# A. Foundation

The first part of this thesis is divided into two sections and builds the foundation of the conducted research. The two sections provide an introduction and an overview of the research background of this thesis.

In the first section (A.I), the motivation for this work is presented. Afterward, the research gaps and research questions are derived and presented, followed by an outline of the structure of this thesis. Then, this thesis is placed in the field of Information Systems (IS) research, the research design is presented, and finally, the contributions of this cumulative thesis are discussed.

The second section (A.II), introduces the context of human-computer interactions. Conversational Agents (CAs) are then described and their application in service encounters is presented. Knowledge of the human-like design of CAs and its underlying theories is then introduced, relating to the ‘computers are social actors’ (CASA) paradigm and social response theory. The section further provides an overview of some of the imperfections associated with conventional CAs and their impact on service perception. Finally, the chapter will close by summarizing the related background knowledge by deriving a prior understanding of human-like design in the context of imperfect CAs, including human-like errors.

# I. Introduction

This section presents the research subject and agenda in the context of this thesis. The first subsection (I.1) highlights the motivational background and relevance of this research endeavor. In the second subsection (I.2), the research gaps are highlighted, and the research questions will be formulated. Then the structural approach of this thesis will be presented (I.3). Afterward the research context and design, as well as an overview of the anticipated contributions are provided (I.4).

## I.1 Motivation

Our world has been suffused with technologies that have profoundly changed how individuals live and interact in private and professional contexts. Specifically, in the service context, individuals are increasingly reliant on technology-based applications to interact with businesses via phone, email, and for instance via web embedded online stores. However, in most cases, interacting with a business means dealing with complex interfaces and entering data manually. Technological advancements in artificial intelligence are addressing this challenge by unlocking a new field of technology-based services that aim to simplify individuals' lives. One prime example of this are Conversational Agents (CAs).

CAs are defined as "software-based systems designed to interact with humans using natural language" (Feine et al. 2019, p.1.) and offer enormous improvements in the interaction between humans and technology. CAs are emergently known in society for interactions through the mode of text (e.g., *Woebot*, *Cleverbot*) or voice (e.g., Amazon's *Alexa*, Google's *Siri*). Furthermore, CAs can be embodied either digitally (e.g., Google's *Duplex*) (Leviathan and Matias 2018) or physically (e.g., Softbank's robot named *Pepper*) (Stock and Merkle 2018). CAs are valuable in various domains, for example in customer services (Verhagen et al. 2014), marketing and sales (Qiu and Benbasat 2009), and human resources (Diederich, Brendel, et al. 2020). Additionally, the technology-based service "employees" can address manifold requests independently and automatically, while overcoming the limitations of a human employee-based service encounter by being more appealing, efficient, and convenient to individuals than real human employees can ever be (e.g., limited time availability and capacity) (Zumstein and Hundertmark 2018). This offers businesses the potential to increase efficacy and reduce costs while increasing customers' service evaluations (Zumstein and Hundertmark 2018). Therefore, CAs are on the verge of overcoming and replacing traditional human-based service encounters and present a particularly interesting technology to be invested in and implemented by businesses (McTear et al. 2016).

One crucial aspect of implementing CAs is to provide them with a natural human-like appearance to enhance user perception (Araujo 2018). In this context, Nass et al. (1994) formulated the ‘computers are social actors’ paradigm (CASA). The paradigm explains how people communicate with media and machines that demonstrate social potential. Strongly related to CASA (Nass et al. 1994), Nass and Moon (2000) revealed that people reciprocate positive behaviors of CAs by behaving similarly in return, (e.g., disclosing personal information, being polite) formulating the social response theory. The theory suggests that humans mindlessly apply social rules and expectations toward computers even though they mindfully know that the computer is not a human (Nass and Moon 2000). One example of this is ELIZA, the first text-based CA (i.e. chatbot) in history (Weizenbaum 1966). ELIZA simulated a psychotherapist to serve its clients. The capabilities of ELIZA’s service were extremely limited, as the computer program was only based on hand-crafted rules to recognize and respond to the client’s requests by text (Shum et al. 2018; Weizenbaum 1966). Consequently, it appeared that the program was just able to respond to domain-specific requests and rephrased questions based on the users’ inputs. The technical imperfection led users to quickly realize that they were not interacting with a real human, but with an artificial computer-based program (Weizenbaum 1966). However, even though ELIZA’s capabilities were limited, users became extremely engaged with the CA. Therefore, up until the point users did realize that ELIZA was, surprisingly, an artificial program and not a human, they trusted personal information and interacted with the CA as one would expect a human therapist to do.

