UNIVERSITY
OF TURKU
Turku School of
Economics

VIRTUAL ASSISTANTS IN CUSTOMER INTERFACE

Information Systems Science, Department of Management and Entrepreneurship

Master's thesis

Author:
Samuli Hiltunen

Supervisor:
Prof. Reima Suomi

21.11.2022
Turku

Master's thesis

Subject: Information Systems Science
Author: Samuli Hiltunen
Title: Virtual assistants in customer interface
Supervisor(s): Prof. Reima Suomi
Number of pages: 61 pages + appendices 1 page
Date: 21.11.2022

This thesis covers use of virtual assistants from a user organization’s perspective, exploring challenges and opportunities related to introducing virtual assistants to an organization’s customer interface. Research related to virtual assistants is spread over many distinct fields of research spanning several decades. However, widespread use of virtual assistants in organizations customer interface is a relatively new and constantly evolving phenomenon. Scientific research is lacking when it comes to current use of virtual assistants and user organization’s considerations related to it.

A qualitative, semi-systematic literature review method is used to analyse progression of research related to virtual assistants, aiming to identify major trends. Several fields of research that cover virtual assistants from different perspectives are explored, focusing primarily on Human-Computer Interaction and Natural Language Processing. Additionally, a case study of a Finnish insurance company’s use of virtual assistants supports the literature review and helps understand the user organization’s perspective. This thesis describes how key technologies have progressed, gives insight on current issues that affect organizations and points out opportunities related to virtual assistants in the future. Interviews related to the case study give a limited understanding as to what challenges are currently at the forefront when it comes to using this new technology in the insurance industry.

The case study and literature review clearly point out that use of virtual assistants is hindered my various practical challenges. Some practical challenges related to making a virtual assistant useful for an organization seem to be industry-specific, for example issues related to giving advice about insurance products. Other challenges are more general, for example unreliability of customer feedback. Different customer segments have different attitudes towards interacting with virtual assistants, from positive to negative, making the technology a clearly polarizing issue. However, customers in general seem to be becoming more accepting towards the technology in the long term. More research is needed to understand future potential of virtual assistants in customer interactions and customer relationship management.

Key words: virtual assistant, chatbot, customer interface, insurance

Pro gradu -tutkielma

Oppiaine: Tietojärjestelmätiede
Tekijä: Samuli Hiltunen
Otsikko: Virtual assistants in customer interface
Ohjaaja(t): Prof. Reima Suomi
Sivumäärä: 61 sivua + liitteet 1 sivu
Päivämäärä: 21.11.2022

Tämä tutkielma tutkii virtuaaliassistenttien käyttöä käyttäjäorganisaation perspektiivistä, antaen käsityksen mitä haasteita ja mahdollisuuksia liittyy virtuaaliassistenttien käyttöönottoon organisaation asiakasrajapinnassa. Virtuaaliassistentteihin liittyvä tutkimus jakautuu monien eri tutkimusalojen alaisuuteen ja useiden vuosikymmenien ajalle. Laajamittainen virtuaaliassistenttien käyttö asiakasrajapinnassa on kuitenkin verrattain uusi ja jatkuvasti kehittyvä ilmiö. Tieteellinen tutkimus joka liittyy virtuaaliassistenttien nykyiseen käyttöön ja käyttäjäorganisaation huomioon otetaviin asioihin on puutteellista.

Tämä tutkielma käyttää kvalitatiivista, puolisystemaattista kirjallisuusanalyysimetodia tutkiakseen virtuaaliassistentteihin liittyviä kehityskulkuja, tarkoituksena tunnistaa merkittäviä trendejä. Tutkimus kattaa useita tutkimusaloja jotka käsittelevät virtuaaliassistentteja eri näkökulmista, keskittyen pääasiassa Human-Computer Interaction- sekä Natural Language Processing -tutkimusaloihin. Lisäksi tutkielmassa on tapaustutkimus suomalaisen vakuutusyhtiön virtuaaliassistenttien käytöstä, joka tukee kirjallisuusanalyysiä ja auttaa ymmärtämään käyttäjäorganisaation perspektiiviä. Tutkielma kuvailee kuinka keskeiset teknologiat ovat kehittyneet, auttaa ymmärtämään tämänhetkisiä ongelmia jotka koskettavat organisaatioita sekä esittelee virtuaaliassistentteihin liittyviä mahdollisuuksia tulevaisuudessa. Tapaustutkimukseen liittyvät haastattelut antavat rajoitetun kuvan kyseisen uuden teknologian käyttöön liittyvistä haasteista vakuutusalalla.

Tapaustutkimus ja kirjallisuusanalyysi osoittavat että virtuaaliassistenttien käyttöönottoon liittyy erilaisia käytännön haasteita. Jotkut haasteet vaikuttavat olevan toimialakohtaisia, liittyen esimerkiksi vakuutustuotteita koskeviin neuvoihin. Toiset haasteet taas ovat yleisempiä, liittyen esimerkiksi asiakaspalautteen epäluotettavuuteen. Eri asiakassegmenteillä on erilaisia asenteita virtuaaliassistentteja kohtaan, vaihdellen positiivisesta negatiiviseen, joten kyseinen teknologia on selvästi polarisoiva aihe. Pitkällä aikavälillä asiakkaiden asenteet teknologiaa kohtaan vaikuttavat kuitenkin muuttuvan hyväksyvämpään suuntaan. Lisää tutkimusta tarvitaan jotta voidaan ymmärtää virtuaaliassistenttien tulevaisuuden potentiaalia asiakaskohtaamisissa ja asiakkuudenhallinnassa.

Avainsanat: virtuaaliassistentti, botti, asiakasrajapinta, vakuutus

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Background

The purpose of this thesis is to examine virtual assistants, a type of chatbot, that have become commonplace on most company's and public organization's websites in recent years. A chatbot could be defined as a software conversational agent that can hold a text-based conversation with a human user in natural language, either for no specific purpose or with a specific task in mind (Chaves & Gerosa 2021, 729). Chatbots used to assist visitors to websites are referred to as virtual assistants in this thesis, as this term is also commonly used on websites that have embedded chatbots used for tasks such as answering simple questions. The research is limited to text-only virtual assistants, so all voice-communication based solutions are excluded from this research.

Open-source databases covering vocabulary and necessary natural language processing logic allow a simple virtual assistant to be implemented with relative ease, but virtual assistants that use more complicated AI are typically more difficult to bring to a working condition with the necessary features for the organization's needs (Khanna et al. 2015, 278). A virtual assistant used by a Finnish company, for example, should be proficient in general conversation but also task-specific vocabulary in Finnish language, creating a need to customize any solutions available on the market to make them useful in practice. Juntunen (2019, 1) found that despite a degree of hype around chatbots, in Finnish enterprises these programs rarely use actual AI, instead being able to answer mostly clear and common questions with the Finnish language being a particular challenge, at least as of 2019.

Some of the perceived advantages afforded by use of virtual assistants are 24/7 availability, personalized interactions with customers and lack of waiting times, in addition to potential cost saving for the organization (Hasal et al. 2021, 2). Each organization using virtual assistants has different basic competencies they need to have met when it comes to customer interactions. The questions a virtual assistant is expected to answer are naturally different in different uses. The roles of virtual assistants are also different on different websites, but using virtual assistants as a first line of contact with visitors on the front page of an organization's website is a growing trend. Motives of a typical visitor to a website are therefore an important consideration, as are attitudes towards chatbot software's ability to understand the visitor.

The insurance company If P&C Insurance collaborated in the making of this thesis by two interviews related to the virtual assistant used at if.fi. This virtual assistant is meant to take care of simple customer service tasks while being reachable 24/7.

1.2 Research problem

Studies related to virtual assistants and chatbots more broadly fall under several different domains of research, mainly Computer Science, Information Systems, Human-Computer Interaction (HCI) and Natural Language Processing (NLP) research. Of these, computer science mainly focuses on technical aspects, for example how to develop better algorithms. Information systems research has studied factors related to user's perception of virtual assistants on one hand and the user organization's adaptation decisions on the other hand. Finally, human-computer interaction research has focused how humans react to different social properties of chatbots. (Gnewuch et al. 2017, 2.)

Research about the technical aspects of chatbots is relatively plentiful, on the other hand there is limited research into how chatbots are perceived and how people interact with them (Chaves & Gerosa 2021, 730). It appears that chatbot-related literature is more focused on the computer science perspective, and not so much on information systems and human-computer interaction perspectives.

This thesis will examine virtual assistants from the perspective of information systems science and closely related domains of research, like Human-Computer Interaction and Natural Language Processing. I will examine positive and negative aspects of using such services from the user organizations perspective. The use of virtual assistants is a relatively new trend and there is added value to be created by examining the topic from the organization's perspective.

1.3 Research questions

The research question and sub-questions are the following:

❖ What challenges and opportunities does an organization face in introducing a virtual assistant to its customer interface?

- What limitations affect virtual assistants?
- What added value can virtual assistants create?

1.4 Structure of this thesis

The methodology chapter will first introduce the frameworks used in the literature review and empirical research processes.

The literature review covers the relevant fields of research covered by this thesis.

Finally, the results chapter presents findings from empirical research while reflecting on the literature review.

Finally, findings are examined in the discussion and conclusion chapter, where limitations of this thesis and future directions of research are also presented.

2 Methodology

2.1 Overview

This thesis covers two relevant fields of research, Human-Computer Interaction and Natural Language Processing, as well as managerial perspectives on the use of virtual assistants in customer interface. While reviewing literature from these fields I will be covering any limiting factors or challenges affecting virtual assistants as well as outlining potential factors that could create added value for an organization's customers.

The literature review gives the reader a picture of what virtual assistants are, how they work and outlines essential developments in the fields of research related to virtual assistants and chatbots, which spans across different research domains as outlined earlier.

Empirical research covers only insurance industry specifically and gives perspectives on the findings of the literature review. Empirical research also provides the reader with an update on how the use of virtual assistants has developed. Virtual assistants or chatbots are not a new research topic, research similar to this have been published in the past years. This research will provide the reader with an update on how this new technology's use has progressed since other similar research studied it by using one company as a case study.

2.2 Qualitative literature review

Snyder's (2019, 333-339) guidelines for literature review as research methodology are used for the literature review chapter. Snyder (2019, 334) outlines three main types of literature reviews: systematic, semi-systematic and integrative. Of these, systematic literature reviews are well suited for quantitative research. Integrative literature reviews aim to facilitate formation of new theoretical frameworks or perspectives by critiquing and assessing current literature on a topic. Semi-systematic literature review, also known as narrative review, is perhaps the most flexible of the three, aiming to create an overview of literature around a topic as well as give the reader a picture of how relevant literature has progressed. (Snyder 2019, 334-335.)

Semi-systematic literature review method is designed for types of topics that have been researched by different groups of researchers representing separate fields of research and are therefore conceptualized in different ways due to different perspectives. These types of topics can hinder a systematic review due to multitude of disciplines related to these topics. Reviewing all literature related to such topics is not possible, instead the scope of semi-systematic literature review must be limited to an extent. Semi-systematic literature can focus on trends of research within a specific field, aiming to create a picture of how research related to a topic has progressed. Semi-systematic review aims to identify and understand every research tradition that is potentially relevant to the topic at hand, then to synthesize them using meta-narratives. (Snyder 2019, 335.)

The semi-systematic review is used because the nature of this research is qualitative and does not seek to create any new frameworks based on the literature. Creating an outline of how literature related to virtual assistant has progressed and what topics have arisen in the literature is essential to create the necessary understanding of the topic for the latter empirical research. Research related to virtual assistants (or chatbots and conversational agents as they are often referred to in literature) spans many distinct fields of research. Semi-structured review is a good method for mapping the state of knowledge related to this topic.

The literature review is used to form propositions that answer the research question of this thesis, outlining potential challenges and opportunities related to virtual assistants.

2.3 Expert interviews

2.3.1 Background

I conducted two expert interviews with persons who have worked with virtual assistants at If P&C Insurance Company. I interviewed Niko Saviluoto and Viivi Inkeroinen from If P&C Insurance Company on 12.5.2022 and 17.5.2022, respectively. Both are digital customer service developers at If, with several years of experience working on virtual assistants. Niko Saviluoto is working at the Commercial (B2B) side of the company, while Viivi Inkeroinen works at the Private (B2C) side.

The interviews were informal in nature, but a pre-planned interview guide based on the literature review directed the course of the interviews.

A framework by Kallio et al. (2016, 2954-2962) for developing quantitative semi-structured interview guides was used to prepare for the interviews. According to this framework, development of an interview guide involves five different phases.

