RESEARCH PAPER

Designing Anthropomorphic Enterprise Conversational Agents

Stephan Diederich · Alfred Benedikt Brendel · Lutz M. Kolbe

Abstract The increasing capabilities of conversational agents (CAs) offer manifold opportunities to assist users in a variety of tasks. In an organizational context, particularly their potential to simulate a human-like interaction via natural language currently attracts attention both at the customer interface as well as for internal purposes, often in the form of chatbots. Emerging experimental studies on CAs look into the impact of anthropomorphic design elements, so-called social cues, on user perception. However, while these studies provide valuable prescriptive knowledge of selected social cues, they neglect the potential detrimental influence of the limited responsiveness of present-day conversational agents. In practice, many CAs fail to continuously provide meaningful responses in a conversation due to the open nature of natural language interaction, which negatively influences user perception and often led to CAs being discontinued in the past. Thus, designing a CA that provides a human-like interaction experience while minimizing the risks associated with limited conversational capabilities represents a substantial design problem. This study addresses the aforementioned problem by proposing and evaluating a design for a CA that offers a human-like interaction experience while mitigating negative effects due to limited responsiveness. Through the presentation of the artifact and the synthesis of prescriptive knowledge in the form of a nascent design theory for anthropomorphic enterprise CAs, this research adds to the growing knowledge base for designing human-like assistants and supports practitioners seeking to introduce them into their organizations.

Accepted after three revisions by the editors of the Special Issue.

S. Diederich (✉) · A. B. Brendel · L. M. Kolbe
Chair of Information Management, University of Göttingen, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany
e-mail: stephan.diederich@stud.uni-goettingen.de

A. B. Brendel
e-mail: abrende1@uni-goettingen.de

L. M. Kolbe
e-mail: lkolbe@uni-goettingen.de

1 Introduction

Technological advances, particularly in machine learning and natural language processing, continue to change the way in which we live, work, and interact with each other, thereby expanding the scope for innovation and automation of human activities (Brynjolfsson and McAfee 2016; Davenport and Kirby 2016). One phenomenon in this wave are conversational agents (CAs), defined as software which users interact with through natural language (McTear et al. 2016). Equipped with increasing capabilities to assist users in a variety of tasks (Maedche et al. 2016; Morana et al. 2017), these agents permeate our private and professional lives (Maedche et al. 2019) in different forms, including digital assistants on smartphones (Chattaraman et al. 2018), chatbots on social media (Xu et al. 2017) or as physically embodied service robots (Stock 2018; Stock and Merkle 2018).

CAs currently attract interest in theory and practice alike due to their potential to provide an enjoyable user

Springer

experience that resembles a human-to-human interaction (Diederich et al. 2019a). From a theoretical perspective, the social cues of such agents, comprising for example the interaction via natural language, the expression of emotions or a human name, trigger social responses by the users (Pfeuffer et al. 2019). These social responses comprise both how users perceive the CA as well as their expectations towards its representation and behavior as posited in Social Response Theory (Nass et al. 1994; Nass and Moon 2000). Different studies highlight that social responses and associated perceptions of anthropomorphism can contribute to a positive user perception of CAs, for example with regard to service satisfaction (Gnewuch et al. 2018; Diederich et al. 2019c), enjoyment (Lee and Choi 2017) or trust (Araujo 2018). However, several studies at the same time indicate that human-like design may lead to undesired negative effects due to feelings of uncanniness (Wiese and Weis 2019), and that a "more is more" approach does not necessarily lead to increased user perception of anthropomorphism (Seeger et al. 2018). Against this background, the Theory of Uncanny Valley (Mori 1970) posits a sharp drop in affinity for human-like artifacts where a user's attention abruptly shifts from the human-like qualities to its inhuman imperfections (MacDorman et al. 2009).

From a design perspective, a variety of social cues for CAs is available to trigger social responses and stimulate perceptions of anthropomorphism (Feine et al. 2019). However, while many of these cues can be incorporated in the design with relatively low effort, such as giving the CA a human name (Cowell and Stanney 2005) or using response delays to simulate thinking and typing of a CA (Gnewuch et al. 2018), sustaining a human-like interaction in an evolving conversation represents a major design challenge. As Følstad and Brandtzæg (2017, p. 41) note, the natural language interface of a CA represents a "blank canvas where the content and features of the underlying service are mostly hidden from the user" and that design to a large extent comprises anticipating user input as well equipping the CA with the ability to provide meaningful responses contingent on what has been communicated (Go and Sundar 2019). In practice, many CAs were discontinued due to their inability to provide meaningful responses and engage in an interactive dialogue (Ben Mimoun et al. 2012), which hinders the usefulness and enjoyment when interacting with such agents compared to systems with graphical interfaces that need to account for a much smaller variety of input (Følstad and Brandtzæg 2017).

Many studies on human-like CA design in IS and HCI are carried out with a focus on selected social cues and by means of experiments (Diederich et al. 2019a) where participants either receive a predefined set of tasks or interact with an actual human in a Wizard-of-Oz setting, thus ensuring responsiveness of the agent during the interaction.

These experiments provide valuable insights regarding the impact of specific social cues. However, with the notable exception of Seeger et al. (2018), they neglect the interplay of the social cues incorporated in the design, in particular with the agent's limited conversational capabilities. Against this background, increasing the social cues incorporated in the design might lead to Uncanny-Valley effects as users start to focus on the inhuman imperfections of the agent related to its responsiveness, ultimately leading to a negative perception. While several studies on CA design, such as Gong (2008) or Gnewuch et al. (2018), discuss potential issues related to uncanniness, such experimental designs with highly structured dialogues or Wizard-of-Oz settings are unlikely to yield strong feelings of uncanniness as adequate, natural responses, of the CA are ensured during the conversation. However, situations in practice where limited conversational capabilities of human-like CAs abruptly induce response failure occur frequently (Følstad and Brandtzæg 2017), potentially leading to the perception of the CA as uncanny. Hence, how the human-like design of a CA can provide utility despite its limited capabilities has yet to be investigated.

With our study, we address this problem and contribute to the knowledge base on anthropomorphic CA design with the following research question: *How can a CA in a professional context be designed to offer a human-like interaction while mitigating feelings of uncanniness due to limited conversational abilities?* Specifically, we bring together prescriptive knowledge for CAs gained mostly in experiments and propose a design for a CA that offers a human-like interaction experience enabled through the combination of social cues with approaches to address the limited responsiveness of present-day CAs. The artifact was created in a design science research (DSR) project over a span of seven months with a large professional services firm and evaluated with a comparative approach to demonstrate that the CA indeed represents an improvement over the extant system with a graphical user interface.

We continue by providing the research background for our study. Then, we introduce our DSR approach and describe the artifact with a focus on design principles as well as present the results from the evaluation. Afterwards, we formulate our design theoretical contribution, state limitations of our work and propose opportunities for future research.