The example of ELIZA illustrates that even though technologies such as CAs are not capable of having feelings, independent intention, or any human motivation, humans elicit social signals and respond to them as they would do with a real human (Nass and Moon 2000). Even though it had long been known that people attribute human-like characteristics to simple objects (e.g., the importance of the relationship with one's favorite teddy bear into adulthood) (Heider and Simmel 1944; Winnicott 1953), technology had not previously been considered a part of this attribution.

Since the first appearance of CAs such as ELIZA, technological advances have continuously improved CAs, (e.g., natural language processing, embedding external knowledge bases) enabling them to handle complex user requests while providing satisfying services (McTear et al. 2016). However, CAs’ capabilities are still limited and succumb to frequent imperfections (e.g., not being able to understand or misinterpreting the users’ input) (Honig and Oron-Gilad 2018; Ben Mimoun et al. 2012). Existing research highlights that this leads to negative service evaluations (Diederich, Lembcke, et al. 2020; Honig and Oron-Gilad 2018; Mirnig et al. 2017; Rossi et al. 2017; Toader et al. 2020), and in extreme cases, the users’ discontinuation of the service (Ben Mimoun et al. 2012). Despite technological progress, it will be impossible to completely avoid or prevent every potential imperfection in the interaction between CAs and humans in the future, just as is the case in the interactions between humans (Go and Sundar 2019; Gregersen 2003; Ben Mimoun et al. 2012; Salem et al. 2013). Imperfections are simply inherent in the nature of life and are a frequent part of interactions with technology (Gregersen 2003; Ben Mimoun et al. 2012; Salem et al. 2013). This renders research on the design of CAs to prevent negative users' responses a thriving topic of interest.

Existing research illustrates the potential of human-like CA design (Araujo 2018; Feine, Gnewuch, et al. 2019; Seeger et al. 2018; Verhagen et al. 2014) to elicit positive user responses toward services (e.g., increased enjoyment (S. Y. Lee and Choi 2017), trustworthiness (Araujo 2018), and service satisfaction (Gnewuch, Morana, et al. 2018)). However, the research also illustrates contradicting effects (e.g., decreased perceived humanness, avoidance, skepticism) (Crolic et al. 2021; Hadi 2019; Seeger et al. 2018; Strait et al. 2015; Wagner and Schramm-Klein 2019). Additionally, prior research shows that human-like design errors of CAs can lead to an increased likeability and perception of humanness by the users (Mirnig et al. 2017; Salem et al. 2013). Hence, identifying knowledge on human-like design is of interest, advancing the need for both practice and research to understand, design, and define the future interaction between technology and humans in services. Therefore, this thesis concerns the design of imperfect CAs to gain an understanding of how a human-like design affects the users' perception and how human-like errors are perceived by users in a service context.

In the research context, the information systems (IS) discipline focuses on socio-technical phenomena. The sub-field of human-computer interaction (HCI) research explores the design and use of technologies at the interface between humans and computers (Gupta 2012). This thesis also contributes to the existing HCI research discourse by investigating the phenomenon of human-like designed imperfect CAs within the service context. Overall, three main aspects are synthesized in this thesis. First, CAs have the potential to replace humans in the near future, a notion complemented by a perspective on CAs' imperfections and human-like errors. Second, the context of human-like CA design draws on kernel theories such as CASA and social response theory. Third, the service context provides implications for research and practice on how to design CAs, specifically with imperfections (see Figure 3). To address them, this thesis primarily focuses on gaining insights into users' affective (e.g., emotions), cognitive (e.g., evaluation of services), and behavioral responses (e.g., treatment reaction) when interacting with CAs.

