user perspective with an investigative approach (Maxwell, 2012; Saunders et al., 2016).

There are scholars suggesting that some research areas are too complicated to investigate using only one type of approach method, where the line between qualitative and quantitative research models are fuzzy and vague. In those cases a mixed method approach is suggested (Tashakkori & Teddlie, 2010; Denzin, 2010). It is argued that by using mixed methods, the researchers can accomplish various purposes by being able to triangulate a phenomenon and explain it from different views, both with data and more in-depth interviews. Haq (2014) states "*knowledge about a social reality can be better accessed and understood from different angles and by adopting both qualitative and quantitative data collection and analysis tools and techniques in the same social enquiry*" (Haq, 2014, s.14). There are several ways and approaches towards mixed methods where the procedures vary in terms of the balance between methods and how to analyse the collected data. The mixed method which was the most applicable to our research is a concurrent mixed method. This method allows the researchers to collect and analyse the data in a parallel fashion. By using a mixed method with a concurrent approach, we were able to extend our statistical data comprehension using more in-depth qualitative data. This allowed for a more efficient addressing of the scope and objective of this study, resulting in something that might have been unobtainable if only one research method was used (Haq, 2014).

The study was carried out in a user-oriented way, applying concepts from the Holtzblatt and Beyer's Contextual Design method (Holtzblatt & Beyer, 2015). This user-centered design process applies to conducted field research, such as interviews and participation, to understand the most pressing needs from stakeholders, leading to an innovative and useful design. Practically, this implied that beyond performing individual and informal interviews with employees, we participated in informal conversations, interviewed experts within the field and attended relevant meetings where nearby topics to the project was discussed, gaining important and necessary company and market insights. As a consequence of basing the project at an AlphaCE office, where daily operations are carried out, several important insights concerning the organisation and the service was obtained, insights that could be implemented into the product development. This local presence is a variable Holtzblatt and Beyer (2015) emphasise to be of great importance when developing a commercial product as it ensures gaining insights about how the users might behave and think as well as potential user needs and core values that should be involved in the product (Holtzblatt & Beyer, 2015).

4.2 Research Course of Action

Early in the project process, considerable time was spent on doing relevant literature research, taking already available chatbot resources and previous studies in the area of trustworthiness into consideration to narrow down our search. Further, time was spent on defining trust in the given context and constructing the evaluative framework for addressing the purpose of the thesis. As the focus was placed on testing and evaluating AI-automated guidance with an emphasis on trust, the aim was to early-on initiate the construction of a prototype chatbot service. Following the literature review, the study continued into researching existing chatbot frameworks and programming languages for developing the initial prototype. Considering the requirement specification, a decision was made to use the already existing chatbot framework Dialogflow for development of the conversation flow, which is explained more in *Section 4.4.1* and in *Appendix A* (Dialogflow, 2019a). Together with Dialoflow working as the frontend, the Google Firebase platform was used as a database and for launching the application. Firebase is further explained in *Section 4.4.2* and in *Appendix A* (Google Firebase, 2019a,b).

4.2.1 Informal Interviews

As a part of using Holtzblatt and Beyer's Contextual Design method, several informal interviews were conducted (Holtzblatt & Beyer, 2015). Interviews in the form of conversations where an exchange of knowledge and information was carried out with employees at AlphaCE. Aside from nearly daily contact and guidance by Maria Mattson Mähl, the project supervisor, the authors had frequent contact with CEO Erik Gustafsson and the company's IT-department. Altogether, this provided great insight on AlphaCE activity and the challenges ahead as well as the position of the Swedish labour market. Furthermore, on a bigger level, what challenges the society as a whole are facing when it comes to education and employment.

4.2.2 Field Studies

In order to gain a better understanding of the Swedish labour market and what challenges it faces in terms of technology and digitalisation, a meeting with a manager and programmer at the IT-department of Arbetsförmedlingen was set. The meeting took place on the 25th of February at the head office of Arbetsförmedlingen IT, and the agenda was to gain insights into the work Arbetsförmedlingen conduct in terms of automation. The meeting provided valuable insights in the field of available data and statistics provided by the agency. Besides from labour market knowledge and insights, the meeting resulted in attaining data in the form of large datasets of different occupations which were used in the schematic mapping in the chatbot.

Between the 4th and 6th of April, the authors participated in the hackathon Hack for Sweden. Hack for Sweden is a government issued idea competition whose primary goal is to increase the awareness and broaden the use of open data in order to benefit society as a whole. The main reason behind participating in such a competition was to be able to solely focus on product development for forty hours. Developing the chatbot so it would work as intended for the large scale tests. The secondary reason behind attending the event was to be able to get in touch with more experts within the field but also people with more in-depth technical expertise. The, within this study created, cahtbot won the Labor Market category, a win that resulted in a lot of new contacts. Contacts that helped with valuable content to the product.

4.2.3 Beta Test

To early on assess and validate the chatbot in terms of functionality and alongside identify and verify system components affecting trust, a beta test was conducted. Beta tests are the last step of testing before commercially distributing a product, or in our case examining the chatbot for other aspects than functionality entities such as trust (Wurangian, 1993). Previous to the distribution of the beta tests, continuous in-house alpha tests were conducted. Alpha tests are conducted and executed within the organisation and are developed to mimic the usage from potential users. When in-house alpha tests no longer can provide any substantial addition of information or ensure a bug-free product, it is according to Dolan and Matthews (1993) recommended to do a more comprehensive beta test. The same scholars also emphasise that beta testing a product serves a great value when the product has a target audience which is very heterogeneous, making it hard to test with only user cases, which is very applicable in our case thus the target audience for a product like the guidance chatbot can be of any age and position in life. Doland and Matthews also state that beta testing is a viable method when every part of the product is not fully understood, which is applicable in the case of identifying relevant trust affecting system components (Doland & Matthews, 1993). The beta test was conducted in accordance with the methodology described by Doland and Matthews (1993) which includes testing the product on a wide demographic audience, using both open and closed questions in the evaluation. The beta test was performed on a total of 13 personally selected participants in different ages and life situations in order to maximise the different use cases. After the test participants had interacted with the chatbot, they were asked to fill out an evaluation form. The form was designed to provide an answer to the fundamental reasons behind the test, enlighten any technical issues and to help to identify components affecting trust. The evaluation form can be found in *Appendix B.1*.

4.2.4 Quantitative Testing

One of the central aspects of this study was to conduct a large-scale quantifiable comparative survey for evaluating identified and selected system components. There are a lot of different methods for data collection within the framework of quantitative testing. Bryman & Bell (2015) emphasises that the selected data collection method should ensure authentic, impartial and relevant data to be gathered, preferably from a primary source. The selected data collection method, evaluation form, enables the authors to manage the data in a way that generates authentic, impartial and relevant data coming from a primary source. A downside with using an electronic-based questionnaire is the uncertainty and under some circumstances difficulty of getting enough answers. On the other hand, online questionnaires serve as a highly cost-effective method to gather primary data in large quantities, if managed and distributed properly (Jones et al., 2008).

After completing the data collection, it was transferred into Microsoft Excel for further analysis. In Microsoft excel primarily four things were prepared; graphs to distinguish relationships between parameters, calculation of Pearson correlation coefficients, F-tests for analysing variance between different versions and t-tests for analysing the mean value between different versions.

Pearson product-moment correlation (r) is calculated on the covariance for the sample population divided by the product of their standard deviations. The interval of the coefficient r is $-1 \leq r \leq 1$ where a value of $r = -1$ indicates a total negative correlation, a value of $r = 0$ indicates no correlation and a value of $r = 1$ indicates a perfect positive correlation between the variables (See table 1) (Asuero et al., 2006). The Pearson product-moment correlation was used to identify similarities within the different versions in the quantitative tests as well as in the beta tests.

Table 1. The table presents interpretation intervals of the correlation coefficient (r) (Ausero et al., 2006).

Size of r	Interpretation
0.90 to 1.00	Very high correlation
0.70 to 1.89	High correlation
0.50 to 0.69	Moderate correlation
0.30 to 0.49	Low correlation
0.00 to 0.29	Little if any correlation

The F-test is a statistical analysis tool using hypothesis to determine if the variances values of two or more groups are different, calculating a ratio between two variances and how far they are distributed from the mean. The model is based upon having a null hypothesis that the variances are the same, $H_0: \sigma_a^2 = \sigma_b^2$, and the alternative hypothesis is that they are different, $H_1: \sigma_a^2 \neq \sigma_b^2$. The goal is to reject or strengthen the null hypothesis, thus implying the different variances are alike or different. In Excel, the F and the $F - critical$ values are automatically calculated. If the calculated variance ratio F is larger than $F - critical$ and the calculated p-value is smaller than 0,05, the null hypothesis is rejected, thus meaning the variances are different (Hosken et al., 2018). In our case, the F-test is used to check whether the variances are the same down to a significance level of 5%, thus verifying that the users haven't answered the different models randomly.

The t-test is used to compare two conditions in order to distinguish if they are significantly different from each other, and to what extent they are. The test compares the data from their means and assumes that the data follows a normal distribution and that the variances are the same, thus the importance of the previous F-test. In the same way as the F-test, the t-test uses a hypothesis where the null hypothesis is that the mean values of the different parameters are the same, $H_0: \mu_a = \mu_b$, and the alternative hypothesis is that they are different, $H_1: \mu_a \neq \mu_b$(Kim, 2015). In Excel, the p-value is automatically calculated, stating the significance on which the null hypothesis can be rejected. If the p-value is lower than 0,05, the null hypothesis can be rejected on a significance level of 5%.

Both hypothesis tests, the F-test and t-test, were used upon the data derived from the large scale quantitative testing. The main goal for conducting both tests was, to some extent, showcase that the results obtained from the different versions differed due to the design changes and not due to coincidence.

4.2.5 Qualitative Testing

As a complementary data collection method and to extend our statistical data comprehension, in-depth qualitative data collection has been performed to more effectively tackle the scope and hypothesis of this study (See table 2). In accordance with Haq (2014) and the chosen research design complementing interview have been held with professional coaches in order to gain another dimension and view of the tests. The chosen approach for the qualitative data gathering was an overwatched think aloud session where the test participants tested the three different versions of the chatbot and discussed everything in every moment. The think aloud method is a well-renowned method mainly used to evaluate usability. It enables in-depth analysis of a product's performance by enlightening everything immediately when occurring instead of afterwards, minimising the loss of information risk which is connected to the person's ability to remember specific details (Maaike & Menno, 2003). According to Guan et al. (2006), the method has a high validity thus the users rarely provide forged or untruthful answers when performing the test (Guam et al., 2006). The think aloud interview was practically executed by letting the coaching experts firstly try the neutral version, then both the biased and opaque version after. The interviews were recorded to facilitate a more accurate result presentation process. Usually, the tests are followed by some sort of overall evaluation where the user gets to rate the usability of the system as a whole (Nielsen & Pernice, 2009). In our case, where this is a complementary data collection and the test participants were not part of the target audience, a decision was made to only use the main part of the think aloud method.

Table 2. The table includes information about the conducted interviews.

Name	Working Title	Date	Duration
Therese Broström	Study and vocational guidance counsellor	2018-05-14	45 min
Mattis Lu	Job Coach	2018-05-15	35 min

4.3 Operationalisation of the Theoretical Framework

To bring legitimacy to the investigation, we chose to stay compliant with suggested research methods from previous researchers (e.g. Lewis et al., 2018). Furthermore, to tackle the ambiguous nature of evaluating TIA, an emphasis was put on proper contextualisation and identifying measurably viable components assumed to affect trust (e.g. Følstad et al., 2018; Hieronymi, 2008; Keren, 2014; Simpson, 2013). The later was done through an iterative phase of literature review, prototyping and beta-testing as accounted for in the previous sections. In terms of contextualisation, weight was put on understanding key concepts of study- and vocational guidance. This was done through informal interviews with coaches and experts and practical market research, accounted for above. Moving forward focus was shifted towards operationalising the theoretical framework, consisting of integrity, biases and ability. This was done by translating how cognitive concepts of trust could be targeted by practical system components suggested to affect them. This was done with careful and suitable alterations to their constituting design choices. As hinted in the theoretical framework we through beta tests and careful consideration of research viability decided to create three system versions. One neutral version which we designed to act as a reference version towards the examined trust affecting system components. One version with an emphasis on affecting perception of integrity through deliberately opacid design choices but equally unbiased as the neutral version. One version with an emphasis on affecting perception of benevolence through deliberately biased design choices but equally transparent as the neutral version. With regards to the ability of the chatbot, we came to the realisation that due to limitations in technical abilities and time restrictions, alterations with regards to system reliability and capability would lie beyond the scope of this study. However, as previous findings strongly weighted the effects on trust due to faults and erroneous behaviour we acknowledge its impact. Furthermore, system predictability was to some extent modified in the opaque system version, testing connections between perceived ability and transparency. Below (See figure 4) is a visual representation of the operationalisation of the theoretical framework, which provides an overview of the modified design choices of the evaluated system components, in their respective system versions. As follows is a more in-depth explanation of the operationalisation of each investigated system component expected to affect trust and its related design choices. A technical explanation of the construction of the chatbot and conversation flow is accounted for in the following chapter.

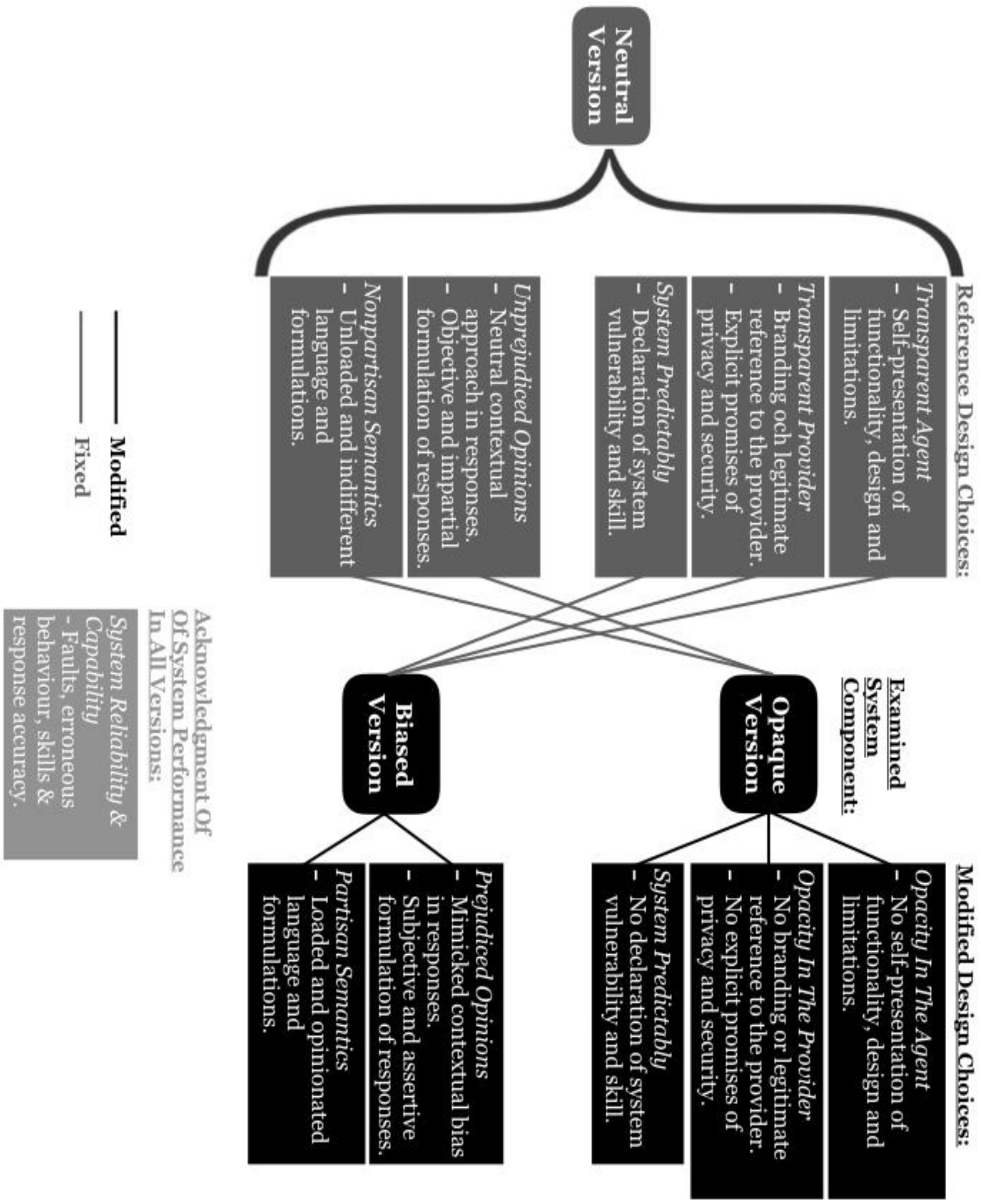

Neutral Version
Reference Design Choices:
Transparent Agent
- Self-presentation of functionality, design and limitations.
Transparent Provider
- Branding och legitimate reference to the provider.
- Explicit promises of privacy and security.
System Predictably
- Declaration of system vulnerability and skill.
Unprejudiced Opinions
- Neutral contextual approach in responses.
- Objective and impartial formulation of responses.
Nonpartisan Semantics
- Unloaded and indifferent language and formulations.
Examined System Component:
Opaque Version
Biased Version
Modified Design Choices:
Opacity In The Agent
- No self-presentation of functionality, design and limitations.
Opacity In The Provider
- No branding or legitimate reference to the provider.
- No explicit promises of privacy and security.
System Predictably
- No declaration of system vulnerability and skill.
Prejudiced Opinions
- Mimicked contextual bias in responses.
- Subjective and assertive formulation of responses.
Partisan Semantics
- Loaded and opinionated language and formulations.
Acknowledgment Of Of System Performance In All Versions:
System Reliability & Capability
- Faults, erroneous behaviour, skills & response accuracy.
Modified
Fixed

Figure 4. A visual representation of the modified design choices related to the respective system component expected to affect trust in the different system versions (Eklund & Isaksson, 2019).

4.3.1 Designing Opacity

As accounted for in the theory we contextualise integrity in chatbots as the users perceived honesty and character of the conversational agent (Følstad et al., 2018). More specifically, we link integrity to transparency on the basis that previous research shows that users tend to trust an AS that is explicit and honest about its nature, functionality, limitations and behaviours (Devitt, 2018). In practice, we dissected transparency in our constructed system with deliberate opacid design choices partly with regards to the agent, e.g. referenced content and partly with regards to the system provider, e.g. branding.

In terms of deliberately decreasing the transparency of our conversational agent, we made a few alterations to the version described as neutral. The major difference is the implementation of proper self-presentation. In the neutral system version, we in accordance with (Mone, 2016; Luger & Sellen, 2016; Kretzschmar et al., 2019) designed Ava to at the beginning of interaction to be upfront about its machine status and about its limitations. Furthermore, we prompted a declaration of the system nature, functionality, design and ability. In contrast, in the opacid version, designed to test the impact of such features on user trust, both self-presentation and system declaration were removed. Instead, a conversation was initiated directly without explanation of the agent's character. Beyond removing proper self-presentation and honesty we throughout the conversation flow made a few adjustments. With reference to previous findings (Brandtzaeg et al., 2019) that using transparent and suitable formulations and language for certain contexts has an impact on user trust, we tweaked certain responses from the neutral version. We altered the dialogue character to be more direct and implicit with regards to motivations behind statements, recommendations and conclusions. For example, instead of saying "*since you would prioritise to go to a party on a free weekend, it seems as if you are a social and outgoing person*", the response would be "*you are a social person*". We strived for a sensation that the system is complex and intricate, excluding logical explanation to why the conversation is behaving the way it is (Castelvecchi, 2016). This was all in favour of testing the vulnerabilities of the user perceiving the system as a "black-box". In relation to decreasing motivations behind statements and assumptions, we hid explanations and references to previous topics when prompting new ones. For example, not justifying and mapping certain topics in relation to the goal of the interaction (Kretzschmar et al., 2019). Lowering the visibility of why the system is doing as it is, limiting the ways for a user to understand the influence and justifications of machine reasoning (Hepenstal et al, 2019).

Secondly, a few adjustments with regards to decreasing the perception of transparency in the service provider was done. For example, an effort for distinguishing between having a reference to a legitimate brand was made. In the opacid version, the access point to the chat was completely unbranded and with no association to a brand or provider. Where as the neutral version was branded with logos and descriptions mimicking a company (Følstad et al., 2018). In association with removing the reference to an actual company providing the chatbot, ensuring statements about the collection and usage of personal data were removed. Removing any guarantee that the chatbot would provide a sufficient level of privacy and security in the certain context (Følstad et al., 2018). The idea was to remove the sensation of having a proper provider to turn to if something went wrong or if the chatbot provided harmful information. Beyond decreasing the brand transparency, explanations and references to responses and claims where removed or altered. In the neutral version, Ava provided a declaration to what extent provided information was backed up by research and evidence (Kretzschmar et al., 2019) and references to factual suggestions. However, in the opacid version, such explanations where removed.

4.3.2 Mimicking Contextual Bias

With reference to previous findings in the theoretical section, we link the benevolence of a conversational agent to its perceived prejudices, motives and personal beliefs (Følstad et al., 2018; Robinette et al., 2015). In practice, we sought to affect the cognitive trust concept of benevolence by mimicking and implementing contextual bias. We basis in that there is a rising concern that the usage of historical data could result in, with or without intention, the prejudices, failings and unfairness that characterises many of society's institutions (Caliskan et al., 2017). More specifically, we considered prejudice and societal expectations to design plausible bias in the conversation flow (Byrnes & Kiger, 1992). By narrowing down the focus group, more on this in following sections, towards to study- and vocational guidance for students attending the last year or graduated high school within 5 years, a more precise contextual bias could be designed (Mittelstadt & Floridi, 2016). For example, the biased version of Ava promoted higher education and put pressure on that professional ambition should be prioritised; "*some want to make money to finance travel and adventure, others want to focus on more important things, like education and experience*". Furthermore, biased Ava provided examples of labour that society historically has perceived as better or of a higher class. Considering the case context and that one of the key elements of engendering trust in social chatbots is their very unbiased nature, the ambition was to examine deliberate biased responses impact on user trust and experience (Fuchs, 2018).

Beyond subjectively designing Ava to have opinions and values based on prejudice in the case context, we made moderate alterations to used language and semantics (Tapus et al., 2007). Although using equally emphatic language as in the other version, biased Ava conversed subjectively. For example, when a user was given a multiple choice question and chooses an alternative, Ava would answer "*Good choice*" instead of "*okay*" or another unloaded synonym as in the neutral system. In this sense, biased Ava used a more anthropomorphic language than other versions, something found to engender trust (Følstad & Brandtzaeg, 2017b). In contrast to considering that human-like semantics might to some extent increase the trust in Ava, we deliberately designed for the risk that partisan semantics might affect the users perceived benevolence of the agent. The language contained loaded and opinionated terms in the case context to investigate the occurrence of uneasy feelings in the user (Følstad & Brandtzaeg, 2017b; Singh, 1999). Due to aiming to resemble a judgmental human advisor, like a "pushy parent", biased Ava also claimed to know what was right and wrong. For example, by using bold phrasing (Caliskan et al., 2017) such as "*I will now interpret and tell you what you enjoy doing*" in contrast to the neutral version; "*Using your input I will now provide suggestions on things you say yourself you would enjoy*".

4.3.3 Developing System Performance

As previously mentioned, system performance as system component for engendering trust from perception of ability was not isolated and altered in a dedicated and modified system version as for integrity and benevolence. However, both by reviewing the previous literature (e.g. (Salem et al., 2015) and by analysis of beta-test results we concluded that the effects on trust due to faults, erroneous behaviour, skills and system predictability needed to be accounted for. Primarily, the focus was to minimise the risk of system failure and major erroneous behaviour as large faults have a significant impact (Moray et al., 2000). This was done by examining certain topics and questions prone to a lot of failures in the beta-test. For example, some questions had to be narrowed down to multiple choice answers. Furthermore, a more considerate design of the conversation flow with a neat "try & catch" structure for "risky" questions was implemented. Taking precaution for that a potential trust relationship would be broken by a single instance of error (Lewis et al., 2018). A central aspect to not changing the conversation flow in the different versions was staying consistent in the "risks" for mistakes or failures. Minimising any significant differences in perceived system ability in the three versions. Beyond striving for an as technically stable conversation flow as possible we accepted that mistakes would most likely occur and that this would affect the subjective assessment of the agent in terms of the trust (Salem et al., 2015). This was done by implementing questions about the perceived technical performance, which could then be accounted for in relation to other evaluated areas.

Another acknowledgement we made through beta-testing and literature review (Lewis et al., 2018) was that the level of perceived skill in the agent would affect the conversation experience and therefore possibly also trust. However, the level of system ability and skill-set in the final versions was rather a result of what was technically viable for our level of development competence and time limitations, then a deliberately chosen skill level. On the previously described quadratic relation between agent competence and trust (See figure 1) it is plausible to assume that Ava was somewhere fairly early on the upgoing curve towards humanly comparable intelligence. Rather prone to suffer from negative effects on trust due to low levels of competence than the risks of exceeding human levels beyond user understanding (Lewis et al., 2018). Furthermore, especially in the biased version of Ava, with a higher level of anthropomorphism, we acknowledged the risk for disappointment in the agent's capability of answering questions thought to be viable to ask (Følstad et al., 2018). Suggesting that Ava could give the impression of being able to handle more complex topics but then not delivering to those expectations. As to such, it was acknowledged that the increase of anthropomorphic characteristics could have a strong impact on trust engendered from technical capabilities (Hancock et al., 2011). This meant evaluating Ava's technical performance, knowing that trust could be affected by perceived human-like properties (Bainbridge, 2008). Managing trust affected by ability, in similar to the above paragraph, included partly working proactively for good system predictability (Lewandowsky, 2000) and partly acknowledging and accounting for its effect in evaluation. System predictability was to some extent altered and lowered in the opacid version of Ava, by removing system description and declaration in the self-presentation. In the other two versions, users were informed about the system nature and what kind of input to avoid in order for the agent to function properly. For example, participants were instructed to not use abbreviations, intricate wording, slang and to overall keep responses simple. The idea was to examine whether information about system vulnerability would result in a smaller impact on the relationship between trust and perceived technical performance (Riley, 1994). However, in all three versions, participants where provided fallback messages with suggestions on how to answer for a valid response in the case of a misunderstanding. Addressing the risk of users feeling unknowing about why failure is occurring and its effect on trust (Lewandowsky, 2000).

4.3.4 Accounting for Surrounding Influences

As stated previously, this study delimited the systematic consideration of trust affected by participants individual expectation and prerequisites. Furthermore, the contextual risk associated with a certain conversation topic was not accounted for in the quantitative part of this study. However, such influences on trust were acknowledged on a qualitative level, in accordance with previous scholars (Hubal et al., 2008; Louwerse et al., 2015; Nass & Brave, 2007; Van Mulken et al., 1999).

Treating trust as a cognitive attitude achieved from personal perception consequently implied dealing with individual interpretations and presumptions (Louwerse et al., 2015; Nass & Brave, 2007). Managing user expectations mainly concerned limiting the focus group towards a homogenous pool of participants where prerequisites and expectations could be considered to be fairly similar (Van Mulken et al., 1999). The behaviour, opinions and characteristics of a person are, according to several studies, highly connected to the surrounding environment which teaches individuals to act and behave in accordance with community norms and standards. Furthermore, by sectioning a population by demographic characteristics such as e.g age, gender, income, political views, a group of people that have more similar aspects of life and are more alike can be distinguished (Lavrakas, 2008, pp.185-186). As potential end-users of a virtual vocational guidance chatbot were considered to be people ranging from the unemployed job seeker to anyone feeling uncertain in life, these findings were used to narrow down a suitable test group. Contextual considerations resulted in the study focusing on a demographic sub-group of users in the same age and education level. More specifically, the quantitative tests were performed on Swedish speaking students attending the last year or graduated high school within the past 5 years, making the age span of 17-25. Beyond homogenizing prerequisites and expectations, a particular focus group allowed for an optimisation of the conversation flow and its evaluation.

Learning from that there seems to be a relationship between the risk associated with a certain context and the threshold for establishing TIA, the particular topics discussed in the conversation with Ava were considered (Lewis et al, 2018). More specifically, as Ava to some extent aspired to have a conversation resembling one had with a human study- and vocational advisor, a similar contextual risk was assumed. On the scale demonstrated in figure 2, a conversation with Ava could be applied somewhere at a "therapeutic" level, not life-changing but definitely a situation where misguidance could have a significant personal impact (Devitt, 2018). Furthermore, Ava at some points asked relatively personal questions. For example, "*is there anything in your life that concerns you especially much right now?*". In the evaluation, participants were then asked both whether they answered truthfully to all questions and about their comfortability answering truthfully. Examining whether certain contexts and conversational topics affected the willingness of sharing sensitive information (Følstad et al., 2018). That is, beyond the effects of transparency, bias and system reliability.