Firstly, prerequisites for using semi-structured interview must be identified. The goal is to evaluate if the semi-structured interview is appropriate as a data collection method for the topic of the research. The semi-structured method is said to be suitable for studying complex issues and people’s opinions on them. (Kallio et al. 2016, 2959.)

Secondly, previous knowledge about the topic needs to be retrieved and used, the goal being to create a predetermined framework for the interview. The necessary knowledge of the subject could be gathered with a literature review and/or complementary empirical research. (Kallio et al. 2016, 2959.)

Thirdly, a preliminary semi-structured interview guide must be formulated. An interview guide is a list of questions that is meant to guide the interview towards the research topic. For a semi-structured interview’s interview guide flexibility is an important attribute. The guide should enable dialogue during the interview, it should also be possible to change the order of the questions during the interview. The goal is to create a tool for collecting interview data by utilizing previous knowledge. The quality of the interview guide has an effect on how the interview plays out, as well as analysis of the collected data. The interview guide can also enable collecting data that allows new concepts to emerge. (Kallio et al. 2016, 2959-2960.)

The interview guide should include two different levels of questions: main themes and follow-up questions. The former covers main content of the research subject and encourages the interviewed person to tell about their experiences and perceptions. The latter is used to help better understand the main subjects. Follow-up questions can be pre-planned or spontaneous and are used to maintain the flow of the conversation and to help collect accurate and optimal information. (Kallio et al. 2016, 2959-2960.)

Fourthly, the interview guide should be pilot tested and any small improvements added. Fifthly, the ready interview guide can be used. (Kallio et al. 2016, 2960-2961.)

2.3.2 Interview guide

The main themes of the interview guide reflect upon my research questions, while focusing specifically on the use of virtual assistants in insurance industry.

Since I conducted two interviews, the first also acted as pilot testing of my interview guide for the other interview.

The interviews are meant to be informal in nature and give space for any unexpected observations that the interview guide does not cover. The interview guide is attached as annex 1.

3 Literature review

3.1 Introduction to chatbots

3.1.1 Background

Chatbots combine the following essential capabilities in order to interact with humans. Firstly, the chatbot must be able to interpret written language in order to decipher the human user's intent, a process referred to as Natural Language Processing (NLP). Interpreting written information can be challenging as the format and style of writing is never constant with different users. This part of the chatbot can be referred to as Language Understanding Component. The chatbot must either express to the user that more information is needed or act upon understood information. Secondly, the chatbot must be able to search an internal database of knowledge it possesses that can be referred to as Knowledge Base, or possibly search for information through the internet. Thirdly, a Response Generating Component must be able to then formulate a response based on the user's intent deciphered by the Language Understanding Component and using data in the Knowledge Base. The response is formatted to resemble human-written text by Natural Language Generation. A Dialogue Management Component keeps track of the intent of the conversation. (Adamopoulou & Moussiades 2020, 379-380.)

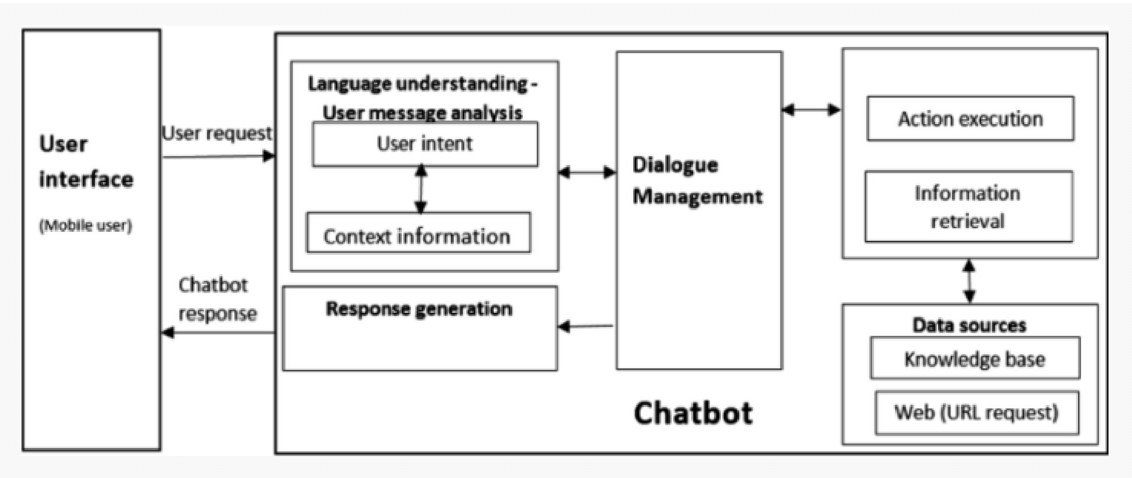

Figure 1: Diagram depicting an example of chatbot architecture (Adamopoulou & Moussiades 2020, 380).

Reshmi & Balakrishnan (2016, 1173) classify a typical chatbot as consisting of three parts, a knowledge base, a chatbot engine that acts as an interface between the bot and its user, and an interpreter program. The interpreter program itself includes an analyser, which analyses text written by human users and utilizes normalization techniques to turn it closer to a standardised format. This reformatted text is then fed to the chatbot engine which in turn searches the knowledge base for a suitable reply to this text. Another part of the interpreter program is a generator, which takes the search result from the chatbot engine and turns it into an appropriate format to be displayed to the chatbots user. (Reshmi & Balakrishnan 2016, 1173-1174.)

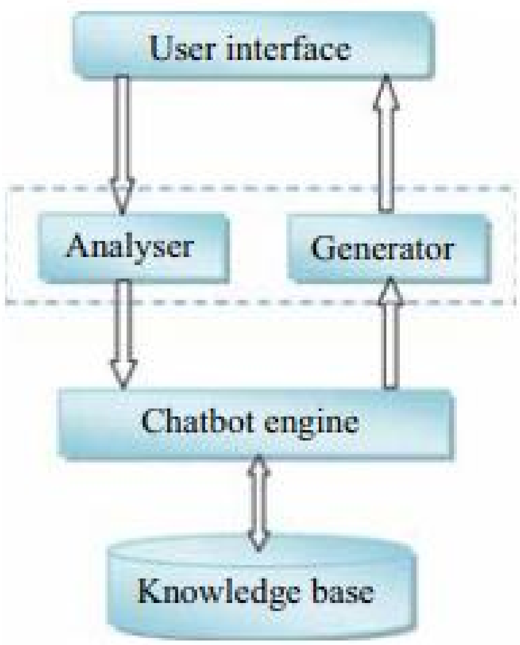

Figure 2: Simplified model of chatbot architecture (Reshmi & Balakrishnan 2016, 1174).

Both figure 1, depicting the model from Adamopoulou & Moussiades, and figure 2, depicting the model from Reshmi & Balakrishnan, feature a knowledge base and user interface as part of the chatbot architecture. The parts between these two elements are different in the two models, however some essential elements of an chatbot can be identified.

The following three sections will further detail the essential functions of chatbots.

3.1.2 Natural Language Processing

Most chatbots work by using Natural Language Processing, comparing user inputs and a knowledge base of expressions to form appropriate responses (Reshmi & Balakrishnan 2016, 1173). Adamopoulou & Moussiades (2020, 377) define Natural Language Processing as a research field within artificial intelligence, with the research being focused on collecting knowledge about use of language to develop techniques that enable computers to understand natural expressions.

3.1.3 Knowledge base

Knowledge base is a database of keywords or expressions and the chatbots responses associated with them (Reshmi & Balakrishnan 2016, 1174).

A chatbots knowledge base can consists of Artificial Intelligence Markup Language (AIML) files, which are a standardized way of programming chatbot behaviour (Reshmi & Balakrishnan 2016, 1174). AIML is designed specifically for programs processing natural language and enables use of existing open-source databases of common questions and answers, making chatbot development process more streamlined than it would be without existing databases (Khanna et al. 2015, 278). Notably, Khanna et al. (2015, 277) consider limited amount of fed-in information as one significant factor limiting chatbot's utility.

3.1.4 Response generating

Adamopoulou & Moussiades (2020) classify three different models the Response Generating Component can use to formulate its response: rule-based, retrieval-based and generative models.

Rule-based model is the original model for first chatbots created and is based around predefined rules concerning text that is fed to the chatbot. Human user's message is analysed for predetermined patterns or keywords and then a suitably predetermined reply is selected, without generating any new text. This type of model does not fare well when there are grammatical mistakes in the text being fed to it. Retrieval-based models are close to rules-based models, except they use Application Programming Interfaces (API's) to retrieve information from other applications, giving these models more possible information for a response. Finally, more complicated generative models generate answers based on both previous and present messages, utilizing machine learning to create new answers independently. Generative models naturally are more complicated to develop than the two other models, but also present potential to a much greater degree. (Adamopoulou & Moussiades 2020, 378-379.)

3.2 Types of chatbots

3.2.1 Overview

Følstad et al. (2019, 149-150) introduces a two-dimensional typology to classify different types of chatbots. The two dimensions are locus of control, either user-driven or chatbot-driven, and duration of relation, either short-term or long-term. Chatbots designed for short-term relation are designed for one-off conversations, this type of chatbot can be found in for example e-retailers and news websites. Chatbots designed for long-term relation are designed to utilize information about the user to achieve a more personalized user experience, because interactions with this type of chatbot are repetitive. These types of chatbots can be on a platform such as Facebook, which can also enable access to some user information for the chatbot. (Følstad et al. 2019, 149-150.)

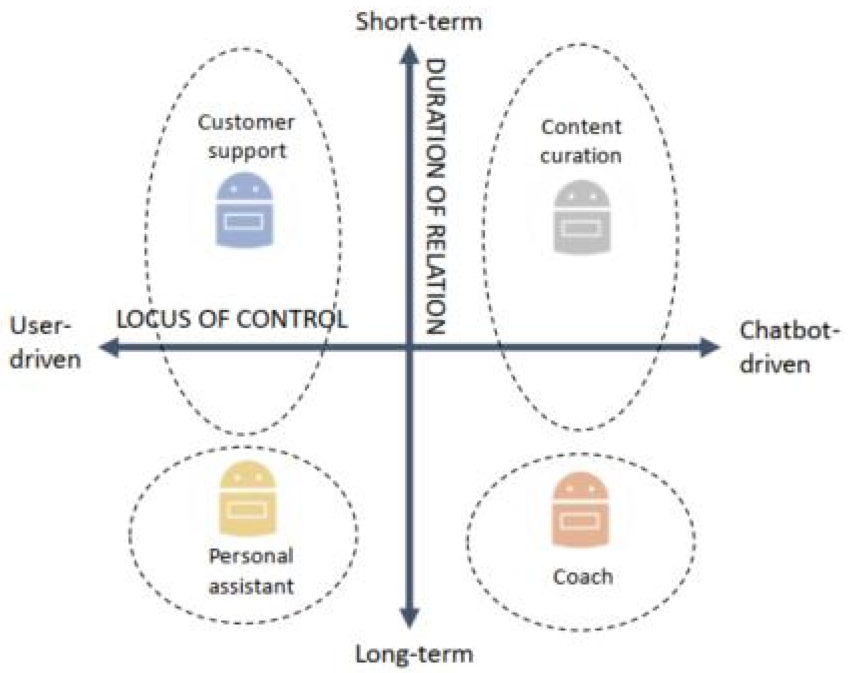

Figure 3: Two-dimensional typology of chatbots (Følstad 2019, 149-150).

Four types of chatbots categories classified by the two-dimensional typology are visible in figure 3. Chatbots could be classified to the following categories: customer support, content curation, personal assistant and coach, the two former are meant for short-term relations and the two latter for long-term relations. (Følstad et al. 2019, 150-151.)

For the purposes of this thesis, the customer support type is particularly interesting. This type of chatbot is designed to give help or advice to a user in a user-driven manner, being typically designed to make initiating the conversation as easy as possible for the user (Følstad et al. 2019, 151-152). A customer support -type chatbot is thereby close to the definition of virtual assistant examined in this thesis.

Chatbots can also be classified based on their purpose in another way. According to Hasal et al. (2021, 2) chatbots made for specific tasks, using company specific data, can be classified as closed-domain chatbots, while open domain chatbots on the other hand can converse about a diverse set of subjects.

Gnewuch et al (2017, 3) classify virtual assistants based two simple dimensions, mode of communication and context, mode of communication means either voice-based or text-based communication, while context is classified in a similar manner as by Hasal et al, into general-purpose and domain-specific.