2 Related Work and Theoretical Foundation

Our DSR project contributes to solving the design problem of crafting anthropomorphic CAs while minimizing the risk of negative perception due to limited conversational capabilities. The design is grounded in two theories on user
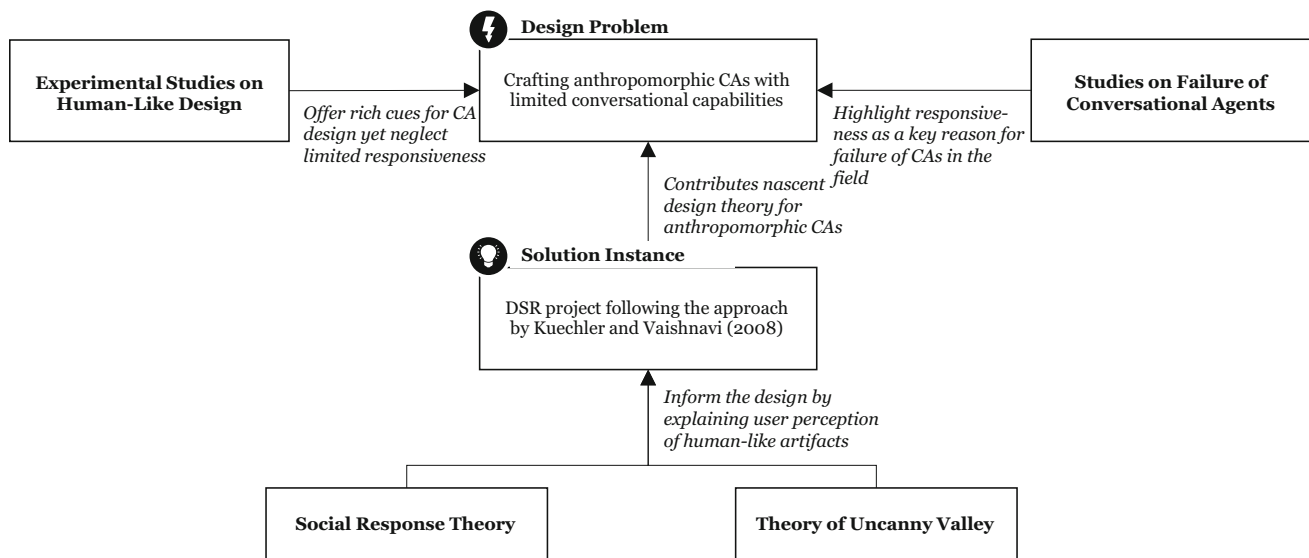
Fig. 1 Research background

perception of human-like artifacts and the project is carried out based on the DSR approach by Kuechler and Vaishnavi (2008). The overall research background is visualized in Fig. 1. In the following, we first describe existing research on human-like CAs and highlight the issue of limited conversational capabilities. Afterwards, Social Response Theory and the Theory of Uncanny Valley are introduced.

2.1 Human-Like Design of CAs and Their Conversational Capabilities

The use of CAs in companies offers the potential to automate and innovate tasks in various application areas. Recent studies have explored text-based CAs, for example in customer service (Wünderlich and Paluch 2017), marketing and sales (Vaccaro et al. 2018), team collaboration (Elson et al. 2018; Toxtli et al. 2018) and human resources (Liao et al. 2018). CAs are typically introduced for two purposes: First, CAs have the potential to provide intuitive access to existing systems via natural language, thus avoiding the need to manually interact with a graphical interface in multiple steps (McTear et al. 2016). Second, CAs can provide the feeling of a human contact in an interaction with a technological artifact (Verhagen et al. 2014). While many studies on CA design suggest crafting a CA as human-like as possible, it is not a design goal per se. For example, Seeger et al. (2017) theorize that the agent's substitution type, i.e. whether the CA substitutes a task previously carried out by a human person or by a computer system, impacts perceived trustworthiness of the agent. In cases where the agent substitutes a human expert, the perceived familiarity with a human-like CA can lead to a more positive evaluation of the CA by the user due the

user's knowledge about the familiar human equivalent (Komiak and Benbasat 2006). In cases, however, where the CA substitutes an existing computer system, a less human-like design could lead be perceived as more useful due to the associated superiority of computers in terms of rationality, reliability, and objectivity (Mosier and Skitka 1996). Thus, increasing the humanness of a CA does not necessarily lead to better perception, but needs to follow a careful consideration of the task that it is intended to fulfill.

Emerging design-oriented work on anthropomorphic CAs investigates the impact of different social cues on user perception, often by means of experiments (Diederich et al. 2019a). According to Seeger et al. (2018), anthropomorphic design comprises three dimensions: A human identity, referring to the representation of the CA, verbal cues, including the choice of words and sentences, and non-verbal cues, comprising the non-verbal communication behavior of the CA. With regard to a human identity, Gong (2008) for example finds empirical evidence in an experiment that agent representations with images which exhibit a higher anthropomorphism increases the social responses shown by the users. Concerning non-verbal cues, Gnewuch et al. (2018) for example explore response times and find that dynamic delays positively impact humanness, social presence, as well as user satisfaction even though they lead to a higher waiting time for the user. Related to the third dimension, verbal cues, Schuetzler et al. (2014) for example found that even modest adjustments to an agents responses with regard to syntax and word variability lead to a more positive evaluation of a CA. Overall, these and similar experiments highlight a variety of social cues that is available to make a CA's appearance and behavior as human-like as possible (Feine et al. 2019).

While extant studies on anthropomorphic CAs provide valuable insights on the impact of selected social cues on user perception, they do not consider the interplay of social cues with each other, except for the work by Seeger et al. (2018), and with further aspects of the agent's design, in particular its limited responsiveness. In these studies, participants either received a set of rather narrowly defined tasks or interacted with an actual human in a Wizard-of-Oz setting to ensure that the agent could provide meaningful responses in the conversation. However, the limitations of present-day CAs with regard to responsiveness are a substantial, practical design issue which often leads to unfulfilled user expectations (Luger and Sellen 2016) and CAs being discontinued (Ben Mimoun et al. 2012). Table 1 provides an overview of exemplary experimental research on CA design and studies that highlight responsiveness as a key issue as well as positions the contribution of this study.

2.2 Social Response Theory and the Theory of Uncanny Valley

A key theory underlying the interaction with and design of IT artifacts with human-like characteristics is Social Response Theory (Reeves and Nass 1996; Nass and Moon 2000). Social Response Theory posits that humans mindlessly respond to social cues from artifacts and apply social rules as well as expectations to anything that demonstrates human-like traits or behavior (Reeves and Nass 1996; Nass and Moon 2000). Nass and Moon (2000) discovered in a set of experiments that humans overuse social categories, such as gender, and social behaviors, such as reciprocity,

and hypothesized that "the more computers present characteristics that are associated with humans, the more likely they are to elicit social behavior" (Nass and Moon 2000, p. 7). The social cues incorporated in an artifact's design lead users to anthropomorphize technology, i.e. to have perceptions of humanness in the interaction. In addition, different studies indicate that the availability of social cues leads to perceptions of social presence [e.g. Gnewuch et al. (2018) or Diederich et al. (2019a, b, c, d)], defined as the sense of human contact embodied in a medium (Gefen and Straub 1997). This is in line with further studies that indicate a positive impact of small adjustments on social presence, such as adding human images or personalized messages, to websites (Gefen and Straub 2003; Cyr et al. 2009). While perceptions of humanness and social presence have been shown to positively impact desired factors, such as trustworthiness (Schroeder and Schroeder 2018), perceived competency (Araujo 2018) or authenticity (Wünderlich and Paluch 2017), the social cues at the same time foster user expectations regarding the artifact's (human-like) characteristics and behavior that need to be accounted for in the design process (Diederich et al. 2019b). In the context of CA design, these perceptions of humanness and social presence often lead to expectations regarding the agent's abilities that are not in line with its capabilities (Luger and Sellen 2016).

A related theory for human-like artifacts is the Theory of Uncanny Valley (Mori 1970), originally from the field of robotics (Fig. 2). The Theory of Uncanny Valley hypothesizes on the relationship between an artifact's humanoid appearance and the emotional responses by humans. The