4.3.5 Choosing Method & Measurements

At the initial stages of this investigation, the focus was placed on finding viable and legitimate methods to frame, isolate and measure components that affect trust in a socially oriented interaction with a chatbot. This process consisted of a literature review, informal interviews and beta-tests to ensure a suitable design for a quantitative study in the particular context. It was there discovered that most refined scales for TIA are based on correlations to the dimensions of interpersonal relations, such as ability, integrity and benevolence (Lewis et al., 2018). However, the existing standardised measurements, such as SATI (Jian et al., 2000), HTC (Madsen & Gregor, 2000), ED (Goillau et al., 2003), did not suit the current context. Instead, inspiration was taken from the recent efforts (Chien et al., 2014; Chien et al., 2015) to develop a general scale for measuring the effects of manipulations of system components expected to affect trust. In practice, we treated trust as an attitude achieved through the perception of a set of cognitive concepts (Lewis et al., 2018). More specifically, we made our analysis of trust in socially-oriented interaction by studying Ava's achieved influence on human attitude, without an explicit cognitive metric. Instead, we relied on situation-based worded questionnaires and interviews to identify and evaluate the distinguished system components. Consequently, we acknowledged criticism given to previous studies, due to their complex modelling and unfalsifiable conclusions (Dekker & Woods, 2002). To address this critique, we provide self-criticism to the choice of method in the section "Research Credibility & Reliability" and through an objective discussion of the applicability of our findings in the general scene. Furthermore, we acknowledge limitations in being able to legitimately backtrack effects on trust to particular design choices and their linkage to specific cognitive aspects (See chapter “Discussion & Analysis” for more on this). Moreover, we applied the method of isolating and altering well-motivated system components expected to affect trust within a single task, performed by a homogenous test group to the best of our abilities (Parasuraman et al., 2008).

Following the selected mixed concurrent research method that was used within this study, the questionnaire was developed both to provide quantifiable results and concurrently to hold a qualitative approach. A brief summary of the different questions asked can be found below (See table 3) or at its verbatim format in *Appendix B.* The assessment was created and distributed through Google forms, thus enabling easy administration and converting the given answers to excel files for further analysis. To maximize the cost-effectiveness of the tests in accordance with Jones et al. (2008), the surveys were created and distributed electronically with the tool Google forms. Links to the Facebook pages together with their accompanying evaluation forms were put into three different emails together with instructions, which can be found in *Appendix B*, and evenly distributed with the help of school staff to students at Lundellska Skolan and Rosendalsgymnasiet in Uppsala. To gain more answers, visits were also made to the schools to present and test the solution live in the classroom. At these occurrences, the three different versions along with the instructions were either evenly distributed by email or on the accessible student intranet.

Table 3. The table illustrates the different questions together with the area of evaluation. The complete and exact evaluation form can be found in Appendix B.

Question	Area of Evaluation	Operationalisation	Type
1-2	*Evaluating & Quantifying Trust* - Test group	Age and Gender	Select one
3-4	*Surrounding Influences* - Experience & Expectations	Previous experience of chatbots Previous knowledge of Artificial Intelligence	Rate 1-10
5-6	*System Performance* - Reliability & Predictability	Technical performance experience Meeting technical expectations	Rate 1-10
7	*Contextual Risk & Effects of opacity, bias and performance* - Trust	Comfortability to answering truthfully	Rate 1-10
8	*Contextual Risk & Effects of opacity, bias and performance* - Trust	Answering all questions truthfully	Yes / No
9	*Effects of opacity, bias and performance* - Trust	Comfortability to answering truthfully in comparison to a conversation with a human advisor	Less - Equal - More
10	*Effects of opacity, bias and performance* - Trust	Trustworthiness of the content in the responses	Rate 1-10
11	*Effects of opacity, bias and performance*	Recurrence of reasons for content not being trustworthy	Free Text

	- Trust		
12	*Effects of opacity, bias and performance* - Trust	Level of consideration of provided content in responses	Rate 1-10
13	*Effects of opacity, bias and performance* - Trust	Recurrence of reasons for content not being considered	Alternatives
14	*Effects of opacity, bias and performance* - Trust	The extent of consideration of the same responses given by a human counselor	Less - Equal - More
15	*Effects of opacity, bias and performance* - Overall attitude	Overall conversation experience	Rate 1-10
16	*Effects of opacity, bias and performance* - Overall attitude	Willingness towards recommending Ava to someone else	Yes / No
17	*Effects of opacity, bias and performance* - Overall attitude	Willingness towards using Ava again	Free Text
18	*Effects of opacity, bias and performance*	Further Input	Free Text

In total, we were able to get 78 participants to test the different versions and answer the evaluation form. Due to having the chat open on Facebook messenger and reachable for everyone, three of the answers were removed due to not fulfilling the demographic restrictions of age group between 17-25 years old, consequently obtaining 75 respondents satisfying the criteria for analysis. Of the 75 responses, an even distribution of 25 responses for every version was obtained.

4.4 Research Credibility & Reliability

As emphasised in the introduction and background of this thesis, there seemed to be a rising demand from practitioners to gain insights about how to increase user trust in automated services. However, from a research perspective, there where a few concerns to consider before being able to add credible, reliable and valid results to the body of literature addressing this subject. We, throughout this investigation, acknowledged self-noted risks with trying to quantify a multidimensional concept in a multidimensional study. Furthermore, we applied the critique given to previous studies to our course of action (Dekker & Hollnagel, 2004; Dekker & Woods, 2002). In hindsight, we consider our way of conduct and our results to serve the practical research purpose in the specific case well. However, before making any claims about the implications of our findings on the generic literature, a few clarifications had to be made.

As stated in the research purpose, the ambition of this investigation was not to isolate and quantify a singular design choice of a component connectable to a concept of TIA. Instead, a broader and more commercially valuable approach was assumed. This was due to the novelty of the idea of trying to develop a study- and vocational guidance chatbot at the hosting company of this master thesis. As no previous investigations about the idea's viability and course of development had been done, it was desirable to conduct a study in a broader perspective. Furthermore, from an academic perspective, the existing literature providing an overview of system components affecting cognitive trust in chatbots was still very sparse. However, a consequence of focusing on several system components with alterations to multiple design choices was the inability to make any conclusions about singular modifications influence. Instead, the achieved results indicate differences in the overall trust due to alterations to one system component as a whole.

As accounted for in the theoretical section, contextualisation is a necessary process when evaluating trust in AS. However, highly contextualised frameworks imply increased difficulty in achieving falsifiable results (Pointon, 2017). Although the findings in this study provide reliable quantitative and qualitative data they suffer from fairly contextualised variables. Limiting their direct applicability in a general setting. Precautions such as including previous findings about individual expectations and the contextual risk were an effort to withhold the legitimacy and transparency of the conducted tests. Although it might be challenging to recreate equivalent contextual circumstances to validate the results of this study, it is plausible to assume that similar findings can be distinguished in a comparable setting.

The quantitative part of this study was founded in that different design choices affect the perception of cognitive trust aspects. Furthermore, we through previous studies assumed that there seemed to be an additive relationship between the amount of manipulation of significant system components and effect on user trust. For example, we assumed that the more biased we designed Ava, the more influence it would have on perceived trust. Although the correlation between the amount of deliberately modified system components and influences on trust where not a part of this study, we had to reason about to what extent we should alter the different versions. More specifically, we had to reason our way to how many and how significant modifications to make in the different system versions for the examined system components to be properly evaluated, without becoming too trivial and obvious. In this aspect, it was important to consider results about differences in perceived trust with regard to the total significance of different alterations to design choices.

The reliability of the study is, according to Kylén (2004), also strongly connected to the amount of participants within the quantitative evaluation, a higher frequency of answers leads to a more reliable result. In our case, the number of participants was limited by the access given by the schools, and therefore, no more than 75 correct participants could be attained. Kylén (2004) also states that the evaluation questions affect the reliability to some extent. By designing the evaluation form without any leading questions strengthens the reliability and validity of the study (Kylén, 2004). When performing hypothesis testing, similar to the F-test and t-test performed within this study, there are mainly two forms of errors that can arise. The Type I error (α) is set and defined by the statistical significance, which in our case is 0,05, meaning that there is a 5% chance that a faulty result will be given. Type I errors occur when the null hypothesis (H_0) is rejected although it is true, therefore also having the name False Positive. The other inaccuracy, the Type II (β) is the error occurring when the null hypothesis is false, but the test fails to reject, also called False Negative. This error is not set or defined but can be managed by changing the number of participants, the more participants, the lower the value of Type II error β. The statistical power (1-β) is the probability of rejecting the null hypothesis when it is indeed false. It is, therefore, a measurement of how likely the test is to detect a real effect given the parameters you have (Chow et al., 2008). The overall goal is to minimise both the Type I and Type II errors. Previous research, Cohen (1988), emphasises that the effects from the Type I (α) errors are up to four times more serious than Type II and therefore he suggests that a statistical power level of 0.8 (β = 0.2) is sufficient. The statistical power is calculated through a formula using the fixed alpha level (in our case 0,05), the sample size (the number of participants) and the effect size (the quantified presence of a result in the population). The formula is a reversible meaning that if a prior statistical power is determined, and the population variables are known, the population size can be calculated. In our case, with population measurements unknown, the approach of performing the test before calculating the statistical power had to be taken (Chow et al., 2008). Thus not rejecting the null hypothesis during the F-tests, only possible Type II errors may have occurred. The statistical power of the F-tests is low on the Neutral and Opaque version indicating that the presence of Type II errors may be high. To reach a power level of 0.8 with the same population statistics, the number of participants would be needed to increase by approximately 200-1000. Between the Neutral and Biased version, a statistical power of between 0.65 and 0.9 was obtained, suggesting that the number of test participants was sufficient enough to minimise the presence of Type II errors (See table B.3.6 in Appendix 3). The calculated statistical power in the t-test (See table B.3.7 in Appendix 3) follows the same schema as the hypothesis tests, reaching a statistical power of 0.8 or higher in the cases when rejecting the null hypothesis. Between the Neutral and Opaque version, the number of participants would have had to increase to approx 60-200 to reach a statistical power of 0.8. The final average result of the statistical power was fort the F-test 0.47 and for the t-test 0.68. In the light of these results we still believe that 75 participants are a good number considering the scope and magnitude of this study, although it would have been preferable with more participants to attain more powerful results.

Another relevant and significant aspect that had to be considered throughout this study and a consequence of relying on text as the communication medium, was the individual interpretation of semantics. The very basis for our investigation was the ability to properly operationalise theoretical findings of design choices suggested to affect trust. For example, in the case of mimicking contextual biases, reaching for different perceptions of benevolence, more philosophical reasoning about what biases really is had to be done. Although we refer to one of the system versions as "neutral", this conversation flow was still in some regards biased. Not only did all systems reflect the values of us as designer's (Zheng & Jarvenpaa, 2019) but also our personal interpretation of the meaning of words and formulations where subconsciously included. Furthermore, Ava makes associations built on generalisations of society in all versions although there is a major difference in how the information is presented. As to such, we address considerations about all system being biased to some extent by distinguishing between, in the context, "irrelevant" biases and Ava's hosting of prejudiced opinions and partisan semantics. A similar analysis of claiming to design opacity had to be done. We acknowledged that increasing transparency through more explanations, clarifications and motivations about Ava might have changed the conversation experience in other, unaccounted, regards. For example, a lower focus on the content relevant to the conversation topic and more responses about the functionality might have decreased the overall experience, influencing ratings about trust.

Beyond considerations to the practical design of the performed tests, considerations about the evaluation had to be made. First and foremost, the questions within the evaluation form were developed with the purpose and research questions in mind, thus, according to Bryman (2011) increases the validity of the research. A conclusion made by previous research is that the concept of trust is subject to personal definition (Hubal et al., 2008; Louwerse et al., 2015; Nass & Brave, 2007; Van Mulken et al., 1999). Therefore, there was a risk that participants would interpret evaluative questions about trust differently. Furthermore, careful consideration of the disposition and design of the evaluation forms had to be done.

5. Data

In the following section, the different types of data used are presented alongside with its origin. The chapter ends with a description of how the authors have worked to stay compliant both with the partner's policies and data protection laws such as GDPR.

5.1 Sources of Data

Data used within this study primarily originates from three different sources: open API data provided by the Swedish public employment service (Arbetsförmedlingen), data collected from various governmental and private administered websites and lastly data provided by our supervisor Maria Mattson Mähl based upon her expertise of the labour market.

5.1.1 Arbetsförmedlingen

The more substantial quantities of data have been collected from the Swedish public employment service (Arbetsförmedlingen) open API's. Arbetsförmedlignen supplies various types of data all connected to the Swedish job market. In the development of the product, two of the provided APIs was used to obtain relevant data, *yrkesvägledning* (vocational guidance) and *yrkesprognoser* (occupational forecast). As can be observed in Appendix C, where examples of the API responses can be found, the responses from both APIs carry a lot of mixed information, much information which were considered irrelevant for this study. Therefore, the APIs were accessed locally using Python, looping through and fetching all provided information. The data was then preprocessed where all relevant information was collected and desirably structured. Python code used for accessing and structuring the API data can be found in Appendix C. The information collected from the two different APIs where primarily;

- Occupation name
    - Short occupation summary
    - Occupation category
    - Preferable personal abilities for that specific occupation
    - SSYK (Unique occupation identification number)
    - A 1-year occupational forecast
    - A 5-year occupational forecast

Even though all information collected from the API's is not directly used in the current version of the product, everything is highly relevant for a more versatile product version in the future. The motivation behind choosing the information presented above is derived from the requirement specification together with the flow chart of the product functionality design.

Besides collecting information from Arbetsförmedlingen trough their open APIs, information was acquired directly from contacts with JobTechdev, the development department at the agency. This data consisted of one file with occupations and one file with competences. Both files had been, by JobTechdev, collected using AI and contained different occupations and competences that at some point had been listed on the open job site Platsbanken. The lists included altogether approximately 70 000 rows of data including different synonyms and regular misspellings.

5.1.2 Manually Collected Data

In order to make the chatbot catch and understand different types of answers and to understand what they imply, an extensive knowledge had to be constructed. A knowledge base, containing simple structured information that could be activated whenever the user writes anything that matches. This knowledge base needed to consist of both the occupations and competencies received from Arbetsförmedlingen but also more trivial inquiries such as different cities, education levels, education orientations etcetera, all the information the user expects the chatbot to know. This information was mainly gathered manually on different websites, both private and governmental administration. The nature of this data collection was quantity over quality, since it was prioritised to be able to cover a considerable amount of possible answers given by the users. This was to limit the occasions when the chatbot was unable to answer correctly rather than ensuring everything was formulated in the precise exact way.

5.1.3 Maria Mattsson Mähl & AlphaCE

As presented in the theory section, previous research emphasises that system performance plays a significant role in creating a reliable and trustworthy experience for the user. In order to ensure a high quality of content and to make sure the responses stay in line with professional vocational guidance, a lot of data and content was collected with the help and expertise of our company supervisor Maria Mattsson Mähl. When developing the product content, Maria contributed with primarily two things; schematic mapping and reply content. Schematic mapping involves mapping the large amount of different user responses into larger generalised groups where sweeping statements can be given. All information and tips Ava responded were either created or proofread by Maria to ensure their correctness and usefulness.

5.2 Data Processing

When it comes to data for the backend of the application, this project solely relied upon open source statistics and data provided by Arbetsförmedlingen available for anyone through their APIs (JobTechdev, 2019). Due to the complexity and size of the callbacks from the APIs, a choice to preprocess and save parts of the callbacks relevant for our work was made, which resulted in the use of static data for the backend. Another underlying reason for downloading the data was to minimise the complexity of the produced code in order to be able to focus on the purpose of the study. The data structuring was performed in Python and resulted in JSON objects that after the process was uploaded to the Firebase platform. The Python code used for extracting the data from the API provided by Artbetsförmedlingen can be found in *Appendix C.*

5.3 Staying GDPR & Integrity Compliant

The data collected and used throughout this project through the product and evaluation could at times be considered as private and include sensitive information. While using the developed products, the user conversations are automatically saved within Facebook Messenger. To protect user integrity, the assigned administrators for the different pages, e.g. the authors, continuously remove conversations manually. The product itself is designed with a scalable mindset which involves some data extraction. AlphaCE intends, in the near future, to extend the product into a phase which requires remembering previous user input to make the experience more versatile. With regards to AlphaCE's future plans, the product has been equipped with a function that saves some user-unique values to a Firebase database which can be enlarged at any time. To protect user integrity, Dialogflow is set to map the data with a unique session ID that only lasts for 30 minutes as well as remove all collected information at the last step of the conversation. To be able to stay compliant both with our partner's policies and laws such as GDPR, this project is anonymised and generalised to a point where no personal user details is presented or can be pointed out. Before answering the questionnaires, the test participants were informed that their responses would be used within in the study but that they would be anonymous. Out of confidentiality reason, this report was somewhat anonymised with regards to the product content, in order to not disclose important product information.

5.4 Ethical Considerations

Following the research objective of conducting a quantitative study on identified and significant system components expected to affect trust, a quantitative evaluation was conveyed through the means of testing different product versions and answering an evaluation form. Since the objective of all developed versions is to provide the user with suggestions and tips, precautions had to be taken to ensure that the test participants did not take the information provided by Ava to sincerely and to base any life-changing decisions upon it. To ensure this, clear instructions were given prior to the testing, instructions describing that the product is a result of a master's thesis project at Uppsala University and that the product wasn't' finished. When starting the evaluation form, the test participants was met by an initial message describing how their answers was going to be handled and used. Permitting the authors to use the result within the master's thesis and at the same time ensuring the privacy of the test participants. In the end, after submitting, the test participants was met with a message stating which version they tried and what was altered followed by a link leading to the neutral version for them to try. The end message also once again stated the nature of the test emphasising that the product is long from done.

6. Construction of the Chatbot

In the following section, the process of creating the chatbot are presented together with the different creative tools used. Due to company restrictions and various financial interests, the different documents used within the design and construction process are only presented in a sweeping fashion, and the actual documents used are not attached to this study.

6.1 Product Development Theory

The development of the chatbot followed the process of new product development often referred to as NPD. The process focuses on bringing an entirely new product from idea to commercialisation while minimising the perceived risks. In general, this established process follows the same steps regardless of what type of product that is being developed or if it is tangible or not (Reid et al., 2016). The NPD process can be categorised into four major groups; *Fuzzy Front-End, Product Design, Product Implementation* and *Fuzzy Back-End*. The first group, *Fuzzy Front-End*, represents actions such as identifying customer or market needs and defining a requirement specification. *Product Design* includes actions where the process goes from an idea into a product that should look and feel in a certain way and at the same time meet the specified requirements. This category also includes the actual development or building of the product, together with all necessary design decisions which follow. The third category, *Product Implementation*, includes the testing phase of the product. Ensuring all specified requirements stated in the Fuzzy Front-End part are fulfilled and that the product works in the intended way. The final stage of the NPD process is the Fussy Back-End. This part includes the commercialisation of the new product and all unspecified actions that follows with it (Reid et al., 2016).

6.2 Structure & Content

Throughout the development of the chatbot, primarily three development tools were used to ensure effectiveness, correctness and efficiency. Development tools that were produced before any actual product development started. The first thing established was a requirement specification containing primarily functional requirements of the product, e.g. topics, content and conversation flow. Some directives on non-functional requirements where given e.g scalability and accessibility but significant variables for the study on trust e.g. security was left for the investigation. The document was composed in close contact together with our supervisor Maria Mattsson Mähl and served as the foundation for the whole product development. In the initial phases of the project when conditions were uncertain about the possibilities and limitations of potential frameworks the requirement specification occasionally changed, both expanding and subtracting. Throughout the process when the opportunities and constraints became more defined, the requirement specification remained unchanged in a more considerable extent.

A schematic mapping diagram of the conversation flow was initially created, partially based upon and in line with the requirement specification. The different questions and their connection within the conversation flow were based upon AplpaCEs own and daily used 7 step coaching methodology. The 7 step methodology is based upon research from renowned sociology researches such as Lev S. Vygotskij and Erving Goffman whose work revolves around how people appear and communicate in fellowship with other human beings, and how to work around the protective barriers that people unintentionally places to hide the underlying obstacle (Goffman, 2014; Vygotskij & Öberg Lindsten, 2001). By starting from a phase of total confusion and no self-esteem or self-confidence, the method aims to step by step relegate the person being coached to a phase of confidence. While the 7 step method starts at the very beginning with e.g. unemployment and then finishes at employment, the product developed was limited to handle a part of that journey. Therefore, only a relevant portion of the 7-step method was implemented, more precisely the fundamental questions; Who am I? What do I want? and What do I know?. These questions acted as a base for the three different themes within the conversation; Personality, External Factors and Motivation, who themselves consisted of other more fundamental and in-depth questions (See figure 5).

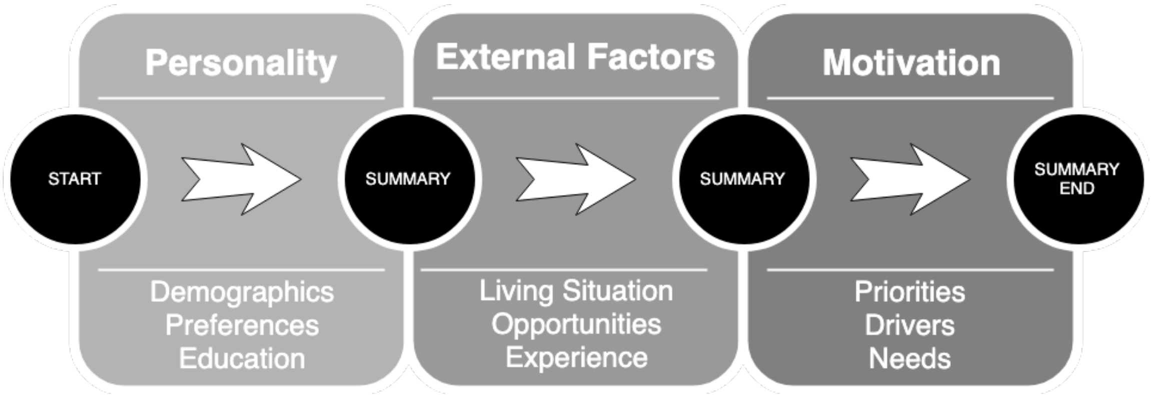


Figure 5. The figure presents a simplified version of the conversation flow. The conversation flow consists of three major parts; Personality, External Factors and Motivation, with a short mirroring between the parts.

Together with the different questions within the conversation flow diagram the purpose of the actual question and what it serves to the experience is stated. Defined in the requirement specification are three reasons behind asking a particular question, incentives that motivate that individual question being asked; mirroring, saving for a later assessment and creation of general and personalised feedback. Mirroring is a well known and used coaching technique which involves using given information, restructure and reformulate before giving it back to the person you are talking to, not providing any analysis och extra information. Saving for later assessment implicated that the information was going to be analysed in a following step, in the in-between themes summary or the end summary (See figure 5). The creation of individual and general feedback means that the answers provided by the user were used to create custom feedback dependant on the answer given, individualising the experience from the user's prerequisites.

The last tool created before initiating product development was a document containing pseudocode of the whole conversation. The document was continuously proofread when developed both by the authors and by the supervisor Maria Mattsson Mähl.

6.3 Software Motivation

In the following subchapter, the different software decisions that have been made during the study are stated together with the motivations behind.

6.3.1 Dialogflow

Throughout the project, the Google-owned human-computer interaction development tool Dialogflow[1] has been used in order to structure the chatbot (Dialogflow, 2019a). Since the main focus of this study is to highlight and evaluate the different key concepts that affect trust in the specific case, a decision was made to use an existing framework instead of building a chatbot from scratch, which would have been immensely time-consuming. The framework is a finished and ready to use development tool based upon natural language conversations. Besides from having an easy-to-use console structure for creating the conversation flow, Dialogflow has built-in tools for analysing user engagement and user patterns which were helpful for the continuous evaluation of the product (Dialogflow, 2019a,g). Dialogflow is highly scalable with easy integration to the Google-owned data storage and deployment platform Firebase, which was an important factor for AlphaCE when starting this project.

To further personalise the chatbot beyond the limits of the Dialogflow console, the framework enables further functionality to be developed in what they call Fulfillment. The fulfillment is code, in our case written in Javascript which can be found in *Appendix C*, that allows Dialogflow to spark backend functions from intent to intent. These business logic functions are deployed as a webhook, a web server endpoint created and hosted by Firebase and enables further functionality by using the information gathered by the natural language processor to trigger back-end functions, to implement rewrites of already defined answers or to generate more dynamic and changeable replies. The fulfillment webhook can be configured using any preferred development environment or by doing as we chose to do, using the inline code editor provided by Dialogflow which automatically deploys the webhook to Firebase (Dialogflow, 2019f).

[1] More technical aspects and descriptions of Dialogflow can be found in Appendix A or at the Dialogflow website https://dialogflow.com/

6.3.2 Firebase

Although Dialogflow is a complex product containing a large number of useful tools when developing a chatbot, in order to save and retrieve input values over an extended time, a third-party database/platform is needed. In this study, we chose to use the Google-owned database and platform Firebase. Since Firebase is in similarity to Dialogflow a Google-owned product the integration between the two different platforms was seamless and worked automatically when the Dialogflow Fulfillment was activated[2] (Google Firebase, 2019a,b). The primary purpose of using Firebase within this project was to enable saving and retrieving entities and answers written by the users in a later step of the chatbot process and use those entities to search the local database. In order to stay compliant with the European data protection laws GDPR, discussed in the previous chapter, all data collected from the chatbot-interacting users were deleted at the last step in the conversation, making sure no data was saved.

6.3.3 Facebook Messenger

The same core motivation behind choosing to work with Dialogflow instead of building the chatbot from scratch was applied to the decision upon the user interface. In order to stay focused on the sole purpose of the study and not spend a substantial amount of time constructing a user interface, a choice was made to use Facebook Messenger as the means for the user to chat with the chatbot. One of Dialogflow's useful functions is quick integration, and it enables smooth integration with several large chat-platforms such as Twitter, Skype, Slack and Facebook Messenger. This integration significantly simplifies the connection between the created agent (chatbot) and any of the message platforms, in our case Facebook Messenger (Dialogflow, 2019i).

6.4 Development

After the completion of the three documents described in the chapter above, the development phase commenced. Roughly, the development phase can be divided into three essential sections. Significant for all the parts is that they were somewhat connected to an evaluation. The first phase is where the initial development started. Notable for this stage where the continuous alpha testing that was present, in-house testing every functionality before moving on. The first phase ended when the alpha testing wasn't sufficient enough, and we needed a more extensive full-scale beta test. As a reaction to the beta testing, the second development phase was commenced. Besides from identifying trust engendering system components the evaluation provided valuable insights on how the users interact with the service and how the overall experience could be enhanced. Both the document containing pseudocode and the conversation flow diagram was changed to represent the new approach. The second development phase ended when a fully functional product had been developed again with the changes

[2] More technical aspects and descriptions of Firebase can be found in Appendix A or at the Firebase website https://firebase.google.com/.

enlightened by the beta test. The third and final development phase started with a fully functional chatbot that performed in the intended way. Phase number three was characterised by the cloning and alteration of the test versions. Before conducting any development, two new pseudo code documents were established containing one biased and one opaque conversation. The clones were then created according to the documents. The third and final development phase ended with three different versions of the chatbot, one neutral version, one biased version and one opaque version.

7. Results & Statistical Analysis

In the following chapter, all empirical results gathered from the various tests are presented. Results gathered from the quantitative and qualitative are presented together but categorized into the different evaluation areas. Since all testing, both evaluation forms and interviews, have been conducted in Swedish, all citations used below has been translated into English. The analysis consists of interesting correlations between questions and their statistical meaning. Elaboration on the implications of the results with regards to research questions are saved for the chapter "Discussion".

7.1 Significant System Components

Only a selection of graphs, relevant to later analysis and discussion, will be presented in the following subchapters, all information and used material connected to the empirical result can be found in *Appendix B*.

7.1.1 Beta Test

After the initial phase of development, a basic structure of the product had been constructed in accordance with the requirement specification and was ready for testing. The main agenda of the beta testing was to gain product specific feedback from a broader demographic audience as well as identify and validate trust affecting system components, highlighted in the previous research. The beta evaluation resulted in an action-plan describing what needed to be done product wise before conducting extensive tests.

The overall satisfaction level of the conversation experience shows an evenly distribution between 4 and 10 on the numerical scale, no evidence suggests a particular statistical distribution other than a uniform distribution. A mean answer of $\mu = 7{,}3$ with a standard deviation of $\sigma = 2{,}08$ show that the overall experience is higher than just a neutral reply of 5. A trend that can be derived is that the overall conversation experience is strongly connected to the technical performance, as can be seen below (See figure 6) where the grade is sorted from high to low. When testing for any correlations by calculating Pearson's correlation coefficient the result of $r = 0.82$, indicates a very high correlation between the responses. The respondents that answer on the lower part of the scale state in the following free-text question regarding technical performance that "*The conversation often broke, Ava didn't understand what I meant, and I didn't understand how to write for Ava to understand*" as well as "*Ended up in an endless loop after the question regarding the number of employers*", implying that their overall bad conversational experience is connected to the perceived technical performance.

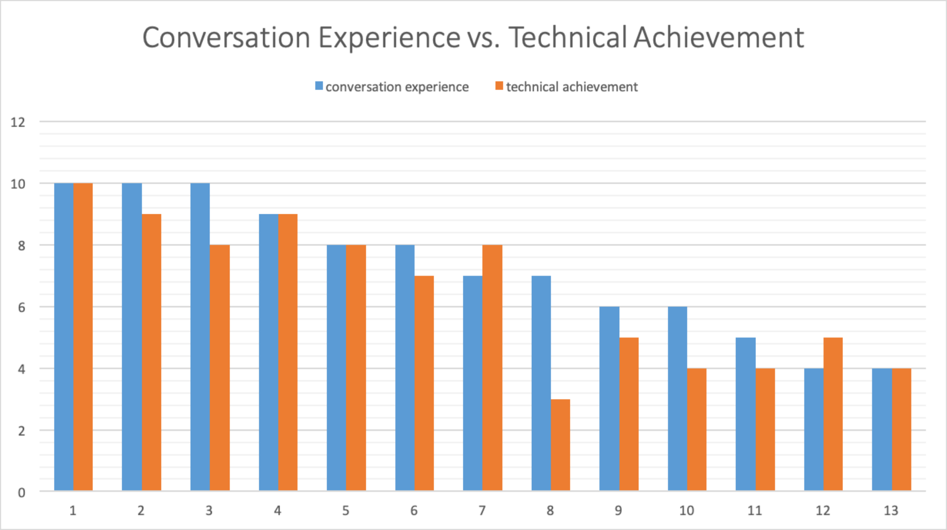


Figure 6. The figure shows the conversation experience in contrast to the perceived technical achievement

Following the connection between the technical performance and the overall experience, similar associations can be found when examining how well the product met the expectations of a study and vocational guidance chatbot. The test participants expectations vary in a pattern similar to the overall experience with a correlation coefficient of $r = 0.84$, indicating a very high correlation between the answers (See figure 7).

