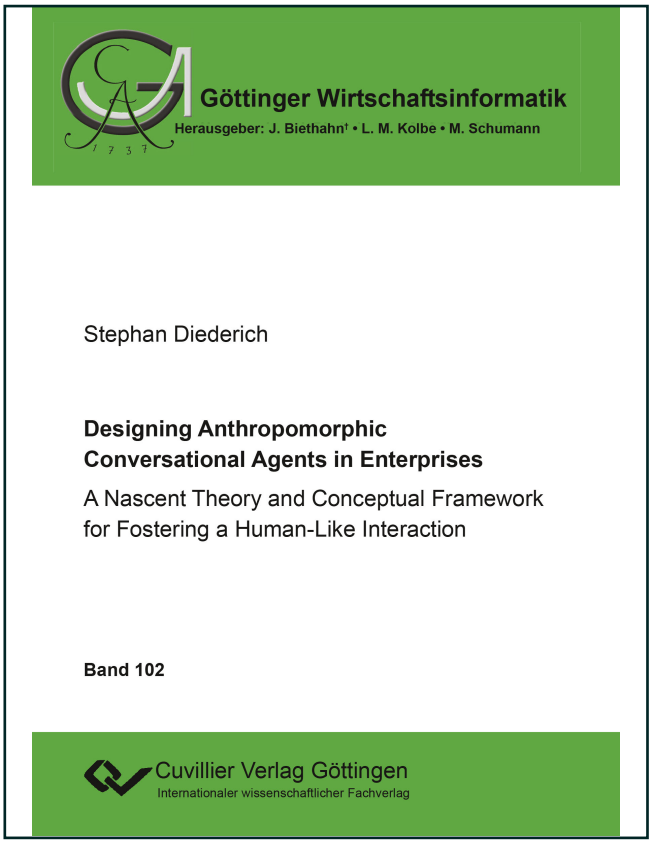
Stephan Diederich (Autor)

Designing Anthropomorphic Conversational Agents in Enterprises

A Nascent Theory and Conceptual Framework for Fostering a Human-Like Interaction



https://cuvillier.de/de/shop/publications/8239

Copyright:
Cuvillier Verlag, Inhaberin Annette Jentzsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen, Germany
Telefon: +49 (0)551 54724-0, E-Mail: info@cuvillier.de, Website: https://cuvillier.de

I. Introduction

The first section (I.1) in this chapter highlights the motivation for and relevance of this research endeavour. Afterwards, the research gaps and questions addressed in this work are presented (I.2) followed be the thesis' structure (I.3), the research context and design (I.4), as well as an overview of anticipated contributions.

I.1 Motivation

> *"What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people."* (Weizenbaum, 1976, p. 371)

In the 1960s, German-American computer scientist Joseph Weizenbaum created ELIZA, the first well-known computer program with which users could interact in natural language. Named after Eliza Doolittle from George Bernard Shaw's play "Pygmalion", the program simulated a Rogerian psychotherapist by rephrasing the user's questions as responses. Originally designed to show the limitations of computers for understanding natural language, the story of ELIZA turned out differently as users developed emotional relationships with the program and interacted with it as if it was an actual human being. For example, the secretary of Joseph Weizenbaum once asked him to leave the room to be able to talk to ELIZA in private. These strong social responses to a computer program with very limited capabilities represent an early example of the emotional connections that humans form with technological artifacts exhibiting human-like characteristics, such as communication via natural language.

A few decades later, Nass and Moon (2000) conducted a series of experiments at Stanford University and found further evidence for what Weizenbaum had described. The researchers discovered that humans over-use social categories in their interaction with computers, such as gender or ethnicity, engage in social behaviors like politeness and reciprocity, and show premature cognitive commitments, for example when a medium is labeled as a "specialist". Nass and Moon formalized their findings in Social Response Theory and the related Computers Are Social Actors (CASA) paradigm, postulating that humans show social responses to computers even though they are aware that machines do not have feelings or intentions. These responses stem from the so-called social cues that computers provide, such as using words for output, interactivity (e.g. responding to human action) or performing human roles. Nass and Moon (2000, p. 7) conclude that "the more computers present characteristics that are associated with humans, the more likely they are to elicit social behavior". Viewing the experience described by Joseph Weizenbaum through the lens of Social Response Theory, it becomes apparent that communication via natural language, though in a very basic form at that time, represents a social cue leading to substantial social responses by humans.