		Context	
		General-Purpose	**Domain-Specific**
Primary Mode of Communication	**Text-based***1	ELIZA (Weizenbaum 1966) Cleverbot, ...	Enterprise-class CAs (Lester et al. 2004; McTear et al. 2016; Shawar and Atwell 2007), IKEA's Anna, ...
	Speech-based*2	Apple's Siri, Amazon's Alexa, Google Now, Samsung's Bixby, ...	SPECIES (Derrick et al. 2011; Nunamaker et al. 2011) In-car assistants (Reisinger et al. 2005); Mercedes-Benz Linguatronic, ...

*1 Text-based: chatbot, chatterbot, dialogue system, etc.
*2 Speech-based: (virtual) personal assistant, digital companion, intelligent/smart agent, etc.

Figure 4: Illustration of virtual assistant types (Gnewuch et al 2017, 3).

Since this thesis is limited to only text-based virtual assistants, the first dimension, mode of communication, is irrelevant. The second dimension, context, is useful in classifying virtual assistants based on their uses. As per figure 4, an example of a text-based, general purpose virtual assistant would be Cleverbot, while an example of text-based, domain-specific would be IKEA's Anna (Gnewuch et al 2017, 3). By this definition the virtual assistants at if.fi are also examples of domain-specific chatbots.

3.2.2 Closed-domain chatbots

Closed-domain chatbots are typically meant for a specific tasks, using proprietary information from the chatbots user organizations, for example frequently asked questions or instructional manuals (Hasal et al. 2021, 2). This makes these types of chatbots particularly interesting for the purposes of this literature review. Janssen et al. (2020, 212) refers to similar type of chatbot as rule-based chatbots, that are commonly used in a variety of different contexts, such as customer service or education, making researching this type of chatbot challenging. These kinds of systems can often be configured using AIML templates. Notably, such systems are limited in the way they can respond to grammatically incorrect questions. Sometimes these limitations can be countered by implementing graphical elements, such as predefined questions for the user to choose from, reducing potential for misinterpretation on the chatbot's behalf. (Janssen et al. 2020, 219-220.) Chatbots can classified into 5 archetypes based on characteristics and areas of application (Janssen et al 2020, 220)

- Goal-oriented daily chatbot (type A)
- Non goal-oriented daily chatbot (type B)
- Utility facilitator chatbot (type C)
- Utility expert chatbot (type D)
- Relationship-oriented chatbot (type E).

Types A-D are predominantly rule-based chatbots with limited ability to adapt to their user's behaviour, while the archetype E has higher ability to adapt its communication to the user. (Janssen et al. 2020, 220-222.)

Different properties of the five chatbot archetypes are presented in figure 5. Types A-D are typically used in closed-domain applications as outlined before, while type E is closer to an open-domain chatbot.

		Label	Goal-oriented daily chatbot	Non goal-oriented daily chatbot	Utility facilitator chatbot	Utility expert chatbot	Relationship-oriented chatbot
		n	24	19	22	29	9
		Archetype	A	B	C	D	E
Intelligence	Intelligence framework	Rule-based systems	88%	79%	77%	76%	
		Utility-based systems	13%	16%	18%	17%	33%
		Model-based systems		5%	5%		44%
		Goal-based systems				7%	
		Self-learning systems					22%
	Intelligence quotient	Only rule-based knowledge	46%	11%	95%	24%	11%
		Text understanding	38%	47%		72%	44%
		Text understanding and further abilities	17%	42%	5%	3%	44%
	Personality processing	Principal self	100%	100%	100%	100%	56%
		Adaptive self					44%
	Socio-emotional behavior	None/low	96%	84%	100%	100%	11%
		High	4%	16%			89%
	Service integration	None		68%	32%	7%	11%
		Single	92%	5%	45%	79%	56%
		Multiple	8%	26%	23%	14%	33%
Interaction	Multimodality	Unidirectional	100%	89%	86%	97%	67%
		Bidirectional		11%	14%	3%	33%
	Interaction classification	Graphical	8%	5%	64%	14%	33%
		Interactive	92%	95%	36%	86%	67%
	Interface personification	Disembodied	79%	68%	86%	72%	11%
		Embodied	21%	32%	14%	28%	89%
	User assistance design	Reactive assistance	75%	95%	91%	86%	
		Proactive assistance	25%	5%	9%	14%	100%
	Number of participants	Individual human partner	96%	100%	95%	93%	100%
		Two or more human participants	4%		5%	7%	
	Additional human support	No	100%	100%	45%	86%	44%
		Yes			55%	14%	56%
	Front-end user interface	App	8%	11%		3%	22%
		Social media	63%	32%	23%	21%	33%
		Collaboration and communication	8%		5%	14%	
		Website	13%	26%	55%	59%	33%
		Multiple	8%	32%	18%	3%	11%
Context	Chatbot role	Facilitator	54%	5%	77%	21%	33%
		Peer		16%			
		Expert	46%	79%	23%	79%	67%
	Relation duration	Short-term relation	75%	89%	86%	100%	44%
		Long-term relation	25%	11%	14%		56%
	Application domain	E-customer service	4%		27%	48%	11%
		Daily life	83%	95%	36%		22%
		E-commerce	4%		27%	7%	
		E-learning		5%			33%
		Finance	4%		5%	38%	
		Work and career	4%		5%	7%	33%
	Collaboration goal	Non goal-oriented	4%	89%	5%		56%
		Goal-oriented	96%	11%	95%	100%	44%
	Motivation for chatbot use	Productivity	17%	11%	32%	7%	56%
		Entertainment	38%	84%	14%		22%
		Social/relational	25%	5%			
		Utility	21%		55%	93%	22%

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Figure 5: Chart of five chatbot arhetypes (Jannsen et al. 2020, 221).

Figure 5 presents results of an analysis of different characteristics distribution between chatbot archetypes, where different characteristics are classified in to three main categories, intelligence, interaction and content. Higher percentages in the chart represent higher probability of a certain archetypes chatbot to match with specific characteristic. (Janssen et al. 2020, 220-221.)

Janssen et al. (2020, 221) made the following observation about the 5 chatbot archetypes, these findings are also present in figure 5. Firstly, in terms of intelligence of the chatbot, closed-domain chatbots typically have little to no ability for socio-emotional behaviour. Secondly, in terms of interaction, closed-domain chatbots are predominantly designed for reactive assistance of its user, instead of proactive. Additional human support is also rare for closed-domain chatbots. Thirdly, in term of context the duration of interaction with a closed-domain chatbots user is typically short. It should be noted that the closed-domain chatbot types in this particular analysis were used most for “daily life” and considerably less for purposes such as electronic customer service and electronic commerce. (Janssen et al. 2020, 221-222.)

3.3 Issues with privacy and data protection

3.3.1 General Data Protection Regulation

Since chatbots work with user data, possibly even learning from this data through the process of machine learning, they might present a problem to data security.

Potential problems with data protection in relation to chatbot usage can, for example, arise with the encryption of communications between a chatbot and its user or verifying that the user is indeed giving information to a legitimate chatbot and not a third party (Hasal et al. 2021, 4-5). Chatbots using some level of machine learning typically record their conversations to aid in developing their quality of service in the future. Stored data about customer interactions can be of significant value as the quality of a machine learning model typically improves the more data it has access to. Information being stored by the chatbot software to a greater extent than data would be stored through other venues of customer service can pose problems with privacy laws such as European Union’s General Data Protection Regulation (GDPR). GDPR requires that data about chatbot interactions be pseudonymized and encrypted, in other words, storing the data in a correct format is important. Some chatbots can share data with third parties, such as Google Analytics, to evaluate different customer types. In such cases, the information being shared and the format of that information is a topic of concern. (Hasal et al. 2021, 6.)

Hasal et al. (2021, 7) specifies rights for individuals pertaining to personal data and explanations for them in table 1.

Table 1: Individual rights pertaining to personal data (Hasal et al. 2021, 7).

Transparency in data procession	Process of collecting personal data must be transparent
Data minimization	Personal data that is processed must be relevant and limited
Purpose limitation	Some personal data can be collected but not further processed
Storage limitation	Personal data must be deleted in a timely matter when it is no longer needed
Right to download	Personal data is available to download on demand
Right to change	Personal data can be changed on request
Right to remove	Personal data can be deleted on request (also known as right to be forgotten)

Deng et al. (2011, 10) points out that for a product, such as chatbot using customer information, to comply with privacy and data security goals, legal requirements have to first be clearly identified and then refined to product requirements. These product requirements should then be integrated to the product design and testing processes taking place for the product (Deng et al. 2011, 10).

To define potential privacy threats that would have to be taken into account in product requirements, privacy threats are classified into different categories. *Information disclosure* for instance in a type of threat where user's personal information become exposed to entities who are not supposed to have it. These types of threats affect data flow from user to system operator as well as data storage associated with the system. *Content unawareness* is another type of threat where a user is unaware of what information might be recorded by a system and therefore expose too much personal information. These threats are related to data storage and how it is managed, the entity storing data is typically responsible for ensuring that necessary consents are provided before storing any data. Finally, *policy and consent non-compliance* is a type of privacy threat where privacy policies are presented to an end user but the system can not guarantee that these policies are actually complied with. These threats affect the entire system (such as customer service chatbot linked to a database of customer data) since data flow, data storage and the entire process of handling private information is responsible for ensuring compliance with privacy policies. (Deng et al. 2011, 10-12.)

To comply with GDPR, chatbots must have "appropriate safeguards" to avoid any encryption-related security issues (Hasal et al. 2021, 7). In other words, GDPR legally requires that sufficient measures are taken to safeguard from privacy issues falling to the information disclosure category of privacy threats. GDPR also includes the requirement that generally personal data should only be stored for as long as necessary for the purposes that its being processed for (Hasal et al. 2021, 6-7). This could present a policy and consent non-compliance risk. Some chatbots might also not ask users consent to their conversational patterns being stored. Compliance might not always be taken into account in chatbot designs, unlike for example websites which ask for a user's compliance to use cookies (Hasal et al. 2021, 8). This would be an example of content unawareness threat to a user's privacy.

Some issues remain open when it comes to chatbots compliance with GDPR. Overall transparency about how data is processed and stored can be lacking, as general basics of chatbots algorithmic structure can be known but details about implementation and knowledge bases used are typically confidential. (Hasal et al. 2021, 7.) Knowledge bases used by more advanced chatbots can be created from collected conversations, further complicating the question of privacy. So called 'big data', collecting and storing all possible information for potential later use, could be viewed as an opposing practice to data minimalization and storage limitation required by GDPR. Right to remove personal data is also complicated with many chatbot solutions, as models using machine learning essentially permanently embed past communications into themselves, making deleting past interactions impossible. GDPR's definition of private data uses possibility of 'profiling' individuals data from several users data, this means that if an individual can be identified in indirect way or otherwise distinguished from a group of users by data, then this data could generally be considered private data by GDPR. (Hasal et al. 2021, 8.)

3.3.2 Design choices to comply with privacy requirements

In order to comply with user's privacy requirements, Deng et al. (2011, 22) outlines the following strategies to solve privacy issues that might affect for example a customer service virtual assistant's user:

- Warning the user of privacy risks, precautions would have to be taken so that especially nontechnical users don't divulge private information.
- Removing, or turning of, features that have high risks or untenable mitigation procedures could be a solution to reduce privacy risks to zero. This approach can help reach a balance between potential privacy risks and user features.
- Countering privacy threats with preventative or reactive technology.

When it comes to technology to counter privacy threats, one possible solution would be to use data mining techniques to identify personal identifiable information provided by a user, then warning the user every time they provide information that comprises a security risk. (Deng et al. 2011, 22-25.)

Kalloniatis et al. (2008, 242) outline a methodology for incorporating necessary privacy requirements into a system design process. Firstly, privacy-related goals have to be elicited with the help of all relevant stakeholders. Secondly, the effect of these privacy goals on organizational processes should be analysed, objective being to identify any impacts these privacy requirements might have on other goals the organization has. Testing how different goals affect different processes help create a map of alternative ways to resolve privacy issues and what affect they have. Finally, affected organizational processes should be modelled and techniques should be selected that best support these modelled processes. (Kalloniatis et al. 2008, 244-245.) Possible activities that could support privacy supporting organizational processes could be for example the strategies outlined by Deng and collegues.

3.4 Chatbot performance metrics

3.4.1 Classifying chatbot performance

Chatbot performance can be measured with different measures relative to the goal these chatbots are intended to fill. According to Przegalinska et al. (2019, 787) the length and structure of conversations chatbots have with customers can be a metric for the quality of service. Another metric is customer retention, how often do customers keep interacting with the bot until their business is concluded. Finally, the ability to provide personalized communication based on customer-specific information is an important, but highly subjective, metric. (Przegalinska et al. 2019, 787.)