Table 1 Overview of exemplary studies on human-like CA design and failure as well as positioning of this study

| Study	Research method	Key results
Gnewuch et al. (2018)	Online experiment with **predefined conversation** (given tasks)	Dynamic response delays of the CA to indicate thinking and typing increase perceived humanness, social presence, and service satisfaction
Gong (2008)	Laboratory experiment with **predefined conversation** (given reply options)	Higher anthropomorphism of agents' facial images increases social response by users regarding to social judgment, homophily, competence, trustworthiness and social influence
Schuetzler et al. (2014)	Laboratory experiment with **predefined conversation** (given tasks)	Modest interactive and responsive communication of the CA significantly increases perceived humanness and engagement of the agent
Burmester et al. (2019)	Laboratory experiment (**Wizard-of-Oz**)	Interactive communication (e.g. remembering user input or suggesting ideas) by the agent and a human name increase the natural feeling in a conversation
Wang et al. (2008) ↓	Laboratory experiment (**Wizard-of-Oz**)	Politeness leads to increased learning performance, providing empirical support for CAs as social actors and emphasizing the need for social intelligence
This study ↑	*Design science research*	*Design of anthropomorphic CAs in a professional context that mitigates negative effects due to limited capabilities*
Ben Mimoun et al. (2012)	Performance analysis and expert interviews	Main reasons for failure of CAs in the field are limitations of the agents to manage and provide meaningful information to complete a user's request
Luger and Sellen (2016)	Semi-structured interviews	User expectations are often not in line with the system's actual capabilities, design should indicate the systems potential and limitations
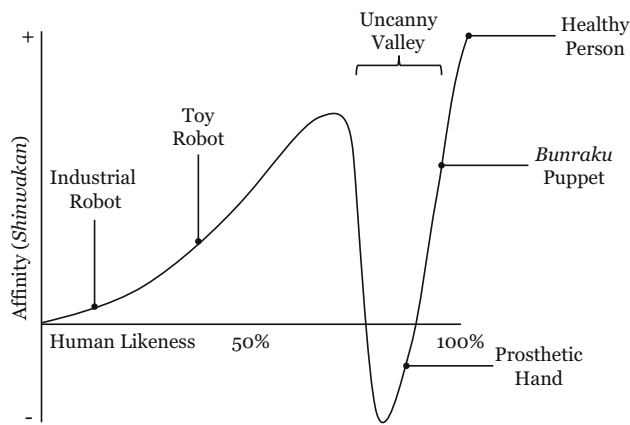
Fig. 2 The Uncanny Valley (Mori et al. 2012)

theory suggests that there is no linear relationship between the degree of human-likeness of an object and positive emotional responses to it, but that a sharp drop in affinity exists at a particular point. MacDorman et al. (2009, p.2) describe this as a shift of attention from the human-like qualities to the aspects that seem to be inhuman; stating that "as something looks more human it looks also more agreeable, until it comes to look so human that we start to find its nonhuman imperfections unsettling". While there is no clear measure or metric for the notion of "affinity" in the context of the Uncanny Valley (Seymour et al. 2018), or "Shinwakan" in the original Japanese wording (Mori et al. 2012), uncanniness is described as negative feelings associated with strangeness of nonhuman imperfections of artifacts (MacDorman et al. 2009).

Against the background of these theories, crafting anthropomorphic CAs represent a substantial design challenge. On the one hand, designers have various social cues at their disposal to make the agent appear human-like and thus benefit from positive effects associated with perceptions of humanness and social presence by the users (Knijnenburg and Willemsen 2016), such as on perceived enjoyment (Qiu and Benbasat 2010). On the other hand, maximizing the human-likeness of an agent increasingly poses the risk of disappointing users and fostering feelings of uncanniness (Ben Mimoun et al. 2012; Luger and Sellen 2016), in particular when the CA does not provide meaningful responses due to its limited conversational capabilities and is thus not able to fulfill user expectations regarding its human-like behavior.

3 Research Approach

Generating knowledge and improving the understanding of a problem through the building and application of a designed artifact is the paradigm that underlies design

science research (Hevner et al. 2004; Hevner 2007). With a fundamental orientation towards problem solving and an engaged relationship between academics and practitioners (Gregory and Muntermann 2014), DSR seeks to build artifacts that solve relevant issues for society, organizations or individuals (Walls et al. 1992) via for instance applying existing kernel theories and deriving design principles (Walls et al. 1992; Iivari 2015). In this research, we address a specific design problem (a human-like CA for simulating job interviews) by building an artifact in a specific context (a professional service firm) and, through the design and evaluation, generate prescriptive knowledge in the form of a nascent design theory (Gregor and Jones 2007) to address a more abstract design problem (designing anthropomorphic CAs with limited conversational capabilities). Our research project is based on the DSR framework by Kuechler and Vaishnavi (2008) and is illustrated in Fig. 3.

We conducted three design cycles. In the first design cycle, we gained an in-depth understanding of the opportunity to innovate the recruiting process through a discussion with a senior HR manager of the case company. We then conducted six semi-structured interviews with members of the HR department and potential job candidates using initial meta-requirements extracted from the literature as a guideline. Specifically, we asked two recruiting specialists with extensive job interview experience from the HR department as well as four job candidates that were preparing for the recruiting process about their requirements for the form and function of a CA in the context of interview preparation. The interviews lasted for 14 to 26 min and the requirements were coded using an iterative approach, which was started with a preliminary list of meta-requirements identified in the literature. The list of codes (meta-requirements) was extended when a new requirement was stated in the interviews. An overview of exemplary codes (meta-requirements) and quotes from the interviews can be found in the Appendix (Table 8; available online via http://link.springer.de). Afterwards, we reviewed further literature on CA design to refine and extend the elicited meta-requirements as well as to formulate preliminary design principles. After the interviews and reading, we had an initial list of meta-requirements (MR1-6) and three preliminary design principles (DP1-3). We then instantiated the principles in an early prototype. After preparing the prototype, we invited seven potential job applicants known to the HR department to interact with the prototype and provide qualitative feedback in a free-form survey. The qualitative feedback was coded with an open and iterative approach where the issues stated by the participants where finally assigned to three categories (see Table 9 in Appendix for the categories and exemplary quotes). The results mainly indicated shortcomings regarding the rather "inhuman" feeling in the conversation
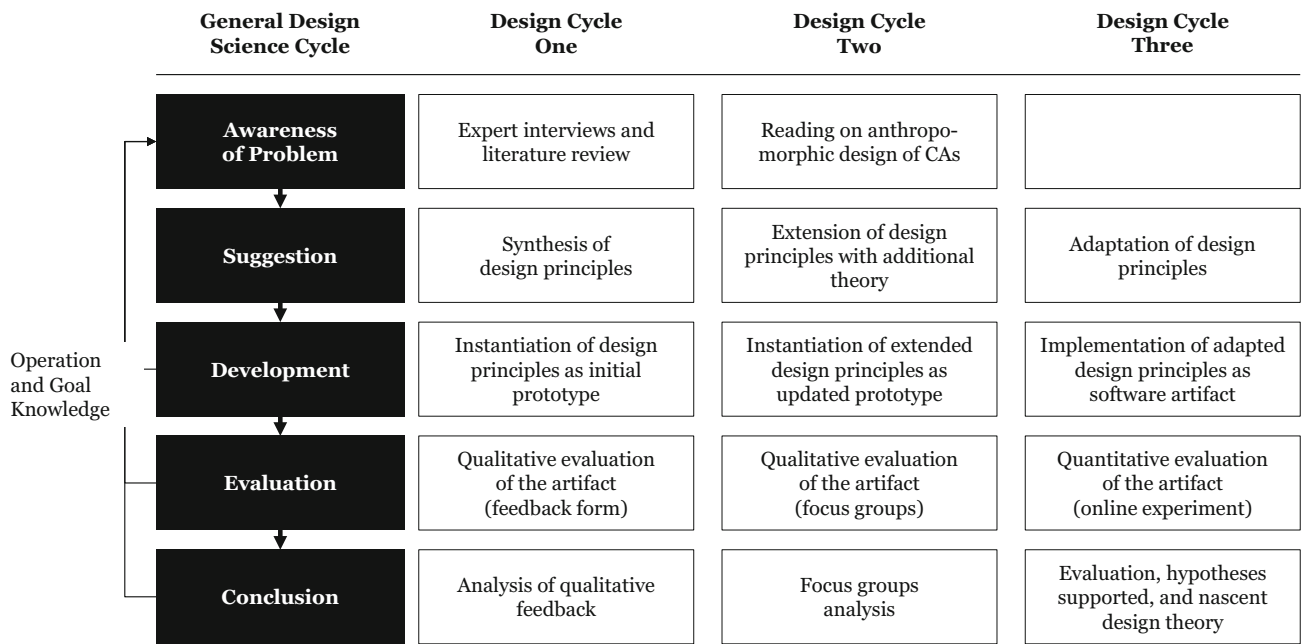
General Design Science Cycle	Design Cycle One	Design Cycle Two	Design Cycle Three
Awareness of Problem	Expert interviews and literature review	Reading on anthropomorphic design of CAs	
Suggestion	Synthesis of design principles	Extension of design principles with additional theory	Adaptation of design principles
Development	Instantiation of design principles as initial prototype	Instantiation of extended design principles as updated prototype	Implementation of adapted design principles as software artifact
Evaluation	Qualitative evaluation of the artifact (feedback form)	Qualitative evaluation of the artifact (focus groups)	Quantitative evaluation of the artifact (online experiment)
Conclusion	Analysis of qualitative feedback	Focus groups analysis	Evaluation, hypotheses supported, and nascent design theory

Operation and Goal Knowledge

Fig. 3 Design cycles and research activities

(for example, one participant stated that "you immediately realize that the same responses are used repeatedly" or that the agent "only understands straight-forward responses, creative replies are not appreciated"), which led us to initiate a second cycle.