> *"As far as the customer is concerned, the interface is the product." (p. 5) – "If you want to create a humane interface, you must have an understanding of the relevant information on how both humans and machines operate (p. 6)."*
>
> *(Raskin 2000)*

## I.2 Research Gaps

The overarching goal of this thesis is to gain a better understanding of users' affective, cognitive, and behavioral responses to imperfect CAs and their human-like design, including human-like errors. Based on the findings, the thesis aims to provide implications for research and practice on how to provide CA-based services in the future. To this end, the research endeavor is split into four distinct parts, each synthesized into one specific research question. In the following, the four questions are derived based on the related literature and the need for research to contribute to the knowledge base of human-to-CA interaction.

In general, CA-based services are technology-based applications. However, from the users' perspective, they are mainly perceived as social encounters (Chaves and Gerosa 2020). This indicates that more than technical capabilities of CAs (e.g., responding accurately, always available) are required to meet the users' expectations (Araujo 2018; Seeger et al. 2018). Hence, even though CA-based services do not provide true human-to-human interaction, as traditional employee-based services would do, CA-based-service encounters aim to provide an interaction as natural as customers expect it to be in human-to-human interaction (Araujo 2018; Liao, Hussain, et al. 2018; M. McTear 2017).

Meeting customers' expectation takes a vital role in services, and therefore also in the interaction with CA-based services (Briggs et al. 2010). One prime example of meeting customer expectations in the interaction with a CA is the ability to design CAs by using so-called social cues (Feine, Gnewuch, et al. 2019). Social cues are used to provide computers the ability to behave with characteristics commonly associated with humans (e.g., avatar, name, gender, voice) (Feine, Gnewuch, et al. 2019). As recent studies reveal, CAs implemented with social cues are associated with increased perceived enjoyment (Diederich, Lembcke, et al. 2020), trustworthiness (Araujo 2018), service quality (Gnewuch et al. 2017), and service satisfaction (Gnewuch, Morana, et al. 2018). Nevertheless, research indicates that it is not recommendable to implement a CA with all possible social cues simultaneously, as it can lead to counterproductive responses potentially mismatching the customers' expectations (Seeger et al. 2018). Hence, it is necessary to find an appropriate and appealing set of social cues that suits the service context and the customers' demands (Feine, Morana, and Maedche 2019; Seeger et al. 2018).

However, many developers still rely on a single human-like design fits all approach when setting up CAs and do not consider the design based on individual requirements and expectations toward the service (Følstad and Brandtzæg 2017). For instance, human-like CA design is expected to be different in the context of customer service (Jain et al. 2018), which benefits from increased feelings of social relatedness and familiarity, compared to a more embarrassing service context (e.g., a CA-based doctor’s guidance on sexually transmitted diseases) (Fadhil and Schiavo 2019; Seeger et al. 2017). Nonetheless, both services may require a rational, reliable, and objective speech style, while the appearance of the CA (e.g., avatar, self-disclosure) may be demanded differently by users. Hence, in some services, a highly socially designed CA might not be the best choice, as this makes customers less likely to entrust themselves to the CA (Mozafari et al. 2021).

As a result, many CAs fail to meet user expectations and are discontinued due to their insufficient or unsuitable design (Følstad and Brandtzæg 2017; Luger and Sellen 2016). In order to capture the complexity of human-like CA design in service encounters and their discourse in research, the first research question of this thesis aims to analyze and synthesize existing knowledge on expectations toward human-like CA design in service encounters:

**RQ1:** How does existing research explore the expectations regarding the human-like design of CAs in different service encounters?