Today, the interaction with technology via natural language attracts strong interest due to advances in natural language processing and machine learning (McTear 2017). CAs are now used in a variety of application domains both in private and professional life (Maedche et al. 2019). Organizations introduce CAs in application domains, such as customer service (Gnewuch et al. 2017; Wünderlich and Paluch 2017), marketing and sales (Baier et al. 2018; Vaccaro et al. 2018), or financial advisory (Dolata et al. 2019). For organizations, CAs represent a particularly interesting technology due to their potential for simulating human-like interaction experience fostered by the anthropomorphic design of such agents (Pfeuffer et al. 2019a).

Early success stories of anthropomorphic CAs in practice underline this potential: “Julie”, the CA of the American railroad company Amtrak, answers more than 5 million service requests per year (NextIT 2018). H&M successfully introduced a virtual shopping assistant that provides individual clothing recommendations (Morana et al. 2017). A further example is “BlueBot” of the Dutch airline KLM, which provides information about flight connections (Vogel-Meijer 2018). In addition to these examples at the customer interface, companies are exploring internal use of CAs, such as for supporting new employees (Liao et al. 2018) or acting as an advisor to project teams (Paredes 2019). Similar to this popularity in practice, CAs are increasingly studied by (IS) researchers (Maedche et al. 2019), often with a focus on human-computer interaction (HCI). Different studies investigate the impact of specific social cues, such as dynamic response delays to simulate thinking and typing of a CA (Gnewuch et al. 2018b) or the representation of the agent with a name and gender (Cowell and Stanney 2005), on user perception in particular with regard to anthropomorphism and further outcome variables like service encounter satisfaction or brand perception.

While most studies place emphasis on the positive effects of anthropomorphic CA design, different researchers indicate the risks of potential negative effects, such as users experiencing feelings of strangeness due to being unsure whether they interact with a machine or a human being (Wünderlich and Paluch 2017). In this context, the Theory of Uncanny Valley (Mori 1970) hypothesizes on the relationship between the anthropomorphic design of an artifact and the emotional responses by humans to it. The theory in short postulates that increasing the degree of anthropomorphism increases the positive emotional responses of humans to the artifact. However, at a point close to achieving human-likeness, the artifact fails to maintain its lifelike appearance due to inhuman imperfections, and triggers strong negative emotional reactions by humans (Mori et al. 2012). In practice, these inhuman imperfections in particular comprise the limited conversational capabilities of CAs compared to humans. For example, IKEA’s agent “Anna” was removed from the company’s webpage in 2016 due to low customer satisfaction. While the CA was specifically designed to offer a human-like interaction, it lacked the necessary conversational capabilities to respond to the wide variety of user

requests about prices, parts, opening hours, and Swedish meals, thus failing to maintain its purposefully designed human-like appearance (Rezvani 2020). Similarly, Ben Mimoun et al. (2012) argue that many CAs are not able to meet high user expectations, fostered by a human-like design, due to their inability to respond to a wide variety of user input and Luger and Sellen (2016) emphasize the gap between user expectations and system capabilities as one of the main reasons for CA failure.

In sum, anthropomorphic design of CAs can both lead to desirable social responses manifested in, for example, increased satisfaction with a company's service or better brand perception, but may also result in negative emotional responses when high expectations of users due to a human-like design are not fulfilled. Thus, from a designer's perspective, the problem arises of how to build anthropomorphic conversational agents that lead to positive social responses and at the same time mitigate the risk of negative emotional reactions particularly due to limited conversational capabilities.

This thesis seeks to address this problem through a set of experiments on specific aspects of a CA's design and a design science research (DSR) project that brings together findings from the experiments and existing prescriptive knowledge for CAs to propose a nascent design theory for anthropomorphic CAs in a company context. Addressing this design problem spans three different research areas (Figure 1). First, the phenomenon of interest are *Conversational Agents*, understood as software with which users interact through natural language. Second, the anthropomorphic design of CAs is investigated in this thesis due to its potential in a company context, drawing on existing theoretical knowledge on *Anthropomorphic Artifacts*. Third and finally, the focus of this thesis is on *Human-Computer Systems Design*, i.e. to provide actionable, prescriptive knowledge on how to build CAs.