Castro et al. (2018, 410) classify quality features that can be used to measure the performance of chatbots into the following three categories:

- *Efficiency*, which measures a chatbots performance, response time and robustness in responding to unexpected events
- *Effectiveness*, which measures a chatbots ability to meet its domain requirements
- *Satisfaction*, which is related to a chatbots ability to create empathy, maintain ethics during interactions and its accessibility resources.

Of these efficiency and effectiveness could be tested using both manual and automated tests. Automated tests could be performed with software such as SoalUI, which tests a wide variety of different functionalities and answer options. Automated testing of quality features could be used to check if a chatbot is compliant with certain specifications, after which further manual testing could be done to ensure usability. (Castro et al. 2018, 410.)

3.4.2 Quality assurance

Radziwill & Benton (2017, 6-7) categorized different attributes of a virtual assistant or chatbot related to its quality into six categories, namely *performance, functionality, humanity, affect, ethics* and *accessibility*. Table 2 illustrates the attributes of a virtual assistant that affect each of these categories (Radziwill & Benton 2017, 6-7).

The first three (performance, functionality, humanity) of these categories have to do with the effectiveness of a solution and are mostly the service providers responsibility, while the latter three (affect, ethics, accessibility) are tied to customer satisfaction and therefore are mostly the responsibility of the party implementing the virtual assistant to its customer interface (Radziwill & Benton 2017, 6).

Table 2: Categories and attributes (Radziwill & Benton 2017, 6-7).

Performance	● Robustness to manipulation ● Robustness to unexpected input ● Avoid inappropriate utterances and be able to perform damage control ● Effective function allocation, provides appropriate escalation channels to humans
Functionality	● Interprets commands accurately ● Use appropriate degrees of formality, linguistic register ● Linguistic accuracy of outputs ● Execute requested tasks ● Facilitate transactions and follows up with status reports ● General ease of use ● Engage in on-the-fly problem solving ● Contains breadth of knowledge, is flexible in interpreting it
Humanity	● Passes the Turing test ● Does not have to pass the Turing Test ● Transparent to inspection, discloses its chatbot identity ● Include errors to increase realism ● Convincing, satisfying, & natural interaction ● Able to respond to specific questions ● Able to maintain themed discussion
Affect	● Provide greetings, convey personality ● Give conversational cues ● Provide emotional information through tone, inflection, and expressivity ● Exude warmth and authenticity ● Make tasks more fun and interesting ● Entertain and/or enable participant to enjoy the interaction ● Read and respond to moods of human participant
Ethics	● Respect, inclusion, and preservation of dignity (linked to choice of training set) ● Ethics and cultural knowledge of users ● Protect and respect privacy ● Non-deception ● Sensitivity to safety and social concerns ● Trustworthiness (linked to perceived quality) ● Awareness of trends and social context
Accessibility	● Responds to social cues or lack thereof ● Can detect meaning or intent ● Meets neurodiverse needs such as extra response time and text interface

3.4.3 Analytic Hierarchy Process

The search for a method to systematically compare the quality attributes of two or several chatbots is apparent in previous literature regarding chatbots quality assurance. To compare chatbots quality, using attributes presented in table 2, Analytic Hierarchy Process (AHP) is proposed. This method entails the following steps.

1. Creating a hierarchy of quality attributes, then select suitable metrics to represent each of these attributes
2. Form pairwise comparisons between quality attributes for one or several product options
3. Form comparison matrices and calculate principal eigenvector of each to assess priorities, both relative and global
4. Combine priorities and calculate inconsistency factors in order to determine which product option best satisfies this hierarchy of quality attributes.

The pairwise comparisons in step 2 are done by weighting different attributes importance relative to each other. Some quality attributes are more critical than others, so prioritization is useful for decision making. For example, performance could be 5 times as important as humanity and likewise humanity 0.2 times as important as performance. Steps 3 and 4 of this process can be done using software. This process could be used, for example, to compare existing and new versions of a chatbot. In this way, quality of new chatbot solutions can be tested and controlled. (Radziwill & Benton 2017, 8-11.)

Table 3: Examples of metrics selected for different quality attributes and study results (Radziwill & Benton 2017, 10-11).

Category	Quality attribute	Metric	Old	New
Performance	Robustness to unexpected input	% of successes	86-92%	91-93%
Performance	Provides appropriate escalation channels	% of successes	80%	100%
Humanity	Transparent to inspection	% of users who correctly classify	100%	100%
Humanity	Able to maintain themed discussion	0 (low) – 100 (high)	Avg. 72	Avg. 85
Humanity	Able to respond to specific questions	% of successes	68%-82%	80%-85%
Affect	Pleasant personality, provides greetings	0 (low) – 100 (high)	Avg. 89	Avg. 96
Affect	Entertaining, engaging	0 (low) – 100 (high)	Avg. 50	Avg. 66
Accessibility	Can detect meaning and intent	% of successes	85%-90%	82%-86%
Accessibility	Responds to social cues appropriately	% of successes	78%	77%

Table 3 illustrates some examples of metrics to represent different quality attributes, as well as numerical data for these metrics from a case study. For example a chatbots robustness to unexpected input, which is part of a chatbots performance as outlined in the categories of table 2, could be measured by % of successes. The exact metrics used can vary based on practical considerations and also include subjective elements. An element that requires subjective rating could be for example considered a successful response when a chatbot encounters an unexpected input. (Radziwill & Benton 2017, 10-12.)

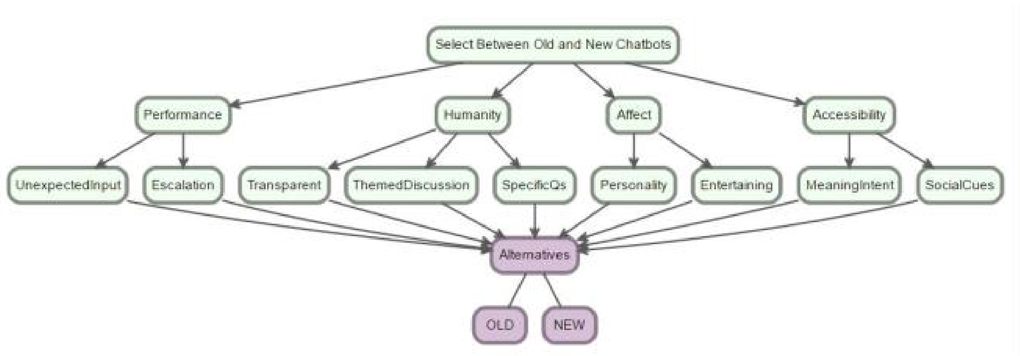

Figure 6: Analytic Hierarchy Process in comparing old and new alternative solutions (Radziwill & Benton 2017, 11).

In figure 6 the process is presented visually, with the top level displaying the goal, the next level showing categories of quality attributes. Below categories are nine quality attributes selected with the outlined prioritization process. Finally, at the bottom two alternatives can be seen. (Radziwill & Benton 2017, 11.)

Table 4 shows the end result of such analysis, helping to choose an alternative that is well suited for satisfying the prioritized quality attributes. In this example the old solution is better suited to meet the priorities used. (Radziwill & Benton 2017, 13.)

Table 4: Final results of AHP process (Radziwill & Benton 2017, 13).

	Weight	OLD	NEW	Consistency
Select Between Old and New Chatbots	100.0%	66.2%	33.8%	18.4%
Accessibility	54.5%	39.1%	15.3%	0.0%
MeaningIntent	47.7%	35.7%	11.9%	0.0%
SocialCues	6.8%	3.4%	3.4%	0.0%
Performance	32.1%	24.6%	7.5%	0.0%
UnexpectedInput	28.1%	21.1%	7.0%	0.0%
Escalation	4.0%	3.5%	0.5%	0.0%
Affect	9.4%	1.6%	7.8%	0.0%
Entertaining	7.8%	1.3%	6.5%	0.0%
Personality	1.6%	0.3%	1.3%	0.0%
Humanity	4.1%	1.0%	3.1%	0.0%
SpecificQs	1.9%	0.3%	1.5%	0.0%
ThemedDiscussion	1.9%	0.5%	1.4%	0.0%
Transparent	0.4%	0.2%	0.2%	0.0%

3.5 Human-Computer Interaction

3.5.1 Overview

Across the various domains of research that have examined chatbots, a common theme has been the need for chatbots to better adapt to human communication to seem more natural and engage the human user better. Achieving this is a technical problem on one hand, but also has much to do with Human-Computer Interaction (HCI). HCI researchers have examined human's reactions to human-like properties of chatbots as well as ways human's behaviour is affected by these properties. (Gnewuch et al. 2017, 2.)

According to Chaves & Gerosa (2021, 729) chatbot technology is changing how businesses interact with their customers. However, problems persist with human satisfaction towards chatbots performance and in particular making humans accept chatbots. These problems might be more social than technical, as people often have expectations of how a chatbot would interact with them. Not meeting people's expectations can in turn lead to disappointment. Examples of expectations a chatbot could fail to meet include conforming to gender stereotypes associated with the bot's purpose and demonstrating positive attitude through text. On the other hand, chatbots that seem too humanized can also be problematic, as they cause expectations to rise excessively. A central problem becomes what chatbot's social competences should be improved to increase human satisfaction in chatbots. In HCI literature, studies focusing specifically on chatbots social abilities are fairly rare. (Chaves & Gerosa 2021, 729-730.)

3.5.2 User's reactions to virtual assistants

Research about the effectiveness of virtual assistants has highlighted numerous practical issues with these systems. For example, Gnewuch et al. (2017, 1-2) conclude that designing virtual assistant services to meet user expectations still faces many challenges. Some virtual assistants have failed to live up to expectations to the point that they have been taken out of use (Gnewuch et al. 2017, 2).

Ben Mimoun et al. (2012, 606) considers lack of reciprocity in the interaction between virtual assistants and their users a major limiting factor, in particular lack of human-like social competences are an issue with users who desire more social and enjoyable interactions. However, virtual assistants that are disconcertingly like humans in their behaviour, but not actually human, might create a negative reaction in its human user (Ben Mimoun et al. 2012, 606). Mori (1970, 33-35) introduced the uncanny valley theory, originally developed for robotics, that suggests interacting with robots that come very close to humans might cause uncomfortable or uncanny feelings in people, while people have more sympathetic views towards robots further away from human standards. Ben Mimoun et al. (2012, 606) considers the uncanny valley theory applicable to virtual assistants, noting that highly human-like virtual assistants are evaluated more critically by users and might provoke unease in users. Ciechanowski et al. (2019, 539-540) support this assertion by concluding that chatbots that are more complex and aim to mimic humans too much tend to be experienced as uncanny more often than simpler, task-oriented chatbots.

Virtual assistants that seem highly human-like to users might also create unattainable expectations, which then lead to credibility and likability issues in the user's eyes. Expectation can create a reference frame that can shape their judgment of the virtual assistant, if this reference frame is then broken by poor performance of the virtual assistant user dissatisfaction ensues. This is referred to as expectancy disconfirmation model. Three ways the perception of virtual assistant's performance can be undermined are outlined. Firstly, the user might feel like the virtual assistant is not listening to what is being written to it. Secondly, virtual assistants might be embedded in a way that feels too intrusive. Thirdly, the process of interacting with the virtual assistant might be too mechanical, for example asking a number of general questions to determine the nature of the user's question. (Ben Mimoun et al. 2012, 606-608.) However, it should be noted that Ben Mimoun's and collegues research is from 2012, so virtual assistants have had time to improve since.

Ciechanowski et al. (2019, 545-547) compared user's emotional reactions to two types of chatbots, one limited to text-only and one with an audio-visual avatar. The study found that users tended to prefer text-based chatbots more, with the avatar-based chatbots having a higher potential to elicit negative emotional reactions (Ciechanowski et al. 2019, 545-547). It could be deduced that text-based chatbots or virtual assistants are therefore less prone to the uncanny valley effect, due to their simplicity and lack of visual or auditory elements.

3.5.3 Attributes of effective communication

According to Chakrabarti & Luger (2015, 6880), the so-called cooperative principles, originally meant to describe properties of effective human to human communication, are also a useful framework for measuring effectiveness of communication between a human and some type of chatbot. Cooperative principles can be summarized by four rules for communication, referred to as Gricean maxims, which are presented below (Chakrabarti & Luger 2015, 6880):

- *quality*: speaker's utterance is the truth as provable by adequate contextual evidence or domain facts
- *quantity*: speaker utterance provides as much information as appropriate, not more or less
- *relation*: speaker's utterance is relevant to the context and the topic of the conversation
- *manner*: speaker's utterance is direct and straightforward, without ambiguity or obfuscation.