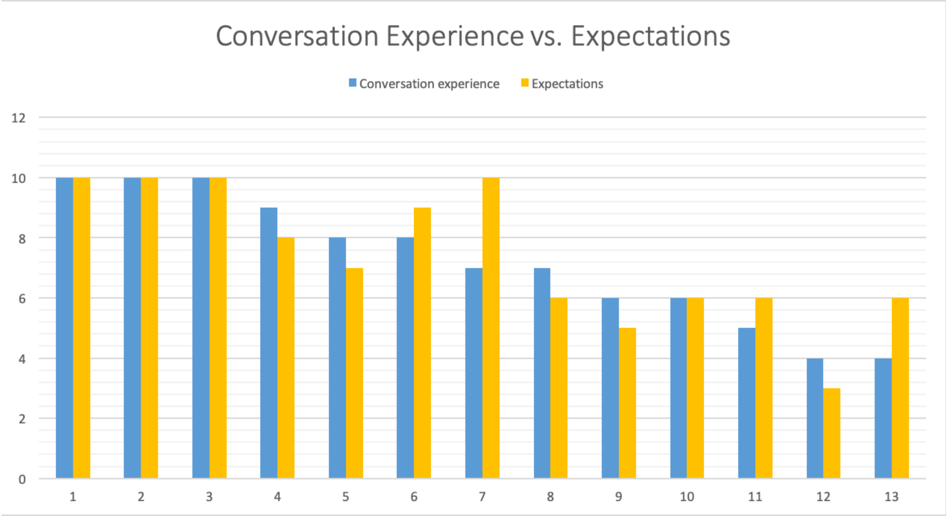


Figure 7. The figure shows the conversation experience in contrast to expectations

The connection and explanation behind why both overall experience, technical performance and expectations follow the same pattern can be derived from two main groups. First is system reliability, where the conversation didn't work correctly. Which in the cases where the score was low have comments such as: "*On the question; What would you say are your skills?, Ava didn't perceive my answers even though I simplified them*" and "*She didn't understand that often. In addition, there were a lot of strange* *questions*", in contrast to comments by participant located on the higher part of the scale, who enlighten minor problems connected to the conversation flow when asked what went wrong. The second identified group is personal customisation, which references to the conversation substance and the tips given out by the product. On the question of why they didn't listen 100% on the recommendations and tips given out, a third of the respondents answered "*I already knew everything*", and another third answered, "*I didn't feel the answers fit on me*".

Overall it can be concluded that the language used was not optimal, the tested product contained misspellings, faulty sentences etcetera. This is highlighted by 2 test participants, where one wrote: "*That there are deficiencies in the language makes it less reliable. Some sentences included misspellings and erroneous sentence constructions.*".

One participant enlighten the lack of motivation behind displayed answers, "*I don't know who Ava is or what sources Ava's facts come from. There is a need for some relationship with Ava as a brand to increase credibility*".

Other interesting observations from the beta tests are that almost everyone, 12 out of 13 or 92%, would consider using the product again in the near future. Although, a majority want the product to be more extensively developed before testing it again. The same amount, 12 out of 13 or 92%, answered "*yes*" or "*maybe*" on the question if they in a future would consider paying for such a service.

7.1.2 Market Analysis & Key Context Findings

The process of disposing the research included intuitive reasoning with regards to interesting areas of research. Learnings from initial labour-market insights and company interactions allowed for a focus on how to develop a first beta version of the chatbot (see 4.2.2). Although being careful not to leadingly focus on one component more than the other before proper beta-testing had been conducted, early discussions with company employees revealed the importance of biases (see 4.2.1). With reference to the envisioned "*natural benefits*" of a HMI in comparison to a human-human dialogue, one of the key components was providing a completely neutral and unbiased entity to reflect personal information with (Brandtzaeg & Følstad, 2017b; Goffman, 2014; Vygotskij & Öberg Lindsten, 2001). This results in a natural identification of focusing on unbiasses as a significant system component for engendering trust. Furthermore, it is found that biases and system transparency has a close linkage, as accounted for in the theoretical section (Zheng & Jarvenpaa, 2019). Moreover, this is shown in the beta test; "*She didn't understand that often. In addition, there were a lot of strange questions*", where perceptions of the system being biased, is at times a result from the lack of explanations to chosen responses and topics.

7.2 Effects of Altered System Components

Only a selection of graphs, relevant to later analysis and discussion, will be presented in the following subchapters, all information and used material connected to the empirical result can be found in *Appendix B*.

7.2.1 Neutral Version

The overall satisfaction level of the conversation experience is in the neutral version higher than in the beta testing. With a mean result value of $\mu = 8{,}2$ along with a moderately even distribution, the results are gathered closely around the mean (See figure 8), which is supported by a standard deviation of $\sigma = 1{,}3$.

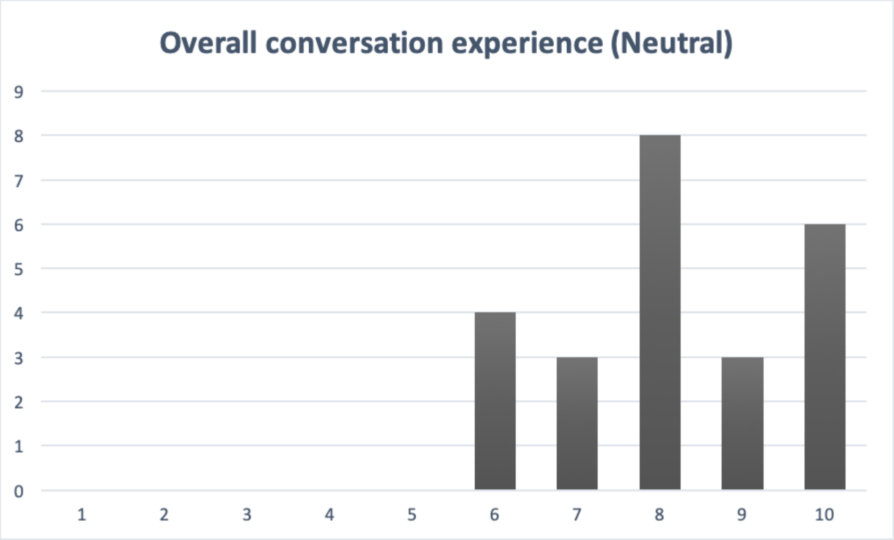


Figure 8. The figure represents the distribution of answers on the question of perceived overall conversation experience.

In contrast to the beta test result, the overall conversation experience does not have the same correlation to the perceived technical performance of the chatbot. Between the variables, there is a correlation coefficient of $r = 0{,}07$ which indicates little if any correlation. The mean value of perceived technical performance is $\mu = 7{,}8$with a standard deviation of $\sigma = 1{,}6$. Overall in the free text sections, a lot fewer comments are made connected to the technical performance of the chatbot in relation to the beta test. In total five remarks with regards to technical performance are stated, including; "*Fix the bug, and then maybe it works great as a vocational coach*" and "*The first group of messages came twice*". The mean value of the perceived technical performance of the test participants leaving comments is $\mu = 5{,}8$, and their overall experience is $\mu = 8{,}2$. Study and vocational guidance counsellor Therese Broström (2019) implies during qualitative interview that "*Errors would probably have been acceptable once. But had it happened repeatedly, there would be bad consequences*" (Broström, 2019). Thus, in similarity with the beta testing, a correlation between the expectations of the chatbot and the overall conversation experience can be identified. With a correlation coefficient of $r = 0{,}58$ indicating a moderate correlation.

When it comes to how willing the user is to answer the questions truthfully, the majority finds it easy to be truthful when chatting with the bot. The results show a mean value of $\mu = 9{,}1$, with a standard deviation of, when asked how comfortable they are to answer truthfully, and when asked whether they responded to every question honestly, 92% reply yes. The participants also state that in a similar conversation with a human study and vocational coach, one of them (4%) would be more comfortable to answer truthfully in relation to the conversation with Ava. 64% of the participants reply that they would be equally comfortable, while 32% say that they would be less comfortable in a human to human conversation.

The perceived reliability of the answers and the replies given by the neutral model has a mean response of $\mu = 8{,}6$ with a standard deviation of $\sigma = 1{,}7$. A majority of the participants reply with a grade 9 or higher. One comment provided from one of the higher answering participants stating; "*I do not understand what can be unreliable. Ava repeated what I wrote and wrote the correct variables back*". Another participant, with a lower reliability answer, stating; "*It is hard to receive completely personalised answers*". Both Terese Broström (2019) and Mattis Lu (2019) emphasise the importance of the mirroring technique when coaching, enabling the person to realise the possibilities themselves. Mattis said; "*It is important from a job search perspective to keep the coaching methodology*" followed by "*When a person comes up with what he/she is supposed to do, it is better rooted than if someone else does it*" (Lu, 2019; Broström, 2019). On the following question of how much they considered the tips provided by Ava, a mixed amount of answers can be distinguished. With a mean of $\mu = 6{,}5$ and a standard deviation of $\sigma = 2{,}5$, the replies are almost uniformly distributed amongst the scale. A low correlation between the reliability and consideration of tips can be identified with a correlation coefficient of $r = 0{,}3$. When the participants were asked why they didn't consider everything Ava said a majority of the respondents, 42%, choose the alternative "*other reason*", not thinking that any of the choices provided are suitable. The second most common answer is "*I already knew everything*" and "*I did not feel that the answers matched me*", receiving 25% of the total responses each. Even though the participants only consider the tips provided with a mean value of $\mu = 6{,}5$, only 24% of the participants would listen more to the suggestions if provided by a human counselor. In total, 20% of the respondents would listen less to the tips provided, and 56% would listen equally to the tips if they were given in a human to human conversation.

In total, 92% of the participants that tested the neutral version would recommend it to their friends. Some of the comments found at the end of the evaluation form stating; "*Ava is easy to use. You do not have to book time or get anywhere. It went fast and smoothly. Convenient to write in Facebook chat!*", "*All the sharp tips for study choices are useful, especially when you know that the chatbot without precedent interprets* *exactly what I wrote and thus becomes more credible.*" and "*Absolutely, if there is an opportunity where I do not know what to do with my life, I may be using Ava*". Mattis Lu (2019) said after testing the neutral version that the conversation was very much like the conversations she has with employment seekers; "*The conversation was empathetic and good, with relevant questions from a coaching perspective. Just what I would ask in my conversations*" (Lu, 2019).

7.2.2 Opaque Version

The overall satisfaction level of the conversation experience is in the opaque version lower than in the reference version but higher than in the biased version. With a mean result value of $\mu = 7{,}6$ along with a moderately even distribution, and with a standard deviation of $\sigma = 1{,}6$. In contrast to the neutral version and more in line with the beta testing, the overall experience correlates to the perceived technical performance (See figure 9). The two variables correlated with a correlation coefficient $r = 0{,}63$, which indicates that a moderate correlation between the variables exists. Also, a high correlation can be identified, with a correlation coefficient of $r = 0{,}87$, between the answers on expectations of the chatbot and the overall conversation experience.

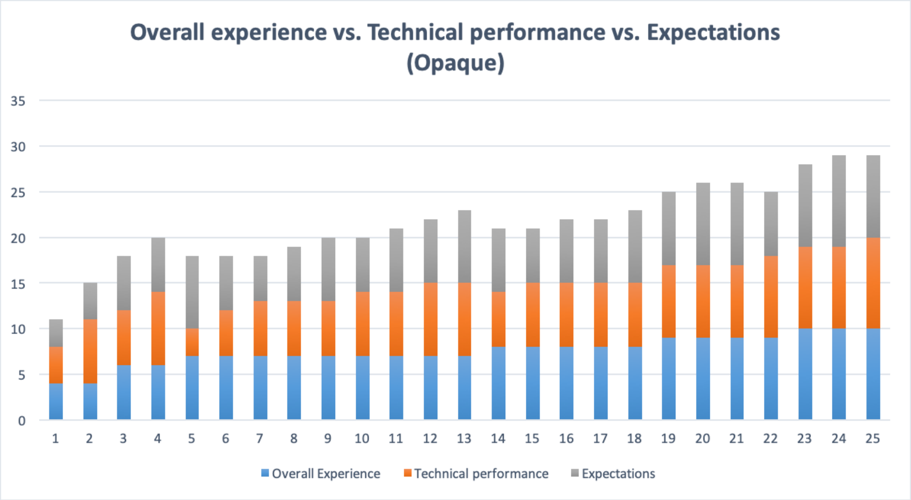


Figure 9. The figure displays the correlation between stated parameters.

On the question of how willing the test participants were to answer the question truthfully the answer is higher than in both the neutral and biased versions. The mean value from the responses is $\mu = 9{,}3$ with a standard deviation of $\sigma = 1{,}2$. Also, 100% of the participants answered that they answered every question truthfully. There are only three participants, 12% of the group, who answer that they would have felt more comfortable to answer the questions more truthfully to a human counselor. Leaving 60% of the respondents feeling equally comfortable and 28% feeling more comfortable answering truthfully in a similar human to human conversation.

The perceived reliability of the answers and the replies given by the opaque version has a mean response of $\mu = 7{,}7$ and a standard deviation of $\sigma = 1{,}7$. The participants who answer in the lower part of the scale stating; "*There was a bit much repetition of what I wrote*", "*Scenarios that Ava painted sometimes felt a bit well-intentioned and impersonal against previous responses. Didn't feel like there was something that could show up soon in one's life*." and "*Some answers felt unreliable and hard to trust*". Participants who answered in the upper part of the scale imply having issues with reliability, answering; "*Uncertainty that the answers are completed and that many receive exactly the same answer as I do, thus the same guidance*" and "*Standardised answers that I already knew*". Connected to the reliability of the content, the overall mean value of how much they considered the tips provided by Ava is lower in comparison to the neutral version, with a mean value of $\mu = 6{,}4$ and a standard deviation of $\sigma = 2{,}0$. In this version, no or little correlation between the reliability and consideration of the tip can be identified. On the following alternative question, the most chosen option is "*I didn't believe the answers*" with 25% of the answers followed by "*other reason*" and "*I already knew everything*" with 20% each. When asked how they would have considered the tips if they were participating in a human to human conversation, 36% of the test participants state that they would consider the tips to a greater extent. Only one person answered that they would consider the tips less in a human to human conversation and a majority, 60% of the participants, would consider the tips equally.

Mattis Lu (2019) feel that this version is the one most similar to the neutral version, since in similarity, it was a coaching conversation instead of an advisory session, saying; "*Reminds a lot about the first one with the difference that it is not as clear with what it does, there is no motivation behind the answers*", which were similar to what Therese Broström (2019) said; "*Got no information at the beginning of how it worked, can be difficult with prior expectations*" (Lu, 2019; Broström, 2019).

A total amount of 84% of the participants that tested the opaque version would recommend using Ava to a friend. In line with the results from the neutral version, the majority of the participants would use the product again, in some cases after additional work has been done to fix minor bugs and content, "*Ava needs further development! Think it needs to be a little more complex, just like us humans*" and "*Ava was good!! But it feels like it needs to collect more data to get a more precise answer*". Other positive comments can be found; "*It worked very well and became almost like a "Wow!" experience when I got an answer. Really cool!*" and "*Good as a simple start in a career guidance process*".

7.2.3 Biased Version

The overall satisfaction level of the conversation experience is in the biased version lower than both the neutral version and the opaque version. The perceived conversation experience has a mean value of $\mu = 6{,}6$ and a standard deviation of $\sigma = 1{,}9$. A moderate correlation can in the biased version be identified between overall experience and perceived technical performance as well as a low correlation between overall experience and expectations, with correlation parameters of $p = 0{,}57$ and $p = 0{,}24$ respectively. The perceived technical performance of the chatbot has a mean value of $\mu = 5{,}4$ and a standard deviation of $\sigma = 1{,}9$, making it the version with the lowest mean out of the three versions. Following the weak trends in comparison, the expectations of the biased version are also the lowest, with a mean value of $\mu = 5{,}4$ and a standard deviation of $\sigma = 1{,}4$.

When asked whether or not the participants felt comfortable to answer all questions truthfully, a mean value of $\mu = 8{,}4$ is attained together with a standard deviation of $\sigma = 2{,}0$. All together, everyone except two, 92%, reply that they answered truthfully on all the questions. Even though such a large part of the population say that they answered all the questions honestly, 20% responded that they would feel more comfortable answering truthfully to the questions if the conversation was with a human. Although, 28% of the population, stating that they would feel less comfortable and 52% state they would feel equally comfortable to answer truthfully in a conversation with a human vocational and guidance coach.

The perceived reliability of the replies given by the biased model has a median response of $\mu = 6{,}2$ and a standard deviation of $\sigma = 2{,}0$, implying that the answers are varying. Some of the participants not replying a ten on the scale commented; "*It didn't understand me*", "*The answers didn't feel personal*" and "*Sometimes Ava formulated herself strangely to the info, and you didn't understand what she was referring to*". A total of three comments are directly referring to a lack of personalisation of the chatbot, wanting Ava to both answer and remember more complicated and complex information. The perceived reliability of the content moderately correlate to the extent of how much participants consider the tips provided during the conversation, with a correlation coefficient of $r = 0{,}54$. Overall, the consideration of tips received have, in relation to the other two models, a low mean value of $\mu = 5{,}2$ and a standard deviation of $\sigma = 2{,}2$. On the following alternative question, the most chosen option is "*other reason*" with 36% of the answers followed by "*I felt the answers didn't fit me*" and "*I already knew everything*" with 20% of the answers each. 40% of the participants who tested the biased version would consider the tips to a greater extent if a human coach provided the tips. The rest, 60% feel they would consider the tips equally in a similar conversation with a human. Worth mentioning is the absence of participants considering the suggestions to be superior when delivered from a chatbot then if provided by a human, a finding present in the other versions. Both Mattis Lu (2019) and Therese Broström (2019) see problems in the way Ava, in this biased version, gives recommendations instead of tips and that in many cases can scare people who do not feel like the information fits. Mattis Lu (2019) said; "*This is more an adviser and not a coach. It knows nothing about me as a person, which is problematic. Providing that advice, there must exist an underlying motivation why*" and Therese Broström (2019) said; "*It provides more suggestions, which can be a bit dangerous when having so little* *background information. It can scare away the people who partly disagree*"(Lu, 2019; Broström, 2019).

A total amount of 68% of the participants would recommend the product to a friend. Participants who would recommend Ava to one of their friends stating; "*Sometimes you don't have the energy to go to the student counsellor, then it is good to have Ava*", "S*ometimes it may feel good not to have to find a study counsellor to talk to and can then simply pull out the phone or computer and write with Ava*" and "*She gave me new insights about things I hadn't reflect upon*". The participants that are not willing to recommend Ava to their friends comment on the following question; "*Maybe, may be good for some purposes*" and "*Maybe, especially if the bot is further developed*", implying that with some further development and bux fixing they could consider to use the product again.

Two test participants comment on the speed of the replies as a negative factor, stating that; "*Ava is extremely quick at replying on my answers, which meant that she was really perceived as a robot, thus more an AI than a human." and "Can answer a little slower if possible*".

One test participant enlighten the language and motivation Ava uses, commenting; "*Sometimes Ava made some long-drawn assumptions, for example with driving license. Instead of saying "a driving license is good to have, you should get one!" Ava could ask you if it is something you can think of to obtain (economic and environmental reasons can affect)",* implying that the language or recommendations used by Ava is not optimal in the given situation. Something Therese Broström (2019) also commented on during the interviews; "*This version is a bit more negative compared to the previous (neutral version); it focuses on what you do not like, which can dissuade people*" (Broström, 2019).

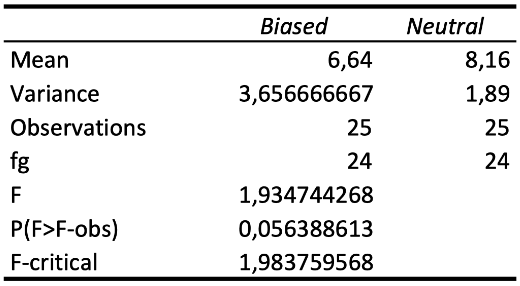
7.3 Final Findings

To summarize the findings made in each individual version of the chatbot, a compilation of the findings can be found below (See table B.3.1 in Appendix), displayed in a systematic and perspicuous order.

Both Therese Broström (2019) and Mattis Lu (2019) emphasized that all conversations within the different versions were similar to the conversations they convey in their everyday job. Mattis Lu said; "*It's clear that this product is for the group of people not sure of what they want to do in their lives. The structure of the chat reminds a lot of the conversations we conduct every day*" (Lu, 2019). Mattis also state that the significant differences between the three versions is that the neutral and opaque version is of a coaching kind and the biased version is more advisory. The difference between the opaque and neutral version is the level of motivation behind the statements, suggesting that the neutral version handles the conversation more transparently. Mattis also points out that a lot of people seeking help often wants to be told what to do, but that isn't the best way to go due to motivation in the later carrying through; "*Often those who I talk to want some sort of answer, they want to be told what to do. ... When a person comes up with what he/she is supposed to do, it is better rooted than if someone else does it*" (Lu, 2019). Both Therese and Mattis agree upon that the preferable version to use within a coach conversation was the neutral (Lu, 2019; Broström, 2019).

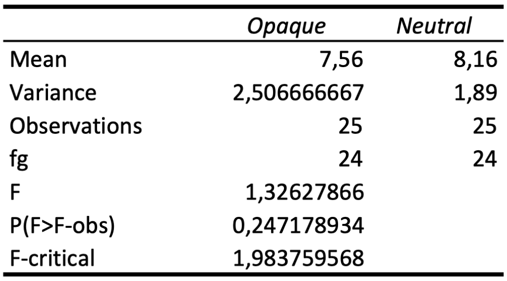
In order to analyze the difference in the variance between the chatbot versions, F-tests are conducted. As can be visualised below (See table 5 and table 6), the calculated F values on the question overall conservation experience are smaller than the critical value F-critic of the distribution. This result, together with the fact that the calculated p-value is larger than our chosen statistical significance value, lead us not to reject the null hypothesis. Following that, we can't with a 95% certainty reject the null hypothesis stating the variances are alike in both comparisons.

Table 5. The table shows conducted F-test on question overall conversation experience for the neutral and biased version.



	Biased	*Neutral*
Mean	6,64	8,16
Variance	3,656666667	1,89
Observations	25	25
fg	24	24
F	1,934744268	
P(F>F-obs)	0,056388613	
F-critical	1,983759568	

Table 6. The table shows conducted F-test on question overall conversation experience for the neutral and opacid version.



	Opaque	*Neutral*
Mean	7,56	8,16
Variance	2,506666667	1,89
Observations	25	25
fg	24	24
F	1,32627866	
P(F>F-obs)	0,247178934	
F-critical	1,983759568	

Further the F-tests that are applied (See Appendix B) on the parameters *Technical Performance* and *Expectations* provide F-values larger than the F-critical together with a p-value larger than $p = 0{,}05$. Following those results, we can't with a 95% certainty reject the null hypothesis stating the variances are alike.

To statistically analyze if the mean values varies between parameters, t-tests are conducted using Microsoft Excel. T-tests were performed on the examined parameters; *Overall Conversation Experience, Technical Performance and Expectations*. The results suggests (See table 8) that the null hypothesis cannot be rejected in any of the cases when comparing the neutral and opaque versions. In the comparison of the neutral and biased version the null hypothesis can be rejected since it has smaller p-values than $0{,}05$ in all cases, implying that the mean values are different between the versions on a significance level of 95%. In the comparison between the opaque and biased version the p-values suggests that the null hypothesis can be rejected on the parameters *Technical Performance* and *Expectations* on a significance level of 95%. Since a p-value of $0{,}07$ is received the null hypothesis can not be rejected on the parameter *Overall Conversation Experience*.

Table 7. The table displays the acquired p-values from the t-test

T - test	N/O*	N/B*	O/B*
Overall conversation experience	0,158	0,002	0,070
Technical performance	0,085	$1{,}45E^{-5}$	0,002
Expectations	0,518	$1{,}42E^{-4}$	0,003

*N=Neutral, O=Opaque & B=Biased

To analyse any correlation between the models as a whole, Pearson's correlation coefficient is calculated on the mean values for every version (See table B.3.1 in Appendix). The means used are those of a numeric kind, and the measurements regarding the previous knowledge are not included in the correlation computation. The result shows that all versions correlate to a high extent between each other. The neutral and opaque version have a correlation coefficient of $p = 0{,}76$. The neutral and biased version have a correlation coefficient of $p = 0{,}83$. And finally, the result from the opaque and biased version correlate with a correlation coefficient of $p = 0{,}83$.

8. Discussion

In the following section the results and analysis presented in the previous chapter are discussed with reference to the theoretical framework, in order to address the purpose and research questions of this study.

8.1 Identifying System Components

The first of the two main research questions of this study is to identify significant system components that affect trust in a service encounter with an AI-powered study- and vocational chatbot. In accordance with the first objective, this was done partly by reviewing existing literature to legitimise method choices and to establish literary grounding for identified system components. Secondly, we conducted a substantial amount of contextual market analysis and iteratively prototype and test a chatbot. To do so we used agreed upon (Gareth et al., 2001; Muir, 1994; Barber, 1983) notions of human TIA to be able to ground and target certain cognitive concepts with distinguished system components. The result is the showcasing of a way of conduct to identify and prioritise trust in the development of a chatbot and as a result a certain emphasis on a few system components. The accumulated outcome of these processes is what allows us to answer the first research question. Our findings both confirm and add to the framework for identifying TIA.

8.1.1 Showcasing a Way of Conduct

As stated in this thesis contributions to science, one of the ambitions was to showcase a method for defining, identifying and evaluating human trust in an AI-Automated service encounter. Therefore, in hindsight, the process described above, of conceptualising, framing and evaluating trust in the certain context is a result in itself. Confirming, to the extent of the usefulness of the results, the practical viability of the way of conduct. Even though this investigation contains room for criticism and concern, enlightened in the method section, the results to some extent empirically confirm that trust can be evaluated through proper contextualisation and concept break down (Lewis et al., 2018; Følstad et al., 2018; Hieronymi, 2008; Keren, 2014; Simpson, 2013). We show that by assuming a contextual viewpoint, it is possible to identify and prioritise amongst the numerous design choices that can be made with regards to establishing trust in dialogue with a chatbot (Devitt, 2018). Furthermore, we validate that design choices with regards to the chosen system components have an effect on the cognitive concepts of trust in user interaction (Muir, 1994; Barber, 1983). In this regard, although we don't contribute to defining a general metric for evaluating TIA (Chien et al., 2014; Chien et al., 2015), we show that chatbot design principles affect ratings of trust, without explicitly defining a social performance goal (Følstad et al., 2018).

8.1.2 Significant System Components That Affect Trust in Ava

The data that supports the identification and definition of the non-functional design aspects we refer to as separate system components gathered in this study is not extensive. It is plausible to assume that there are several more significant components that affect the cognitive concepts of TIA in the certain case. However, we reason that there is enough empirical notion of recurring empirical emphasis on certain system components, also hinted to affect trust in the past literature, to establish a significant focus for a quantitative study.

A lack of motivation to provided response was displayed during the beta test; "*I don't know who Ava is or what sources Ava's facts come from. There is a need for some relationship with Ava as a brand to increase credibility*". A result that emphasise transparency, more specifically in the form of a brand or identification for the product (Følstad et al., 2018). Although this is the only explicit empirical result from beta testing that highlight transparency, other findings indirectly support its importance. For example, a finding such as; "*She didn't understand that often. In addition, there were a lot of strange questions*" supports the negative implications of the perceived lack of integrity (Devitt, 2018). Similar findings (See section 7.7.1 Beta Test) both implicitly highlight transparency and more explicitly the ability of Ava as important system components for engendering trust. The results display the importance of designing an AS that is perceived to be trying its best and that takes responsibility for its actions. Motivating a quantitative study on the effects on trust by alterations to the translucency of its actions (Følstad et al., 2018). Furthermore, supporting the connection of such design differences to the perception of the systems honesty, motives and character (Lewis et al., 2018).

Beyond hinting about the importance of system transparency most results from initial beta testing emphasise the influences on trust due to the perceived ability of the system. There is a clear correlation between the overall conversation experience and perceived technical performance (See figure 5). For example, participants imply that "*The conversation often broke, Ava didn't understand what I meant, and I didn't understand how to write for Ava to understand*". In accordance with previous findings, the influences on perceived ability seem to be damaged by the lack of system reliability (Salem et al., 2015; Moray et al., 2000). The same findings support a negative impact on user experience due to the lack of system predictability (Lewandowsky, 2000). Motivating a quantitative study on how the impact of erroneous behaviour can be affected by being more or less explicit about system vulnerabilities (Riley, 1994). Furthermore, our qualitative results support influences on trust due to shortcomings of the system capability. Both in terms of being able to process input expected to be manageable and the ability to discuss more customised topics (Lewis et al., 2018). Collectively these finding are acknowledged to support system performance as a significant component for the perceived ability of the AS. Moreover, supporting a deeper investigation on its impact on trust as a function of the perceived reliability, skills and accuracy (Hoffman et al., 2013).