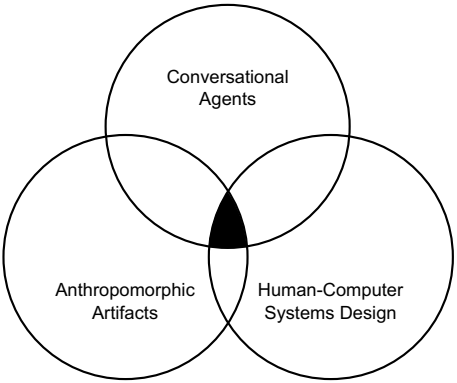


Figure 1: Research Areas addressed in this Thesis

I.2 Research Gaps and Research Questions

As outlined above, this thesis seeks to explore the anthropomorphic design of CAs in a company context. For this purpose, it focuses on four research questions that build on each other, and, together, address the overarching question of how to build human-like CAs. In the following, each question is derived by a specific gap in the knowledge base and afterwards the interaction of the questions is visualized in a research framework.

First, research on CAs – despite their emergence already several decades ago – has recently gained substantial interest in different disciplines, such as computer science (McTear 2017), social psychology (Krämer 2010), and also information systems (Schuetzler et al. 2014), covering topics prevalent in these disciplines, like architectural components of CAs in computer science (Sarikaya 2017) or resilience in social psychology (de Visser et al. 2016). In IS research, in particular scholars focusing on HCI, increasingly explore CAs for example with regard to agent authenticity (Wünderlich and Paluch 2017), privacy concerns (Saffarizadeh et al. 2017), or trust (Schuetzler et al. 2014). Similarly, recent HCI research in computer science studies topics like first impressions in human-CA encounters (Cafaro et al. 2016) or the effects of facial similarity on user responses to CAs (Vugt et al. 2010). In addition to these research foci, studies on the design of and interaction with CAs cover different forms of CAs, such as physically (Stock and Merkle 2018) or virtually (Araujo 2018) embodied CAs, and communication via speech (Purington et al. 2017) or text (Vaccaro et al. 2018). Furthermore, studies focus on CAs in different application domains from private life, such as health (Kim et al. 2013; Meier et al. 2019) or education (Hobert 2019; Winkler et al. 2019), as well as in professional contexts like customer service (Feine et al. 2019b; Wünderlich and Paluch 2017), collaboration (Seeber et al. 2019; Toxtli et al. 2018), or sales (Bertacchini et al. 2017; Vaccaro et al. 2018).

In short, the knowledge base on the design of and interaction with (anthropomorphic) CAs covers a wide variety of topics, forms of CAs, and application domains and, due to the recently increased research interest on natural language software and resulting number of studies, is dispersed. Hence, in order to accomplish the overarching research objective of this thesis, there is a need to first organize the existing body of knowledge, synthesize existing prescriptive knowledge in this area, and identify research gaps. Following the idea of “(re-)constructing the giant” as termed by vom Brocke et al. (2009), the first question aims to structure and assess existing knowledge on the design of CAs:

RQ 1: *What is the status quo of research on the design of and human interaction with conversational agents in a company context?*

Second, while CAs currently attract large interest in research (Diederich et al. 2019a; McTear 2017) and practice (Jacobs et al. 2017) alike, many CAs fell behind expectations (Luger and Sellen 2016) and disappeared in the past due to flaws related to their design (Ben Mimoun et al. 2012). In their assessment of 80 agents on French commercial websites, Ben Mimoun et al. (2012) investigated reasons for failure of CAs. The authors suggest that CAs in particular fall behind expectations due to limited conversational capabilities and argue that a mismatch between the human-like design of many CAs in practice and their actual capabilities results in negative user reactions (Ben Mimoun et al. 2012). Luger and Sellen (2016) similarly emphasize that user expectations are often substantially out of line with the systems' conversational capabilities and features.

In this context, Følstad and Brandtzæg (2017, p. 41) argue that the natural language interface in comparison to a graphical interface "is to a much greater degree a blank canvas where the content and features of the underlying service as mostly hidden from the user, and where the interaction is more dependent on the user's input". As a consequence of this wide variety of potentially unanticipated user input and the limited capabilities of present-day CAs, situations where the CA fails to provide a meaningful reply are likely to occur (Følstad and Brandtzæg 2017). Thus, in order to design CAs that are able to offer a human-like interaction in a professional context, there is a need to first understand the impact of limited conversational capabilities on user perception and to study how negative effects can be prevented or, at least, alleviated:

RQ 2: *What is the impact of limited conversational capabilities of CAs on user perception and how can negative effects be mitigated?*

The research question can be decomposed into two sub-questions. The first sub-question focuses on the (negative) impact of limited conversational capabilities of present-day CAs manifested in failure to provide a meaningful response in a professional service interaction:

RQ 2.1: *How does failure to provide a meaningful response influence user perception of anthropomorphic CAs in a professional encounter?*