Of these maxims, the one concerning quality could be considered an objective measure since it can be verified from the data available, while quantity, relation and manner are subjective measures. (Chakrabarti & Luger 2015, 6880-6881.)

Based on these maxims, it can be concluded that a message sent by a virtual assistant used for some pragmatic purpose should be true, contain just the necessary amount of information, be relevant to the conversation preceding it and be easily understandable without any ambiguity.

3.5.4 Measuring effective communication

The following seven performance metrics can be used to evaluate the effectiveness of chatbots communication with its user in a customer service scenario (Chakrabarti & Luger 2015, 6880-6882):

- Percentage of follow-up questions (asked by the chatbot)
- Number of coherent conversation turns
- Percentage of subjectively successful resolutions
- Grice’s quality maxim
- Grice’s quantity maxim
- Grice’s relation maxim
- Grice’s manner maxim.

The first three of these metrics, as well as the quality maxim, can be evaluated based on a transcript of a conversation between a chatbot and its user, as well as suitable answers to utterances based on relevant domain knowledge to the conversation. (Chakrabarti & Luger 2015, 6880-6882.)

Percentage of follow-up questions refers to number of questions a chatbot needs to ask its user to finish a customer service interaction successfully, and how many of these questions the chatbot asks. Number of questions needed is based on the situational context of the interaction. Percentage of follow-up questions is defined as such that if, for example, four distinct follow-up questions need an answer and the chatbot asks three of them, this metric is 75%. (Chakrabarti & Luger 2015, 6881.)

Number of coherent conversation turns refers to utterance-exchange pairs that consist of a user's utterance and the chatbots response to them. A number of utterance-exchange pairs can be in a coherent chain and logically related to the users need, but at some point the response from the chatbot might be out of place, breaking the chain of logical messages between the chatbot and its user. Number of coherent conversation turns refers to how many of these coherent question-answer pairs are in succession. Coherence of course is subjective and decided based how this metric is used. (Chakrabarti & Luger 2015, 6881-6882.)

Finally, percentage of successful resolutions simply refers to what percentage of interactions with users ended with the user's questions answered by the customer service chatbot. The last three of these metrics, quantity, relation and manner maxims are evaluated using subjective evaluations. (Chakrabarti & Luger 2015, 6882.)

In conclusion, the seven evaluation metrics are clearly different in nature, as the first three of them could be given numerical values based on a transcript of an interaction, albeit a rating has a subjective element. The metric about quality based on the quality maxim by Grice could be considered a boolean metric, as a message from the chatbot is either factually correct or not. The subjective metrics on the other hand could be rated for example on a 1-5 scale. The results from these seven metrics could then be combined into a total scoring for an interaction between a chatbot and its user.

3.6 Anthropomorphism

3.6.1 Overview

Anthropomorphism, in the context of this thesis the attribution of human-like qualities to a virtual assistant or chatbot, is a topic that touches on several fields of research, like HCI, psychology, marketing and computer science (Schanke et al 2021, 738).

Nass et al. (1996, 675-676) first introduced a paradigm known as "computers are social actors" to the field of HCI in the 1990s. This paradigm is still commonly mentioned in newer HCI literature. Meaning of this paradigm is that people typically revert to social norms and identity a computer (or more specifically, a chatbot) as a social actor when presented with typical social patterns such as engaging in dialogue and turn taking (Schanke et al. 2021, 738).

Schanke et al. (2021, 739-740) studied anthropomorphism in chatbots specifically by analysing user's reactions to specific social features a chatbot might utilize: social presence, communicative delay and humor.

Social presence is a measure of how a chatbot is presented to a user and how interactive is its user interface. If a chatbot has a human name for example, it has higher level of social presence. The greeting a chatbot gives it user at the start of an interaction can also increase social presence. Language used by the chatbot can also increase perception of social presence, for example by using particularly polite, informal or social language. (Schanke et al. 2021, 739.) Holtgraves et al. (2007, 2163–2174) studied how variables in a chatbot's use of language can affect user's perception of the chatbots human-like 'personality' characteristics. For instance, a chatbot that uses its human conversational partners name in the interaction is perceived to be more human-like, engaging and polite than a chatbot that doesn't. This points in the direction that small changes in use of language can affect perceived personality traits of a chatbot significantly. (Holtgraves et al. 2007, 2170.) Features supporting perception of social presence, such as human-like name, introduction and specific kind of language used could cause an anthropomorphic perception towards the chatbot by the user, thus supporting user engagement. These features could also result in inflated user expectations for the interaction, which can potentially result in user frustration. Perception of increased social presence can also be created by features of the chatbots chat interface itself, for example timestamps and read receipts. According to theory of social information processing, these types of cues can become relevant in an interaction in the absence of typical face-to-face cues. In general, features that increase perception of social presence can lead to higher sociability between a chatbot and its user, but also potentially hamper a task-oriented approach causing interactions to last longer. Chatbots are often used as self-service technology due to their perceived convenience and quickness. Optimizing chatbots for social presence could have a counterproductive effect if it causes interaction times to increase, because longer interaction times might negate the perceived quickness of chatbots and reduce customer satisfaction. On the other hand, higher perceived social presence can increase customers willingness to stay engaged with a chatbot until their matter is successfully resolved, potentially saving time from human-to-human customer service channels. (Schanke et al. 2021, 739-740.)

Another method that could potentially create anthropomorphic perception are slight communication delays. This means the chatbot is capable of practically instant response to its human user’s messages, but delays replying in order to seem more human-like, as humans take some time to type out a message. Much like with features enhancing perceived social presence, communication delays could work towards different outcomes. On one hand these delays could enhance levels of trust and lead to better user satisfaction, but on the other hand these delays could impede on the interaction by slowing it down and possibly cause the user to not resolve the problem at hand. (Schanke et al. 2021, 740.)

Humor can also support anthropomorphic perception, but could also have counterproductive effects on user experience, just like social presence or communication delays. Humor is highly context dependent, and thereby requires some nuance when it is used. It is unclear if humor generally increases user satisfaction or if it hinders the user’s experience. Further research is needed to better understand effect of humor on interactions. (Schanke et al. 2021, 740.)

3.6.2 Conversion and offer sensitivity

Araujo (2018, 187-188) found that using design features like using human name and human-like language increased user’s perception of human-likeness or anthropomorphism.

Schanke et al. (2021, 736) studied effects of anthropomorphic features in customer service chatbots on transaction conversion, the ratio of successful human-chatbot interactions and total interactions, and business outcomes more broadly. In general, anthropomorphic features can affect social outcomes of chatbot-human interactions, for instance the users trust towards the chatbot. Customers trust towards a chatbot and satisfaction with a service provider are considered mediating factors that affect transaction outcomes when chatbots are used in the customer interface. However, even if customers have a high level of trust towards a chatbot, they might still perceive inefficiencies at the same time, potentially leading them to a different service provider. Nevertheless, anthropomorphic features tend to have positive effects on successful transaction outcomes. (Schanke et al. 2021, 747-748.) Feine et al. (2019, 138) generally agrees with the proposition that human users tend to respond socially to chatbots that express social cues like small talk, gender or social gestures. Additionally, there is evidence to suggest humans respond subconsciously to any social cues from chatbots, no matter how simple or rudimentary these cues might be (Heine et al. 2019, 139).

Anthropomorphic features can also increase a customer's *offer sensitivity*. Offer sensitivity in this context means that when sufficient degree of anthropomorphism is present in a chatbot, human users of a customer service chatbot begin to scrutinize offers made by the chatbot more carefully than they would if the chatbot expressed lesser anthropomorphism. Reasons for this could be that offers received from humans are perceived as having more potential for price gouging or that they are considered more inconsistent, thereby justifying further scrutiny and attention by a human user. (Schanke et al. 2021, 738.) Schanke et al. (2021, 749) concludes that in general, anthropomorphism positively affects overall transaction conversion when human users interact with chatbots. However, increased offer sensitivity can be an unintended consequence of anthropomorphism, one that could have either positive or negative business consequences (Schanke et al. 2021, 749).

3.6.3 Request compliance

Adam et al. (2020, 428) studied the effect of chatbots verbal anthropomorphic design cues on *user's request compliance* in customer self-service situations. User's request compliance refers to customer's willingness to comply with requests from the chatbot. Customers in request compliance situations typically lack time to understand, evaluate and reply to requests, therefore lacking time to make rational decisions. This causes individuals to typically resort to heuristics, learned rules for decision making, to decide their options. The degree of anthropomorphism expressed by a chatbot might make customers more inclined to transfer behaviours learned in human to human interactions to interactions with chatbots. Additionally, users are often motivated to be consistent with past behaviour from the perspective of other social actors. User's self-consistency can be a factor in request compliance situations, as users might make seemingly irrational decision to stay consistent with earlier interactions. (Adam et al. 2020, 430-432.)

Customers willingness for consistency could be advantageous as chatbots giving an impression of social presence can be focused on user's heuristics learned from not only interactions with chatbots, but also human to human interactions. One concrete example of anthropomorphic design cues working to increase customers request compliance is customers complying with a request for service feedback. (Adam et al. 2020, 430-432.) Customer behaviour can also be modified by compliance techniques, such as presenting small requests followed by more significant request, utilizing customers tendency to be consistent with earlier actions. Customers can act first and then form their attitudes based on their actions, favouring their previous course of action and thus positively affecting their future attitude towards said course of action. These types of techniques originating from marketing could be effective because chatbot can be unconsciously treated as social actors. (Adam et al. 2020, 432.) These findings seem to imply that chatbots with anthropomorphic features have some underutilized potential in customer self-service situations, particularly in offering customers additional product and services or asking for feedback that would become useful data.

Adam et al. (2020, 432) experimented with the direct effects of chatbot’s anthropomorphic design cues (ADC’s) on user compliance in customer self-service situations. The framework for the experiment can be seen in figure 7. Social presence expresses user’s perception of the chatbot as a social actor and was measured with a yes/no question to a chatbots user. ADC’s used were features of a customer service chatbot like identity, small talk and empathy. The same experiment also studied direct effect of “foot-in-the-door technique” (FITD technique), an example of compliance techniques, on user compliance, as well as mediating and moderating effects of social presence on user compliance. FITD technique’s effectiveness was measured by requesting a larger favour from a customer (filling out a survey), either directly or after asking for a smaller favour (rating an interaction 1-5), to determine whether asking for a small favour makes a customer more likely to comply with larger requests. Both ADC’s and FITH technique were found to have direct positive effect on user compliance, supporting the notion that humans can attribute human-like behaviours and characteristics to customer service chatbots as the relationship between companies and consumers gradually becomes dominated by technology. Additionally, social presence was found to be a significantly mediating factor driving the effect of ADC’s on user compliance. The effect FITD technique has on user compliance was found to not be moderated by social presence. (Adam et al. 2020, 432-437.)

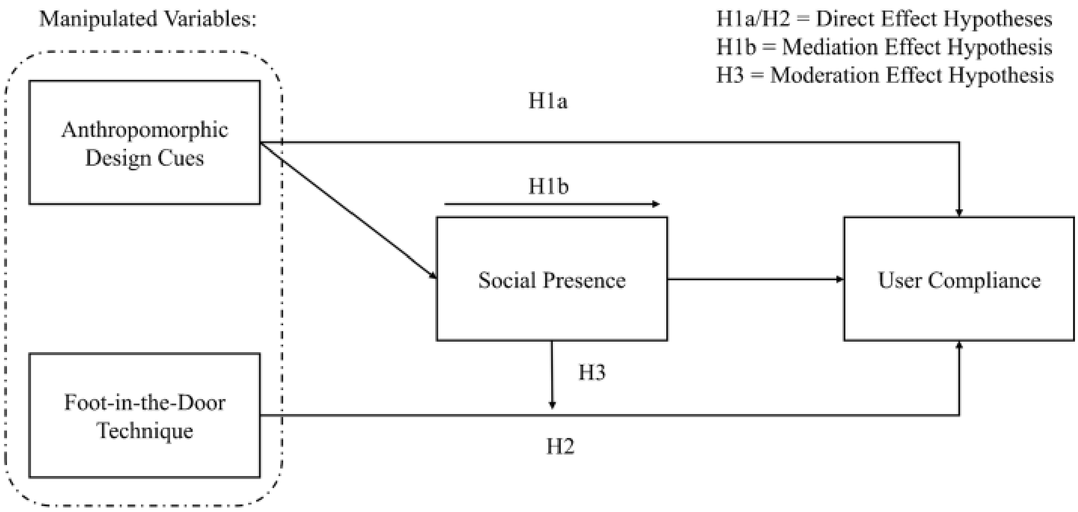

Figure 7: Framework for experiment on user compliance (Adam et al. 2020, 433).