Lastly, contextual circumstances highlight unbiasses as a significant system component to investigate with regards to trust in the particular case. This is partly due to empirical labour market insights and informal interviews with job coaches but also by considering the methods and findings that inspired the conversation flow (Goffman, 2014; Vygotskij & Öberg Lindsten, 2001). With a vision of further development, we consider the likeliness that some level of contextual bias could be included if proper precautions are not taken (Caliskan et al., 2017). Therefore, it is, in addition, intuitively motivated to study the effects on trust as a function of changes to the biases of the agent. Furthermore, it can be found that biases and system transparency has a close linkage in the previous literature (Zheng & Jarvenpaa, 2019). Spiking the interest for evaluating the effects of differences in the altruistic characterisation of the conversational agent in accordance with previous studies (Følstad et al., 2018). All together, we consider there to be significant motivation for a quantified evaluation of the effects on the perceived prejudice, motives and beliefs through alteration of design choices connected to contextual biases (Robinette et al., 2015).

8.2 Evaluating System Components

The second of the two main research questions of this study is to evaluate how alterations to design choices related to the identified system components affect the targeted cognitive concepts of TIA. In accordance with the second objective, this was done in a quantitative study where deliberately altered design choices in different conversation designs where tested. The results display that design alterations do have an effect on the subjective evaluation, confirming to different extents the identified components significance in establishing trust. Although no explicit conclusions can be made with regards to singular design choices effect on a particular cognitive trust concept, interesting implications can be distinguished. Findings that both confirm and add to the framework for evaluating TIA.

8.2.1 Containing Effects of System Performance

Similarly to the beta test, we in the quantitative testing confirm previous results on the effects on trust due to shortcomings in system performance (Salem et al., 2015; Moray et al., 2000; Lewis et al., 2018; Følstad et al., 2018; Lewandowsky, 2000; Hancock et al., 2011). Our findings show an overall lower effect on trust due to a more technically stable conversation than in the beta test. As a result of the course of action established after the beta test, a sharp decrease in the effects of lacking system performance can be acknowledged. In beta testing a clear correlation between the overall conversation experience and perceived technical performance could be distinguished. Whereas in the neutral version evaluated in the quantitative test, no such relationship is notable. Moreover, the overall satisfaction level of the conversation experience in the neutral version was higher than in beta testing, supported by both standard deviation and qualitative evaluations. A result that strengthens that the most prominent and significant effects due to faults and erroneous behaviour were moderated. Furthermore, the individuals that did comment on weaknesses in system performance still have a high overall conversation experience. A finding that supports that the technical issues that arose during the conversation weren't critical for the whole experience. A plausible explanation which follows the theory (Moray et al., 2000) is that the mistakes were not that critical, resulting in a smaller effect on the subjective evaluation.

Overall the results from the quantitative testing on the neutral version display a lower effect on evaluation due to the perception of ability. However, in comparison to the other evaluated system components, we don't quantify its effects with design alterations to a dedicated system version. Meaning that there is still a risk that the performance could be a dominant component affecting trust in all versions evaluated in the quantitative study. However, since the results display a lower effect due to system performance, accounted for above, its impact on trust is, to some extent, mitigated. With regards to the review of the other two versions, this increases the reliability that the distinguished effects are related to other system components than performance. Furthermore, since the risk for faults and erroneous behaviour is the same in all, more valid conclusions with regards to the other evaluated system components can be made.

8.2.2 Consequences of System Opacity

One of the opaque versions most distinguishable and interesting findings is that the correlation noted in beta testing but not in the neutral version, between the overall experience and the perceived technical performance (See figure 9), is shown again. A plausible explanation is that by removing proper self-presentation at the beginning of the conversation (Kretzschmar et al., 2019) even small mistakes and erroneous behaviour result in a bigger impact on user trust (Riley, 1994). Both the quantitative correlation and qualitative evaluations imply that the lack of transparency decreases trust due to insufficient system predictability (Luger & Sellen, 2016). Supporting that a lower understanding of system behaviour and mistakes results in a lower subjective evaluation of experience (Lewandowsky, 2000). There is a correlation between perceived system performance and overall experience in the biased version although not as strong as in the opaque version. This implies that there could be other reasons why mistakes and erroneous behaviours have a different impact on the respective versions. However, in the biased version, the correlation lacks support by qualitative data and is most likely a symptom of other alterations. Supported by the lowest overall experience of all versions.

Secondly, the experienced reliability of the responses is considerably lower than in the neutral version. Qualitative results imply that this is partly due to the lack of personalised answers, suggesting that the perceived capability of Ava is lower in the opaque version than in the neutral version. This finding support that the decrease of system transparency increases the probability of mismatches in user expectations and system capabilities (Luger & Sellen, 2016). Suggesting that the decrease of transparency not necessarily gives a direct effect on the perceived integrity but rather in this case on the ability. Consequently, fewer people considered the content that Ava provided when maintaining an opaque system. Results that display an effect from complex and intricate reasoning to why the conversation is behaving the way it is (Castelvecchi, 2016; Hepenstal et al., 2019). Furthermore, in the opaque version, the most common reason for not considering a response was the lack of belief that is was true. A lack of belief is likely a result of the absence of branding and information to what extent the responses is backed up by research and evidence (Følstad et al., 2018; Kretzschmar et al., 2019). Although the experienced reliability of responses was even lower in the biased version, qualitative evaluations suggest that this is due to other reasons than in the opaque version. In the biased version, the main reason for not considering responses was due to unspecified reason and the qualitative data rather support it as a symptom from an overall lower conversation experience.

8.2.3 Consequences of Contextual Biases

One of the more interesting questions with regards to the effect of contextual bias is the willingness to answer truthfully to all questions asked by Ava. Previous research shows that interaction with a conversational agent could have beneficial characteristics over interpersonal interaction due to its unbiased nature. Both the neutral version and the opaque version display significantly higher percentages on comfortability to answer truthfully. Furthermore, the biased version has the lowest number of comfortability to answer a human coach instead. This finding supports the fact that the threshold for answering truthfully is higher, knowing that the agent might judge or value the answer (Følstad & Brandtzaeg, 2017b). Moreover, qualitative results from the interview with a job coach support that the fear of thinking or saying stupid or silly things is increased in the biased versions. Implying a significant negative impact on user trust and experience (Fuchs, 2018).

Secondly, the biased version has the highest percentage of participants saying they would consider the same tips to a greater extent if it was a human coach that provided them. Furthermore, in comparison to the neutral and opaque version, no one considered the conversation to be superior to have with Ava then with a human counsellor. It is possible that participants conversing with biased Ava might experience a closer resemblance to a human dialogue, supported by a few qualitative results. However, as this, in theory, might increase trust (Tapus et al., 2007; Følstad & Brandtzaeg, 2017b) the mimicking of contextual bias rather implies negative effects in this case. Potential benefits of empathic and expressive characterisation (Tapus et al., 2007) by alterations to the used language are outweighed by the negative consequences on trust due to reflections of prejudiced regularities latent in the context. Both participants and interviewed experts highlighted uneasy feelings in the user due to the use of loaded terms and formulations (Følstad & Brandtzaeg, 2017b; Singh, 1999).

8.2.4 Noise from Surrounding Influences

Although several interesting and significant effects on trust can be backtracked to the alterations made in the respective system versions, it is likely that some differences are due to surrounding influences. Both in the beta and quantitative tests we confirm (Louwerse et al., 2015; Nass & Brave, 2007; Van Mulken et al., 1999; Hubal et al., 2008) that there is a correlation between individual expectations and user experience (See figure 6 & 8). The correlation is not as distinct in the beta test, however, our findings suggest that people with a lower overall experience also felt that the product met their expectations to a lower extent. Moreover, the mean expectations amongst participants in the respective system versions differed. As to such, we confirm that TIA needs to be put in the perspective of how well users expectations are met (Van Mulken et al., 1999).

Furthermore, results from the neutral version to some extent, emphasise effects due to the contextual risk. For example, in all versions, a significant amount of participants did not consider the majority of the responses. In above this is discussed as an effect of the lack of transparency or biases respectively, however, it could also be that the conversation topic has a reasonably high required threshold for establishing trustworthiness (Devitt, 2018). Therefore, we with recommendation from previous scholars acknowledge the weight of our results in the light of the effects on trust due to surrounding and individual influences (Louwerse et al., 2015; Nass & Brave, 2007).

8.2.5 Significance & Final Implications of Results

Admitted repeatedly throughout this thesis is the difficulty to make explicit conclusions about what effect on trust are due to what system alterations. However, the benefit of conducting a multidimensional study is the ability to display several implications. In the previous sections, we have made an effort to derive certain results to particular design choices and more legitimately to evaluated system components. However, the main analysis from our results that should be made is that the identified system components do have a considerable effect on trusting Ava as a whole.

The null hypothesis in the F-test imply the variances are different and can't be rejected on a significance level of 5%, leading us to believe that the variances are the same to a 95% certainty in the compared correlations made in the different versions. Practically this supports that all participants have answered somewhat close to the mean. Implying that differences in trust acknowledged in the evaluations is an effect of the altered design choices and not sparse and varying answers. Moreover, the fact that we can't reject the null hypothesis in any of the conducted F-test strengthens the results obtained from the conducted t-tests. The p-values obtained from the parameter comparison (See table 7) between the neutral and opaque version is larger than 0,05, implying that we can reject the null hypothesis testing if the parameters are the same. However, when comparing the opaque and the biased version, there is a difference on the desired significance level in parameters Technical Experience and Expectations but we are still unable to reject the null hypothesis on parameter Overall Conversation Experience. Although, in the parameter comparison between neutral and biased version, all obtained p-values are smaller than 0,05, allowing us to reject the null hypothesis and claim that the compared means are different on a significance level of 5%. In addition to the qualitative data, we can with certainty conclude that the overall conversation experience is the highest in the neutral version as implied in the theoretical framework. Secondly, we distinguish larger effects on trust due to contextual bias than from the increase of opacity (see table B.3.1 in Appendix B).

Lastly, an important analysis to make is that our results support that evaluating TIA is intricate and that its effects on certain cognitive concepts are hard to isolate. An example is that the results display a clear correlation between perceptions of Ava's ability and the decrease in transparency. Furthermore, qualitative responses support the decrease of perceived ability due to contextual biases. Meaning that although alterations to the transparency and unbiasses of Ava meant to affect perceptions of integrity and benevolence respectively, the consequences in trust to some extent manifest as differences in experienced ability. Spillovers on all cognitive notions of TIA highlighted in this study, due to alterations of a system component targeted towards a particular concept, support that trust is best treated as an overall psychological attitude (Gareth et al., 2001). The break down of trust into different cognitive concepts is what allowed for a more systematic distinguishment of significant system components. However, in alignment with previous literature and our theoretical framework (McKnight et al. 1998), we agree upon that it is the accumulated merge of alterations to design choices in the distinguished system components that collectively contribute to the effects on trust in Ava.

8.3 Suggestions for Practitioners

In the purpose of this thesis, we state that isolating and quantifying a singular design choice and making a comparative analysis, was not a main objective. It is plausible to assume that such a study would have provided more crisply defined and clear quantitative results with higher reliability. However, the comprehensive and explorative study explained in this thesis holds the ambition to find several implications directly valuable for practitioners and to provide suggestions for further research in a growingly significant subject.

Although both beta testing and final quantitative testing contain several pressing concerns with regards to user experience, as shown in the results, the overall experience was still seemingly good. More specifically, despite a considerate amount of technical shortcomings in the beta test, 92% of the participant would consider using the product again in the near future. These results support the expectancy of automated service encounters in social areas where computers not yet have been able to compete with the trust associated with humanly monitored processes (Fearon, 2018). Furthermore, our qualitative results from participants indicate significant demand for automatisation in socially oriented interactions; "*Ava is easy to use. You do not have to book time or get anywhere. It went fast and smoothly. Convenient to write in Facebook chat!*", "*All the sharp tips for study choices are useful, especially when you know that the chatbot without precedent interprets exactly what I wrote and thus becomes more credible.*". These examples of user demand support the driving forces for that new technology will soon be used in more high-risk socially oriented service encounters (Friedman et al., 2007; Fitzpatrick et al., 2017). Furthermore, 92 % in the beta test answered "yes" or "maybe" on whether they could consider paying for such a service. Numbers that to some extent confirm the service market potential if further developed for the hosting company of this master thesis. Moreover, job coach Mattis Lu (2019) stated that after testing the neutral version that the conversation was very much like the conversations she has with employment seekers. Adding further interest to automatic study- and vocational guidance as it would allow for larger scale possibilities in terms of distribution and service value.

Lastly and most importantly, in the quantitative tests of Ava, 92% of the personas that tested the neutral version would recommend it to their friends in comparison to 84% and 68% in the other two versions. Furthermore, 96% answered that they would be equally comfortable or less comfortable answering truthfully to a human counsellor in the neutral version but only 88% and 80% in the opaque and biased version respectively. Findings that support that establishing knowledge and being able to make predictions about what design choices will be accepted and which will not and why will be imperative for practitioners (Juma, 2016; Leung et al., 2018). Our results show that beyond considering the ability of AS machine learning algorithms, service developers will have to recognise integrity and benevolence as critical concepts to engender consumers trust (Hemment, 2018). Furthermore, our findings emphasise that trust is what bridges the gap between human perception of characteristics and abilities of automation and the individual's intentions to use and rely on a service (Lee et al., 2004).

Provided our implications for practitioners a few suggestions are provided. Firstly, service providers must prioritise system performance and efficiency. One of the key determinants for establishing user trust in chatbots and therefore their likelihood in using them is achieving reliable and efficient provision. We have repeatedly through literature and empirical findings showed the effect on trust due to shortcomings in technical performance and system capability. A fundamental part of automation is providing a service which is consistently perceived as superior to a humanly supervised option. Something achieved through proper development. Secondly, we recommend in accordance with our results to be transparent about the chatbot features and limitations. It is critical that the chatbot clearly communicates its capabilities and limitations to the user through proper self-presentation and overall translucent responses. This aids in managing user expectations and experiences, providing a lower dissonance between beliefs and reality. Beyond transparency in the agent, we recommend leveraging user's trust in the service brand. By providing a legitimate reference to a brand, a higher priority of security and privacy are displayed, increasing trust in topics with high contextual risk as a result. Lastly, we want to highlight the mitigation of contextual biases. The subject of machine prejudice is a pressing concern and without precaution, it is likely that virtual assistants could host unwanted societal biases. However, we note that the effects of biases on user trust may have big differences depending on its significance in certain contexts. Although, we highlight trust decreasing tendencies due to contextual biases in this particular case, advocating careful precaution for practitioners in the future.

9. Conclusion

In the following section the conclusions of this thesis are listed without discussion or references to specific results. Rather they act as summary to the outcomes from the results, analysis and discussion. This chapter and study is ended by recommendations to future researchers.

9.1 List of Conclusions

- In this thesis, we have showcased that trust in an AI-automated service encounter can be studied using contextualisation and concept breakdown with affecting system components.

- Through the way of conduct described above, we strengthen the fact that Transparency, Unbiasses and System Performance are significant system components, that according to this studies definitions, affect the cognitive concepts of trust in automation.

- Although this investigation makes sparse claims about singular design choices effect on particular concepts of trust, we show that the increase of system opacity has consequences on trust due to the lack of system predictability. Secondly, we show that system transparency correlates to the reliability of responses provided by the developed chatbot. Moreover, we show that the lack of system transparency has spillovers on other cognitive concepts than integrity. More specifically, that it decreases trust due to lower perceived ability.

- With regards to mimicking contextual bias, we empirically support that bias has an overall significant negative effect of user trust. More specifically, we strengthen that the increase of the agent's prejudice and partisan opinions leads to a decrease in the comfortability to answer truthfully. Secondly, we support that by using language that reflects prejudiced regularities latent in the context, the willingness to speak to a chatbot rather than a human is lowered.

- The significance of our results is strengthened partly by the fact that we contain system performance as an affecting system component on trust. Secondly, due to that, surrounding influences and individual perception are accounted for.

- Lastly, the quantitative results show, with support from statistical significance tests, that of the evaluated system components contextual biases have the most considerable overall negative effect on trust. Although several concerns are highlighted as a consequence of the increase of opacity.

- Finally, we support the value and importance for both practitioners and future research of identifying and evaluating trust in automation.

9.2 Suggestions for Further Research

Provided the outcome and character of this investigation, a few suggestions for further research are provided.

Firstly, we recommend studies to be made with regards to delimited but related fields to this investigation. Namely, research that systematically makes an empiric comparison between trust in Human-Human-Interaction and Human-Machine-Interaction in a similar social context as the one in this case.

Secondly, additional findings on how trust is affected by providing the ability to chat through other communication mediums than text, such as images and voice would have great value in the particular service context.

Lastly, in order to reliably distinguish the effect of specific design choices targeted towards the different cognitive concept of trust, clinical and one-dimensional quantitative and in depth qualitative studies are required.

References

Published References

Abbass, H.A., Scholz, J. & Reid, D.J. (2018). Foundations of Trusted Autonomy Studies in Systems, Decision and Control. Springer International Publishing, Cham.

Adams, D., Bryant, J. & Webb, R,D. (2001). Trust in teams: Literature review. Technical Report Technical Report CR-2001-042, Report to Defense and Civil Institute of Environmental Medicine. Humansystems Inc.

Allen & Hamilton, Booz. (1980). New products management for the 1980s. New York, NJ: Author.

Allport, G. W. (1935). Attitudes. In A Handbook of Social Psychology (pp. 798-844). Worcester, MA, US: Clark University.

AlphaCE. (2019). Om oss, url: https://www.alphace.se/om-oss (2019-03-11).

Arrow, J. (1974). The Limits of Organization. Fels Lectures on Public Policy Analysis. W. W. Norton and Co. New York.

Asuero, A.G., Sayago, A. & González, A.G. (2006). The Correlation Coefficient: An Overview. Critical Reviews in Analytical Chemistry, vol. 36, no. 1, pp. 41-59.

Bainbridge, W.A. Hart, J. Kim, E.S. & Scassellati, B. (2008). The effect of presence on human-robot interaction. In RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication, pages 701–706. IEEE.

Barber, B. (1983). The logic and limits of trust . Rutgers University Press.

Bitner, M. J., & Wang, H. S. (2014). Service encounters in service marketing research. In Rust & Huang (Eds.), Handbook of Service Marketing Research (pp. 221-243). Cheltenham: Edward Elgar.

Brandtzaeg, B.P., Følstad, A., Haugstveit, M.I. & Skjuve, M. (2019). "Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction", Human Technology, vol. 15, no. 1, pp. 30-54.

Bryman, A. (2011). Samhällsvetenskapliga metoder. Malmö: Liber.

Bryman, A. & Bell, E. (2015). Business Research Methods, 4th edition, Oxford, Oxford University Press Inc

Byrnes, D.A. & Kiger, G. (1992). Common bonds: Anti - Bias Teaching in a Diverse Society. Association for Childhood Education International, Wheaton, MD.

Caliskan, A., Bryson, J.J., & Narayanan, A. (2017). “Semantics derived automatically from language corpora contain human-like biases.” Science 356.6334 : 183-186.

Castelvecchi, D. (2016). Can we open the black box of AI?. Nature 538, 7623, 20. Chien, S., Lewis, M., Hergeth, S., Semnani-Azad, Z. & Sycara, K. (2015). Crosscountry validation of a cultural scale in measuring trust in automation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 59, pages 686–690. SAGE Publications.

Chien, S., Semnani-Azad, Z., Lewis, M. & Sycara, K. (2014). Towards the development of an inter-cultural scale to measure trust in automation. In International Conference on Cross-Cultural Design, pages 35–46. Springer.

Chow S., Shao J. & Wang H. 2008. *Sample Size Calculations in Clinical Research*. 2nd Ed. Chapman & Hall/CRC Biostatistics Series.

Ciechanowski, L., Przegalinska, A., Magnuski, M. & Gloor, P. (2018). In the shades of the uncanny valley: An experimental study of human–chatbot interaction, Future Generation Computer Systems url: https://doi.org/10.1016/j.future.2018.01.055.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Connelly, B.L., Miller, T. & Devers, C.E. (2012). Under a cloud of suspicion: trust, distrust, and their interactive effect in interorganizational contracting. Strat. Manage. J. 33 (7), 820–833

Connelly, B.L., Crook, T.R., Combs, J.G., Ketchen, D.J. & Aguinis, D.J. (2015) Competence- and integrity-based trust in interorganizational relationships: Which matters more? J. Manage.

Dekker, S. & Hollnagel, E. (2004). Human factors and folk models. Cognition, Technology & Work 6 (2), 79–86.

Dekker, S.W.A. & Woods, D. (2002). Maba-maba or abracadabra? progress on human-automation co-ordination. Cognition, Technology & Work 4 (4), 240–244.

Denzin, N.K. (2010) Moments, mixed methods, and paradigm dialogs. Qualitative Inquiry.

Devitt, S.K. (2018). Trustworthiness of Autonomous Systems. Hussein A., Scholz, J. & Reid, J.D. (red). Foundations of Trusted Autonomy. Department of Information Sciences, University of Pittsburgh, Pittsburgh, PA, USA. Robotics Institute School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Dialogflow (2019a), Docs, url: https://dialogflow.com/docs (2019-03-21)

Dialogflow (2019b), Agents overview, url: https://dialogflow.com/docs/agents (2019-04-21)

Dialogflow (2019c), Intents overview, url: https://dialogflow.com/docs/intents (2019-04-21)

Dialogflow (2019d), Entities overview, url: https://dialogflow.com/docs/entities (2019-04-21)

Dialogflow (2019e), Contexts overview, url: https://dialogflow.com/docs/contexts (2019-04-21)

Dialogflow (2019f), Fulfillment overview, url: https://dialogflow.com/docs/fulfillment (2019-04-21)

Dialogflow (2019g), Training and Analytics overview, url: https://dialogflow.com/docs/training-analytics (2019-04-21)

Dialogflow (2019h), Dialogflow SDKs, url: https://dialogflow.com/docs/sdks (2019-04-21)

Dialogflow (2019i), Integrations, url: https://dialogflow.com/docs/integrations (2019-04-21)

Dixon, M., Freeman, K. & Toman, N. (2010). Stop trying to delight your customers. Harvard Business Review 88(7/8), 116-122.

Dolan, R.J. & Matthews, J.M. (1993). Maximizing the utility of customer product testing: Beta test design and management. The Journal of Product Innovation Management, vol. 10, no. 4, pp. 318-330.

Fearon & Maglio, P.P. (2018), Page 77, Handbook of Service Science, Volume II, Springer, S.l.

Fitzpatrick, KK., Darcy, A. & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Mental Health 4(2). DOI: 10.2196/mental.7785.

Følstad A., Nordheim C B. & Bjørkli C A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In: Bodrunova S. (eds) Internet Science. INSCI 2018. Lecture Notes in Computer Science, vol 11193.

Følstad, A. & Brandtzæg, P B. (2017a). Chatbots and the new world of HCI. interactions 24(4), 38- 42. DOI: 10.1145/3085558.

Følstad, A. & Brandtzaeg, P. B. (2017b). Why people use chatbots. In Proceedings of the International Conference on Internet Science, pp. 377-392, Cham, Switzerland: Springer. DOI: 10.1007/978-3-319-70284-1_30.

Friedman, B., Khan Jr, P.H. & Howe, D.C. (2007). Trust online. Communications of the ACM 43(12), 34-40. DOI: 10.1145/355112.355120.

Fryer, L. & Carpenter, R. (2006). Bots as language learning tools. Language Learning & Technology 10(3). DOI: 10125/44068.

Fuchs, D.J. (2018). The Dangers of Human-Like Bias in Machine-Learning Algorithms." Missouri S&T’s Peer to Peer 2, (1).

Gareth, R., George, J. & George, J M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. Academy of management review 23(3), 531–546.

Gaucher, D., Friese, J. & Kay, A.C. (2011). Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality. Journal of Personality and Social Psychology Vol. 101, No. 1, 109 –128.

Goffman, E. (2014), Jaget och maskerna: en studie i vardagslivets dramatik, 6. uppl. edn, Studentlitteratur, Stockholm.

Goillau, C., Boardman, K.M. & Jeannot, E. (2003). Guidelines for trust in future atm systems-measures. EUROCONTROL, the European Organization for the Safety of Air Navigation.

Google Firebase (2019a), Cloud Functions for Firebase, url: https://firebase.google.com/docs/functions (2019-04-01)

Google Firebase (2019b), Firebase Realtime Database, url: https://firebase.google.com/docs/database (2019-04-01)

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. CHI '06 Proceedings of the SIGCHI conference on Human Factors in computing systems. Montréal, Québec, Canada: CHI'2006.

Gupta, S. & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. Marketing Science, 25(6), 718-739.

Hancock, P.A., Billings, D.R., Schaefer, K.E. & Chen, J.YC. (2011) De Visser, J, Ewart. Parasuraman, Raja. A meta-analysis of factors affecting trust in human-robot interaction. Human (5):517–527, 44.

Haq, M. (2014). A comparative analysis of qualitative and quantitative research methods and a justification for use of mixed methods in social research. Annual PhD Conference, University of Bradford Business School of Management.

Hawley, K. (2012). Trust, distrust and commitment. Noûs 48 (1), 1–20 (2014). Wiley Periodicals, Inc.

Hemment, D. (2018). Trust in Invisible Agents. Leonardo, vol. 51, no. 5, pp. 450-450.

Hepenstal, S., Kodagoda, N., Zhang, L., Paudyal, P. & William W., B.L. (2019). Algorithmic Transparency of Conversational Agents. In Joint Proceedings of the ACM IUI 2019 Workshops, Los Angeles, USA, March 20, 2019 , 11 pages.

Heung-Yeung, S., Xiaodong, H. & Di, L. (2018). From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots. Microsoft Corporation.

Hieronymi, P. (2008). The reasons of trust. Australas. J. Philos. 86 (2), 213–236.

Hoffman, R.R., Johnson, M. & Bradshaw, M.J. (2013). Trust in Automation. Florida Institute for Human and Machine Cognition Al Underbrink, Sentar. IEEE Computer Society.

Holland J.L. (1997) Making Vocational Choices: A theory of vocational personalities and work environment. Odessa, FL: Psychological Assessment Resources Inc.

Holtzblatt, K. & Beyer, H. (2015). Contextual design: evolved, Morgan & Claypool. San Rafael, California.

Hosken, D.J., Buss, D.L. & Hodgson, D.J. (2018), "Beware the F test (or, how to compare variances)", Animal Behaviour, vol. 136, pp. 119-126.

Hubal, R., Fishbein, D., Sheppard, M., Paschall, M., Eldreth, D. & Hyde, C. (2008). How do varied populations interact with embodied conversational agents? Findings from inner-city adolescents and prisoners. Computers in Human Behavior, 24, 3, 1104-1138.

Jian, J., Bisantz, A.M. & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics 4 (1), 53–71.

JobTechdev (2019), API, url: https://jobtechdev.se/ (2019-03-20)

Jones, S., Murphy, F., Edwards, M. & James, J. (2008). Doing things differently: advantages and disadvantages of Web questionnaires, Nurse Researcher, 15, 4, pp. 15- 26

Juma, C. (2016). Innovation and its enemies: Why people resist new technologies. New York: Oxford University Press.

Keren. (2014). Trust and belief: a preemptive reasons account. Synth. Int. J. Epistemol. Methodol. Philos. Sci. 191 (12), 2593–2615.

Kim, P.H., Ferrin, D.L., Cooper, D. & Dirks, K.T. (2004). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence- versus integrity-based trust violations. J. Appl. Psychol. 89 (1), 104–118.

Kim, T.K. (2015), "T test as a parametric statistic", Korean Journal of Anesthesiology, vol. 68, no. 6, pp. 540-546.

Kretzschmar, K., Manzini, A., Pavarini, G., Tyroll, H. & Singh, I. (2019). Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. Biomedical Informatics Insights Volume 11: 1–9.

Kylén, J. (2004). Att få svar: Intervju, enkät, observation. Stockholm: Bonnier Utbildning.

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., De Keyser, A. (2017). "Service encounter 2.0": An investigation into the roles of technology, employees and customers. Journal of Business Research, 79, 238-246.

Lavrakas, P.J. (2008;2012). Encyclopedia of survey research methods, SAGE Publications, Thousand Oaks, Calif.

Lee, J.D. & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 46 (1):50–80.

Lerch, J., Prietula, J.M. & Kulik, T.C. (1997) The turing effect: The nature of trust in expert systems advice. In Expertise in context, pages 417–448. MIT Press.

Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A. & Bhogal, R.S. (1997). The persona effect: affective impact of animated pedagogical agents. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, pages 359–366. ACM.

Leung, E., Paolacci, G., & Puntoni, S. (2018) Man versus machine: Resisting automation in identity-based consumer behavior. Journal of Marketing Research.

Lewandowsky, S., Mundy, M. & Tan, G. (2000). The dynamics of trust: comparing humans to automation. Journal of Experimental Psychology: Applied 6 (2), 104.

Lewis, M., Sycara, K. & Walker, P. (2018). The Role of Trust in Human-Robot Interaction. Hussein A., Scholz, J. & Reid, J.D. (red). Foundations of Trusted Autonomy. Department of Information Sciences, University of Pittsburgh, Pittsburgh, PA, USA. Robotics Institute School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.

Louwerse, M., Graesser, A., Lu, S. & Mitchell, H. (2005). Social cues in animated conversational agents. Applied Cognitive Psychology, 19, 6, 683-704.

Luger, E. & Sellen, A. (2016). Like having a really bad PA: The gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16; pp. 5286–5297). New York, NY, USA: ACM.

Lundgren, A. (2012). ATT VÄLJA YRKE - faktorer som påverkar valet. Linnéuniversitet. Institutionen för pedagogik, psykologi och idrottsvetenskap.

Luo, Y. (2002). Contract, cooperation, and performance in international joint ventures. Strat. Manage. J. 23 (10), 903–919.

Lyons, J.B. & Stokes, C.K. (2011). Human–human reliance in the context of automation. Human Factors: The Journal of the Human Factors and Ergonomics Society.