In the second cycle we added two meta-requirements (MR7-8) to address the issues described in the interviews regarding the rather mechanical nature of the interaction. After further reading of CA literature, in particular on anthropomorphic design, we added a fourth design principle. Then, we instantiated the design principle with a set of anthropomorphic cues identified in the literature in an updated prototype. We evaluated the prototype with two focus groups, one consisting of four representatives of the HR department and one consisting of three members of the so-called talent pool of the company that contained promising job candidates known from marketing events. During the focus group sessions, we discussed and noted strengths and weaknesses of our adapted design. While the updated prototype was in general perceived as positive by both focus groups, a lack of context-dependent support, guidance and agent responsiveness was reported during the interview simulation. Consequently, we engaged in a third design cycle in which we adapted our design principles to account for context-specific fallback handling and guiding users, such as by providing suggestions or hints in the conversation. Furthermore, we used dialogue data from the previous cycle to improve the CA's capabilities. A visualization of the evolution of our design can be found in the appendix.

4 Artifact Description and Evaluation

Throughout the design cycles, we gained an in-depth understanding of the opportunity for innovation, elicited eight meta-requirements and derived four design principles. Then, we instantiated the design principles, formulated testable propositions, and evaluated the artifact.

The overall motivation for our DSR project stemmed from the idea to provide a new tool that supports applicants in their interview preparation at the professional services company. At the case company, the candidates, mostly recent graduates from university, apply for consultant positions and have to undergo a larger recruiting process comprising several interviews. As these job interviews are standardized and case-study based, which is common for companies of that size and in that industry, applicants can prepare themselves through practicing online case studies. These cases involve the structuring of a business problem, estimating and calculating numbers, and presenting as well as defending the solution and take usually half an hour for completion. Existing training systems typically consist of Q&A forms with a transparent structure and multiple-choice questions. While those systems can be helpful to understand the basic course of interviews, they lack realism due to their obvious structure and do not offer the feeling of a personal interaction like in a human dialogue. Against this background, we considered an anthropomorphic text-based CA as a promising opportunity to improve existing solutions in this application domain (Gregor and Hevner 2013).
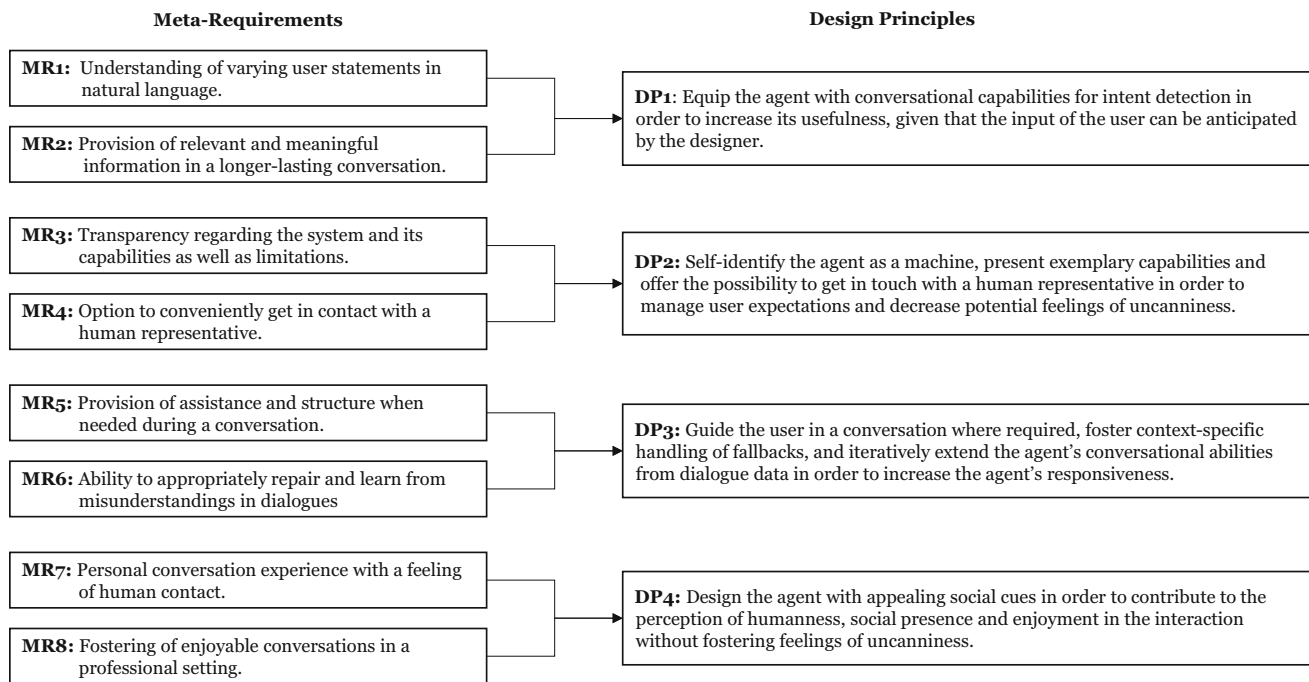
Meta-Requirements   **Design Principles**

MR1: Understanding of varying user statements in natural language.

MR2: Provision of relevant and meaningful information in a longer-lasting conversation.

DP1: Equip the agent with conversational capabilities for intent detection in order to increase its usefulness, given that the input of the user can be anticipated by the designer.

MR3: Transparency regarding the system and its capabilities as well as limitations.

MR4: Option to conveniently get in contact with a human representative.

DP2: Self-identify the agent as a machine, present exemplary capabilities and offer the possibility to get in touch with a human representative in order to manage user expectations and decrease potential feelings of uncanniness.

MR5: Provision of assistance and structure when needed during a conversation.

MR6: Ability to appropriately repair and learn from misunderstandings in dialogues

DP3: Guide the user in a conversation where required, foster context-specific handling of fallbacks, and iteratively extend the agent's conversational abilities from dialogue data in order to increase the agent's responsiveness.

MR7: Personal conversation experience with a feeling of human contact.

MR8: Fostering of enjoyable conversations in a professional setting.

DP4: Design the agent with appealing social cues in order to contribute to the perception of humanness, social presence and enjoyment in the interaction without fostering feelings of uncanniness.

Fig. 4 Meta-requirements and design principles

4.1 Meta-Requirements and Design Principles

We identified meta-requirements (MR) that comprise the fundamental conversational capabilities of anthropomorphic CAs as well as approaches to address the agent's limited responsiveness, both by mitigating and handling situations where the agent is not able to provide a meaningful reply, and a human-like interaction experience. To address these meta-requirements, we formulated four design principles (DP), as visualized in Fig. 4. In the context of this study, we consider design principles not to lead to a certain effect in a deterministic manner but rather consider them as opening up potential for action (Chandra et al. 2015). Our approach for design principle formulation thus follows the suggestions by Chandra et al. (2015) and Seidel et al. (2017) to incorporate material- and action-oriented information as well as, if relevant, boundary conditions stemming from user characteristics or implementation settings.