Adam et al. (2020, 438) concludes that potential for shaping relationships between companies and customers using anthropomorphic features of chatbots is largely unexplored in relevant literature, but for few exceptions. Additionally, the experiment seems to suggest that humans can acknowledge chatbots as a source of persuasive messages. No claims are made whether chatbots can be more or less persuasive as humans, but it appears human user's degree of compliance with customer service chatbots is dependent on the techniques used in the interaction. These finding also lead to an assertion that fooling customers to believe they are interacting with a human customer service agent is not necessary, instead strategies can be implemented to make chatbots more likable using anthropomorphic cues. (Adam et al. 2020, 438.)

3.6.4 Message interactivity

Findings by Go & Sundar (2019, 314) point out that a chatbots ability to provide *message interactivity* is an essential factor in determining behavioural and attitudinal outcomes in human-chatbot interactions. An experiment testing simple chatbots with either high or low message interactivity showed that chatbots with high message interactivity were perceived as having higher social presence, contingency and perceived similarity to humans. It is asserted that simple changes in chatbots conversational styles could increase perceived contingency and social presence, these perceived properties can then contribute to positive attitudes and better outcomes in human-chatbot interactions. Additionally, the experiment showed that visual cues in chatbot's user interface, such as a simple human-like icon for the chatbot, have a compensatory effect on perceived message interactivity and vice versa. (Go & Sundar 2019, 314.)

3.6.5 Business outcomes

Araujo (2018, 183) studied the connection between anthropomorphic features in chatbots and company-related outcomes, like customer's attitudes towards and satisfaction with a company, as well as possible emotional connection consumers might feel with a company after interacting with a customer service chatbot. A distinction was made between *mindful anthropomorphism* and *mindless anthropomorphism*. The former was measured by inquiring chatbot users whether they perceived a chatbot to be more human-like or machine-like and natural or unnatural. The latter was measured by asking users to evaluate chatbot's features like likableness, sociability and friendliness. (Araujo 2018, 186.)

The study found that chatbot's anthropomorphic design cues have no significant effect on customer's attitudes towards a company that utilizes said chatbot. Same applies to customer's satisfaction with a company. However, the study found that anthropomorphic design cues do have a significant effect on customer's emotional connection with a company. (Araujo 2018, 187-188.)

Results also indicate that users perceive anthropomorphism differently depending on whether a chatbot is introduced as an artificial intelligence or not. Users tended to perceive lower level of mindless anthropomorphism if the chatbot was framed as an AI, compared to the chatbot not initially framing itself as an AI. (Araujo 2018, 188.) This could indicate that framing a chatbot as AI when initiating an interaction could have certain downsides.

The finding that anthropomorphic design cues can affect emotional connection acts as early evidence to suggest that chatbots designed to mimic humans can be a positive influence on relationship building between a business and its customers (Araujo 2018, 188). However, more research is needed to further understand the relation between anthropomorphism and emotional connection.

It can be concluded that anthropomorphic design cues and message interactivity can lead to positive customer outcomes when implemented in customer service chatbots. These effects on customer outcomes in encounters with customer service chatbots do not seem to reflect on customer's views on company's that utilize these chatbots. Additionally, there is evidence to suggest anthropomorphic design features have potential in the customer service context that warrants further research to fully understand.

3.7 Natural Language Processing

3.7.1 Overview

Natural Language Processing, or NLP, could be considered a field within computer science concerned with learning, understanding and producing natural language (Aleedy et al 2019, 1). NLP is considered a large research field that has seen much progress over the last decades, however some of the fundamental challenges in this field remain the same as they were as research into NLP began some 50 years ago (Barriere 2016, 1).

Rule-based natural language processing is an early method of NLP focused on small sets of example sentences, this method relies heavily on manual programming (Mitchell 1995, 10052).

Statistical natural language processing methods were originally an application of distributional analysis, a research field which reached prominence in the 1950s. These methods were based on the discovery that language can be deciphered by analysing distributional pattens of linguistic entities. Figure 8 demonstrates a "conventional" or rules-based NLP method in the a column and an example of statistical NLP methods in the b column. (Mitchell 1995, 10052-10053.)

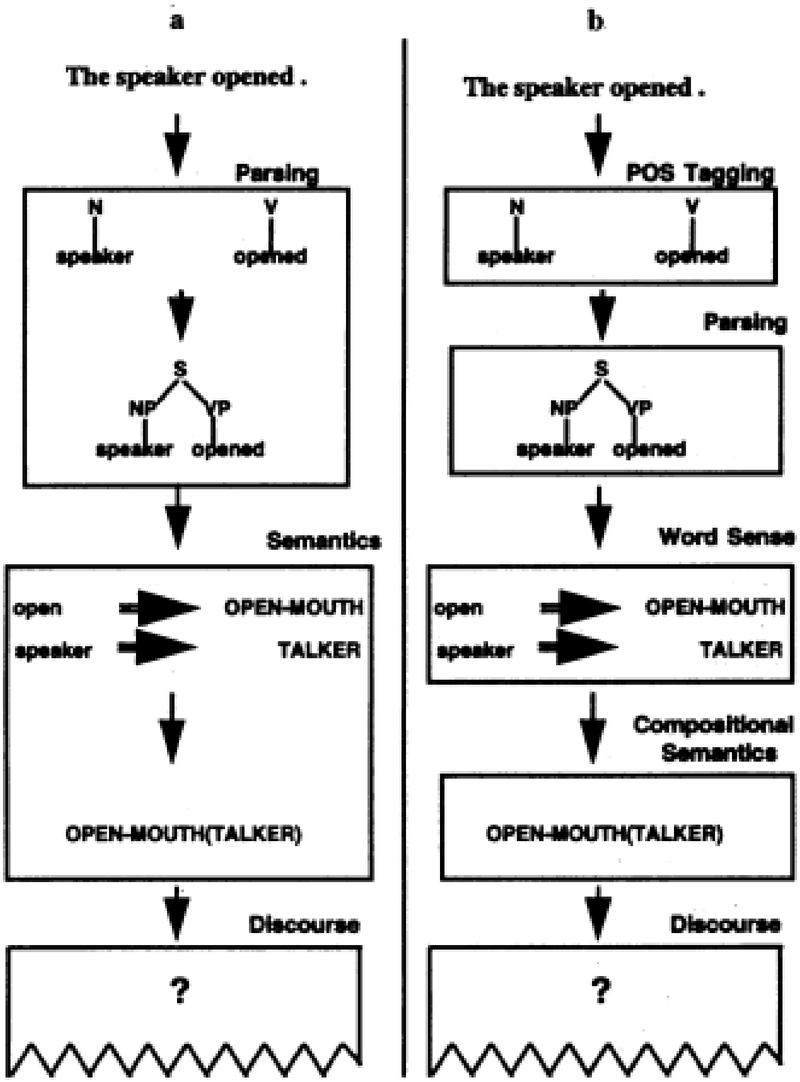

Figure 8: Illustration of conventional NLP process (a) and statistical NLP process (b) (Mitchell 1995, 10053).

This early model of statistical NLP from 1995 contains several phases to achieve usable understanding of a user utterance. The first phase is POS tagging, meaning part-of-speech tagging, followed by stochastic parsing, word sense and finally compositional semantics. (Mitchell 1995, 10053.)

Part-of-speech tagging means mapping incoming words in a way that puts each word into correct context as part of the sentence or piece of text that it is a part of. It is essential to distinguish words that could potentially be several different parts of speech. In many languages, such as English, it is common to use words in multiple parts of speech. For example, the word “can” could be used either as modal, verb or noun. POS tagging will attempt to classify each word based on its role using the context of the utterance fed into it. Classifying each words function correctly is essential so that correct grammatical structure can be assigned for semantic analysis later on. (Mitchell 1995, 10053-10054.)

In stochastic parsing the goal is to assign grammatical structure to a sentence correctly to the maximum extent possible. This means subdividing a sentence to the grammatical parts that constitute it by, for example, analysing similar sentences in a database of example sentences. In the 1990's, algorithms performing stochastic parsing faced a few major problems. Firstly, the algorithm has to consider a large number of different parameters when it is attempting to classify words into different categories, for example, if there is n number of different categories the algorithm has to consider n^3 number of parameters. Secondly, an algorithm might get mis-trained by the training data, example sentences, used to make it. (Mitchell 1995, 10054.)

Word-sense means making sense of what particular words mean in their context. Finally, compositional semantics form a coherent picture of each word's meaning in context, so the sentence can be understood. In the field of compositional semantics, statistics are an important tool. Statistical tools can be used, for example, to classify a verb's meaning correctly based on context. (Mitchell 1995, 10057-10058.)

3.7.2 Natural Language Understanding

Related to NLP is the field of *Natural Language Understanding* or NLU. NLU could be defined as transforming text to deeper representation so that logical reasoning may occur, a task that could be considered to fall in the realm of artificial intelligence more than NLP (Barriere 2016, 1).

Knowledge-based NLU is focused on intrinsic denotative meanings that are associated with natural language text, as opposed to using metrics such as word co-occurrence. Knowledge-based NLU aims to detect semantics expressed in a subtle manner. This means analysing concepts that do not directly express relevant information, but are linked implicitly to concepts that do express relevant information. There is a wide spectrum of issues related to knowledge-based NLU systems. (Cambria et al. 2016, 1.)

One such issue related to knowledge-based NLU has to do with Open Knowledge Extraction or OKE, which means extracting information from natural text and presenting it in a format that is machine readable. This is accomplished using unsupervised open-domain techniques. There is lack of established standards for converting natural language into machine readable format, so that open domain knowledge bases could be utilized to fullest potential. (Gangemi et al. 2015, 33-34.)

According to McShane (2017, 43-45) there are the following misconceptions about NLU that pose challenges to NLU as an independent field of research from NLP. Firstly, automating language understanding is often perceived to be challenging and expensive, so research to this subject has dwindled. Secondly, there exists a perception that knowledge-based Natural Language Understanding methods have little potential, instead so-called statistical NLP methods have taken a central role in research. Statistical NLP methods fall under NLP research, causing NLU-specific research, such as researching knowledge-based NLU methods, to have little focus. The author proposes that if research into knowledge-based NLU methods had received even a part of the resources devoted to research on statistical NLP systems, the research field of NLU might be much more advanced in present day. Thirdly, NLU in general is often considered to be an extension or subfield of NLP. NLU could also be considered to be more related to robotics than NLP research. The reason for this is that just like in the field of robotics, NLU work in specific domain, attempting to accomplish specific tasks when the agent has the required knowledge and reasoning ability. It is proposed that the term NLU should only be used when discussing deep understanding of text. Fourthly, NLP and NLU are sometimes considered competing research field, while they should be considered complimentary. Lastly, NLU, especially in the sense of deep understanding of language, is often considered an unrealistic goal. In general, NLU necessitates an integrated approach to processing text, where text can be analysed with multiple language analysis tools to achieve good results, that would be difficult to achieve with task-driven or method-driven approaches commonly used in NLP. (McShane 2017, 43-46.)

3.8 Cloud platforms for chatbot development

3.8.1 Market solutions

Adamopoulou & Moussiades (2020, 380) specified several major cloud platforms using machine learning that allow developers to create chatbots that are able to process natural language as an alternative to programming chatbots from the ground up. Examples of such platforms are Google’s DialogFlow, Facebook’s Wit.ai and Microsoft’s Luis.

One example in the Finnish market of cloud-based platforms for building or customizing chatbots is the software company Giosg's Bot Builder, which is advertised as a code-free way to create chatbots (Giosg 2021).

Zubani et al. (2022, 1) conducted a case study into different cloud-based platform that enable virtual assistants to be developed relatively rapidly. Most of the services available in the market are geared towards recognizing user's *intent*, thus they could be classified as intent-based chatbots. These services provide the necessary Natural Language Understanding capability to interpret user intentions. In addition to *intent*, these services use parameters present in a chat such as names, numbers or quantities, referred to as *entities*, to form replies to their users that aim to help their users. (Zubani et al. 2022, 5.) No prior knowledge of NLU is generally required with many market solutions. Many market solutions are marketed as fast to deploy and need small amount of examples instead of large training datasets. (Zubani et al. 2022, 3.)