Maaike, V.D., & Menno, D.J. (2003). Exploring two methods of usability testing: Concurrent versus retrospective think-aloud protocols. The Shape of Knowledge, 285-287.

Madsen, M. & Gregor, S. (2000). Measuring human-computer trust. In 11th australasian conference on information systems, volume 53, pages 6–8. Citeseer.

Maxwell, J.A. (2012) Qualitative research design: An interactive approach. Vol. 41 Sage publications, London.

Mayer, R., Davis, J.H. & Schoorman, F.D. (1995). An integrative model of organizational trust. Academy of management review 20 (3), 709–734.

McKnight, D. H., Cummings, L.L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. Academy of Management Review, 23, 473–490.

Mittelstadt, B.D. & Floridi, L. 2016, The ethics of biomedical big data, Springer, Switzerland.

Mone, G. (2016). The edge of the uncanny. Communications of the ACM, 59(9), 17–19. url: https://doi.com/10.1145/2967977.

Moray, N., Inagaki, T. & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. Journal of Experimental Psychology: Applied 6 (1), 44.

Muijs, D. (2010). Doing quantitative research in education with SPSS. Sage. London.

Muir, B. (1994). Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37 (11), 1905–1922.

Najam-us-Sahar. (2015). Impact of Personality Type on Job Productivity. J Hotel Bus Manage 2016, 5:1 DOI: 10.4172/2169-0286.1000119.

Nass, C. & Brave, S. (2007). Wired for Speech: How voice activates and advances the Human-Computer Relationship. The MIT Press, Cambridge, US.

Nave, C. & Camerer, M.M. (2015). Does oxytocin increase trust in humans? A critical review of research. Perspect. Psychol. Sci. 10 (6), 772–789.

Nielsen, J. & Pernice, K. (2010). Eyetracking web usability. California: Nielsen Norman Group.

Parasuraman, R., Raja Sheridan, R. & Wickens, C.D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making, 2 (2):140–160.

Pointon, Matthew (2017) Searching for a contextualised framework to inform testing methodology in the mobile arena. Doctoral thesis, Northumbria University.

Race, R. (2008), 'Literature review', in Given, LM (ed.), The sage encyclopedia of qualitative research methods, SAGE Publications, Inc., Thousand Oaks, CA, pp. 488-489

Reid, M., Hultink, E.J., Marion, T. & Barczak, G. (2016), "The impact of the frequency of usage of IT artifacts on predevelopment performance in the NPD process", Information & Management, vol. 53, no. 4, pp. 422-434.

Riley, V.A. (1994). Human use of automation PhD thesis, University of Minneapolis.

Robinette, P., Howard, A. & Wagner, A.R. (2015). Timing is key for robot trust repair. In International Conference on Social Robotics, pages 574–583. Springer.

Robinette, P., Li, W., Allen, R., Howard, A.M. & Wagner, A.R. (2016). Overtrust of robots in emergency evacuation scenarios. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 101–108. IEEE, Salem, M., Lakatos, G., Amirabdollahian, F. & Dautenhahn, K. (2015). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, pages 141–148. ACM.

Saunders, M., Lewis, P. & Thornhill, A. (2016). Research Methods For Business Students, n.p.: Harlow, Pearson Education

Simpson, E. (2013). Reasonable trust. Eur. J. Philos. 21 (3), 402–423

Singh, M.P. (1999). A Social Semantics for Agent Communication Languages. Department of Computer Science North Carolina State University.

Surprenant, C.F. & Solomon, M.R. (1987). Predictability and personalization in the service encounter. Journal of Marketing, 51(2), 86-96.

Tapus, A., Mataric, M. & Scassellati, B. (2007). Socially assistive robotics [grand challenges of robotics]. IEEE Robotics & Automation Magazine 14 (1), 35–42.

Tashakkori, A. & Teddlie, C. (2010) Sage handbook of mixed methods in social & behavioral research. Sage. London.

Van Mulken, S., André, E. & Müller, J. (1999). An empirical study on the trustworthiness of life-like interface agents. In Proceedings of the HCI International '99 (the 8th International Conference on Human-Computer Interaction).

Vinyals, O. & Le, Q. (2015). A neural conversational model. arXiv preprint arXiv:1506.05869.

Vygotskij, L.S. & Öberg Lindsten, K. (2001) Tänkande och språk, Daidalos, Göteborg.

Wurangian, N.C. (1993). Testing - Alpha, Beta. OCLC Systems & Services: International digital library perspectives, vol. 9, no. 3, pp. 40-42.

Xu, A., Liu, Z., Guo, Y., Sinha, V. & Akkiraju, R. (2017). A new chatbot for customer service on social media. In Proceedings of CHI' 17, pp. 3506-3510. New York, NY: ACM. DOI: 10.1145/3025453.3025496.

Zheng, J.F. & Jarvenpaa, S.L. (2019). Negative Consequences of Anthropomorphized Technology: A Bias-Threat-Illusion Model. Proceedings of the 52nd Hawaii International Conference on System Sciences.

Interviews

Broström, T. Study and vocational guidance counsellor at AlphaCE Coaching & Education, May 14, 2019, Uppsala, AlphaCE Coaching & Education, Personal Interview.

Lu, M. Job Coach at AlphaCE Coaching & Education, May 15, 2019, Uppsala, AlphaCE Coaching & Education, Personal Interview.

Appendix A

In depth software description

Below follows a functionality description of Dialogflow and Firebase. More in-depth information regarding the different services can be found on the websites https://dialogflow.com and https://firebase.google.com.

A.1 Dialogflow

In Dialogflow, agents act as the top Natural Language Understanding module that enables your app, product or service to understand input text or spoken words and translate them into different kinds of actions. The agent can contain one or several functionalities that are activated whenever a user says or writes something that triggers an underlying intent within the agent. Intents, that usually represents one conversation turn, define how the conversation within the agent work by predefining examples of what a user can say together with what the intent should give in return. The intent, when triggered, delivers this pre-defined response back to the user. The answer can be provided in the form of text, verbal acknowledgement or webhook response. As an example, one could create an agent that automatically orders pizza with different intents for logging size, topping, delivery address etc. and that will use webhook response in order to validate and place the order at the local pizza shop. The different intents would then be activated upon different words or sentences defined by training phrases (See figure A1.1)(Dialogflow, 2019b,c).

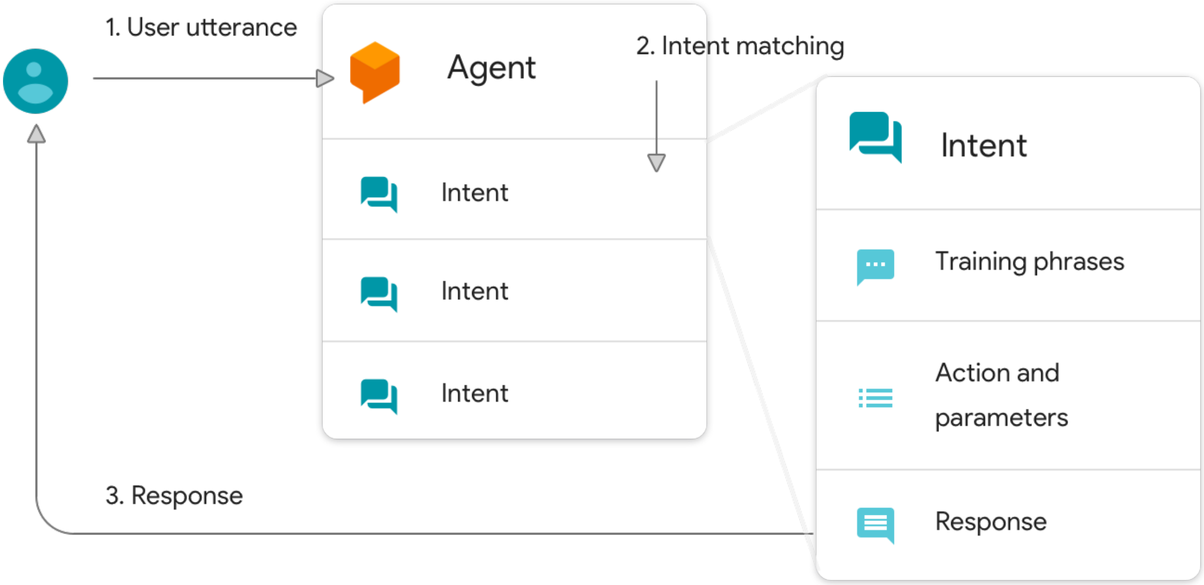


Figure A.1.1. The figure shows a user interaction with a agent (Dialogflow, 2019b).

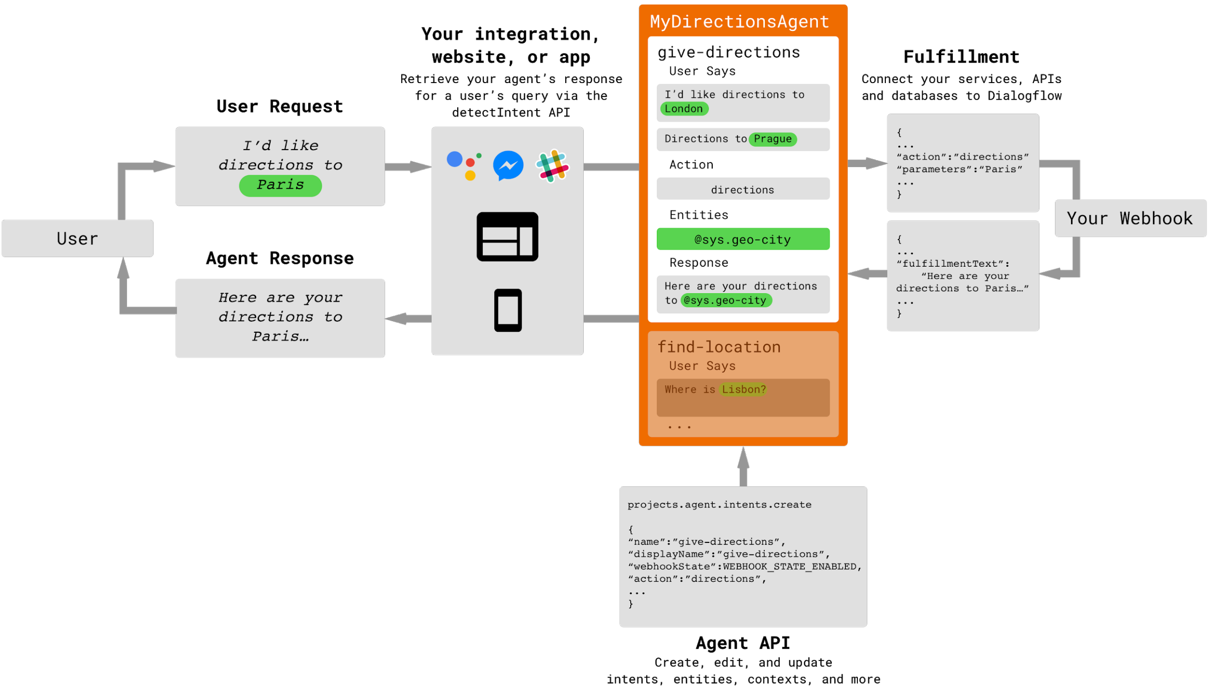
In order for Dialogflow to be able to distinguish and deduce valuable information from the conversation, it utilises built-in entities to pick out specific pieces of data from the natural language input. The intents enable your designed agent to understand the overall motive behind an input while entities enable the intent to pick out specific information mentioned by the user. These entities could be anything from colours to amounts and units. There are various built-in entities that could be used for identifying keywords that the developer doesn't need to define themselves. Supplementary further product specific entities can also be created in the cases when the agent needs to extract specific information from the conversation that isn't previously defined within Dialogflow. These new entities require the developer to define all possible entries possible that should activate the entity (Dialogflow, 2019d).

To be able to control the path of a conversation Dialogflow uses contexts. The active context represents the present state of the conversation and allows information to be saved and retrieved between intents. Can be activated or terminated at any exit or entry from an intent. This enables the developer to direct the conversation flow thus the intents can be set to only activate if a specific set of input och output contexts are present. If the example with the pizza store is applied again the context could define a certain path in your order, e.g. you need to choose a size before the toppings. In this specific case, the contexts also work as a memory, gathering your whole order before sending it to the local pizza shop. The contexts are saved within the Dialogflow conversation as JSON objects which holds all parameters and context lifespan (Dialogflow, 2019e).

To enable the Dialogflow agent to create dynamic responses on an intent to intent basis it uses fulfillment. The fulfillment is code deployed as a webhook, a self-created and hosted web server endpoint, that enables the agent to call different types of business logic. The fulfillment allows the developer to use information gathered by the NLP in the conversation to trigger back-end actions or dynamic responses. The fulfillment enables the developer to create or activate specially defined actions dependant on the user input, e.g generation of a dynamic response or placing a pizza order. Anytime the developer wants the agent to interact with some third party api or deliver a special response, the webhook and fulfillment need to be activated. When the user interacts with the agent in a way that activates an intent with enabled fulfillment, Dialogflow executes, with a JSON object an HTTP POST to the webhook carrying all data from the intent. The webhook then performs the stated tasks before responding back to Dialogflow with instructions for what should be done next, e.g return a special response to the user or create/update/delete an attached context (Dialogflow, 2019f). How the fulfilment interacts with all other presented parts of the Dialogflow framework can be visualised below (See figure A.1.2*)*.

Dialogflow provides a built-in tool for refinement of agents using its user interaction logs. By analysing the records, the developer can use the interaction data to train and refine the agent further, making it grasp a broader span of responses by automatically adding more training phrases to the intent. The developer can also leverage conversation data gathered on their own by uploading it to the training tool, thus making the agent more reliable. The logs can also be used to gain a broader understanding of user interactions that can be used to improve the agent's performance regarding design and conversation flow. Dialogflow also provides an analytics tool that might help developers to assess the overall performance of the agent. The tool lets the developer, in a tangible and interacting way, analyse potential bottlenecks within the system, places where the system faults in a more significant extent (Dialogflow, 2019g).

Dialogflow enables various one-click integration tools for the developed agent allowing it to be available on multiple platforms with minimal effort. The different integration options range from other NLP and assistant platforms such as Google Assistant, Amazon Alexa and Microsoft Cortana to several popular messaging platforms such as Facebook Messenger, Whats app, Slack and Twitter (Dialogflow, 2019i).


User
User Request
I'd like directions to Paris
Agent Response
Here are your directions to Paris…
Your integration, website, or app
Retrieve your agent's response for a user's query via the detectIntent API
MyDirectionsAgent
give-directions
User Says
I'd like directions to London
Directions to Prague
Action
directions
Entities
@sys.geo-city
Response
Here are your directions to @sys.geo-city
find-location
User Says
Where is Lisbon?
. . .
Fulfillment
Connect your services, APIs and databases to Dialogflow
Your Webhook
Agent API
Create, edit, and update intents, entities, contexts, and more

```
{
...
"action":"directions"
"parameters":"Paris"
...
}
```

```
{
...
"fulfillmentText":
    "Here are your
directions to Paris…"
...
}
```

```
projects.agent.intents.create

{
"name":"give-directions",
"displayName":"give-directions",
"webhookState":WEBHOOK_STATE_ENABLED,
"action":"directions",
...
}
```

Figure A.1.2. The figure shows the Software Development Kit (SDK) which exhibits the development tools that are available for Dialogflow (Dialogflow, 2019h).

A.2 Google Firebase

The created webhook fulfillment code, written in the inline editor in Dialogflow gets automatically deployed and stored at the Cloud Functions for Firebase. The Cloud Functions for Firebase enables the product to run the backend code in response to events triggered by HTTP requests. All code is saved within the Google cloud and operates in a constantly managed environment. The JavaScript code is deployed on the servers and Firebase automatically scale the computing resources to match the usage patterns of every specific product, making it cost effective. After deployment, the functions can be reached and executed using a simple HTTP request which the servers constantly listen for, returning the requested values. If the workload decreases or increases rapidly, Firebase will automatically scale the number of server instances needed to run the application making it highly applicable for large differences in user involvement (Google Firebase, 2019a).

The Firebase Realtime Database is a NoSQL cloud database which enables realtime data storage and synchronisation for every client. The data is stored in JSON format and the same realtime database is shared with every client regardless of what kind of platform the application is built on. The Firebase Realime Database uses synchronisation instead of HTTP requests which enables the devices to be automatically updated within milliseconds whenever there is a change in the data. The Firebase Realtime Database enables the developer to build and manage collaborative applications. Because the data is persisted locally, the application continues to work when offline, synchronizing the local data changes and merging any conflicts when regaining connection again. To manage the database security rules, the Firebase Realtime Database provides an expression-based language called Security Rules. The Security Rules define how the data stored within Firebase should be structured and how the data may be read and written. Firebase also provides an authentication service, and together with the Security Rules, the developer can define who has access to the data and what the user can do with it (Google Firebase, 2019b).

Appendix B

B Evaluation Forms

Below the evaluation forms for the two different tests are presented together with accompanying instructions that were sent to the test personas. In the last subchapter one can find notes from the overwatched think aloud sessions.

B.1 Beta Test

Text sent on Facebook Messenger 6/3 -19:

Grattis! Du har blivit utvald för att delta i ett beta-test av den tjänst som jag och Fred Isaksson har utvecklat i samband med vårt examensarbete.

Testet går ut på att du kommer få chatta med vår virtuella yrkesvägledare, Ava, här på facebook. Vid konversationens slut så får du fylla i en utvärdering. Det spelar ingen roll om du själv inte har ett behov av yrkesvägledning eller är arbetssökande för tillfället, det viktiga är din generella upplevelse av tjänsten.

Dina svar på Ava's frågor är konfidentiella och ingen personspecifik information kommer att sparas. Utvärderingen är anonym. Testet beräknas ta maximalt 30 minuter, ink utvärdering, men pga av säkerhetsmässiga skäl så kommer testet enbart gå att göra mellan 18-20 imorgon torsdag den 7/3. Om du har möjlighet att delta under den angivna tiden, svara "Jag deltar" på följande meddelande. Om du inte har möjlighet att göra testet under denna tid så kan du svara på det här meddelandet med en tid som passar dig så försöker vi lösa en annan tid.

Vid 18 imorgon, 7/3, kommer du få en länk till chatten med tillhörande instruktioner på hur du påbörjar och slutför testet.

STORT TACK FÖR DIN MEDVERKAN!

Allt gott.

//Fred & Joakim

Text sent on Facebook Messenger 6/3 -19:

Hej! Vi har valt att hålla testet öppet lite längre och Ava kommer att vara tillgänglig att prova mellan nu-22 idag. Svara "klar" på detta meddelande när du är färdig med testet. Var vänlig att INTE likea eller rekommendera sidan.

Länk till facebook-sidan:

Hyperlink to Facebook Messenger

1) Tryck på länken till sidan.

2) Tryck på knappen "Skicka meddelande"

3) Tryck på "kom igång" i chattfönstret som poppat up.

4) Vänta på att Ava välkomnar dig. Det kan dröja några sekunder.

5) Öppna konversationen i messenger för bästa upplevelse.

6) Läs välkomstmeddelandet noga.

7) Svara på Ava's frågor och njut av konversationen!

8)Konversationen är färdig när Ava svarar att konversationen är slut.

9) Gå tillbaka till denna chat och följ länken till utvärderingen.

10)Fyll i utvärderingen.

11) Färdig.

12)TACK!

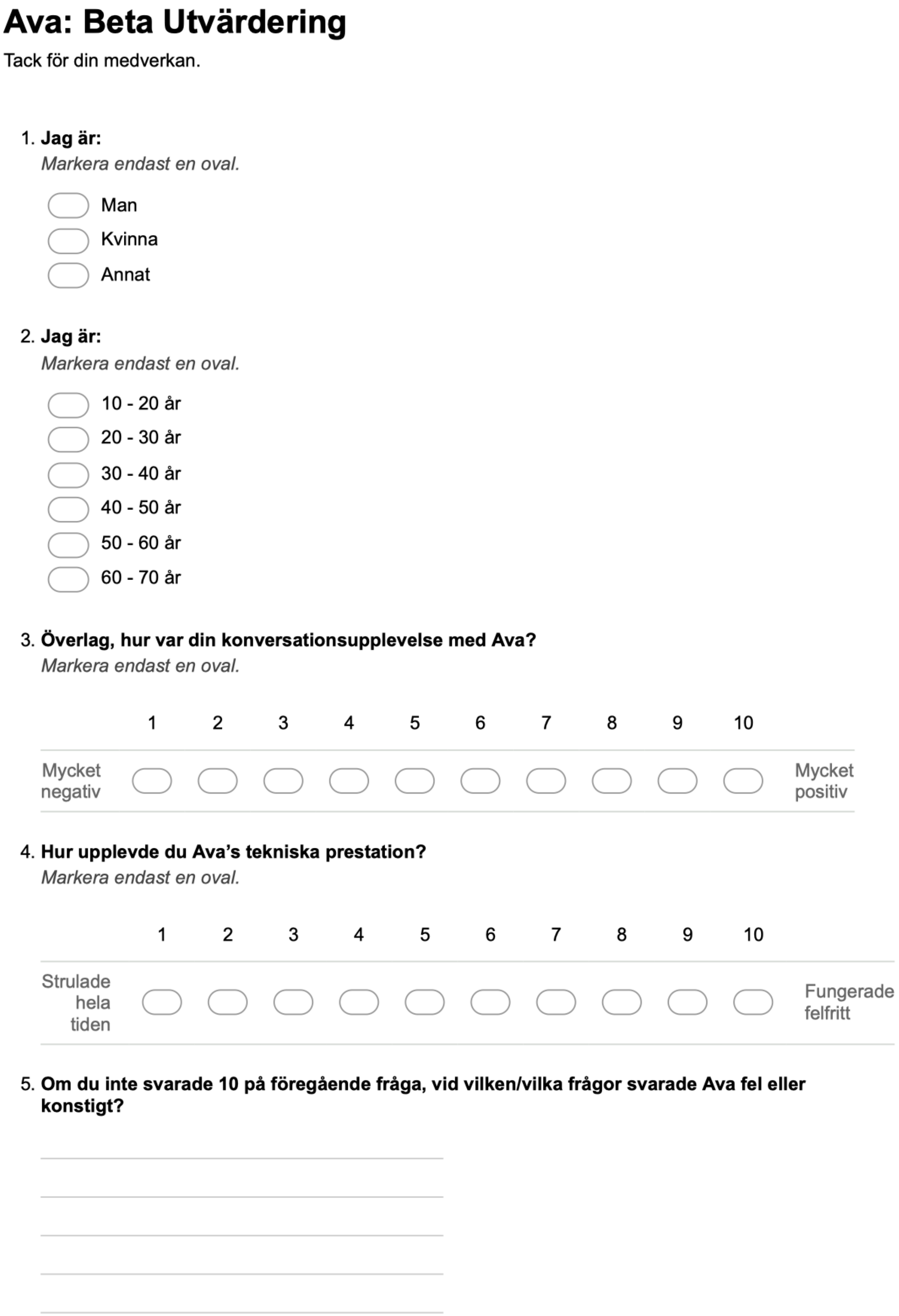
Om något inte skulle fungera eller Ava låser sig helt så kan du prova att starta om konversationen genom att trycka på "ta bort konversationen" och sen börja om på steg 1). Om du inte kommer till konversationens slut efter några försök så avsluta testet och skriv vart du fastna i utvärderingen.

Länk till utvärdering:

Hyperlink to Evaluation form

Ha så roligt!

Evaluation form for beta testing:



Ava: Beta Utvärdering

Tack för din medverkan.

1. **Jag är:**

   Markera endast en oval.

   ○ Man
   ○ Kvinna
   ○ Annat

2. **Jag är:**

   Markera endast en oval.

   ○ 10 - 20 år
   ○ 20 - 30 år
   ○ 30 - 40 år
   ○ 40 - 50 år
   ○ 50 - 60 år
   ○ 60 - 70 år

3. **Överlag, hur var din konversationsupplevelse med Ava?**

   Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Mycket negativ	○	○	○	○	○	○	○	○	○	○	Mycket positiv

4. **Hur upplevde du Ava’s tekniska prestation?**

   Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Strulade hela tiden	○	○	○	○	○	○	○	○	○	○	Fungerade felfritt

5. **Om du inte svarade 10 på föregående fråga, vid vilken/vilka frågor svarade Ava fel eller konstigt?**

6. **Om du hade några, hur väl mötte Ava dina förväntningar av en virtuell yrkesvägledare i form av chatbot?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Långt under mina förväntingar	○	○	○	○	○	○	○	○	○	○	Långt över mina förväntningar

7. **Vilka delar av konversationen tyckte du mest om? Vilka delar kändes mest värdefulla och relevanta?**

8. **Vilka delar av konversationen tyckte du minst om? Vilka delar kändes onödigt komplicerade eller irrelevanta?**

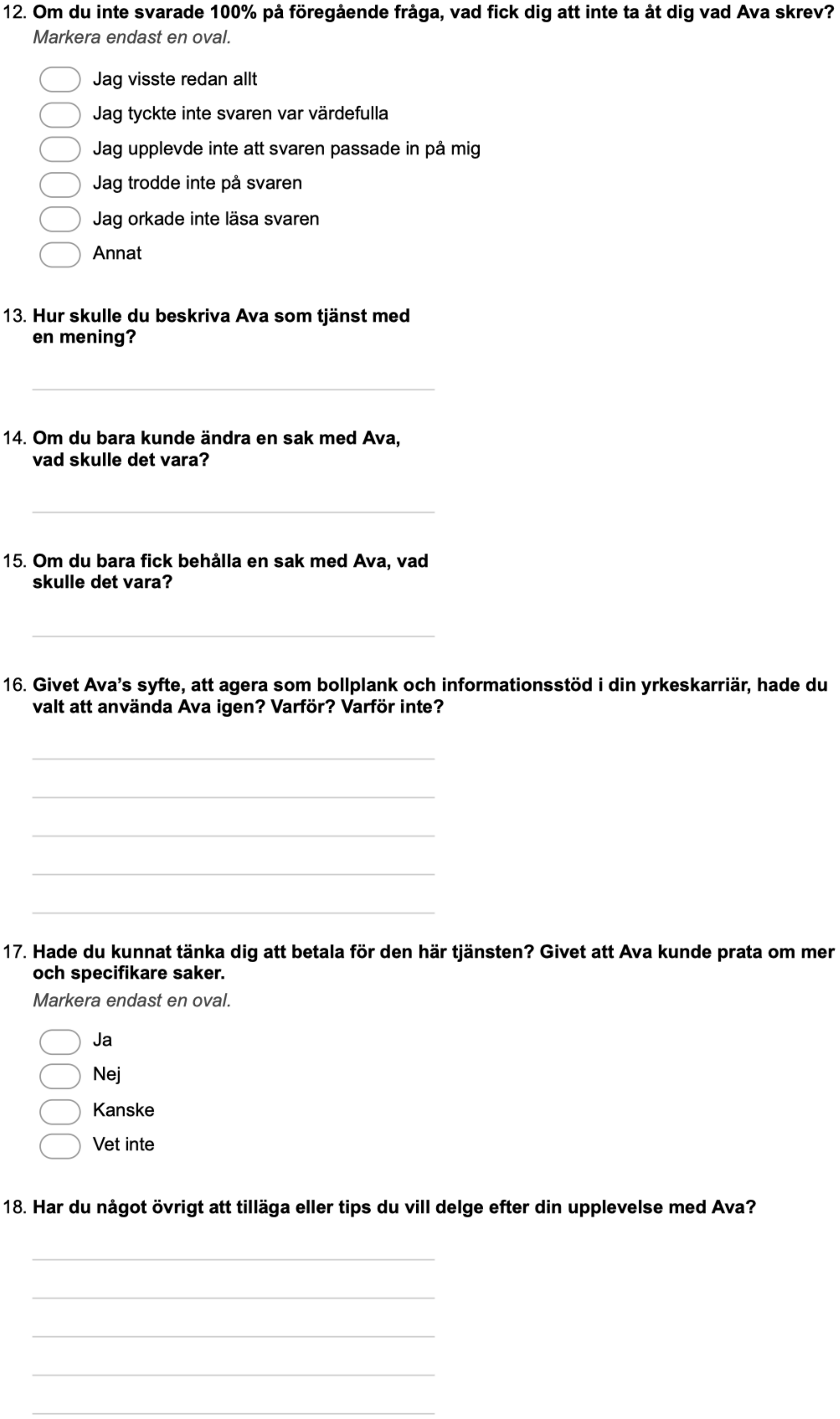
9. **Hur pålitligt upplevde du innehållet i frågorna och svaren som Ava skrev?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Mycket opålitligt	○	○	○	○	○	○	○	○	○	○	Mycket pålitligt

10. **Om du inte svarade 10 på föregående fråga, vad var det som du inte upplevde som pålitligt?**

11. **Hur mycket tog du åt dig av Ava's tips och rekommendationer?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	100%



12. **Om du inte svarade 100% på föregående fråga, vad fick dig att inte ta åt dig vad Ava skrev?**
Markera endast en oval.

○ Jag visste redan allt
○ Jag tyckte inte svaren var värdefulla
○ Jag upplevde inte att svaren passade in på mig
○ Jag trodde inte på svaren
○ Jag orkade inte läsa svaren
○ Annat

13. **Hur skulle du beskriva Ava som tjänst med en mening?**

14. **Om du bara kunde ändra en sak med Ava, vad skulle det vara?**

15. **Om du bara fick behålla en sak med Ava, vad skulle det vara?**

16. **Givet Ava's syfte, att agera som bollplank och informationsstöd i din yrkeskarriär, hade du valt att använda Ava igen? Varför? Varför inte?**

17. **Hade du kunnat tänka dig att betala för den här tjänsten? Givet att Ava kunde prata om mer och specifikare saker.**
Markera endast en oval.