Although CAs have gained traction in recent years and significant efforts have been made to make CAs more reliable, errors and failures are common and inevitable (Coles et al. 1995; Gregersen 2003; Honig and Oron-Gilad 2018; Kantowitz and Sorkin 1983; Makarem et al. 2009). Imperfect CAs occur frequently due to technical faults or interaction errors (Honig and Oron-Gilad 2018). Technical failures are caused either by hardware (e.g., embodied CA not able to move) or software deficiencies (e.g., interruption of CAs service, misunderstanding users’ input). Interaction errors refer to issues that may arise due to a particular social environment (e.g., social norms) or natural human-like errors (e.g., grammar) caused by the users (Honig and Oron-Gilad 2018; Reason 1990). Existing research in this context has a strong focus on technical imperfection and its mitigation (e.g. Larivière et al. (2017), Perez-Marin and Pascual-Nieto (2011)) and indicates that a CA’s imperfections regularly implicate negative consequences such as reducing the users’ perception of competence (Cha et al. 2016; Ragni et al. 2016; Salem et al. 2015), trustworthiness (Desai et al. 2012; Law et al. 2017; de Visser and Parasuraman 2011), reliability (Ragni et al. 2016; Salem et al. 2015), and intelligence (Bajones et al. 2016; Ragni et al. 2016; Takayama et al. 2011). However, research reveals that imperfect CAs can also lead to an increased perception of humanness (Salem et al. 2013), likeability (Mirnig et al. 2017), and perceived familiarity (Gompei and Umemuro 2015).

Similar observations are evident in human-to-human interaction, as imperfect human behavior can lead to higher likeability, increasing the attractiveness of people when they make a mistake (Aronson et al. 1966). Making mistakes is essential for learning and an inherent part of each individual's life (Gregersen 2003; Kantowitz and Sorkin 1983). Errors associated with text interpretation and processing frequently appear in human-based service interaction (Banovic et al. 2017; Tang et al. 2005). The probability, frequency, and other aspects of errors are socially dependent on the individual and can change based on the exogenous conditions of a situation (e.g., emotional distractions like being stressed or disappointed) (Goldstein et al. 2007).

Ultimately, research has shown that people prefer to interact with humans rather than with technology even though the latter is known to demonstrate a higher level of perfection (Dietvorst et al. 2014; Longoni et al. 2019; Schmitt 2020). Therefore, by emulating the human-like imperfect behavior (e.g., occasionally performing human-like errors), CAs can be potentially identified as a real human and possibly lead to increased service perception by users (Araujo 2018; Gnewuch, Morana, et al. 2018; Nass and Moon 2000). In this context, prior research by Salem et al. (2013) found that CA imperfection (embodied CA gesture errors) increased the perceived humanness. Although Mirnig et al. (2017) did not find differences in people's perception of a CA's human-like appearance, they revealed that the CA's imperfection can lead to an increased likability by the users. Furthermore, intentionally embedding human-like imperfection to design CAs to be more human-like and improve perception is also not a novel practice. Developers of the chatbot *Eugene Goostman*, which portrayed a thirteen-year-old boy not able to speak English well, intentionally included human-like errors to convince users that they were interacting with a real human being (Trazzi and Yampolskiy 2018). Ultimately, Google's *Duplex*, a voice-based CA that assists to make an appointment, is designed with human-like imperfection. During the interaction, *Duplex* uses the interjection "uh", to imitate a space-filler which in human-to-human interaction is a sign of poor communication skills (Trazzi and Yampolskiy 2018).

Building on the findings from research and the practical use of imperfect CAs, this thesis proposes that the natural occurrence of human-like imperfections (e.g., humans making an error) implemented in a CA-based service may offer potential benefits to individuals' perceptions and evaluations of the service. Therefore, the second research question is formulated as follows:

**RQ2:** How do human-like errors of CAs influence the user's perception in a service encounter?

The third research question explores in greater detail the disadvantages of human-to-CA interaction, in relation to the CAs' frequently occurring technical imperfections. Thereby, the human-like CA design is intended to evoke positive emotions to improve the user's perception of the service, even if it is imperfect.

There are many ways in which CAs are imperfect and fall short of users' expectations, for instance not understanding or misinterpreting the users' responses, potentially preventing a successful service interaction (Diederich et al. 2021; Honig and Oron-Gilad 2018; Toader et al. 2020). The imperfection is mainly caused by natural language processing problems (e.g., limited vocabulary) and machine learning deficiencies (e.g., learning and usage of insults) (Brandtzæg and Følstad 2018; Honig and Oron-Gilad 2018). Ultimately, insufficient communicative functions, such as the CA not replying in a comprehensible or relevant manner, are recurring issues (Følstad and Brandtzæg 2017). While avoiding or preventing every potential error is impossible to meet customer expectations (Go and Sundar 2019; Ben Mimoun et al. 2012), conducting research on how to design CAs to prevent negative user emotions is an important area of research.