The second sub-question aims to explore how to mitigate response failure. Specifically, CAs can be designed to provide preset answer options in the form of suggestions in the conversation to lead users to (conveniently) select one of the given options instead of thinking of and formulating a manual reply in natural language (Brandtzæg and Følstad 2018). However, as such answer options resemble elements from graphical interfaces and may be detrimental to the natural feeling in the interaction, the impact of such design elements on user perception needs to be investigated:

RQ 2.2: *How do preset answer options of text-based CAs influence user perception?*

Complementary to understanding the limited conversational capabilities of CAs, the potential of anthropomorphic CAs, as an emerging technology in a professional context, needs to be explored. Initial studies on anthropomorphic CA design investigate how to make CAs appear human-like by means of different social cues, such as dynamic response delays to indicate thinking and typing by the CA (Gnewuch et al. 2018b), a human name (Araujo 2018), or the use of self-references in a conversation (Schuetzler et al. 2018b). These studies find positive effects of the specific social cues incorporated in the agent's design on the perception of humanness of the CA. Furthermore, several studies indicate a positive impact on further outcome variables, such as service satisfaction (Gnewuch et al. 2018b), brand perception (Araujo 2018), trust (Cowell and Stanney 2005; de Visser et al. 2016), or information disclosure (Pickard et al. 2016). However, different authors indicate (potential) negative effects, for example regarding privacy concerns (Sohn 2019) or feelings of uncanniness when users are unsure whether they interact with a computer program or an actual human (Wünderlich and Paluch 2017). Against this background, the design and perception of an anthropomorphic CA requires careful evaluation (Seeger et al. 2018). In sum, recent research on anthropomorphic CAs provides valuable prescriptive knowledge on how to design human-like agents in a company context with a variety of social cues (Feine et al. 2019a) and how such designs influence user perception.

However, two aspects of human-like CAs, which are of particular relevance in a company context, are yet to be explored: the design of empathetic CAs and the design of persuasive CAs. While modeling empathy in CAs has been studied in HCI research for several years, in particular for physically embodied agents (Leite et al. 2013; Niewiadomski and Pelachaud 2010), it has not been investigated in a company context despite its relevance, for example in customer service (Larivière et al. 2017). Similarly, combining prescriptive knowledge from persuasive system design (Oinas-Kukkonen and Harjumaa 2009) with anthropomorphic CAs to craft persuasive agents can be useful for example in a marketing and sales context (Hanus and Fox 2015) yet has not been addressed in CA research so far. Hence, the third question is formulated as follows:

RQ 3: *How can empathetic and persuasive CAs be designed and how do such designs influence user perception?*

The third research question consists of two sub-questions related to the design of empathetic (1) and persuasive (2) anthropomorphic CAs in a professional context. First, while recent studies explore different social cues for anthropomorphic design of CAs in a company context, the main critique for CAs, in particular at the customer interface, remains that the technology dehumanizes the interaction as it "cannot provide the empathy a human agent can provide" (Yan et al. 2013, p. 7). According to Larivière et al. (2017), recognizing a conversation partners emotional state is considered one of the main

factors that distinguishes service provision by humans and technology. Thus, there is a need to explore design approaches that serve to overcome such existing social limits of (conversational) technology, particularly in a company context (Frey and Osborne 2017). One approach to recognize emotional tonality in natural language communication is sentiment analysis, which automatically extracts information about the human's emotional state (Liu 2012). Combining sentiment analysis with adaptive responses in a CA's design has the potential to emulate empathy and thus contribute to the perception of humanness of an agent. Thus, the first sub question of RQ3 seeks to explore a sentiment-adaptive CA design and its impact on user perception:

RQ 3.1: *How does empathetic behavior of a CA based on sentiment-adaptive responses affect customer perception compared to a non-empathetic agent?*

Second, the design of CAs with effective persuasive capabilities can improve the potential of CAs in a company context, particularly in marketing and sales to influence customer beliefs, such as about products, and (purchase) behavior (Hanus and Fox 2015). While the importance of persuasion capabilities has long been recognized for human sales agents (Friestad and Wright 1999) and various design elements exist for persuasive system design (Oinas-Kukkonen and Harjumaa 2009), combining anthropomorphic design and persuasive system design remains to be explored. As many elements known from persuasive design rely on the system to behave like a human, such as praise, suggestions, or social roles (e.g. identifying the system as an expert for a specific topic area), bringing the two design approaches together seems to be promising endeavor. Hence, the second sub question of RQ3 focuses on the design of persuasive CAs to influence individual beliefs, exemplary in the context of sustainable, shared mobility:

RQ 3.2: *How can persuasive and human-like design of conversational agents positively shape individual beliefs?*

The formulated research questions aim to assess the status quo of CA research and organize existing prescriptive knowledge (RQ1), to better understand the impact of limited conversational capabilities and how to mitigate negative effects on user perception (RQ2), as well as to explore empathy and persuasive capabilities as two specific aspects of CA design with particular relevance in a company context (RQ3). After this improved understanding of the overarching design problem, the last step remains to build on findings from the previous research questions and existing know-how from the knowledge base for CA design to design, implement and introduce the artifact in the field as well as to formulate a nascent design theory (Gregor and Jones 2007). Thus, the fourth and final research question is formulated as follows:

RQ 4: *How can a CA in a professional context be designed to offer a human-like interaction while mitigating negative effects due to limited conversational capabilities?*

Figure 2 visualizes the four research questions and their interaction in the research framework for this thesis.

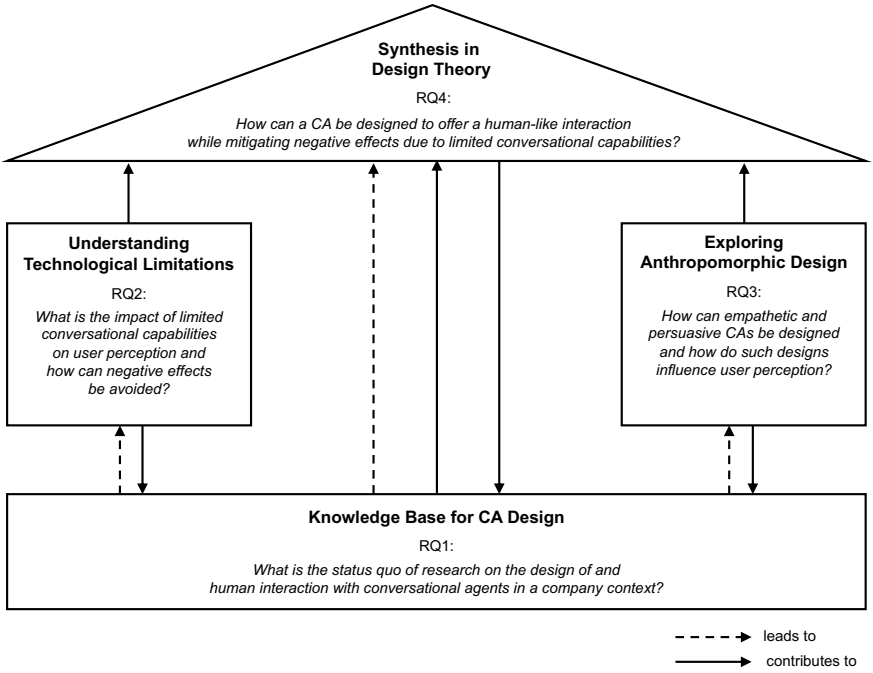


Figure 2: Research Framework

I.3 Structure of the Thesis

This dissertation is cumulative in nature and consists of three parts. Part A lays the foundation for this work by motivating the research endeavor (A.I.1) and highlighting the research gaps as well as research questions (A.I.2) followed by the thesis' structure (A.I.3). Afterwards, the research context and design (A.I.4) and anticipated contributions (A.I.5) are presented. The next chapter (A.II) provides an overview of CAs and their application in enterprises, introduces Social Response Theory and the Theory of Uncanny Valley as two key theories for the design of and interaction with anthropomorphic artifacts, as well as outlines existing conceptual and empirical studies on human-like design of conversational agents.

Part B represents the main body of this thesis, comprising six studies on the topic of CAs in a company context and their human-like design that address the formulated questions and contribute to closing the selected gaps in the knowledge base (Table 1).