○ Ja
○ Nej
○ Kanske
○ Vet inte

18. **Har du något övrigt att tilläga eller tips du vill delge efter din upplevelse med Ava?**

B.2 Quantitative Testing for Trust

Text sent to students 12/4 -19:

Hej!

Här kommer ett bra tips i ansökningstider. Chatta med Ava, en virtuell studie- och yrkesvägledare!

Fred Isaksson och Joakim Eklund som skriver examensarbete på civilingenjörsutbildningen i system i teknik och samhälle på Uppsala Universitet behöver din hjälp att testa Ava. Det enkelt, roligt och värdefullt både för dem och för dig.

Såhär gör du:

Använd helst en dator, men din telefon går även bra.

Tryck på den här länken som tar dig till Ava's Facebook sida:

Hyperlink to Facebook chat

Tryck på knappen "Skicka meddelande"

Tryck på "kom igång" i chattfönstret som poppat upp eller skriv "hej".

Öppna konversationen I "Messenger", ett större chatfönster för bästa upplevelse.

Vänta på att Ava välkomnar dig. Det kan dröja några sekunder.

Läs välkomstmeddelandena noga. Viktigt!

Njut av konversationen!

SUPERVIKTIGT. Fyll i utvärdering på länken nedan. Kommer även i slutet av konversationen!

Hyperlink to Evaluation form

Tack för din medverkan!

Evaluation form for quantitative testing:

Ava - Utvärdering

Tack för att du tar dig tid och svarar på denna utvärdering.

Genom att svara på följande formulär godkänner jag deltagandet i ett test av den virtuella studie- och yrkesvägledaren Ava. Härmed intygas också att mitt deltagande får användas i Fred Isaksson och Joakim Eklunds examensarbete på civilingenjörsutbildningen i system i teknik och samhälle på Uppsala Universitet. Detta är under förutsättningen att mina personliga svar anonymiseras och inte kan härledas tillbaka till mig som enskild person. Vidare förtydligas min rätt att närsomhelst avbryta testet och att all sparad information raderas.

Vid andra frågor vänligen kontakta: avachat@alphace.se

1. **Jag är:**

   Markera endast en oval.

   ○ Man
   ○ Kvinna
   ○ Annat

2. **Jag är:**

   Markera endast en oval.

   ○ 14 - 16 år
   ○ 17 - 19 år
   ○ 20 - 22 år
   ○ 23 - 25 år
   ○ 26 eller äldre

3. **Jag har:**

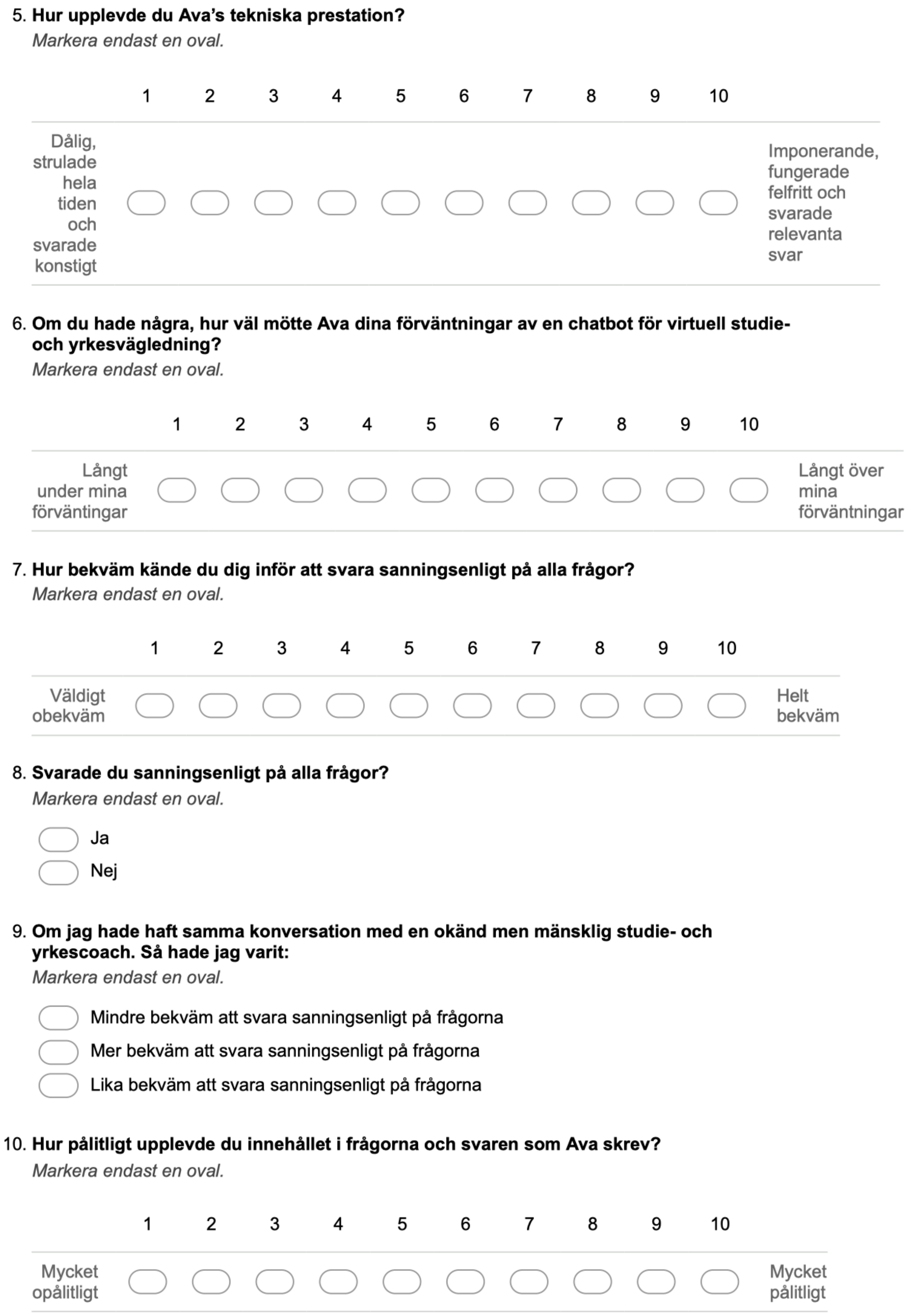
   Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Aldrig prövat en chatbot förut	○	○	○	○	○	○	○	○	○	○	Mycket erfarenhet av chatbottar

4. **Min kunskap inom programmering av Artificiell Intelligens är:**

   Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Obefintlig	○	○	○	○	○	○	○	○	○	○	Stor



5. **Hur upplevde du Ava's tekniska prestation?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Dålig, strulade hela tiden och svarade konstigt	○	○	○	○	○	○	○	○	○	○	Imponerande, fungerade felfritt och svarade relevanta svar

6. **Om du hade några, hur väl mötte Ava dina förväntningar av en chatbot för virtuell studie- och yrkesvägledning?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Långt under mina förväntingar	○	○	○	○	○	○	○	○	○	○	Långt över mina förväntningar

7. **Hur bekväm kände du dig inför att svara sanningsenligt på alla frågor?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Väldigt obekväm	○	○	○	○	○	○	○	○	○	○	Helt bekväm

8. **Svarade du sanningsenligt på alla frågor?**
Markera endast en oval.

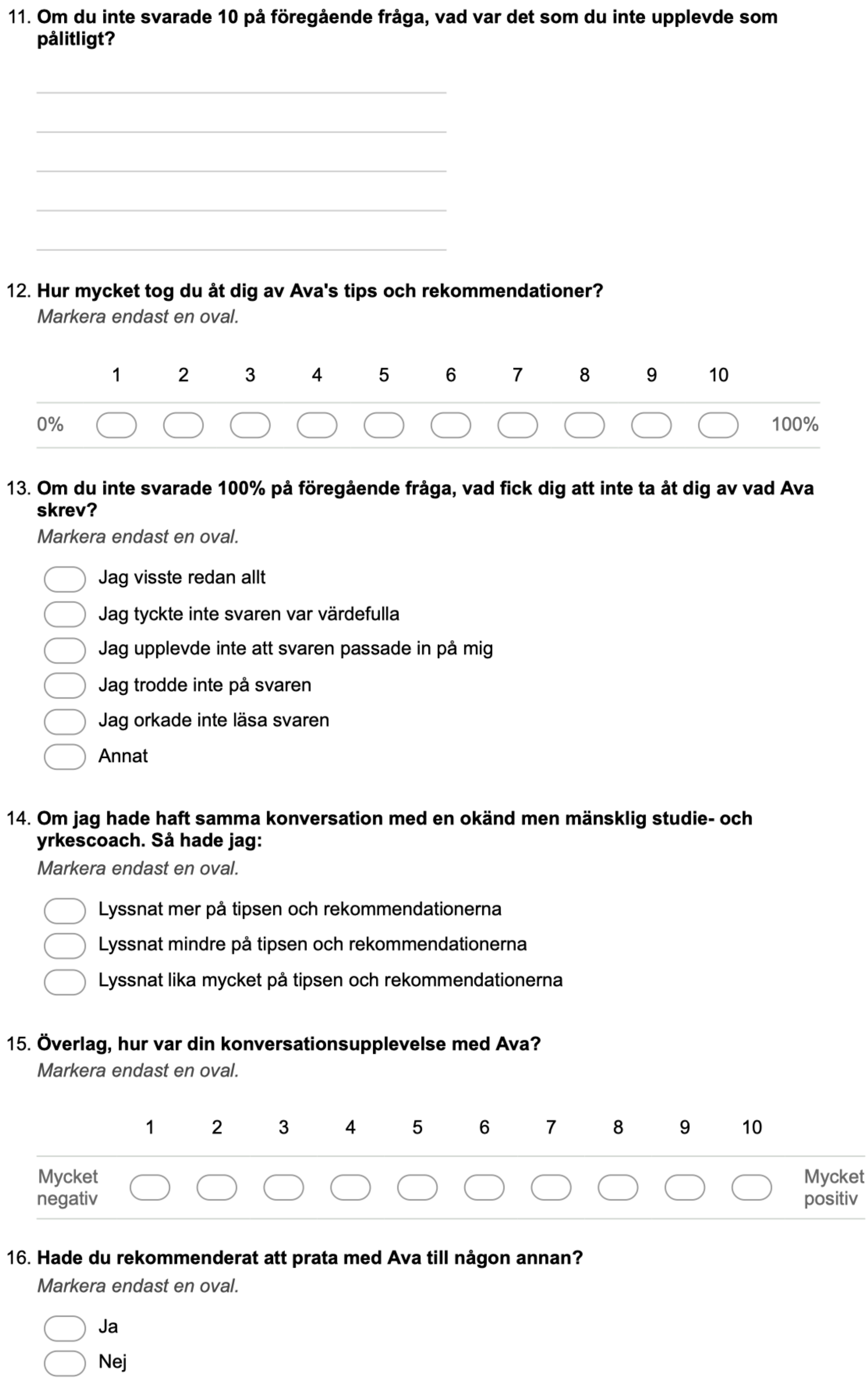
○ Ja
○ Nej

9. **Om jag hade haft samma konversation med en okänd men mänsklig studie- och yrkescoach. Så hade jag varit:**
Markera endast en oval.

○ Mindre bekväm att svara sanningsenligt på frågorna
○ Mer bekväm att svara sanningsenligt på frågorna
○ Lika bekväm att svara sanningsenligt på frågorna

10. **Hur pålitligt upplevde du innehållet i frågorna och svaren som Ava skrev?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Mycket opålitligt	○	○	○	○	○	○	○	○	○	○	Mycket pålitligt



11. **Om du inte svarade 10 på föregående fråga, vad var det som du inte upplevde som pålitligt?**

____

____

____

____

____

12. **Hur mycket tog du åt dig av Ava's tips och rekommendationer?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	100%

13. **Om du inte svarade 100% på föregående fråga, vad fick dig att inte ta åt dig av vad Ava skrev?**
Markera endast en oval.

○ Jag visste redan allt
○ Jag tyckte inte svaren var värdefulla
○ Jag upplevde inte att svaren passade in på mig
○ Jag trodde inte på svaren
○ Jag orkade inte läsa svaren
○ Annat

14. **Om jag hade haft samma konversation med en okänd men mänsklig studie- och yrkescoach. Så hade jag:**
Markera endast en oval.

○ Lyssnat mer på tipsen och rekommendationerna
○ Lyssnat mindre på tipsen och rekommendationerna
○ Lyssnat lika mycket på tipsen och rekommendationerna

15. **Överlag, hur var din konversationsupplevelse med Ava?**
Markera endast en oval.

	1	2	3	4	5	6	7	8	9	10	
Mycket negativ	○	○	○	○	○	○	○	○	○	○	Mycket positiv

16. **Hade du rekommenderat att prata med Ava till någon annan?**
Markera endast en oval.

○ Ja
○ Nej

17. **Givet Ava's syfte, att agera som bollplank och informationsstöd i din studie- och yrkeskarriär, hade du valt att använda Ava igen? Varför? Varför inte?**

18. **Har du något övrigt att tilläga eller tips du vill delge efter din upplevelse med Ava?**

19. **Fyll i din mailadress för att vara med i utlottningen av biobiljetter**

Tillhandahålls av

Google Forms

B.3 Tables & Graphs From Quantitative Testing

In the following appendix chapter, graphs and tables produced from the results are presented.

Material From Beta Test

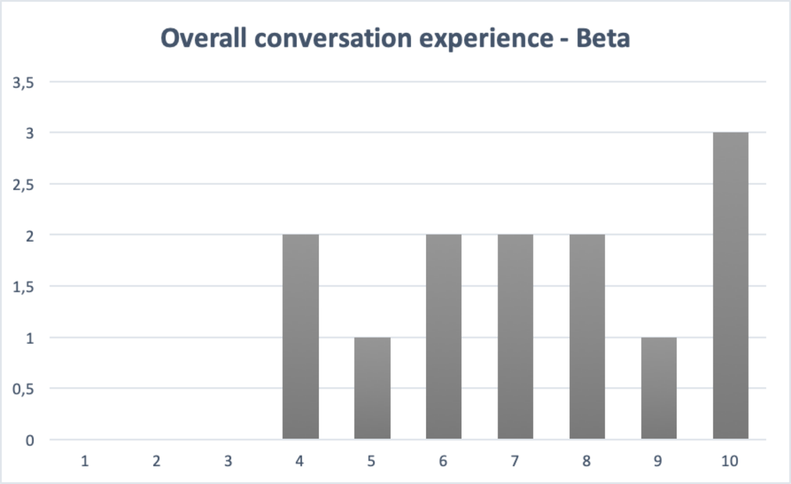


Figure B.3.1. Overall conversation experience from beta test

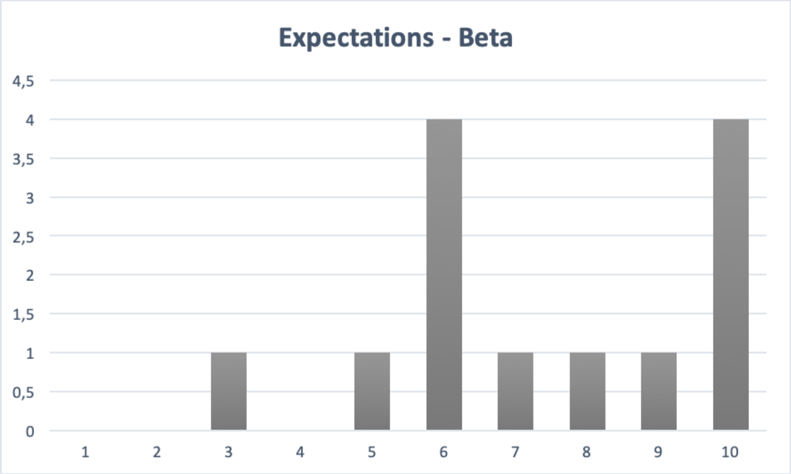


Figure B.3.2. How well Ava met expectations of a vocational guidance chatbot from beta test

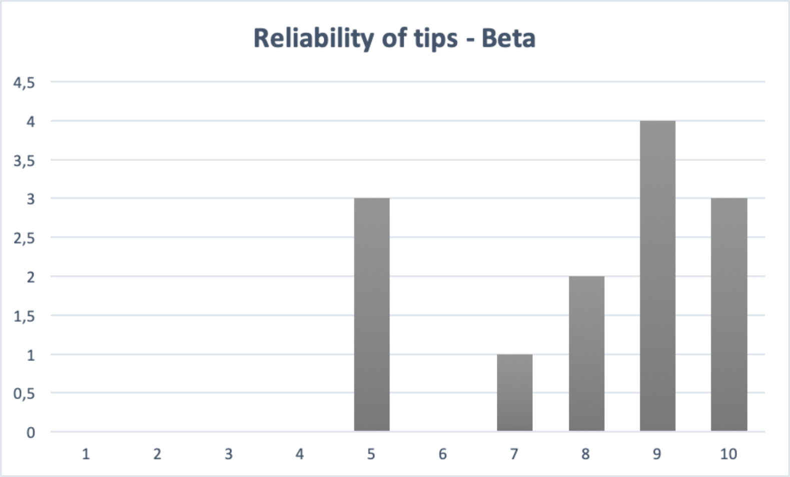


Figure B.3.3. Reliability of the tips presented in the conversation in beta testing

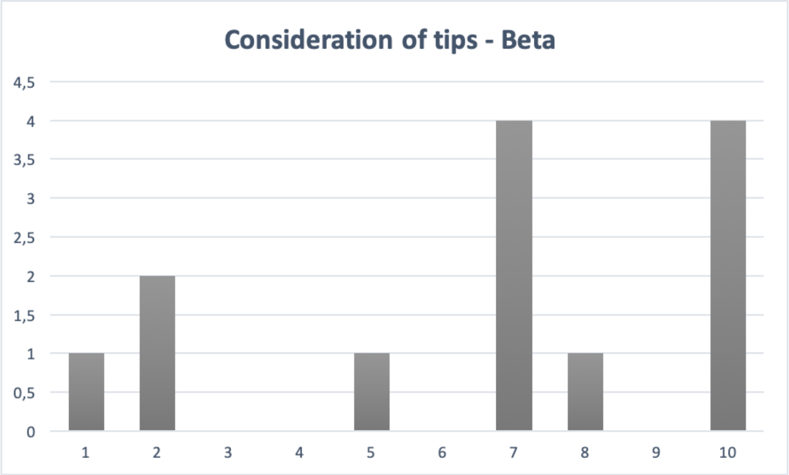


Figure B.3.4. Consideration of tips provided in beta test

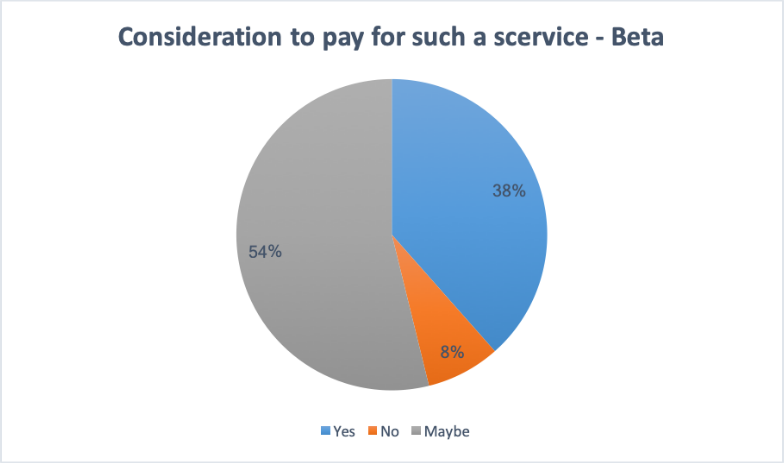


Figure B.3.5. The willingness to pay for such a product from beta test

Material Neutral Version

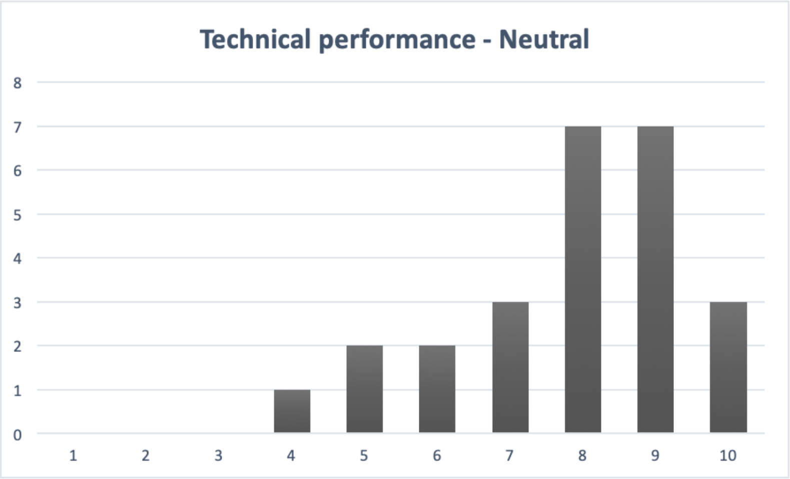


Figure B.3.6. Technical performance in Neutral version

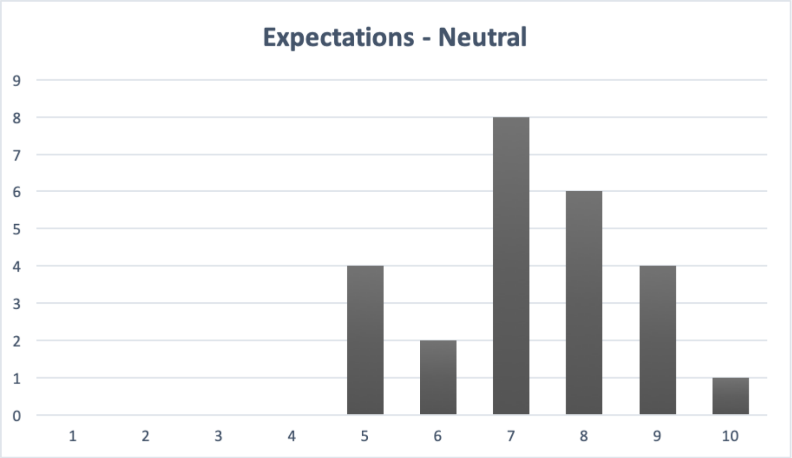


Figure B.3.7. How well Ava met expectations of a vocational guidance chatbot from Neutral version

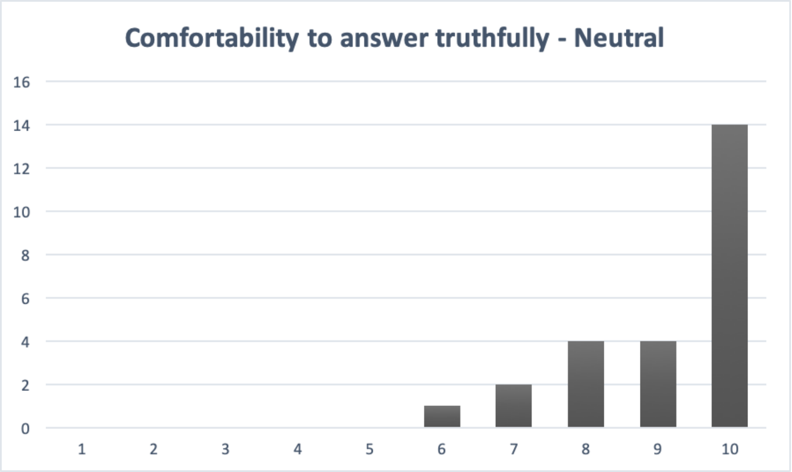


Figure B.3.8. The level of comfortability to answer questions truthfully in neutral version

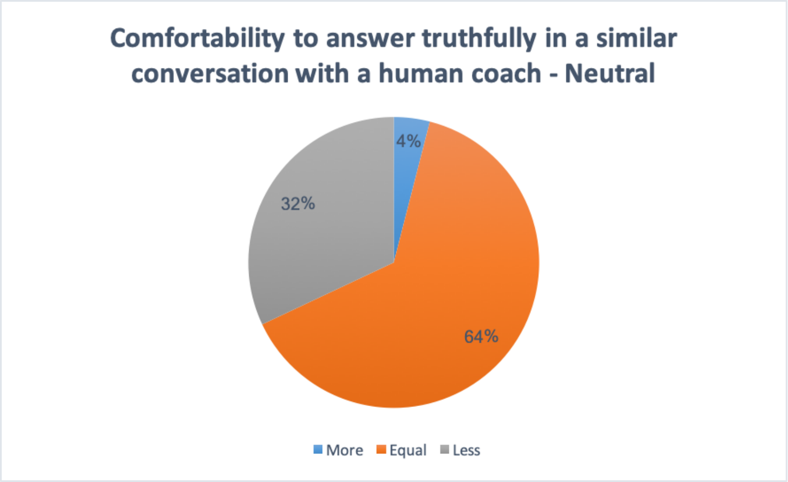


Figure B.3.9. The comfortability to answer truthfully in a human to human conversation in neutral version

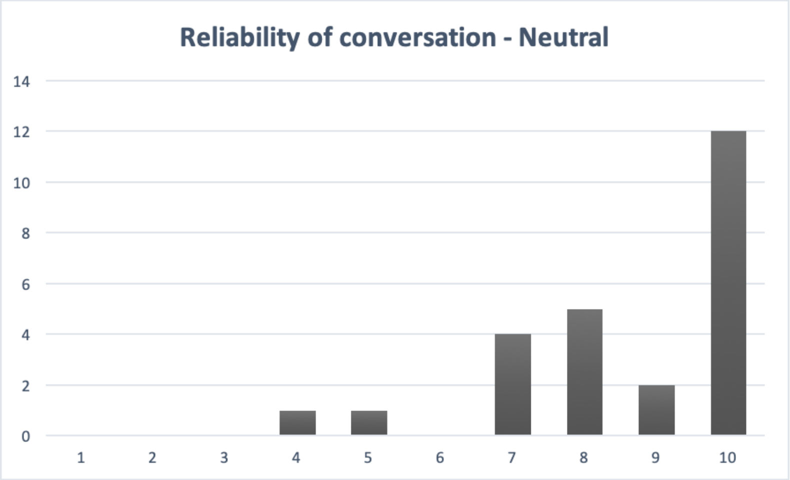


Figure B.3.10. The reliability of the conversation information and tips neutral version

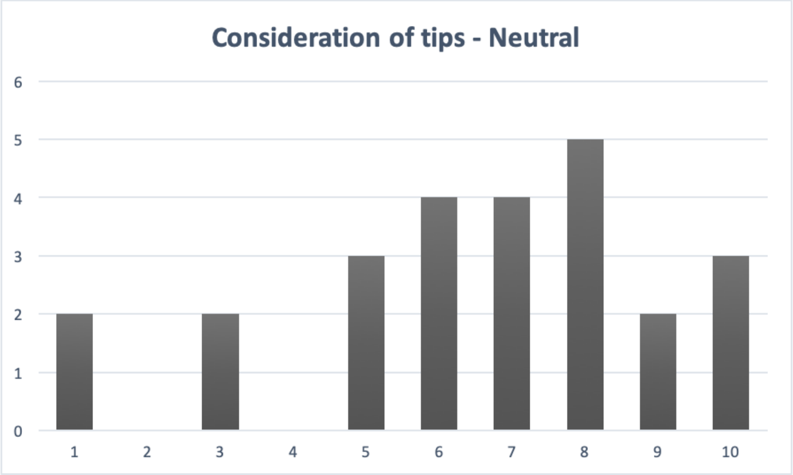


Figure B.3.11. The level of consideration to the tips provided neutral version

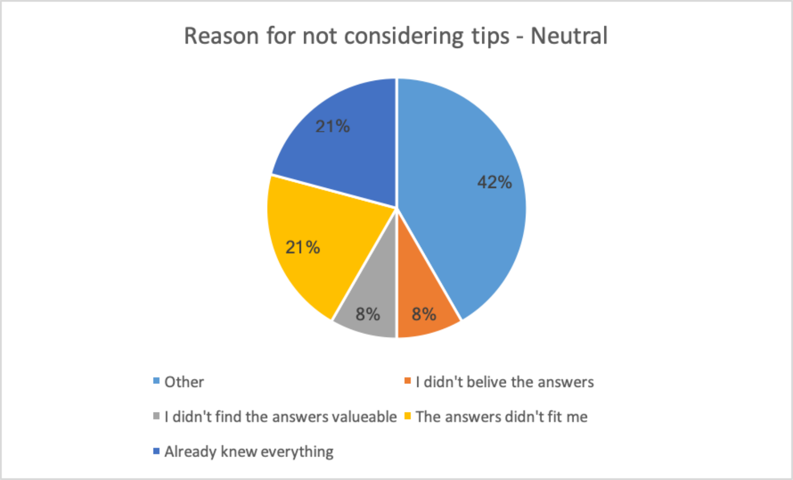


Figure B.3.12. Reason for not considering all the tips provided neutral version

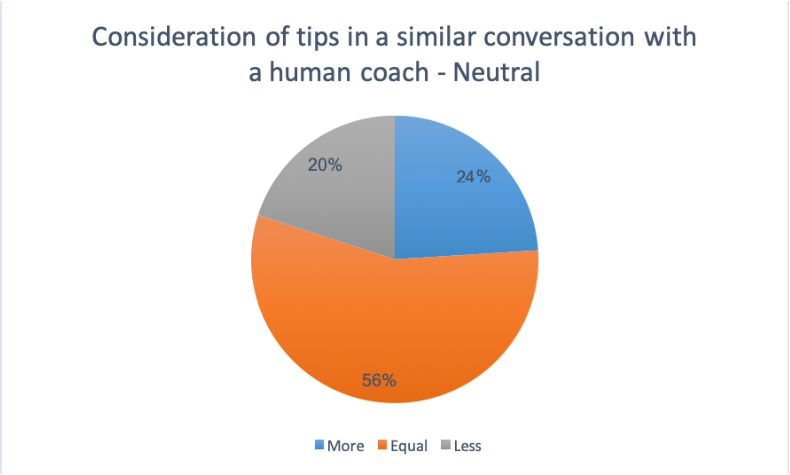


Figure B.3.13. Consideration of tips in similar Human to Human conversation neutral version