The case study evaluated four most used cloud-based platform's performance in real world use. The platforms evaluated were IBM's Watson, Google's DialogFlow, Microsoft's Luis and Facebook's Wit.ai. All of these platforms have a web-based user interface for building chatbots and API support, however, the platforms differ in how they process intent and entities. Watson and DialogFlow also provide software development kits (SDK's) to a significant degree with their solution, allowing developers of new chatbots to use different programming languages or have limited programming ability in general. All of the services in question have ability to store recognized intents and entities from conversations and reuse them, in effect using machine learning. (Zubani et al. 2022, 5-6.) It was found that the evaluated cloud-based solutions were typically able to ambiguously predict the user's intention, on average it was found the user's intention was identified in 79% of cases. A common feature among the platforms studied is that when the chatbot is not able to discern the user's intent, it produces a disambiguation message that gives the user a list of probable intentions to choose from, this leads to a successful conclusion of interaction for majority of interactions. The study asserts that modern cloud-based platforms for chatbot development have noticeably reduced times required to develop chatbots, primarily due to reduction in effort for data collection and preprocessing. (Zubani et al. 2022, 16-18.)

3.8.2 SaaS

Cloud-based solutions sold to a virtual assistant's user organization using a software-as-a-service or SaaS model could have potential benefits to their user organizations. The user organization does not need to develop or maintain any software with the SaaS model, as the solution runs in the cloud. For instance, such a solution can consist of a system core, web client/mobile client and modules that can communicate with each other, creating a modular solution to each user's needs. The system core in such a model works as the virtual assistant's source for access to relational databases and file systems, as well as manages various modules that expand the functionality of the virtual assistant. Modules in such a solution can have functions such as formulating and providing answers to questions, recording interactions between a virtual assistant and its user, as well as helping the user organization of the virtual assistant manage the system. (Kuznar et al 2016, 285-288.)

According to Radziwill & Benton (2017, 6) when a SaaS model is used, the service provider that provides the virtual assistant platform typically is responsible for demonstrating the effectiveness of the virtual assistant solution, while the implementer (user organization) of the virtual assistant is responsible for ensuing customer satisfaction with the solution.

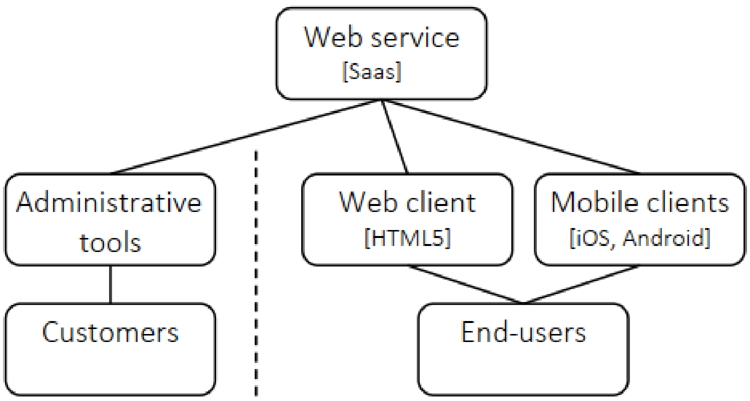

Figure 9: SaaS solution for virtual assistant (Kuznar et al. 2016, 285).

3.9 Chatbots in insurance industry

3.9.1 Overview

Riikkinen et al. (2017, 1145) has researched customers value creation from the perspective of the insurance industry specifically. In the paper, three shifts are laid out that affect the insurance industry. Firstly, chatbots are increasingly being used in service settings, raising interest in understanding how they could create value. Secondly, the academic perspective on value is shifting. Value used to be considered an outcome, but is increasingly thought of as a process, creating a need to understand how customer's own processes create value and how can these processes be supported. Thirdly, exploiting data to create value for customers is increasingly researched, as data can not only be used for company's own processes but to help its customers in creating value. (Riikkinen et al. 2017, 1147.)

Related to these shifts is the idea that customer data can not only be used for tasks like building customer risk profiles or identifying most profitable customers, but also for helping the customer with facilitating the customer's value-creating processes (Riikkinen et al. 2017, 1150). It is not relevant what specific products or services a client purchases, as the client is ultimately purchasing resources and processes to support its own value creation (Riikkinen et al. 2017, 1149). The research points out the following main points. Firstly, chatbots can help create value from the customers perspective, by using customer's data not only to the benefit of the company hosting a chatbot but to facilitate customer's own value creating processes. Second, companies can influence customers value creating processes by providing them with a new resource, namely the chatbot. (Riikkinen et al. 2017, 1155-1156.)

In the following section, the role of customer data in customers value creating process if further explained.

3.9.2 Role of customer data in customer's value creation

Grönroos & Ravald (2011, 15) introduce the concept of reciprocal value creation in the relationship between a customer and service provider. Reciprocal value creation means that both sides are able to create value for themselves as a result of their interaction. In this process, a service provider should help its customer in creating value for itself, while naturally also creating value for the service provider itself to stay viable. (Grönroos & Ravald 2011, 16.)

The role of customer data in customer relationship management is evolving, notably customer data is increasingly being used externally to provide support to the customer instead of being used only internally, the following four themes describing how customer data is used (Saarijärvi et al. 2013, 589):

- Customer loyalty could be improved by reconfiguring customer data that a firm has in its disposal.
- Differentiation in the markets could be achieved by reconfiguring customer data. A firm can set itself apart from competition and even attract new customers, in the process a firm can position itself more as a service company.
- Firm values are also related to reconfiguring customer data, as it can be a way for a company to express its values and carry out its strategy.
- Firm image could also be affected by reconfiguring customer data, making customers perceive the firm with a more positive perspective due to additional services that customer data enables.

Reconfiguring customer data in this context means shifting from thinking of customer data as an internal asset to an asset that could be shared with a customer, helping the customer create added value and taking appropriate steps to use customer data in this way. (Saarijärvi et al. 2013, 589-593.)

Research by Saarijärvi et al. (2014, 529) further elaborates on this emphasis on customer data as a part of customer's value creation process. When customer data is used for customers value creation, the role of the firm moves from being a passive facilitator more to taking a role as active supporter of the processes that create value for a customer (Saarijärvi et al. 2014, 529). Saarijärvi et al. (2014, 531) calls this concept of reverse use of customer data in service-based business models, in essence meaning converting customer data to relevant information that can have a direct impact on customers value creating process.

4 Results

Interviews covered the general topics planned in the interview guide. The following chapters will go through findings from the interviews in the categories relevant to this research while reflecting upon the literature review.

4.1 Virtual assistants at If

4.1.1 Overview

The interviews gave insight into how virtual assistants are used at If. The Finnish branch of If has three separate virtual assistants that are used for different purposes, one for customer service on the B2C side, another for customer service on the B2B side and the third for internally supporting customer service persons. As of about two years ago, all of these virtual assistants are based on the same platform, provided by a Norwegian company boost.ai. Before that, the B2B and B2C sides of the company had separate "siloed" chatbot solutions in use, that were being developed independently from each other.

The use of virtual assistants used to be more apparent to the customers visiting If's website, as the chat window for the virtual assistant would automatically pop up to any visitor's screen on the front page of the website. However, the visibility of the virtual assistant has been recently reduced so that a visitor has to select the virtual assistant from a symbol on the periphery of the front page. The reason for this, according to one of the experts interviewed, is that virtual assistants are presented to the customer as one method of contact among others, other methods being contacting human customer service representatives through various means. The performance of the virtual assistants is therefore not behind this change. As pointed out in the literature review, poor performance of chatbots or virtual assistants has in some instances prompted these services to be removed from use altogether, or their use might have been curtailed (Gnewuch et al. 2017, 1-2).

4.1.2 Human-Computer Interaction

Use of virtual assistants is limited by issues related to Human-Computer Interaction, as the literature review points out.

The interviews made it clear that virtual assistants are a polarizing issue for customers. Some customer segments accept interacting with a virtual assistant, even preferring its benefits, like 24/7 availability to customers. On the other hand, some customer segments are very conservative and do not like interacting with them. Generally speaking, younger customers seem to exhibit more positive attitudes towards virtual assistants, while older customers hold more negative attitudes. No concrete numbers came up when it comes to the correlation between age and positive/negative attitudes towards virtual assistants in the interviews. However, it became clear that the experts perceive correlation between younger age of users and more positive attitudes towards virtual assistants, and conversely between older age and more negative attitudes. It was also pointed out that a relevant number of exceptions to this rule exist. One of the experts emphasized that the evolving demographics of the customer base support using virtual assistants going forward, with the portion of customers willing to interact with virtual assistants constantly growing.

These findings appear to be related to problems outlined in the literature review, where some customers might not accept interacting with a virtual assistant. It is possible that the virtual assistants are lacking in anthropomorphic features and therefore also lacking in engagement with some customer segments. Another explanation for lack of engagement with some customer segments is that anthropomorphic features might be ineffective for these customer segments because these customers might not consider virtual assistants as social actors. It is possible older customer segments specifically might be less willing to engage with the virtual assistants because they are willing to engage with a social actor but don’t consider virtual assistants as social actors.

4.1.3 Natural Language Processing

Natural Language Processing was covered in the interviews by first enquiring how is the responsibility for creating necessary vocabulary data shared between the user of the virtual assistants and the software provider. If has full responsibility for creating the vocabulary used by the virtual assistants. The software provider for If’s virtual assistants, the Norwegian boost.ai, only provides a platform on which If then trains the virtual assistant to meet its requirements.

Interviews also revealed that firstly, the use of virtual assistants in other If countries, Sweden, Norway, Denmark and the Baltic counties, is clearly behind Finland. Notably, virtual assistants are only used in the customer interface by the Finnish branch of If. Secondly, projects are underway in other If countries to start using virtual assistants in the customer interface. The Swedish branch of If has recently started using a virtual assistant in an internal support role for customer service persons, with good success. Interestingly, each If branch is developing virtual assistants from the ground up, including developing the Natural Language Processing ability for each branches virtual assistants to meet unique needs. The Finnish virtual assistants are not used as a platform to build upon for other countries Natural Language Processing ability because the insurance products and languages spoken are too different. As remarked by one of the experts interviewed, translating the vocabulary from one language to another does not work in practice, particularly because of different insurance products and different customer needs between countries. Therefore, each country's branch is developing a registry of vocabulary tailored to their specific needs.

Based on these findings, different requirements and products for different countries branches, along with language related challenges, appears to be a significant challenge for scalability of virtual assistant technology in multinational companies.

Another issue that come up with the interviews is the difficulty of developing the virtual assistant's natural language understanding using customer feedback. As is typical for similar bots, the virtual assistants window allows the customer to react to each individual message from the virtual assistant by "thumb-up" and "thumb-down" symbols or written feedback. However, so called false positives, where a customer gives sarcastic feedback indicating satisfaction in the virtual assistant, but is not satisfied, are a notable issue. This kind of feedback would be problematic if the virtual assistant used some type of machine learning to improve its language processing and reply process. The virtual assistants at If do not use any type of machine learning due to issues with the reliability of customer feedback. Instead, conversations with customers are manually read and rated using tools provided by boost.ai in order to improve the virtual assistants Natural Language Processing ability over time. The rating tools are closely related to the attributes of effective communication covered in section 3.5.3.

4.1.4 Liability issues

The interviews inquired about any issues related to liability between software provider (boost.ai) and the virtual assistant's user (If) in situations where the virtual assistant gives either wrong information to the customer or causes a missed opportunity. An example of the former would be a situation where the customer does not submit a claims application in a situation when doing so would be appropriate due to misleading advice, while an example of the latter would be a situation where a customer does not take a necessary insurance because of lacking information provided by a virtual assistant.

The virtual assistants at If are built in-house from the ground up, with the software provider only contributing a platform on which the virtual assistants run on and tools to rate interactions and improve the responses over time. Therefore, the software provider does not bear any liability when it comes to interactions with customers.

In terms of claims related issues, it became clear in the interviews and through personal work experience that If and other insurance companies are cautious when it comes to any customer enquiries, through any channels, about claims issues. Insurance companies typically seek to avoid giving any advice that could be interpreted as taking a stance on whether an insurance claim submitted by a customer would be denied or approved. The virtual assistants do not answer any claims related questions at all, instead offering to transfer the chat to a human customer service representative.

In conclusion, the interviews did not uncover any liability related challenges when it comes to the relationship between If and its software provider for the virtual assistant.

5 Discussion and conclusion

5.1 Overview

Virtual assistants have become commonplace in recent years. Notably, time and effort needed to begin use of a virtual assistant in an organization's customer interface seems to have been reduced in recent years.

Practical issues seem to limit use of virtual assistants in some contexts. Multinational companies can experience scalability issues if virtual assistants are introduced in different countries and/or branches. This presents management an information systems dilemma, whether to develop siloed solutions for differing requirements or use the same platform while separating development efforts to a degree to meet differing requirements.