Table 1: Overview of Studies included in this Thesis

No.	Outlet	Status	Ranking[1]	Section	RQ	Contribution
1	Journal of the Association for Information Systems	Revision (1st round)	A	B.I.	1	Overview of the status quo of CA research with a focus on IS and HCI as well as synthesis of existing prescriptive knowledge
2	European Conference on Information Systems (2020)	Accepted (conditio-nally)	B	B.II.	2	Understanding of the impact of response failure as a key technological limitation of CAs on user perception
3	European Conference on Information Systems (2019)	Published	B	B.II.	2	Assessment of the impact of preset answer options to mitigate response failure including changes in user perception
4	International Conference on Information Systems (2019)	Published	A	B.III.	3	Exploration of a design approach for empathetic CAs based on sentiment-adaptive responses and assessment of user perception
5	International Conference on Information Systems (2019)	Published	A	B.III.	3	Exploration of the combination of anthropomorphic and persuasive CA design and assessment of user perception
6	Business and Information Systems Engineering	Forth-coming	B	B.IV.	4	Nascent design theory for anthropomorphic CAs derived from the design and evaluation of a CA for a professional service firm

Part C summarizes and synthesizes the findings of this thesis. Afterwards, the implications for research and practice are presented followed by the limitations of this work and opportunities for future research.

Figure 3 visualizes the structure of this thesis.

1) According to VHB-JOURQUAL 3

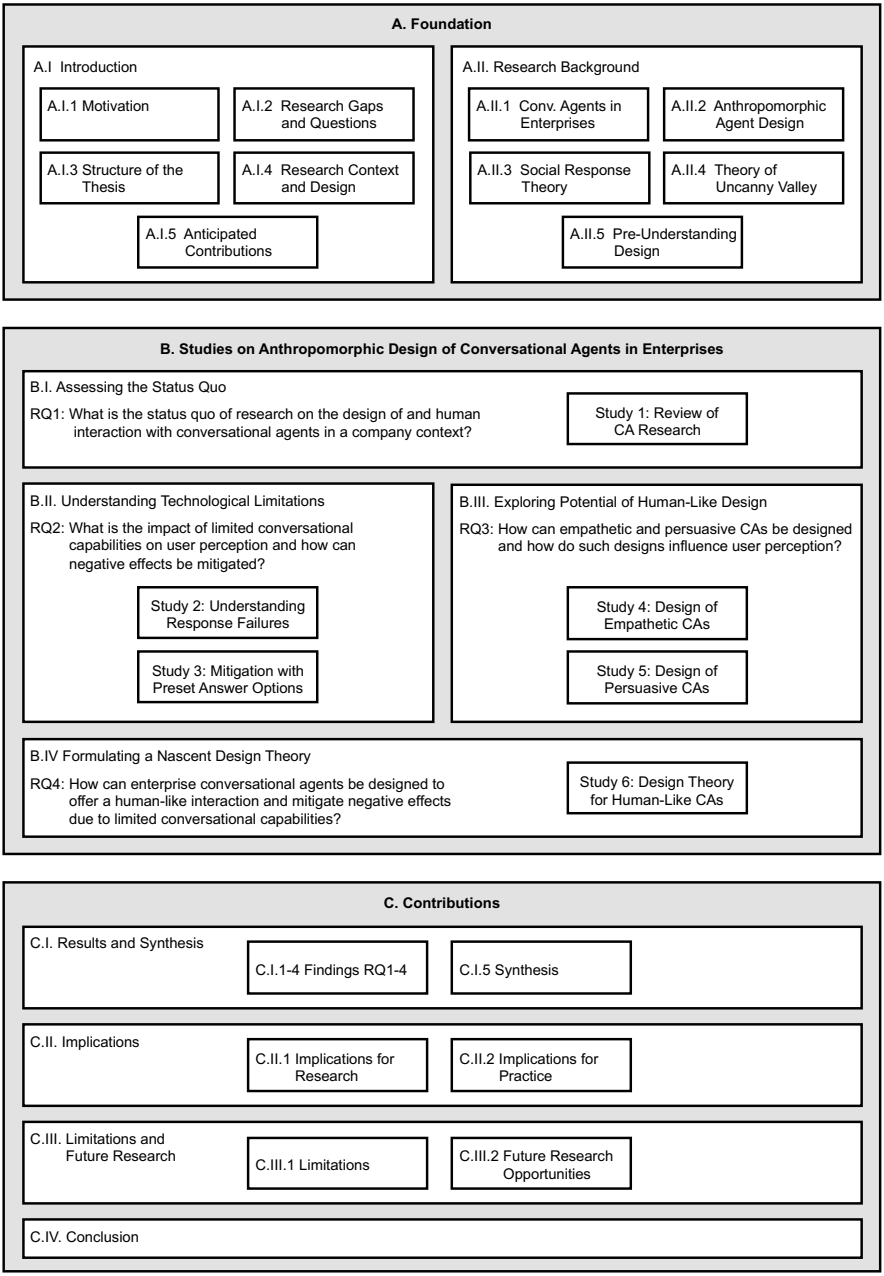


Figure 3: Structure of this Thesis