Material Opaque Version

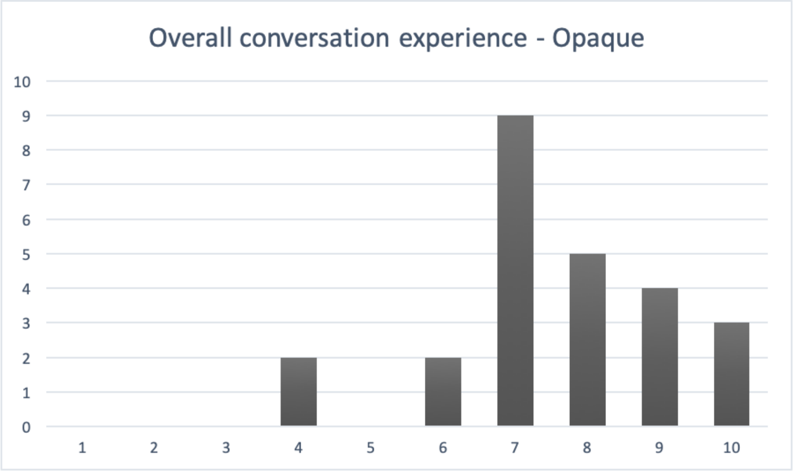


Figure B.3.14. Overall conversation experience opaque version

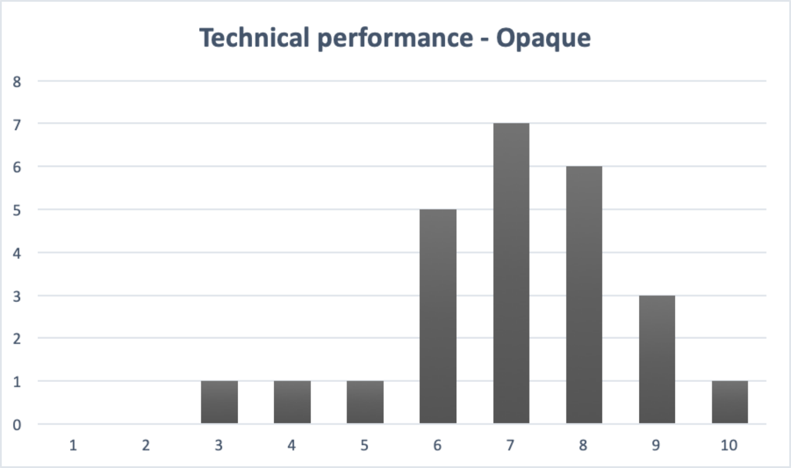


Figure B.3.15. Perceived technical performance opaque version

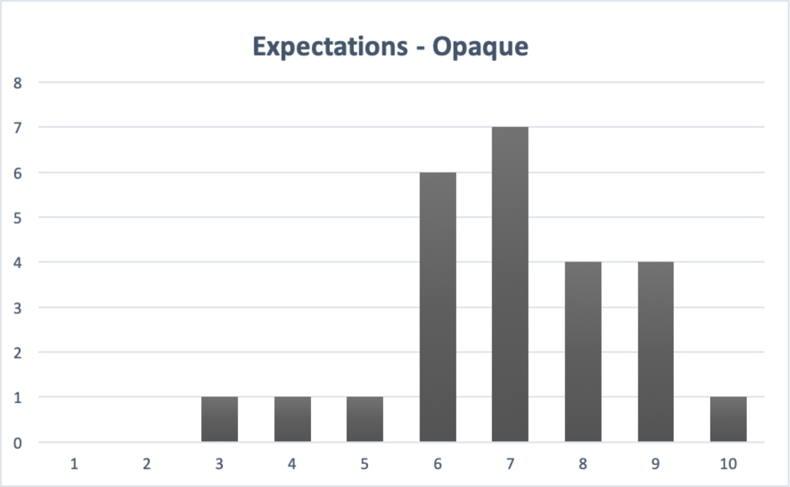


Figure B.3.16. How well Ava met expectations of a vocational guidance chatbot from opaque version

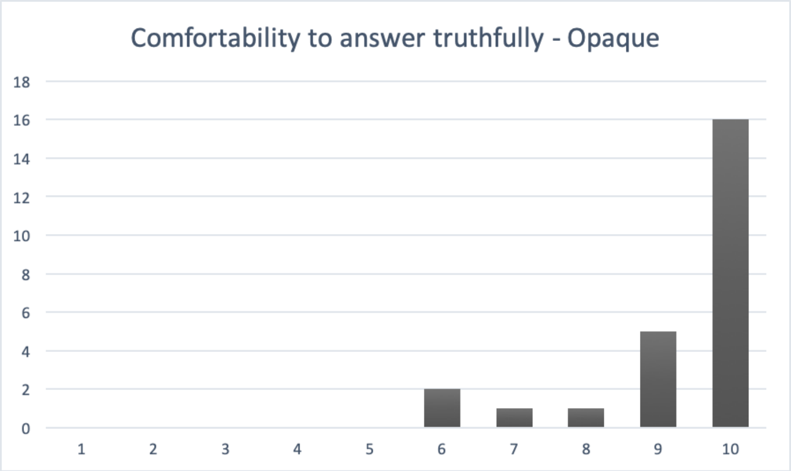


Figure B.3.17. Comfortability to answer the questions truthfully in opaque version

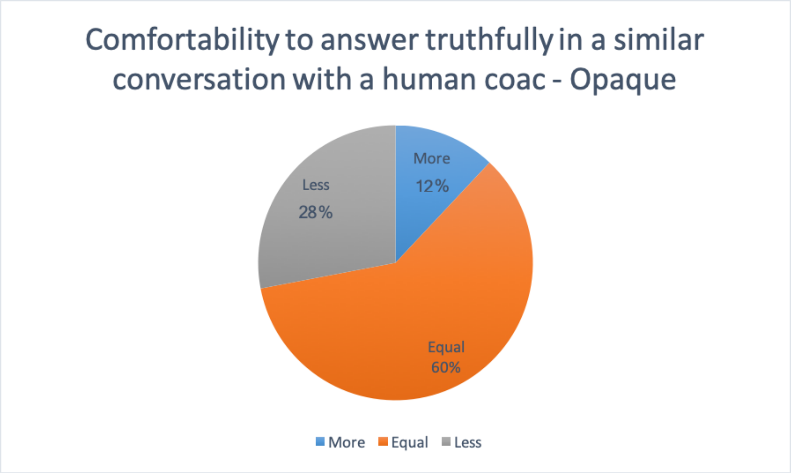


Figure B.3.18. The comfortability to answer truthfully in a human to human conversation in opaque version

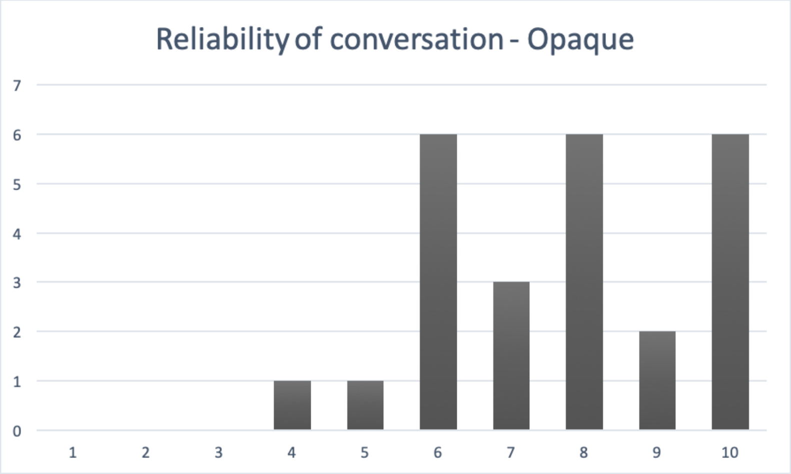


Figure B.3.19. Perceived conversation reliability opaque version

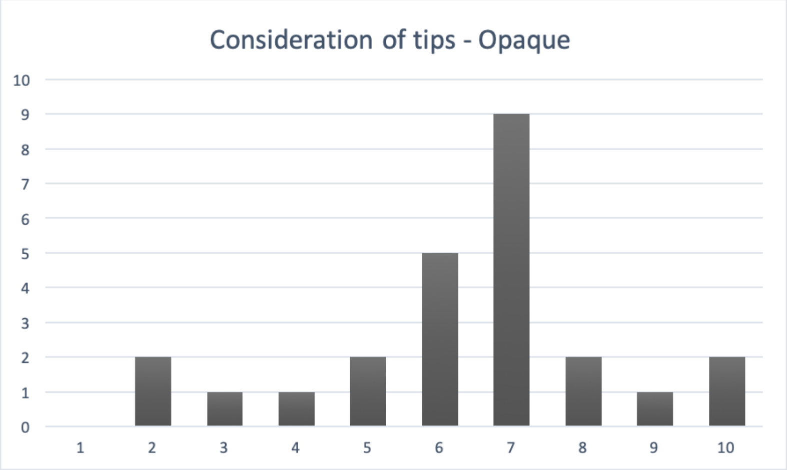


Figure B.3.19. Consideration of tips in opaque version

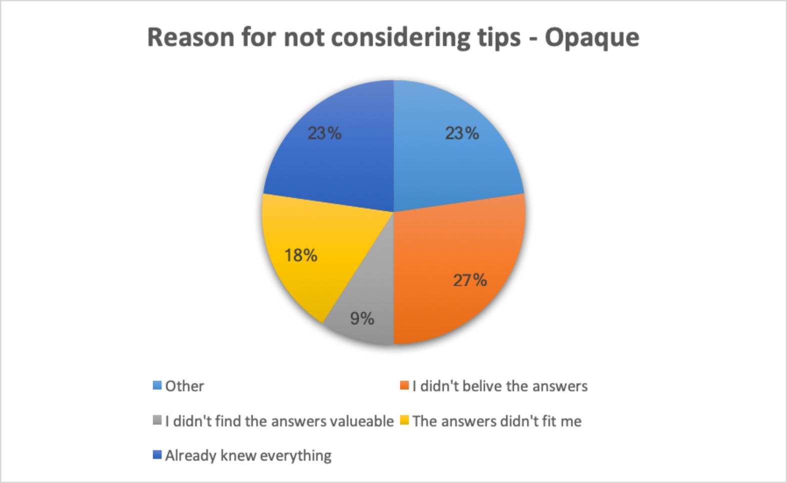


Figure B.3.20. Reason for not considering all the tips provided opaque version

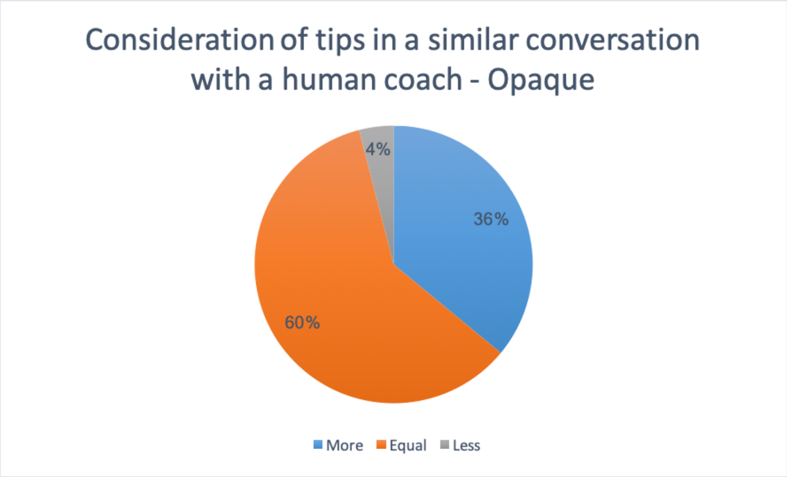


Figure B.3.21. Consideration of tips in similar Human to Human conversation opaque version

Material Biased Version

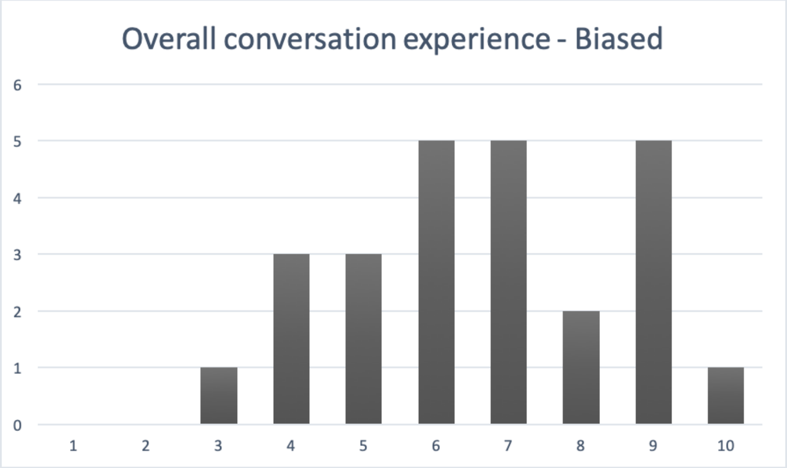


Figure B.3.22. Overall conversation experience biased version

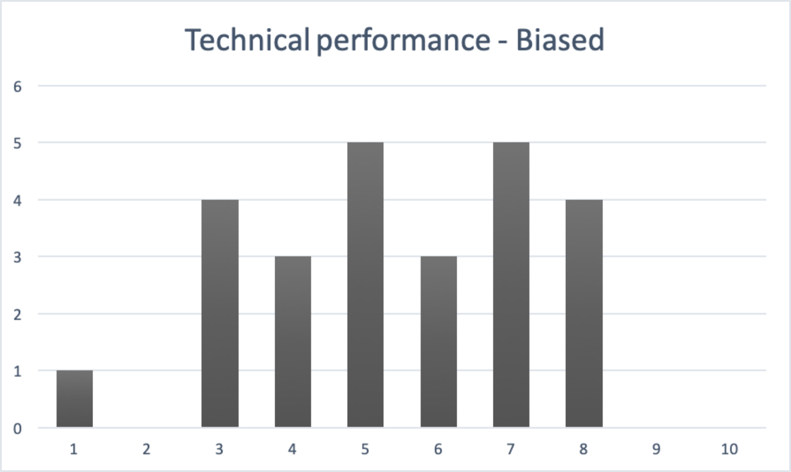


Figure B.3.23. Perceived technical performance biased version

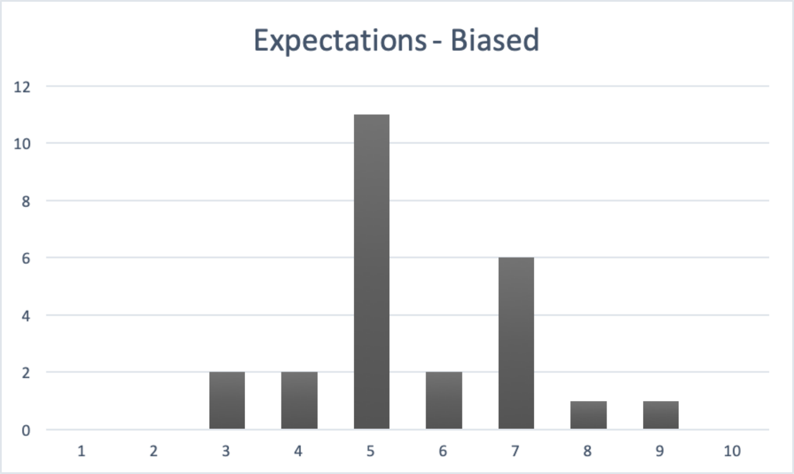


Figure B.3.24. How well Ava met expectations of a vocational guidance chatbot from biased version

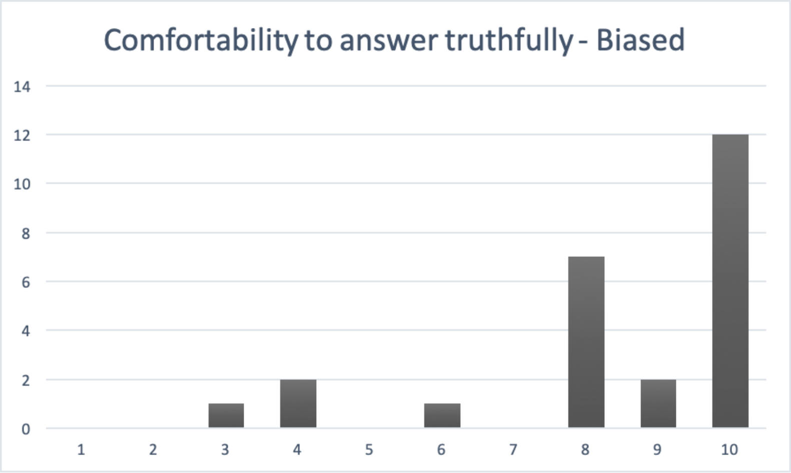


Figure B.3.25. Comfortability to answer truthfully biased version

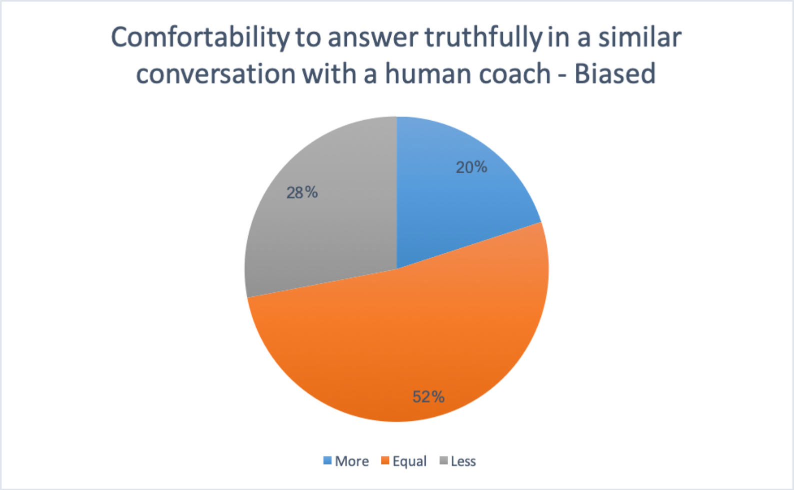


Figure B.3.26. The comfortability to answer truthfully in a human to human conversation in opaque version

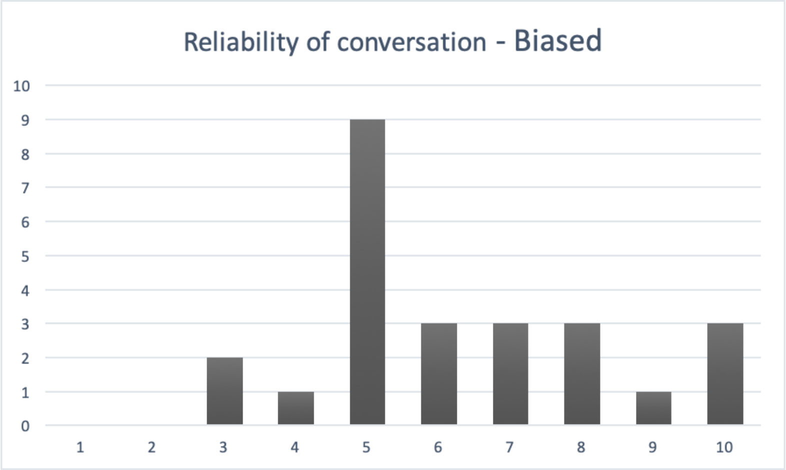


Figure B.3.27. Perceived reliability of the tips in the conversation biased version

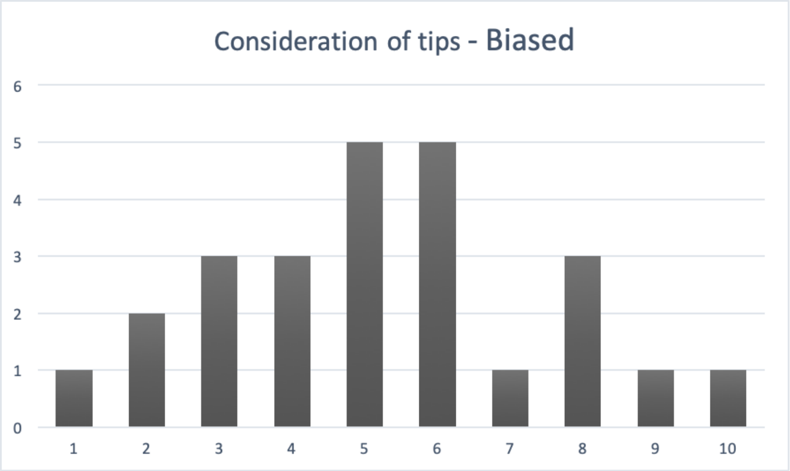


Figure B.3.28. Personal consideration of tips of biased version

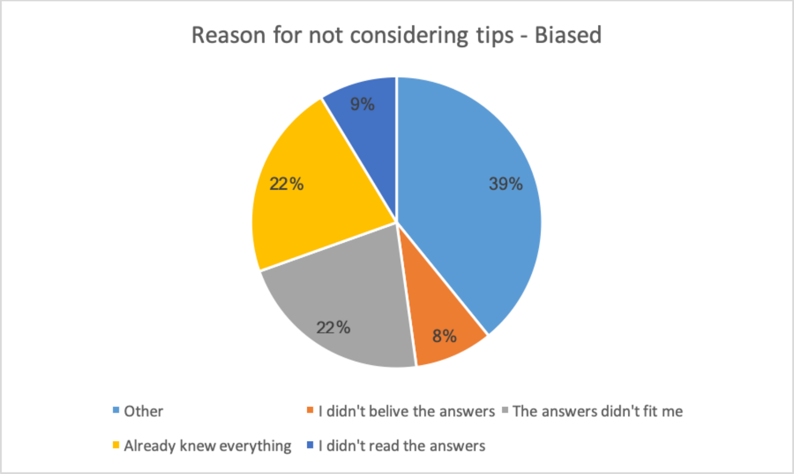


Figure B.3.29. Reason for not considering tips biased version

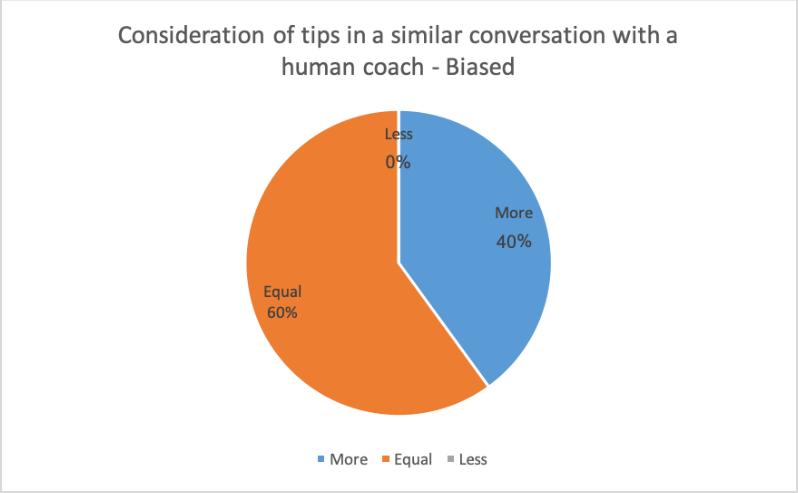


Figure B.3.30. Consideration of tips in similar Human to Human conversation biased version

Material all Versions

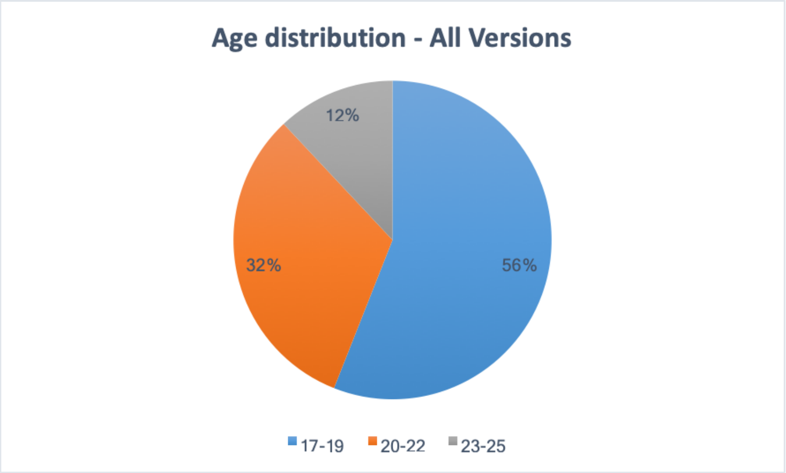


Figure B.3.31 - Age distribution among test personas

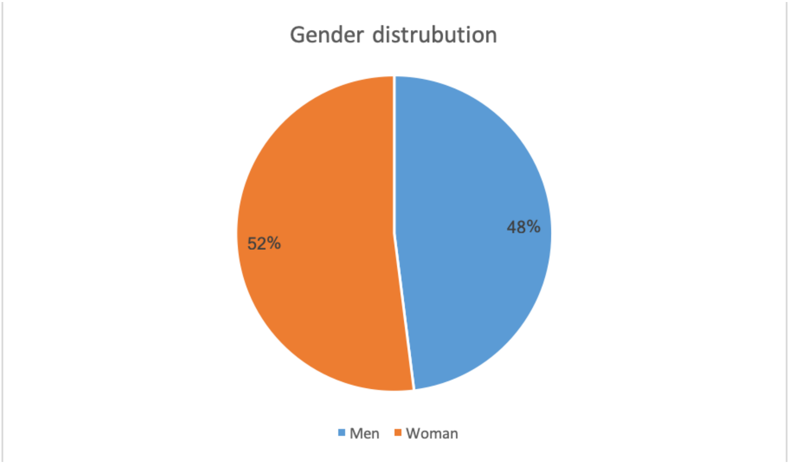


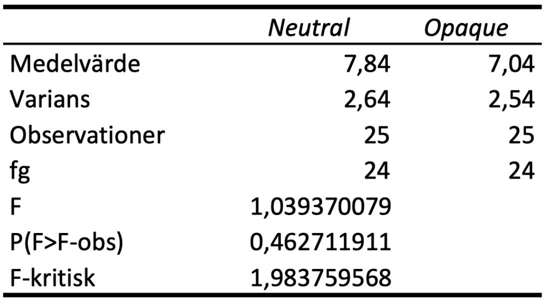
Figure B.3.32. Gender distribution between all test personas

Table B.3.1. The table shows measurements derived from the empirical findings.

Evaluation form question	Neutral	Opaque	Biased	Value
Overall conversation experience.	8,2	7,6	6,6	Mean
	1,3	1,6	1,9	Standard deviation
Previous experience of Chatbots	4,4	3,3	3,6	Mean
	2,3	2,1	2,3	Standard deviation
Prior knowledge of AI.	3,4	2,3	2,3	Mean
	2,3	1,8	1,9	Standard deviation
Perceived technical performance	7,8	7,0	5,4	Mean
	1,6	1,6	1,9	Standard deviation
How well the expectations were met.	7,3	7,0	5,6	Mean
	1,4	1,6	1,4	Standard deviation
Comfortability to answer truthfully.	9,1	9,3	8,4	Mean
	1,2	1,2	2,0	Standard deviation
Did you answer truthfully (yes)	92%	100%	92%	Percentage (yes)
Equal or less ability to answer truthfully in human interaction.	96%	88%	80%	Percentage (Equal or Less)
Perceived reliability of the given answers and tips.	8,6	7,7	6,2	Mean
	1,7	1,7	2,0	Standard deviation
Personal consideration of tips.	6,5	6,4	5,2	Mean

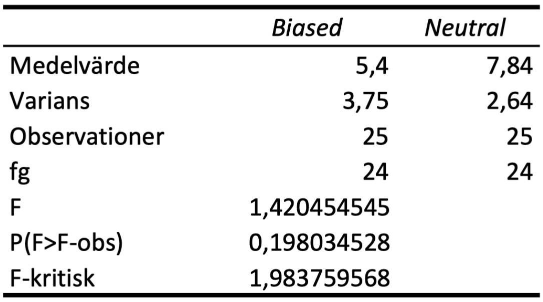
	2,5	2,0	2,2	Standard deviation
Equal or less consideration to the tips in same human to human interaction	76%	64%	60%	Percentage (Equal or Less)
Recommend Ava to a Friend (yes)	92%	84%	68%	Percentage (yes)

Table B.3.2. The table below illustrates f-test on parameter technical experience between neutral and opaque version.



	Neutral	*Opaque*
Medelvärde	7,84	7,04
Varians	2,64	2,54
Observationer	25	25
fg	24	24
F	1,039370079	
P(F>F-obs)	0,462711911	
F-kritisk	1,983759568	

Table B.3.3. The table below illustrates f-test on parameter technical experience between neutral and biased version.



	Biased	*Neutral*
Medelvärde	5,4	7,84
Varians	3,75	2,64
Observationer	25	25
fg	24	24
F	1,420454545	
P(F>F-obs)	0,198034528	
F-kritisk	1,983759568	

Table B.3.4. The table below illustrates f-test on parameter expectations between neutral and opaque version.

	Opaque	*Neutral*
Medelvärde	7	7,28
Varians	2,666666667	1,96
Observationer	25	25
fg	24	24
F	1,360544218	
P(F>F-obs)	0,228166922	
F-kritisk	1,983759568	

Table B.3.5. The table below illustrates f-test on parameter expectations between neutral and biased version.