To this extent, Sheehan et al. (2020) reported that imperfect CAs (e.g., not resolving errors) are sufficient to reduce perceived humanness, implicating that perceived humanness may be an important criterion for improving users' service satisfaction. Diederich et al. (2021) share this perspective and illustrate that response failures have a negative effect on a CA's humanness leading to lower service satisfaction. Besides, research illustrates that an effective human-like design can trigger positive emotions and prevent negative ones (Brandtzæg and Følstad 2018; Seeger et al. 2018). Furthermore, inducing positive emotions has been reported to elicit more engaged and friendly behavior, leading to positive thinking and appraisal of a situation (Forgas 2002).

In the service context, human employees affecting customers with positive emotions have crucial importance in the service evaluation and customer behavior toward the employees (Huang and Dai 2010; Lechner and Paul 2019; Manthiou et al. 2020; Mattila and Enz 2002). Similar effects have been shown for CA-based solutions (Baylor 2009; Boulic et al. 2017; Lisetti et al. 2004; Potdevin et al. 2021; Roussou et al. 2019). For instance, Boulic (2017) and Niewiadomski and Pelachaud (2007) show how the facial expressions on a CA's avatar can improve user perception (e.g., enjoyment, emotional connection, and dialogue quality). Therefore, designing a CA to be more human-like seems an effective approach to mitigating a CA's imperfections in service encounters (Diederich et al. 2021; Honig and Oron-Gilad 2018).

At present, however, little is known if a human-like CA design can induce positive emotions toward the interaction with an imperfect CA-based service. Therefore, designing CAs to elicit positive user emotions that lead to improved service satisfaction is an important topic for practice and research. The third research question sheds a light on this subject:

**RQ3:** How does the human-like design of an imperfect CA influence the user's emotions and service satisfaction?

Besides the potential positive emotions induced by a human-like but imperfect CA-based service, research and practice reveal that users react with frustration and aggression toward imperfect CA-based services (e.g., De Angeli et al. 2006; De Angeli and Brahnam 2008; Seering et al. 2020).

In social psychology, the study of human behavior, it is argued that aggression can manifest itself through the occurrence of frustration caused by events hindering an individual's goals (Dollard et al. 1939). In human-to-CA interaction, a service encounter that is unable to provide its offered services (e.g., because it does not understand the users' input) may cause such a situation. An example of this is *BabyBot*, a CA that is designed like a child and continuously develops its language and social behavior abilities based on self-learning algorithms. The users insulted the CA when there was no response or when *BabyBot* reacted differently than expected (Seering et al. 2020). This aggressive behavior can cause several undesirable consequences, such as the CA being unable to self-learn, the reduction in the service's turnover time, and the increase in data traffic with related increased computation time (Chin and Yi 2019; Seering et al. 2020; Zemčík 2021). In extreme cases, this can lead to the CA learning and reacting to user input with racist and offensive language, as the practical example of Microsoft's *Tay* demonstrated. (Zemčík 2021).

While research has found that the human-like design of CAs is associated with increased perception of the CA by users (Araujo 2018; Diederich, Max Janßen-Müller, Brendel, and Morana 2019; Gnewuch, Morana, et al. 2018), some studies report contradicting effects showing that CAs equipped with social cues negatively affect the users' perception (Crolic et al. 2021; Hadi 2019). Crolice et al. (2021) for instance suggest that it is important in the service context to carefully design CAs to be human-like while taking into consideration the users' emotions toward them. Hadi (2019) implies that a human-like design might be beneficial for overall service satisfaction, even though users may occasionally be frustrated. Seeger et al. (2018) identify that human-like design has a dark side and can lead to negative user perceptions, proposing to implement an appealing set of social cues. Therefore, the fourth research question aims to better understand what leads users to aggressive behavior toward CAs and which role human-like CA design as part of human-to-CA interaction plays in the users' perception and behavior:

**RQ4:** What leads to the aggressive behavior of users toward Conversational Agents and what role does a human-like CA design play in this context?