MR1-2 and DP1 refer to the essential conversational abilities of the agent. As user input in a natural language interaction has a much higher variety than input in graphical user interfaces (Følstad and Brandtzæg 2017), the agent needs to be able to accurately detect the intent in a user's statement, given that it can be anticipated by the designer. This variety comprises both the different intents with which a user approaches the agent as well as different formulations for the same intent. After detecting the intent, the agent needs to provide a meaningful response, for

example through integration in business systems where the requested information is stored (Gnewuch et al. 2017) or by directly embedding information in the processing logic. Understanding a person's intent and providing a reply that fits the conversational context and contains relevant information contingent on what has already been communicated (Go and Sundar 2019) are fundamental requirements for text-based CAs and essential for the agent to be a useful tool for the user (Følstad and Brandtzæg 2017). Against this background, one interviewee described that the agent needs to be able to "logically connect what has been communicated in the conversation" and must not "forget everything that has been said and just offer the same reply". Thus, we formulate DP1 to provide the agent with capabilities to detect a user's intent and provide meaningful responses in an evolving conversation.

MR3-4 and DP2 describe a need for transparency regarding the agent's capabilities and limitations as well as the possibility to conveniently contact an actual human person in case it encounters a request that it is not able to complete. As users anthropomorphize CA, they form expectations towards the system that are similar to humans instead of computer systems and which substantially differ from the agent's actual capabilities (Dzindolet et al. 2003). Consequently, the design of the agent should reveal the system's capabilities throughout the interaction (Luger and Sellen 2016). In this context, one participant highlighted that the agent should "clearly delineate areas in which it can provide as good replies as possible". In addition,

creating transparency whether the user interacts with a human or machine (Wünderlich and Paluch 2017) and self-disclosure of the CA (Saffarizadeh et al. 2017) were found to positively impact the perception of CAs and mitigate feelings of uncanniness. Against this background, one interviewee stated that he appreciates if an anthropomorphic system sympathetically states that "it is a computer but it also has certain human characteristics". DP2 thus comprises the self-disclosure of the CA as a machine, the presentation of exemplary capabilities, as well as the possibility to get in touch with a human representative in situations where the agent fails to provide a meaningful response in order to decrease potential feelings of uncanniness.

MR5-6 and DP3 address the ability of the agent to provide structure where needed as well as to recover from misunderstandings to contribute to agent responsiveness. Due to the varying user input, the agent needs to be able to guide the conversation towards a specific goal, for example by suggesting responses (Diederich et al. 2019b) or creating transparency for the conversation flow (Gnewuch et al. 2017) in order to avoid situations of limited responsiveness. In addition, the agent should provide context-specific assistance to the user (Maedche et al. 2016). For case-study based interviews, this includes assisting the user with calculations by indicating how close the user's estimate is to the correct value or repeating important information for the solution. Furthermore, as misunderstandings are always possible in dialogues, the agent needs to be able to recover, for example by clarifying a statement or asking for reformulation, and be iteratively trained to learn from conversation data over time (Følstad and Brandtzæg 2017).

MR7-8 and DP4 refer to the anthropomorphic design and comprise meta-requirements concerning the feeling of a personal contact as well as enjoyment. The agent is supposed to offer an experience that resembles a human-to-human dialogue to elicit positive effects associated with humanness and social presence (Araujo 2018). Regarding a human interaction experience, one interviewee emphasized that the agent should for example adequately provide positive, motivating feedback (e.g. "you can do it") and thus contribute to "taking away the fear for the actual recruiting day". Furthermore, due to the non-linear relationship between human-like design and user affinity towards an artifact as postulated in the Theory of Uncanny Valley (Mori 1970), the fourth design principle emphasizes the need to find an appealing combination of social cues (Seeger et al. 2018), which in combination with the agent's limited conversational capabilities is able to foster a human-like interaction. As the agent will most likely encounter unexpected user input at some point in a conversation (Følstad and Brandtzæg 2017), a high degree of humanness can also increase user expectations and lead to

substantial disappointment if the agent fails provide a meaningful, relevant response (Ben Mimoun et al. 2012). Thus, DP4 addresses a combination of social cues that balances the need for a human conversation experience with the actual conversational capabilities of the agent (Gnewuch et al. 2017). In addition, the agent should foster an enjoyable conversation, for example by using praise (Wang et al. 2008; Diederich et al. 2019d) as well as polite statements (Mayer et al. 2006).

4.2 Implementation of the Artifact

We built the artifact and instantiated the design principles using Google Dialogflow and a custom-built web interface (Fig. 5). Dialogflow provided the natural language processing capabilities, in particular for intent detection, while the web interface was developed to provide convenient access. We collaborated with the HR department to better understand a case-study based interview and to design the conversation, in particular to model the different intents with which users approach the agent (DP1) on Google Dialogflow. The agent was designed to self-identify itself and highlight exemplary capabilities (DP2). To address the high variability of input and the reported lack of guidance, we created fallback responses that fit the conversational context and implemented guidance to help the user to arrive at the solution (e.g. by indicating errors in calculations or repeating assumptions), as well as extended the capabilities from dialogue data (e.g. by using unanticipated user input as training phrases). We further implemented suggestions to guide the user and increase the responsiveness of the agent in parts of the conversation where user input varied substantially (DP3).

To address the requirements for an enjoyable, and professional interaction similar to a human-to-human conversation, we selected a set of social cues for anthropomorphic design (Table 2) and organized them using the design framework by Seeger et al. (2018). Due to the non-linear relationship between the human-like design and positive emotional responses (Mori et al. 2012), we first reflected on the desired human traits the users described and then purposefully chose cues that we expected to support these characteristics. After identifying the cues in the extant literature, we aligned the selected cues with HR and marketing staff of the company.

5 Evaluation

Every design cycle was accompanied by an evaluation of the artifact. Drawing on the FEDS framework proposed by Venable et al. (2016), we selected a Human Risk & Effectiveness strategy for our project due to the major
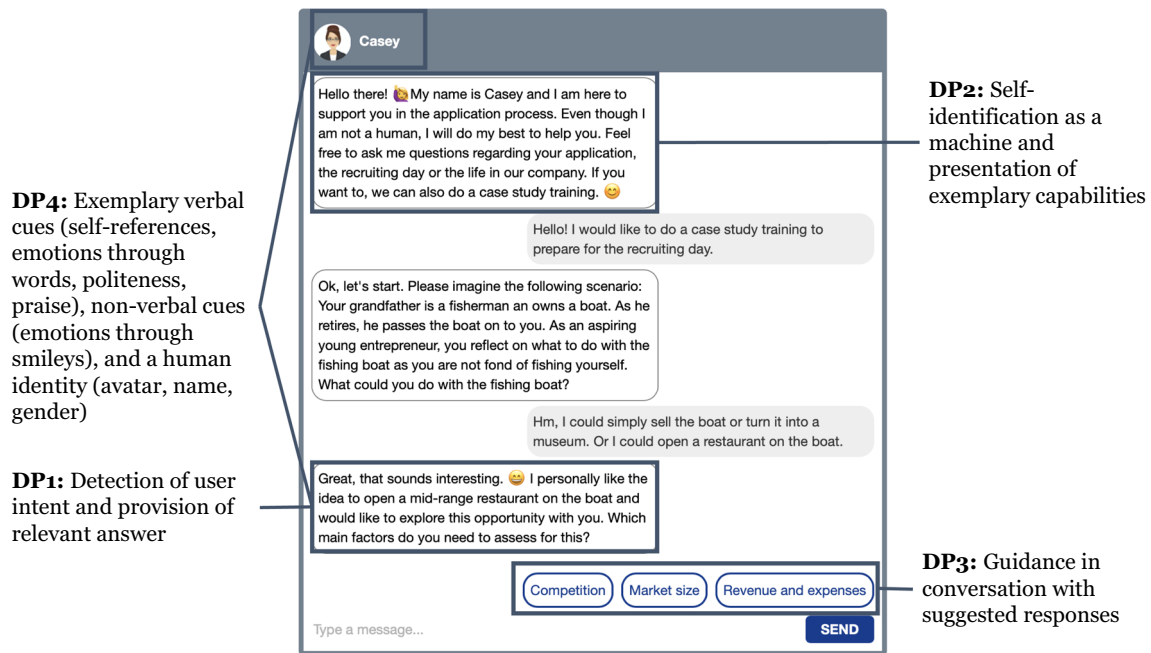
DP4: Exemplary verbal cues (self-references, emotions through words, politeness, praise), non-verbal cues (emotions through smileys), and a human identity (avatar, name, gender)

DP1: Detection of user intent and provision of relevant answer

DP2: Self-identification as a machine and presentation of exemplary capabilities

DP3: Guidance in conversation with suggested responses

Casey

Hello there! 👋 My name is Casey and I am here to support you in the application process. Even though I am not a human, I will do my best to help you. Feel free to ask me questions regarding your application, the recruiting day or the life in our company. If you want to, we can also do a case study training. 🙂

Hello! I would like to do a case study training to prepare for the recruiting day.