	Biased	*Neutral*
Medelvärde	5,6	7,28
Varians	2,166666667	1,96
Observationer	25	25
fg	24	24
F	1,105442177	
P(F>F-obs)	0,404019741	
F-kritisk	1,983759568	

Table B.3.6. The table below illustrates the calculated statistical power (%) together with the number of participants needed to achieve a power of 0.8 for the F-test

F - test	N/O*	N/B*
Overall conversation experience	16% (192)	65% (34)
Technical performance	22% (127)	90%
Expectations	5% (<700)	83%

*N=Neutral, O=Opaque & B=Biased

Table B.3.7. The table below illustrates the calculated statistical power (%) together with the number of participants needed to achieve a power of 0.8 for the t-test

t - test	N/O*	N/B*	O/B*
Overall conversation experience	31% (93)	94%	52% (49)
Technical performance	42% (63)	99%	89%
Expectations	11% (300)	99%	91%

*N=Neutral, O=Opaque & B=Biased

Appendix C

C Code and API response

C.1 Code for Data Structuring

Python code for retrieving the desired data from Arbetsförmedlingen APIs and structuring it in a preferable way.

```
import json
import requests

###############   AREA   ###############

idNum = int(1)
stopp = int(40)
arbeten={}

while idNum <= stopp:
    try:
        formedlingen =
requests.get('http://api.arbetsformedlingen.se:80/af/v2/forecasts/occupationalAr
ea/forcastsRefs/list/%d' %idNum)
        data = formedlingen.json()
        arbeten[data[0]['occupationalAreaName']] =
data[0]['occupationPrognosisRefs']
        idNum= idNum + 1
    except:
        idNum= idNum + 1

###############   ABILITIES   ###############

search_id = int(1060)
beskrivning = {}

while search_id <= 1500:
    try:
        response =
requests.get('https://apier.arbetsformedlingen.se/yrkesinfo/publik/vagledning/v1
/yrken/%d?client_id&client_secret' % search_id)
        apidata = response.json()

        beskrivning[apidata['namn']]={}
        shortDesc = apidata['kortSammanfattning']
        kategorier = apidata['formagor']['detaljer']
        counter = int(0)
        formogor=[]
        for object in kategorier:
            formogor.append(kategorier[counter]['kategori'])
            counter=counter+1

        beskrivning[apidata['namn']]['kortSammanfattning'] = shortDesc
        beskrivning[apidata['namn']]['formagor'] = formogor
        search_id = search_id + 1

    except:
        search_id = search_id + 1

###############   FUTURE & EDUCATION  ###############

rubrik=list(arbeten)
for object in rubrik:
    x = int(0)
    for yrken in arbeten[object]:
        try:
            ssyk_no=int(yrken['ssyk'])
            response =
requests.get('http://api.arbetsformedlingen.se:80/af/v2/forecasts/occupationalGr
oup/longTerm/%s' %ssyk_no)
            apidata = response.json()
            dictObject=apidata[0]
            longTerm=dictObject['assessment5Year']
            arbeten[object][x]['assessment5Year'] = longTerm

            x=x+1
        except:
            x=x+1

###############   PUT TOGETHER  ###############

for object in rubrik:
    y=int(0)
    for yrken in arbeten[object]:
        try:
            besk = beskrivning[yrken['heading']]
            arbeten[object][y]['kortSammanfattning'] =
besk['kortSammanfattning']
            arbeten[object][y]['formagor'] = besk['formagor']
            y=y+1
        except:
            arbeten[object][y]['kortSammanfattning'] = None
            arbeten[object][y]['formagor'] = None
            y=y+1

###############   SAVE AS OUTFILE   ###############
```

```
with open('framtid.json', 'w') as outfile:
    json.dump(arbeten, outfile, indent=2)
```

Python code to structure the final JSON object.

```
import json
import requests

finishedJSON = { 'yrken' :{}, 'mapReduce':{}, 'field': {}}
finishedJJSSOONN= {}

with open('framtid_SV.json') as file:
       data = json.load(file)

for object in data:
    finishedJSON['field'][object]={}
    x = int(0)
    for yrke in data[object]:
        omrade = str(object)
        ssyk = yrke['ssyk']
        namn = yrke['heading']
        finishedJSON['yrken'][namn]= {}
        sammanfattning = yrke['kortSammanfattning']
        framtid = yrke['assessment5Year']
        finishedJSON['field'][object][x]= namn
        finishedJSON['yrken'][namn]['ssyk']= ssyk
        finishedJSON['yrken'][namn]['field']= omrade
        finishedJSON['yrken'][namn]['Sammanfattning']= sammanfattning
        finishedJSON['yrken'][namn]['framtid'] = framtid
        finishedJSON['yrken'][namn]['formagor'] = {}
        formagor = yrke['formagor']

        y= int(0)
        for word in formagor:
            try:
                finishedJSON['yrken'][namn]['formagor'][y] = word
                y=y+1
                if word not in finishedJSON['mapReduce']:
                    finishedJSON['mapReduce'][word] = {}
                    finishedJSON['mapReduce'][word][0] = namn
                else:

finishedJSON['mapReduce'][word][len(finishedJSON['mapReduce'][word])] = namn
            except:
                print('fel')
                y=y+1
        x=x+1
```

```
with open('finishedJSON.json', 'w') as outfile:
   json.dump(finishedJSON, outfile, indent=2)
```

C.2 Code for Backend Fulfillment

Javascript code for enabling backend functionality. The code is edited due to confidentiality reasons.

```
'use strict';

const functions = require('firebase-functions');
const {WebhookClient} = require('dialogflow-fulfillment');
const {Text, Card, Image, Suggestion, Payload} = require('dialogflow-
fulfillment');
const admin = require('firebase-admin');

process.env.DEBUG = 'dialogflow:debug'; // enables lib debugging statements
const https = require('https');

admin.initializeApp({
  credential: admin.credential.applicationDefault(),
  databaseURL: '---CONFIDENTIAL---',
});
var db = admin.database();

exports.dialogflowFirebaseFulfillment = functions.https.onRequest((request,
response) => {
  const agent = new WebhookClient({ request, response });
  console.log('Dialogflow Request headers: ' + JSON.stringify(request.headers));
  console.log('Dialogflow Request body: ' + JSON.stringify(request.body));

///////////   STYRKOR   //////////////

  function saveStrengths(agent){
            const context_obj = agent.getContext('conversationpurpose');
            var orginal = context_obj.parameters.personalSkills;
            context_obj.parameters.Styrkor = orginal;
            agent.setContext(context_obj);
  }

  function rearangeStr(agent){
            const context_obj = agent.getContext('conversationpurpose');
            var object=[];
            var a = context_obj.parameters.personalSkills;
            var b = context_obj.parameters.Styrkor;
            var z = 0;
            for (var y in a){
                        if(object.includes(a[y])){
```

```
            z = z + 1;
      }else {object.push(a[y]);}
            }
            for (var x in b) {
        if (object.includes(b[x])) {
            z = z + 1;
        }else{object.push(b[x]);
        }
      }
    context_obj.parameters.personalSkills = object;
    agent.setContext(context_obj);

  }
////////////////   REWRITE STRENGTHS   ////////////////

  function rStrengths(agent){
    var context_obj = agent.getContext('conversationpurpose');
    var lista = context_obj.parameters.personalSkills;

    for (var index in lista){
            if (lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
       }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
       }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }else if(lista[index] === '---CONFIDENTIAL---'){
            lista[index] = '---CONFIDENTIAL---';
        }
      }
      var contextt = agent.getContext('demographics');
         contextt.parameters.Skills = lista;
              agent.setContext(contextt);
  }

////////////////   AGE   ///////////////////

function changeAge(agent){

  const context_obj = agent.getContext('conversationpurpose');
  agent.clearContext('conversationpurpose');
  var original = context_obj.parameters.age;

  if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }else if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }else if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }else if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }else if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }else if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }else if(original === '---CONFIDENTIAL---'){
    context_obj.parameters.age = '---CONFIDENTIAL---';
  }
  agent.setContext(context_obj);

}
///////////////    FREE TIME    ////////////////
  function remapFree(agent){

             const context_obj = agent.getContext('conversationpurpose');
             agent.clearContext('conversationpurpose');
             var original = context_obj.parameters.freeTimePriority;

             if(original === '---CONFIDENTIAL---'){
                          context_obj.parameters.freeTimePriority = '---
CONFIDENTIAL---';
             }else if(original === '---CONFIDENTIAL---'){
            context_obj.parameters.freeTimePriority = '---CONFIDENTIAL---';
             }else if(original === '---CONFIDENTIAL---'){
      context_obj.parameters.freeTimePriority = '---CONFIDENTIAL---';
             }else if(original === '---CONFIDENTIAL---'){
            context_obj.parameters.freeTimePriority = '---CONFIDENTIAL---';
             }else if(original === '---CONFIDENTIAL---'){
            context_obj.parameters.freeTimePriority = '---CONFIDENTIAL---';
             }else if(original === '---CONFIDENTIAL---'){
            context_obj.parameters.freeTimePriority = '---CONFIDENTIAL---';
             }else if(original === '---CONFIDENTIAL---'){
            context_obj.parameters.freeTimePriority = '---CONFIDENTIAL---';
  }
  agent.setContext(context_obj);
 }

////////////   WORK TIME PRIORITY   ////////////////

  function wTimePriority(agent){
            const context_obj = agent.getContext('conversationpurpose');
            var original = context_obj.parameters.workTimePriority;

    if(original === '---CONFIDENTIAL---'){
            context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }else if(original === '---CONFIDENTIAL---'){
        context_obj.parameters.workTimePriority = '---CONFIDENTIAL---';
    }
    agent.setContext(context_obj);
  }

////////////   PERSONAL SUMMERY   ////////////////
  function persSummery(agent){

    const context_obj = agent.getContext('conversationpurpose');
    const context = agent.getContext('demographics');
    var occup, nameo, freeTime, workTime, edlevel, field, age, first_abilities,
second_abilities;
        let str = agent.session;
        var splited = str.split("/");
        var sessionId = splited[4];

             occup = context_obj.parameters.occupation;
             age = context_obj.parameters.age;
             nameo = context_obj.parameters['given-name'];
             freeTime = context_obj.parameters.freeTimePriority;
        workTime = context_obj.parameters.workTimePriority;
        edlevel = context_obj.parameters.levelOfEducation;
        field = context_obj.parameters.degreeField;
        first_abilities = context.parameters.Skills;
        context_obj.parameters.Styrkor = first_abilities;
             second_abilities = first_abilities.pop();

             var abiliities = context_obj.parameters.personalSkills;

            if(occup && nameo && age && freeTime && workTime && edlevel && field
&& first_abilities && second_abilities){
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);

    }else if(occup && age && freeTime && workTime && edlevel && field &&
first_abilities && second_abilities){
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);

    }else if(occup && freeTime && workTime && edlevel && first_abilities &&
second_abilities){
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
    }else{
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
    }

    try{
      db.ref('users/' + sessionId).set({
        CITY: true,
        FIELD: field,
        JOB: true,
        STYRKA: abilities
      });
    }catch(error){console.log(error);}
}
////////////   LIVING CIRCUMSTANCE    /////////////

function livingC(agent){

  const context_obj = agent.getContext('conversationpurpose');
  agent.clearContext('conversationpurpose');
  var object = context_obj.parameters.livingCircumstance;

  if(object === '---CONFIDENTIAL---'){
            context_obj.parameters.livingCircumstance = '---CONFIDENTIAL---';
  }else if(object === '---CONFIDENTIAL---'){
            context_obj.parameters.livingCircumstance = '---CONFIDENTIAL---';
  }else if(object === '---CONFIDENTIAL---'){
       context_obj.parameters.livingCircumstance = '---CONFIDENTIAL---';
  }else if(object === '---CONFIDENTIAL---'){
            context_obj.parameters.livingCircumstance = '---CONFIDENTIAL---';
  }

  agent.setContext(context_obj);
}

////////////   SKIP FORWARD IF 0 PREV. EMPLOYEES   /////////////

  function skipIntents(agent){
    const context_obj = agent.getContext('conversationpurpose');

    var pEmployeesNo = context_obj.parameters['previousEmployers.original'];
    var pEmployees = context_obj.parameters.previousEmployers;
    const ExternafaktorerSummering = {'name': 'ExternaFaktorerSummering',

                                      'lifespan': 10
                                     };
    const ExternafaktorerSummeringCR = {'name': 'ExternaFaktorerSummeringCR',
                                      'lifespan': 1,
                                     };
    var orginal = context_obj.parameters.previousEmployers;

    if(orginal === '---CONFIDENTIAL---'){
              context_obj.parameters.previousEmployers = '---CONFIDENTIAL---';
    }else if(orginal === '---CONFIDENTIAL---'){
        context_obj.parameters.previousEmployers = '---CONFIDENTIAL---';
    }else if(orginal === '---CONFIDENTIAL---'){
        context_obj.parameters.previousEmployers = '---CONFIDENTIAL---';
    }else if(orginal === '---CONFIDENTIAL---'){
              context_obj.parameters.previousEmployers = '---CONFIDENTIAL---';
    }else if(orginal === '---CONFIDENTIAL---'){
        context_obj.parameters.previousEmployers = '---CONFIDENTIAL---';
    }

   if(pEmployeesNo === 0 || pEmployees === '---CONFIDENTIAL---'){
      agent.setContext(ExternafaktorerSummering);
      agent.setContext(ExternafaktorerSummeringCR);
      agent.add(pEmployees);
      agent.add(`---CONFIDENTIAL---`);
   }
   agent.setContext(context_obj);

  }
  ////////   LATEST JOB   /////////

  function latestJob(agent){
            const context_obj = agent.getContext('conversationpurpose');
    var job = context_obj.parameters.jobRoles;

    if(job){
            agent.add(`---CONFIDENTIAL---`);
    }else{
            agent.add(`---CONFIDENTIAL---`);
    }
    agent.add(`För: `);
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
  }

  /////////   EXTERNA FAKTORER SUMMERING   /////////

  function extSummery(agent){
                        const context_obj =
agent.getContext('conversationpurpose');
                        var cit, nameo, liveC, driveL, comuteC, prevEm, jobRole,
workMotiv;

            let str = agent.session;
      var splited = str.split("/");
      var sessionId = splited[4];

            cit = context_obj.parameters.city;
       nameo = context_obj.parameters['given-name'];
            liveC = context_obj.parameters.livingCircumstance;
       driveL = context_obj.parameters.driversLicense;
            comuteC = context_obj.parameters.commutingConsideration;
            prevEm = context_obj.parameters.previousEmployers;
            jobRole = context_obj.parameters.jobRoles;
            workMotiv = context_obj.parameters.workMotivator;

  if(cit && nameo && liveC && driveL && comuteC && prevEm && jobRole &&
workMotiv){
     agent.add(`--CONFIDENTIAL---`);
  }else if(cit && nameo && driveL && comuteC && prevEm && jobRole && workMotiv){
     agent.add(`---CONFIDENTIAL---`);
  }else if(cit && nameo && liveC && driveL && comuteC && prevEm){
     agent.add(`---CONFIDENTIAL---`);
  }else{
             agent.add(`---CONFIDENTIAL---`);
  }

  if(jobRole && workMotiv){
    agent.add(`---CONFIDENTIAL---`);
  }else if(workMotiv){
    agent.add(`---CONFIDENTIAL---`);
  }else{
    agent.add(`---CONFIDENTIAL---`);
  }

    agent.add(`---CONFIDENTIAL---`);
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));
    agent.add(new Suggestion(`---CONFIDENTIAL---`));

    if(jobRole){
    try{
      db.ref('users/' + sessionId).update({
        CITY: cit,
             JOB: jobRole
     });}catch(error){console.log(error);}  }
  }

  //////////////   REWRITE PREFERENCE   //////////////
```

```
function wPreferences(agent){

          const context_obj = agent.getContext('conversationpurpose');
          var object = context_obj.parameters['workPreference.original'];

          if(object === '---CONFIDENTIAL---'){
     context_obj.parameters.workPreference = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.workPreference = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.workPreference = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.workPreference = '---CONFIDENTIAL---';
  }
  agent.setContext(context_obj);
}

/////////////    REWRITE CONCERN     //////////////

function rConcern(agent){
          const context_obj = agent.getContext('conversationpurpose');
          var object = context_obj.parameters['concern.original'];

  if(object === '---CONFIDENTIAL---'){
     context_obj.parameters.concern = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.concern = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.concern = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.concern = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.concern = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.concern = false;
  }
  agent.setContext(context_obj);
}

///////////   REWRITE DREAM   /////////

function rDream(agent){
const context_obj = agent.getContext('conversationpurpose');
          var object = context_obj.parameters['dreamPursuit.original'];

  if(object === '---CONFIDENTIAL---'){
     context_obj.parameters.dreamPursuit = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
     context_obj.parameters.dreamPursuit = '---CONFIDENTIAL---';
  }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.dreamPursuit = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.dreamPursuit = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.dreamPursuit = '---CONFIDENTIAL---';
    }
    agent.setContext(context_obj);

  }

  //////////   REWRITE DUTIES   //////////

  function rDuties(agent){
      const context_obj = agent.getContext('conversationpurpose');
            var object = context_obj.parameters['naturalDuties.orginial'];

    if(object === '---CONFIDENTIAL---'){
       context_obj.parameters.naturalDuties = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.naturalDuties = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.naturalDuties = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.naturalDuties = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.naturalDuties = '---CONFIDENTIAL---';
    }else if (object === '---CONFIDENTIAL---'){
       context_obj.parameters.naturalDuties = '---CONFIDENTIAL---';
    }
    agent.setContext(context_obj);
  }

  ///////////////    MOTIVATION SUMMERY   ////////////////

  function motSummery(agent){
            const context_obj = agent.getContext('conversationpurpose');
            var name, concern, concernOrg, workP, dreamP, dreamR, natD,
workPOrg;

    name = context_obj.parameters['given-name'];
    concern = context_obj.parameters.concern;
    concernOrg = context_obj.parameters['concern.original'];
    workP = context_obj.parameters.workPreference;
    workPOrg = context_obj.parameters['workPreference.original'];
    dreamP = context_obj.parameters.dreamPursuit;
    dreamR = context_obj.parameters.dreamRole;
    natD = context_obj.parameters.naturalDuties;

    if(concernOrg === '---CONFIDENTIAL---'){
      if(workP && workPOrg && dreamP && dreamR && natD){
             agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);

    }else if(workP && workPOrg && dreamP && natD){
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
    }
  }
  else{
    if(workP && workPOrg && dreamP && dreamR && natD){
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);

    }else if (workP && workPOrg && dreamP && natD){
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
      agent.add(`---CONFIDENTIAL---`);
    }
  }
}

////////////    FINAL PLAN    ///////////////

function finalPlan(agent){
  const url_plats = '---CONFIDENTIAL---';
  const image_jobb = '---CONFIDENTIAL---';
  const url_plugg = '---CONFIDENTIAL---';
  const image_plugg = '---CONFIDENTIAL---';
  const url_jobb_plugg ='---CONFIDENTIAL---';


  const context_obj = agent.getContext('conversationpurpose');

  var abilities = context_obj.parameters.personalSkills;
  var name = context_obj.parameters['given-name'];
  var first_abilities = context_obj.parameters.Styrkor;
  var second_abilities = first_abilities.pop();
  var plugga = context_obj.parameters.Styrkor;
  var stad = context_obj.parameters.city;
  var vidare = context_obj.parameters.furtherEducation;
  var occu = context_obj.parameters.occupation;
  var latestJob = context_obj.parameters.jobRoles;
  var dreamJob = context_obj.parameters.dreamRole;
  var edu = context_obj.parameters.levelOfEducation;


  var possibleResponse = [
    `---CONFIDENTIAL---`,
    `---CONFIDENTIAL---`,
    `---CONFIDENTIAL---`,
    `---CONFIDENTIAL---`
```

```
];

var pick = Math.floor(Math.random() * possibleResponse.length);
var response = possibleResponse[pick];
agent.add(response);

if(occu === '---CONFIDENTIAL---'){
        if(edu === '---CONFIDENTIAL---' || '---CONFIDENTIAL---'){
      if(vidare !== '---CONFIDENTIAL---'){
        agent.add(`---CONFIDENTIAL---`);
        agent.add(`---CONFIDENTIAL---`);

         let card = new Card({
           title: `Vidare Studier`,
           imageUrl: image_plugg,
           text: `---CONFIDENTIAL---`,
           buttonText: 'Vidare Studier',
           buttonUrl: url_plugg
         });
         agent.add(card);
      }else if(latestJob){
        agent.add(`---CONFIDENTIAL---`);
        agent.add(`---CONFIDENTIAL---`);
        let card = new Card({
          title: `Yrkesguiden`,
          imageUrl: image_jobb,
          text: `---CONFIDENTIAL---`,
          buttonText: 'Yrkesguiden',
          buttonUrl: url_jobb_plugg
        });
        agent.add(card);
      }else {
        agent.add(`---CONFIDENTIAL---`);
        agent.add(`---CONFIDENTIAL---`);
        let card = new Card({
          title: `Yrkesguiden`,
          imageUrl: image_jobb,
          text: `---CONFIDENTIAL---`,
          buttonText: 'Yrkesguiden',
          buttonUrl: url_jobb_plugg
        });
        agent.add(card);
      }
   }
}else {
  agent.add(`---CONFIDENTIAL---`);
  agent.add(`---CONFIDENTIAL---`);

  let card = new Card({
    title: `JobScanner`,
    imageUrl: image_jobb,
        text: `---CONFIDENTIAL---`,
        buttonText: 'JobScanner',
        buttonUrl: url_plats
      });
      agent.add(card);
    }
    const utv_url = '---CONFIDENTIAL---';
    const utv_img = '---CONFIDENTIAL---';
    let caard = new Card({
        title: `Utvärdering Gymnasieelever`,
        imageUrl: utv_img,
        text: `---CONFIDENTIAL---`,
        buttonText: 'Till Utvärdering',
        buttonUrl: utv_url
      });
    agent.add(`!!GLÖM INTE ATT SVARA PÅ UTVÄRDERINGEN!!`);
    agent.add(caard);
  }

  ////////////////    FUNCTION HANDLER   //////////////////
  // Run the proper function handler based on the matched Dialogflow intent name
  let intentMap = new Map();
  intentMap.set('Personlighet - Summering', persSummery);
  intentMap.set('Personlighet - Fritidsprioritering', remapFree);
  intentMap.set('Personlighet - Arbetsprioritering', wTimePriority);
  intentMap.set('Personlighet - Utbildningsnivå', rStrengths);
  intentMap.set('Personlighet - Ålder', changeAge);
  intentMap.set('Personlighet - Förmågor Från Dig', saveStrengths);
  intentMap.set('Personlighet - Förmågor Från Någon Annan', rearangeStr);
  intentMap.set('Externa Faktorer - Levnadssituation', livingC);
  intentMap.set('Externa Faktorer - Tidigare Arbetsgivare', skipIntents);
  intentMap.set('Externa Faktorer - Senaste Yrkestitel', latestJob);
  intentMap.set('Externa Faktorer - Summering', extSummery);
  intentMap.set('Motivation - Viktigaste Arbetsfaktorn', wPreferences);
  intentMap.set('Motivation - Orosmoment', rConcern);
  intentMap.set('Motivation - Drömsysselsättning', rDream);
  intentMap.set('Motivation - Naturlig Roll', rDuties);
  intentMap.set('Motivation - Summering', motSummery);
  intentMap.set('THE END', finalPlan);
  agent.handleRequest(intentMap);

});
```

C.3 API response from Arbetsförmedlingen

The JSON output below shows an example of a response from the yrkesvägledning API provided by Arbetsförmedlingen.

```
{
    "id": 1345,
    "amsOccupationId": 303,
    "namn": "Brevbärare",
    "kortSammanfattning": "Ser till att posten kommer fram i tid och till rätt
person, ett yrke för den som gillar högt tempo.",
    "sammanfattning": "Brevbärare arbetar med att sortera och dela ut post. Det
är viktigt att posten kommer fram i tid och till rätt person. Därför måste
brevbäraren vara bra på att hantera sin tid och kunna arbeta självständigt
eftersom man ofta arbetar ensam. Utdelningen sker numera oftast i bil eller med
moped, men ibland även med cykel eller till fots. Därför kan det vara bra att
vara rörlig i jobbet.",
    "arbetsuppgifter": "Brevbärare börjar jobba tidigt på morgonen med att
sortera dagens post. Därefter ger man sig ut på sin postrunda och delar ut
posten. När posten från morgonen är utdelad återvänder man till kontoret för att
sortera den del av morgondagens post som redan kommit. \n\nBrevbäraren delar ut
post ensam och är ute i alla väder. Idag är det vanligast att man kör någon form
av fordon, en bil eller moped när man delar ut posten. Därför är körkort ofta
ett krav när man anställs. Men det är också vanligt att posten delas ut med
hjälp av cykel eller genom att gå med en kärra, särskilt i tätorternas
stadskärna. \n\nTill postterminalerna kommer post som ska sorteras för att
fortsätta till olika delar av landet eller till olika utdelningsområden. Att ta
hand om post som ska eftersändas ingår också i arbetet.\n\nPostsorterare
organiserar posten före sortering och sköter brevsorteringsmaskinen. Det mesta
av posten sorteras med hjälp av maskiner, men de brev som maskinen inte klarar
av att hantera sorteras manuellt. Postsorterare kör truck när de lastar och
lossar post.",
    "arbetsmiljo": "Arbetet kan innebära tunga lyft och man rör på sig mycket.
Arbetet utförs delvis utomhus.",
    "arbetstid": "Brevbärare har fasta arbetstider. Det mesta sorteringsarbetet
pågår eftermiddagar, kvällar och nätter. En postsorterare arbetar ofta på
schemalagda arbetstider. Inom en del av de postdistributionsföretag kan arbetet
även göras under helgen.",
    "arbetsplats": "",
    "internationellaMojligheter": null,
    "formagor": {
        "beskrivning": "Förmågor som brevbärare och postsorterare behöver ha
eller utveckla",
        "detaljer": [
            {
                "text": "Brevbärare kan vara ett rörligt arbete, där det kan
vara viktigt att ha en normal fysisk förmåga.",
                "kategori": "God fysik"
            },
            {
                "text": "Det är viktigt att kunna hålla koncentrationen och vara
noggrann även när man arbetar under tidspress.",
                "kategori": "Koncentrationsförmåga"
            },
            {
                "text": "Det dagliga arbetet sker tidvis i högt tempo eftersom
posten måste sorteras färdigt och delas ut i tid.",
                "kategori": "Stresstålighet"
            },
            {
                "text": "Kunna kombinera olika rörelser samtidigt är viktigt vid
till exempel postsortering.",
                "kategori": "Koordinationsförmåga"
            },
            {
                "text": "Brevbärare och postsorterare behöver kunna producera
tillfredsställande resultat på kort tid.",
                "kategori": "Resultatinriktad"
            }
        ]
    },
    "utbildningar": {
        "beskrivning": "Man måste ha fyllt 18 år för att arbeta som brevbärare.
Den som ska köra ut post med bil eller moped behöver körkort.",
        "detaljer": [
            {
                "text": "",
                "kategori": "Inga krav på formell utbildning"
            }
        ]
    },
    "utbildningsvag": {
        "beskrivning": "Yrket har inga formella utbildningskrav, men det kan
vara bra att ha gått en gymnasieutbildning.",
        "visaValideringstext": false,
        "utbildningsvagkategorier": [
            {
                "kategori": "Gymnasieutbildning",
                "kategoriText": "Ibland ställs krav på gymnasieutbildning som
grund."
            },
            {
                "kategori": "Företagsförlagd utbildning",
                "kategoriText": "Brevbärare utbildas internt i samband med att
de anställs."
            }
        ]
    },
    "intresseprofil": {
        "primar": "C",
        "sekundar": "R",
        "tertiar": null
    },
    "yrkesgrupperOchBenamningar": [
        {
            "ssyk": "4420",
            "namn": "Brevbärare och postterminalarbetare",
            "taxonomiId": 4420,
            "taxonomiTyp": "SSYK4"
        }
    ],
    "liknandeYrken": [
        {
            "id": 1182,
            "namn": "Budbilsförare"
        },
        {
            "id": 1278,
            "namn": "Tidningsbud"
        },
        {
            "id": 1340,
            "namn": "Handpaketerare"
        }
    ],
    "lankar": [
        {
            "namn": "Postnord",
            "url": "http://www.postnord.com/sv",
            "kategori": "Extern information"
        },
        {
            "namn": "Bring",
            "url": "http://www.bring.se",
            "kategori": "Extern information"
        },
        {
            "namn": "Gymnasieinfo.se",
            "url": "http://www.gymnasieinfo.se/",
            "kategori": "Gymnasieutbildningar"
        }
    ],
    "yrkesomraden": [
        "Transport"
    ],
    "metadata": {
        "version": 1,
        "senastUppdaterad": "2018-11-19T14:21:43"
    },
    "bilder": [
        {
            "lank":
"https://www.arbetsformedlingen.se/webdav/images/yrkesbeskrivning/3295647-
postbil.png ",
            "sammanfattning": "En brevbärare delar ut post.",
            "primar": true,
            "normbrytandeKon": false,
            "normbrytandeEtnicitet": false,
            "normbrytandeFunktionsnedsattning": false
        }
    ]
}
```

The JSON output shows an example of a response from the yrkesprognoser API provided by Arbetsförmedlingen.

```
{

   "forecastOccupation": "Brevbärare och postterminalarbetare",
    "ssyk": [
      "4420"
    ],
    "releaseDate": "2019-02-07",
    "assessmentNow": 3,
    "assessmentNowText": "Balans",
    "assessment1year": 3,
    "assessment1yearText": "Balans",
    "assessment5Year": 2,
    "assessment5YearText": "Stor konkurrens",
    "occupationalAreaId": "19"
 }
```