Factors that hinder introduction and use of virtual assistants to an organization's customer interface could be categorized as either social or technical.

Social factors involve customer's preconceptions and attitudes towards virtual assistants. Virtual assistants that are appear human-like or anthropomorphic might cause users expectations to rise initially as they interact with the bot, potentially leading to later disillusionment if the virtual assistant can not meet these expectations. Organizations should consider what level of anthropomorphism balances positive and negative customer reactions. On the other hand, some customers prefer not to interact with virtual assistants at all, but human users seem to become more accepting towards virtual assistants over time.

Technical factors are often related to virtual assistant's NLP abilities. Availability of third party platforms to run virtual assistants on is higher for more common languages. In the Finnish market, language-related issues appear to be a particular concern. In theory, machine learning can be used to make training of virtual assistants more efficient, but as empirical research revealed machine learning is sometimes not used on purpose, instead much slower manual methods might be preferred for training NLP ability.

In the insurance industry, virtual assistants can face certain unique challenges related to liability issues. Customers might inquire about personal insurances or claims cases as they interact the virtual assistant. In such cases, problems can arise with liability if the virtual assistant gives any product recommendations or implies any specific stance on the insurance company's behalf on a claims case. Insurance companies might purposefully limit the role and potential of virtual assistants as customer service agents due to such issues.

If a virtual assistant is integrated to an insurance company's CRM system and other relevant databases, GDPR compliance can present challenges. If a virtual assistant can disclose customer data, customers would have to log in using strong identification (like online banking) to plausibly ensure compliance with GDPR. Additionally, if the virtual assistant runs on a third-party platform using machine learning, sharing of potentially identifying data with said third party should be carefully considered.

5.2 Limitations

This research offers a limited perspective on real world use of virtual assistants, empirical research only covered one company in the insurance industry. Expert interviews were conducted with two persons involved with the development and quality assurance of If's virtual assistants and their perspectives might be limited, particularly in terms of managerial perspectives.

Literature review covered relevant fields of scientific research, however due to the scale of this thesis only limited picture of relevant trends within these field of research was formed.

5.3 Future research

Fields of research like HCI and NLP have a long history of research spanning decades, however, some emerging topics related to virtual assistants warrant further future research to better understand the potential virtual assistants have to offer.

Effects of anthropomorphism is still poorly understood in the context of customer service virtual assistants. Human user's perception of social presence, for example, is a topic that has only in recent years been linked with disciplines like marketing techniques. As understanding of what creates perception of anthropomorphism or social presence in virtual assistants improves, research can further shift to potential added value virtual assistants are creating as customer service or sales agents. Virtual assistants appear to be increasingly considered agents in managing an organization's customer relations, but this role is still poorly understood.

More broadly, virtual assistant's role as part of utilizing vast amounts of customer data, so called "big data", is progressing with information systems research. Research examining how customer data could be used specifically by virtual assistants is limited as of the writing of this thesis. However, trends seem to be pointing in the direction of increasing use of customer data.

Further research is needed to understand what potential using customer data, for example information about a customer's insurance coverage, has when combined with emerging topics like high perceived anthropomorphism.

Machine learning is becoming more relevant in virtual assistant platforms currently in the market, theoretically enabling virtual assistants to better adapt to customer's demands with lesser expertise required on the hosting organization's part. Issues related to transparency and liability issues seem require further research to understand, as many market platforms seem to be lacking in general transparency about how customer data is collected, stored and handled. Liability issues arising from use of virtual assistants as well as the question of how liability is shared between organizations using virtual assistants and third party providers of software seems to be emerging issues that not sufficiently covered by exiting research.

5.4 Conclusion

Using virtual assistants in customer interface is at an in intersection of many scientific fields of research. Disciplines like HCI and NLP have progressed and emerging topics like anthropomorphism of virtual assistants in customer service context and machine learning to enhance language learning have appeared in recent years.

As the empirical research pointed out, the role of virtual assistants in organizations customer interface and management of customer relations is constantly evolving but also challenged by practical issues. For example, reliability of customer feedback is a clear challenge hindering use of machine learning and slowing down development and quality assurance of virtual assistants.

Trends in scientific literature seem to reveal many opportunities of virtual assistants, for example, added value could be created if virtual assistant take a stronger role as not only customer service but sales agents and managers of customer relations. Future research will reveal what practical issues emerge if use of virtual assistants continues to be more relevant going forward.

References

Adam, M. – Wessel, M. – Benlian, A. (2021) AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, Vol. 31 (1), 427–445.

Adamopoulou, E. – Moussiades, L (2020) An overview of chatbot technology. <https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31#Sec6>, retrieved 14.9.2021.

Aleedy, M. – Shaiba, H. – Bezbradica, M. (2019) Generating and Analyzing Chatbot Responses using Natural Language Processing. *International Journal of Advanced Computer Science and Applications*, Vol. 10 (9), 60-68.

Araujo, T. (2018) Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, Vol. 85 (1), 183-189.

Barriere, C. (2016) *Natural language understanding in a semantic web context*. Springer International Publishing, New York, NY.

Ben Mimoun, M. – Poncin, I. – Garnier, M. (2012) Case study - embodied virtual agents: An analysis on reasons for failure. *Journal of Retailing and Consumer Services*, Vol. 19 (1), 605–612.

Cambria, E. – Björn, S. – Yuqin, X. – Bebo, W. (2016) New avenues in knowledge bases for natural language processing. *Knowledge-Based Systems*, Vol. 108 (1), 1-4.

Castro, F. – Tedesco, P. – Alves, H. – Quintine, J. P. – Steffen, J. – Oliveira, F. – Soares, R. – Santos, A. L. M. – Silva, F. (2018) Developing a corporate chatbot for a customer engagement program: A roadmap. In: *Proceedings of the 14th International Conference on Intelligent Computing*, 400-412.

Chakrabarti, C. – Luger, G. F. (2015) Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics. *Expert Systems with Applications*, Vol. 42 (1), 6878–6897.

Chaves, A. – Gerosa, M. (2021) How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human-Computer Interaction*, Vol. 37 (8), 729–758.

Ciechanowski, L. – Przegalinska, A. – Magnuski, M. – Gloor, P. (2019) In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, Vol. 92 (1), 539-548.

Deng, M. – Wuyts, K. – Scandariato, R. – Preneel, B. – Joosen, W. (2011) A privacy threat analysis framework: Supporting the elicitation and fulfilment of privacy requirements. *Requirements Engineering*, Vol. 16 (1), 3-32.

Feine, J. – Gnewuch, U. – Morana, S. – Maedche, A. (2019) A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, Vol. 132 (1), 138-161.

Følstad, A. – Skjuve, M. – Brandtzaeg, P. B. (2019) Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. In: *Proceedings of International Conference on Internet Science*, 145-156.

Gangemi, A. – Recupero, D. R. – Mongiovi, M. – Nuzzolese, A. G. Presutti, V. (2015) Identifying motifs for evaluating open knowledge extraction on the Web. *Knowledge-Based Systems*, Vol. 108 (1), 33-41.

Botit, Giosg. <https://www.giosg.com/fi/ominaisuudet/bot-builder>, retrieved 5.12.2021.

Gnewuch, U. – Morana, S. – Maedche, A. (2017) Towards designing cooperative and social conversational agents for customer service. In: *Proceedings of the International Conference on Information Systems (ICIS)*, 1-13.

Go, E. – Sundar, S. S. (2019) Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, Vol. 97 (1), 304-316.

Grönroos, C. – Ravald, A. (2011) Service as business logic: implications for value creation and marketing. *Journal of Service Management*, Vol. 22 (1), 5-22.

Hasal, M. – Nowaková, J. – Ahmed, S. – Abdulla, H. – Snášel, V. – Ogiela, L. (2021) Chatbots: security, privacy, data protection, and social aspects. *Concurrency and Computation*, Vol. 33 (19), 1-13.

Holtgraves, T. M. – Ross, S. J. – Waywadt, C. R. – Han, T. L. (2007) Perceiving artificial social agents. *Computers in Human Behavior*, Vol. 23 (1), 2163–2174.

Janssen, A. – Passlick, J. – Rodriques-Cardona, D. – Breitner, M. (2020) Virtual assistance in any context - A taxonomy of design elements for domain-specific chatbots. *Business Information Systems,* Vol. 62 (3), 211–225.

Juntunen, M. (2019) *Automaatio, tekoäly ja chatbotit asiakaspalvelussa : Case-tutkimus: organisaatioiden näkemyksiä*. Master's thesis. University of Turku, Turku.

Kallio, H. – Pietilä, A-M. – Johnson, M. – Kangasniemi, M. (2016) Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, Vol. 72 (12), 2954-2965.

Kalloniatis, C. – Kavakli, E. – Gritzalis, E. (2008) Addressing privacy requirements in system design: The PriS method. *Requirements Engineering*, Vol. 13 (3), 241-255.

Khanna, A. – Pandey, B. – Vashishta, K. – Kalia, K. – Pradeepkumar, B. – Das, T. (2015) A study of today's A.I. through chatbots and rediscovery of machine intelligence. *International Journal of u- and e- Service, Science and Technology*, Vol. 8 (7), 277-284.

Kuznar, D. – Tavcar, A. – Zupancic, J. (2016) Virtual assistant platform. *Informatica*, Vol. 40 (1), 285–289.

McShane, M. (2017) Natural Language Understanding (NLU, not NLP) in Cognitive Systems. *The AI Magazine*, Vol. 38 (4), 43-56.

Mitchell, M. (1995) New trends in natural language processing: Statistical natural language processing. *Colloquium Paper*, Vol. 92 (1), 10052-10059.

Mori, M. (1970) The uncanny valley. *Energy*, Vol. 7 (4), 33–35.

Nass, C. – Fogg, B. J. – Moon, Y. (1996) Can computers be teammates? *International Journal of Human-Computer Studies*, Vol. 45 (1), 669 – 678.

Przegalinska, A. – Ciechanowski, L. – Stroz, A. – Gloor, P. – Mazurek, G. (2019) In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, Vol. 62 (1), 785-797.

Radziwill, N. – Benton, M. (2017) Evaluating Quality of Chatbots and Intelligent Conversational Agents. *Software Quality Professional*, Vol. 19 (3), 1-13.

Reshmi, S. – Balakrishnan, K. (2016) Implementation of an inquisitive chatbot for database supported knowledge bases. *Sadhana*, Vol. 41 (10), 1173–1178.

Saarijärvi, H. – Grönroos, C. – Kuusela, H. (2014) Reverse use of customer data: Implications for service-based business models. *Journal of Services Marketing*, Vol. 28 (7), 529–537.

Saarijärvi, H. – Karjaluoto, H. – Kuusela, H. (2013) Customer relationship management: The evolving role of customer data. *Marketing Intelligence & Planning*, Vol. 31 (6), 584-600.

Schanke, S. – Burtch, G. – Ray, G. (2021) Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, Vol. 32 (3), 736–751.

Shawar, B. – Atwell, E. (2007) Different measurements metrics to evaluate a chatbot system. In*: Proceedings of the NAACL'07 Workshop: Academic and Industrial Research in Dialog Technologies,* 89-96.

Shyam, S. – Saraswathi, B. – Jeeyun, O. – Haiyan, J. – Hyang-Sook, K. (2016) Theoretical importance of contingency in human computer interaction: Effects of message interactivity on user engagement. *Communication Research*, Vol. 43 (5), 595–625.

Snyder, H. (2019) Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, Vol. 104 (1), 333-339.

Zubani, M. – Sigalini, L. – Serina, I. – Putelli, L. – Gerevini, A. E. – Criari, M. (2022) A performance comparison of different cloud-based natural language understanding services for an Italian e-learning platform. *Future Internet*, Vol. 14 (2), 1-20.

Annex 1 Interview guide

The following topics were meant to direct the two interviews:

- The timeline of using virtual assistants at the company.
- Any differences in the different virtual assistant used by the company (the company uses separate virtual assistants for private customers and small & medium enterprises).
- Natural Language Processing related challenges in virtual assistant development.
- Measuring customer satisfaction.
- Effectiveness of solving customers problems, ratio of successful or unsuccessful resolutions to customer interactions.
- Liability issues with respect to the provider of the virtual assistant solution.
- Any missed opportunities in sales/claims matters related to the virtual assistant.
- Use of machine learning now or in the future for training the virtual assistant.
- Visibility of the virtual assistant on the front page of the company's website and its role among other methods to contact the company.
- The role of the virtual assistant in directing the customer to digital services.
- Any expectations regarding future development of the virtual assistant.