Ok, let's start. Please imagine the following scenario: Your grandfather is a fisherman an owns a boat. As he retires, he passes the boat on to you. As an aspiring young entrepreneur, you reflect on what to do with the fishing boat as you are not fond of fishing yourself. What could you do with the fishing boat?

Hm, I could simply sell the boat or turn it into a museum. Or I could open a restaurant on the boat.

Great, that sounds interesting. 😊 I personally like the idea to open a mid-range restaurant on the boat and would like to explore this opportunity with you. Which main factors do you need to assess for this?

(Competition) (Market size) (Revenue and expenses)

Type a message...   SEND

Fig. 5 Instantiated design principles (material translated)

Table 2 Social cues incorporated in the artifact

Desired traits	Dimensions and cues for anthropomorphic design		
	Human identity	Verbal cues	Non-verbal cues
Essential human-like representation and behavior	Name (Cowell and Stanney 2005)	Turn-taking (Gong 2008)	Dynamic response times (Gnewuch et al. 2018)
	Avatar (Gong 2008)	Syntax and word variability (Seeger et al. 2018)	Blinking dots
	Gender (Nunamaker et al. 2011)	Self-references and self-disclosure (Schuetzler et al. 2018)	(de Visser et al. 2016)
Supportive demeanor		Empathy and support (McQuiggan and Lester 2007)	Praise through emoticons (Wang et al. 2008)
		Praise through words (Diederich et al. 2019d)	
Enjoyable and professional		Personal introduction and greeting (Cafaro et al. 2016)	Politeness through emoticons (Mayer et al. 2006)
		Politeness through words (Mayer et al. 2006)	

design risks stemming from the user perception of the artifact. We implemented the evaluation strategy with two formative evaluations in the first two cycles by means of qualitative feedback and focus groups (Fig. 3), focusing on selected aspects of our design (mechanical nature of the interaction in cycle one, lack of context-dependent support and limited agent responsiveness in cycle two). Afterwards, we conducted a summative evaluation by means of an online experiment with two goals. First, the experiment was intended to show that a human-like CA was perceived as more useful and enjoyable than the extant training

system, i.e. that anthropomorphic design in this context is of value to the user. Second, the experiment aimed to evaluate whether the human-like design and approach to address the limited responsiveness does actually lead to perceptions of humanness and social presence as well whether it induces increased feelings of uncanniness.

For the purpose of the experiment, the HR department provided a list of members of their talent pool, comprising potential applicants that were known from marketing events. Overall, we invited 226 members via e-mail of which 72 participated in the experiment (response rate

31.8%). Participation in the experiment took around 25 min per participant without compensation. The sample consisted of 18 female participants (25%) and the average age was 24.9 years (min = 21 years, max = 36 years). In the following, we present the hypotheses we formulated, the design of the experiment and measures, as well as the results from the evaluation.

5.1 Derivation of Constructs and Testable Propositions

In line with the suggestions by Gregor and Jones (2007), we formulate testable propositions for our proposed design. We follow the idea that these propositions can exhibit a comparative logic similar to "if a system or method that follows certain principles is instantiated then it will work, or it will be better in some way than other systems or methods." (Gregor and Jones 2007, p. 327). In our context, the propositions aim to validate that our proposed anthropomorphic CA design works better than the existing training system with a graphical user interface.

First, the objective of the DSR project was to design a CA that helps the candidates to prepare for their job interviews at the company. Thus, the overall utility of the CA with its human-like appearance is defined by the extent to which this design is perceived as more useful for the interview preparation than the existing online systems. The usefulness of a CA in a natural language interaction to a large extent depends on its ability to understand a user's request and provide a meaningful reply contingent on what has already been communicated (Wünderlich and Paluch 2017; Go and Sundar 2019) as reflected in the DP1. Against the background of the agent's proposed conversational capabilities and the overall idea that an anthropomorphic CA is suitable to better simulate a job interview than the extant training system with a graphical user interface, we thus hypothesize:

H1 If a CA follows the proposed design, then it is perceived as more useful than the extant system with a graphical user interface.

Complementary to this utilitarian perspective, we consider enjoyment as a relevant hedonic variable as indicated in MR8. Enjoyment is characterized by its non-goal orientation, that is the pleasure users perceive in the use of a system per se (Junglas et al. 2013). As the proposed CA design is intended to contribute to an enjoyable user experience (DP4) through an appealing combination of social cues (Liao et al. 2018), such as praising the user where adequate or offering a personal introduction, we hypothesize that the interaction with the CA is perceived as more enjoyable than with the extant system that does not contain such cues:

H2 If a CA follows the proposed design, then it is perceived as more enjoyable than the extant system with a graphical interface.

Third, we hypothesize that the CA exhibits a higher level of humanness and social presence as the extant training system due to the rich social cues as posited in Social Response Theory. While the idea that users anthropomorphize a CA more than the existing online training system might seem obvious at first glance, the cues might also be detrimental as users could focus more on the inhuman imperfections instead of its human-like qualities (MacDorman et al. 2009) as indicated in the Theory of Uncanny Valley. Thus, in line with the suggestions by Seeger et al. (2018), we propose that the selected combination of social cues (DP4) in our design provides an appealing human-like experience and fosters feelings of social presence:

H3 If a CA follows the proposed design, then it yields a higher perceived humanness than the extant system with a graphical interface.

H4 If a CA follows the proposed design, then it yields a higher feeling of social presence than the extant system with a graphical interface.

Finally, the human-like design could lead to unintended perception of the artifact as uncanny due to the non-linear relationship between human-likeness and affinity (Mori 1970). As the selected social cues are intended to foster a higher level of perceived humanness and associated user expectations regarding the CA's behavior, they could also lead users to focus on its nonhuman imperfections (MacDorman et al. 2009). Against this background, we suggest that the deliberate selection of social cues depending on the desired human traits (DP4) in combination with self-identification of the CA as a machine and the presentation of exemplary capabilities leads to the CA having a low level of uncanniness:

H5 If a CA follows the proposed design, then it exhibits a low level of uncanniness.

5.2 Experimental Design and Measures

The propositions for our design were tested by means of an online experiment with a between-subjects design (Boudreau et al. 2001). Participants were invited to interact with a new tool to support their preparation for the recruiting day and were assigned either to the extant online training system with the graphical user interface (control) or the CA (treatment). Both the CA and the extant online training system included the same set of questions for the case-study based job interview. Participants in the control

Table 3 Constructs, items, and factor loadings

Constructs and items	Loadings	Scale and source
Perceived usefulness (α = .937, CR = .937, AVE = .714)		
The tool would enable me to faster prepare for the interviews	.882	7-point Likert scale (Davis 1989)
The tool would improve my performance in the interviews	.762	
The tool would increase my productivity in interview preparation	.824	
The tool would increase my effectiveness in interview preparation	.823	
The tool would make it easier for me to prepare for the interviews	.880	
The tool would be useful to prepare for the interviews	.891	
Enjoyment (α = .901, CR = .907, AVE = .709)		
I think the interaction with the tool is interesting	.864	7-point Likert scale (Koufaris 2002)
I think the interaction with the tool is enjoyable	.820	
I think the interaction with the tool is exciting	.887	
I think the interaction with the tool is fun.	.793	
Social presence (α = .841, CR = .838, AVE = .564)		
I felt a sense of human contact with the tool	.685	7-point Likert scale (Gefen and Straub 1997)
I felt a sense of personalness with the tool	.716	
I felt a sense of sociability with the tool (dropped)	(0.457)	
I felt a sense of human warmth with the tool	.787	
I felt sense of human sensitivity with the tool	.810	
Uncanniness (α = .847, CR = .843, AVE = .645)		
I perceived the tool as eerie	.677	7-point Likert scale (MacDorman et al. 2009; Tinwell and Sloan 2014)
I perceived the tool as inhuman-like (dropped)	(.599)	
I perceived the tool as strange	.804	
I perceived the tool as unappealing	.911	

condition interacted with the extant training system in a multiple-choice manner while participants in the treatment condition interacted via natural language. After participation in the training, the job candidates completed a survey in which we measured how candidates perceived the interaction.