In summary, Figure 1 illustrates the dissertation's research framework, which highlights the four derived research questions and their interaction. To this end, the objectives of this research are summarized in four steps. First, this thesis aims to explore the expectations toward CA design across different service contexts based on literature analysis. Second, this research examines how users perceive a CA-based service that relies on human imperfection (e.g., a CA that makes human-like errors). Third and fourth, this thesis aims to better understand the users' affective, cognitive, and behavioral perception of human-like CA design in a technically imperfect CA service setting (e.g., CA misinterprets the individuals' input) with a focus on positive emotions (e.g., joy, happiness) and negative emotions (e.g., frustration, aggression). A detailed theoretical background to all relevant topics is provided in section A.II.

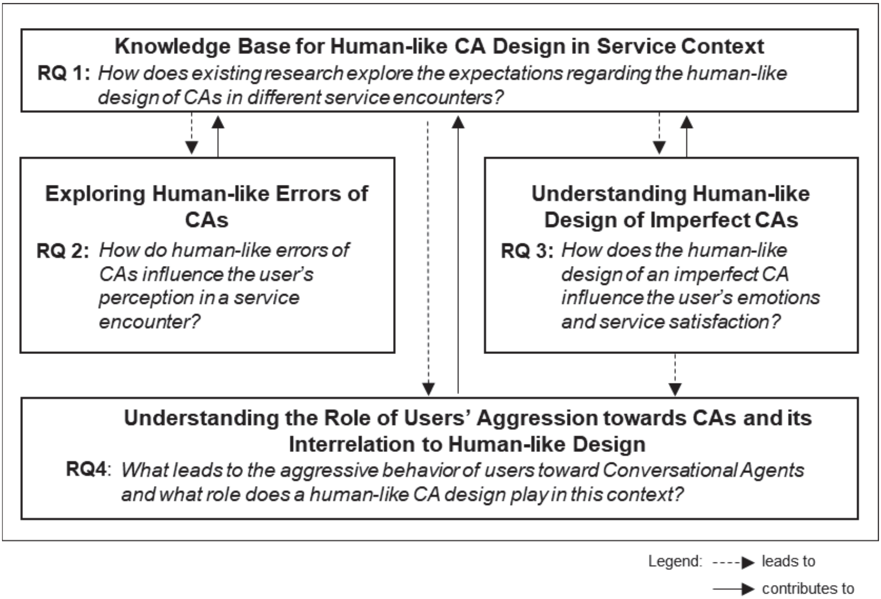
Knowledge Base for Human-like CA Design in Service Context
RQ 1: How does existing research explore the expectations regarding the human-like design of CAs in different service encounters?
Exploring Human-like Errors of CAs
RQ 2: How do human-like errors of CAs influence the user's perception in a service encounter?
Understanding Human-like Design of Imperfect CAs
RQ 3: How does the human-like design of an imperfect CA influence the user's emotions and service satisfaction?
Understanding the Role of Users' Aggression towards CAs and its Interrelation to Human-like Design
RQ4: What leads to the aggressive behavior of users toward Conversational Agents and what role does a human-like CA design play in this context?
Legend:
leads to
contributes to

**Figure 1: Research Framework**

## I.3 Structure of this Thesis

This cumulative dissertation is divided into three main sections (see Figure 2) and builds on four independent studies. Part A provides the foundation by motivating this research's purpose (A.I.1). Afterward, the research questions are derived (A.I.2). Then, the structure of the thesis is explained (A.I.3), followed by the research positioning and research design (A.I.4). The introductory chapter ends with a description of the expected contributions (A.I.5). Chapter A.II expands the research background by introducing the context of Human-Computer Interaction (A.II.1). Then, Conversational Agents are introduced, complemented by a review of their service purpose (A.II.2). Section A.II.3 provides knowledge about human-like CA design and its related theories, namely the 'computers are social actors' paradigm and social response theory. The chapter is outlined by providing knowledge of imperfection including a view toward service perception, specifically focused on CAs (A.II.4). Finally, the findings from the research background