We adapted established instruments from previous studies that correspond to our hypotheses. Table 3 shows the constructs, items, and factor loadings as well as Cronbach's α, composite reliability (CR) and average variance extracted (AVE) for perceived usefulness, enjoyment, social presence, and uncanniness. To measure perceived humanness, we asked participants to indicate how very inhuman-like to very human-like they perceived the tool on a 9-point semantic differential scale as similarly done by Holtgraves and Han (2007). Three items were dropped from the analysis due factor loadings lower than .60 as proposed by Gefen and Straub (2005). All constructs showed sufficient Cronbach's α (larger than .80), CR (larger than .80) and AVE (larger than .50) with respect to the levels proposed by Urbach and Ahlemann (2010).

5.3 Results

The survey data was analyzed by means of descriptive statistics and one-sided t-tests for the comparative evaluation of the artifact. First, we checked for variance homogeneity. The Levene tests indicated unequal variance for perceived usefulness, thus we used Welch's t test for the analysis. The remaining constructs exhibited equal variance and were analyzed using Student's t-tests. Our data indicated that participants indeed perceived the agent as more useful (H1), more enjoyable (H2), more human-like (H3), and socially present (H4) than the extant online training system. A one sample t-test against the fixed value of 3 showed that the CA exhibited a low level of uncanniness (H5). Additionally, no significant difference for the ratings for uncanniness between the anthropomorphic CA and the extant system with the graphical user interface, for which one would naturally expect a low level of uncanniness as it contains only very few social cues, was found. Table 4 shows the main results of the summative evaluation.

Table 4 Descriptive statistics and t-test results

Condition			t-value (df)	p-value	Result
Control (n = 35)		Treatment (n = 37)			
Perceived usefulness					
Mean	4.52	5.05	− 1.73 (64.3)	.045	H1
SD	1.44	1.12			supported
SE	0.24	0.18			
Enjoyment					
Mean	4.10	4.79	− 2.25 (70)	.014	H2
SD	1.37	1.25			supported
SE	0.23	0.21			
Humanness					
Mean	3.60	4.76	− 2.63 (70)	.005	H3
SD	1.97	1.75			supported
SE	0.33	0.29			
Social presence					
Mean	2.54	3.38	− 3.49 (70)	.001	H4
SD	0.94	1.08			supported
SE	0.16	0.18			
Uncanniness					
Mean	3	1.82	− 8.22 (36)	< 0.001	H5
SD	*(fixed value)*	0.87			supported
SE		0.14			
Mean	1.96		0.69 (70)	.247	
SD	0.88				
SE	0.15				

SD standard deviation, *SE* Standard error

First row for uncanniness depicts one sample *t* test of the CA against fixed value of 3, second row shows the *t* test comparing control and treatment group

Table 5 Evolution of agent responsiveness across design cycles

Design cycle	Avg. minutes per interaction	Avg. fallbacks per interaction	Avg. fallbacks per minute	Completion rate (%)
1	33.6	10.80	0.53	20.0
2	28.7	10.00	0.42	42.9
3	22.6	1.16	0.06	89.2

5.4 Evaluation of Agent Responsiveness

In addition to the evaluation of user perception of the designed artifact, we analyzed the changes in responsiveness of the agent throughout the three design cycles. Using data provided by Google Dialogflow on the successful detection of user intents in the conversations as well as the use of fallback responses in case no intent was matched to the user's query, we investigated the impact of our design adaptions (Table 5).

We observed a decreasing interaction time and fallbacks per interaction and minute across the cycles. In particular, our design adaptation in the last cycle, where we added more specific user guidance in the conversation by means of context-specific hints as well as selected response suggestions and improved the agent's conversational abilities based on training with existing dialogue data, led to a substantial increase in agent responsiveness. Furthermore, the percentage of users successfully completing the interview training increased from 42.9 to 89.2%.

6 Discussion

The design science research project presented in this study aimed to address the design problem of crafting

anthropomorphic conversational agents in a professional context that have limited conversational capabilities as given in present-day technology. In the following, we discuss implications of our results for designing anthropomorphic CAs and summarize the generated prescriptive knowledge in the form of a nascent design theory.

6.1 Implications for Anthropomorphic Conversational Agent Design

Our research contributes to the knowledge base for anthropomorphic CA design by proposing and evaluating a design for a human-like agent in a professional context that leverages existing prescriptive knowledge on social cues and at the same time mitigates potential detrimental effects on user perception due to the limited conversational capabilities of present-day natural language technology. Thus, it contributes to overcoming limitations of existing experimental IS and HCI studies on anthropomorphic CA design [e.g. Schuetzler et al. (2014); Gnewuch et al. (2018) or Burmester et al. (2019)] as well as to addressing the issue of limited conversational capabilities of CAs in practice (Ben Mimoun et al. 2012; Luger and Sellen 2016).

Specifically, the anthropomorphic design proposed in this study has been shown to foster a human-like interaction experience while mitigating and addressing response failures. In line with research on CA design flaws (Ben Mimoun et al. 2012; Luger and Sellen 2016), the insights from the evolution of our design throughout the three cycles emphasized the importance of meaningful responses by the agent in a conversation, in particular as users expect the agent to conform to human conversation behavior due to the rich social cues as posited in Social Response Theory (Reeves and Nass 1996; Nass and Moon 2000). While anticipating user input and conversation flows remains a substantial design challenge due to the open nature of natural language interaction (Følstad and Brandtzæg 2017), our proposed design addresses the currently limited agent responsiveness and thus fosters user perception of social presence and humanness while avoiding feelings of uncanniness. The combination of DP2 (self-identification and presentation of capabilities to manage user expectations as well as providing contact to an actual human person in situations of response failure) and DP3 (offering structure by providing response suggestions and transparent conversation flows as well as using context-specific fallbacks and iteratively training the agent from emerging dialogue data) represents an efficient approach to mitigate as well as address limited conversational capabilities of present-day CAs. In this context, our evaluation of agent responsiveness showed a considerable progress from design cycle 2 to cycle 3 after we added and instantiated DP3, which reduced the number of fallback responses per

minute (0.42 to 0.06) as well as more than doubled the percentage of successfully completed interactions (42.9 to 89.2%). Thus, the use of response suggestions in situations where input varied substantially and often led to fallback replies, in combination with context-specific fallbacks, allowed to successfully steer the conversation in a way that the CAs limited conversational capabilities are (in most cases) not revealed and a user's attention is not drawn to the inhuman imperfections of the agent, avoiding potential feelings of uncanniness related to the Uncanny Valley (Mori et al. 2012).

Furthermore, the evaluation of our design showed that a human-like CA in the specific context of interview preparation is perceived more positively by users than the extant training system with a graphical user interface. Specifically, users perceived the designed CA as more useful and enjoyable than the existing system. The designed artifact can thus be more abstractly considered as an improvement, representing a new solution for a known problem (Gregor and Hevner 2013) in a specific application domain. Drawing on the idea that anthropomorphic design is not a goal per se, but beneficial for tasks typically attributed to actual humans (Seeger et al. 2017), we argue that a human-like design in this context is particularly useful as the task at hand (conducting a job interview training) consists of human-to-human interaction.

6.2 Towards a Nascent Design Theory

We presented a situated instantiation in the form of an artifact and formulated more general knowledge in the form of constructs, design principles and testable propositions. Table 6 summarizes these contributions using the components suggested by Gregor and Jones (2007).

7 Limitations and Opportunities for Future Research

Our research exhibits four main limitations and offers opportunities for future studies on anthropomorphic CA design. First, we selected a comparative approach for the evaluation, which allowed to evaluate the artifact as a whole in comparison to the extant training system with a graphical user interface. Under consideration of the different DSR genres suggested by Peffers et al. (2018) and our research objective to formulate a nascent design theory, we position our work in the genre of "IS Design Theory" (Gregor and Jones 2007) rather than an "Explanatory Design Theory" (Baskerville and Pries-Heje 2010) where a systematic manipulation of design variables is favorable. Against this background, our evaluation was suitable to demonstrate that the designed artifact indeed represents an improvement over the status quo (extant training system

Table 6 Nascent design theory for anthropomorphic and communicative enterprise CAs

Component	Description
Purpose and scope	The objective of the text-based enterprise CA is to offer the feeling of a human contact in a professional dialogue while minimizing feelings of uncanniness, particularly due to limited conversational abilities
Constructs	Six constructs were derived that are relevant for the design: Usefulness (Davis 1989), enjoyment (Koufaris 2002), humanness (Holtgraves and Han 2007), social presence (Gefen and Straub 1997), uncanniness (MacDorman et al. 2009; Tinwell and Sloan 2014), and agent responsiveness
Principles of form and function	Through a review of CA literature and interviews, we derived four theoretically grounded DPs, which were qualitatively and quantitatively evaluated:
	DP1: Equip the agent with conversational capabilities for intent detection in order to increase its usefulness, given that the input of the user can be anticipated by the designer
	DP2: Self-identify the agent as a machine, present exemplary capabilities and offer the possibility to get in touch with a human representative in order to manage user expectations and decrease feelings of uncanniness
	DP3: Guide the user in a conversation where required, foster context-specific handling of fallbacks, and iteratively extend the agent's abilities from dialogue data in order to increase the agent's responsiveness
	DP4: Design the agent with appealing social cues in order to contribute to the perception of humanness, social presence and enjoyment in the interaction without fostering feelings of uncanniness
Artifact mutability	The CA improves over time as unanswerable user input is monitored and used to continuously extend the agent's conversational abilities
Testable propositions	Following a comparative logic that CAs often replace extant systems with graphical interfaces, we formulate five propositions to evaluate the design:
	H1: If a CA follows the proposed design, then it is perceived as more useful than the extant system with a graphical user interface
	H2: If a CA follows the proposed design, then it is perceived as more enjoyable than the extant system with a graphical interface
	H3: If a CA follows the proposed design, then it yields a higher perceived humanness than the extant system with a graphical interface
	H4: If a CA follows the proposed design, then it yields a higher feeling of social presence than the extant system with a graphical interface
	H5: If a CA follows the proposed design, then it exhibits a low level of uncanniness
Justificatory knowledge	The overall design is grounded in Social Response Theory (Nass and Moon 2000), the Theory of Uncanny Valley (Mori 1970), and related work on CAs
Principles of implementation	We provide an example of how to instantiate the design in the form of an artifact, particularly the selected cues can be used in future CA designs
Expository instantiation	Two artifacts are presented with different instantiations (prototype with 3/4 DPs, major version with 4/4 DPs which was released by the case company)

with graphical user interface) in the sense of Gregor and Hevner (2013) including a higher level of utility manifested in the constructs. However, it does not allow to explain the impact of single design principles on user perception and performance of the CA. Notwithstanding, the positive impact of adding and instantiating DP3 to address the limited responsiveness can be observed in our analysis of dialogue data (Table 5). Thus, we suggest future studies to investigate the impacts of the three remaining design principles (DP1, DP2, DP4) on user perception of the CA. Additionally, the evaluation did not include varying degrees of anthropomorphism of the CA but focused on a specific combination of social cues as shown in Table 2 and its interplay the agent's conversational capabilities. Thus, we propose to adapt the nascent design theory in future studies and craft anthropomorphic

CAs with different variations of social cues and evaluate changes in user perception.

Second, with regard the CA's responsiveness, our evaluation highlighted a positive effect of adding user guidance in situations where response failure occurs as well as using context-specific fallback handling and continuous training from dialogue data (DP3) to address the limited conversational capabilities of the agent. As failure to provide a meaningful reply in a conversation represents a major design issue for anthropomorphic CAs, we suggest to further investigate the impact of response failure on user perception the agent, for example by deliberately altering the number of fallback replies in a (professional) conversation and measuring changes in user perception with regard to humanness, social presence, and uncanniness of

the agent as well as to systematically explore different fallback replies.

Third, the summative evaluation of our artifact is based on a sample size of 72 participants. The participants in this evaluation can be considered suitable as they represent actual potential job applicants from the case company's talent pool. However, despite the statistically significant results for the hypotheses tests, the evaluation of our design could be strengthened by increasing the sample size with further participants.

Fourth, our measurement approach exhibits two main limitations. First, we measured perceived humanness with a single item as done in other studies on anthropomorphism (e.g. MacDorman (2006) and Holtgraves and Han (2007)). As described by Bartneck et al. (2009) alternative, multi-item measurement instruments for anthropomorphism exist, such as the six items used by Powers and Kiesler (2006) that could further increase consistency and reliability of the results. Second, we collected demographic information of the participants but did not gain information on further contextual aspects that could influence user perception of the anthropomorphic CA. For example, experience with CAs or the task at hand could have a (moderating) effect on for example perceived humanness or usefulness of the agent. Thus, future studies on anthropomorphic CAs could explore which contextual factors related to the user or the given task impact the perception of the agent.

8 Concluding Remarks

Anthropomorphic conversational agents continue to gain substantial interest in companies to automate and innovate different tasks while providing the feeling of a human contact in the interaction. However, the limited conversational capabilities of present-day CAs often lead to situations where a meaningful response cannot be provided by the agent, which abruptly remind users that they are actually interacting with a machine and thus are detrimental to a human-like interaction experience and associated positive effects.

The present study contributes to solving this problem by designing and evaluating an artifact as well as formulating a nascent design theory for anthropomorphic CAs in a professional context that allows to benefit from a human-like interaction experience while mitigating and addressing situations in which the agent's limited conversational capabilities come to light. We invite researchers and designers to apply, evaluate and extend the proposed design theory to improve our understanding of how to craft human-like technological artifacts while deliberately minimizing negative effects due to the limited capabilities of machines.

References

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. Comput Hum Behav 85:183–189

Bartneck C, Kulić D, Croft E, Zoghbi S (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robot 1:71–81

Baskerville R, Pries-Heje J (2010) Explanatory design theory. Bus Inf Syst Eng 2:271–282

Ben Mimoun MS, Poncin I, Garnier M (2012) Case study-embodied virtual agents: an analysis on reasons for failure. J Retail Consum Serv 19:605–612

Boudreau M-C, Gefen D, Straub D (2001) Validation in information systems research: a state-of-the-art assessment. Manag Inf Syst Q 25:1–16

Brynjolfsson E, McAfee A (2016) The second machine age: work, progress, and prosperity in a time of brilliant technologies. Norton & Company, New York

Burmester M, Schippert K, Zeiner KM, Platz A (2019) Creating positive experiences with digital companions. In: Proceedings of the ACM CHI conference on human factors in computing systems. Glasgow, pp 1–6

Cafaro A, Vilhjalmsson HH, Bickmore T (2016) First impressions in human-agent virtual encounters. ACM Trans Comput Interact 24:1–40

Chandra L, Seidel S, Gregor S (2015) Prescriptive knowledge in IS research: conceptualizing design principles in terms of materiality, action, and boundary conditions. In: Proceedings of the Hawaii international conference on system sciences (HICSS), pp 4039–4048

Chattaraman V, Kwon W-S, Gilbert JE, Ross K (2018) Should AI-based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. Comput Hum Behav 90:315–330

Cowell AJ, Stanney KM (2005) Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. Int J Hum Comput Stud 62:281–306

Cyr D, Head M, Larios H, Pan B (2009) Exploring human images in website design: a multi-method approach. Manag Inf Syst Q 33:539

Davenport TH, Kirby J (2016) Just how smart are smart machines? MIT Sloan Manag Rev 57:21–25

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q 13:319–340

de Visser EJ, Monfort SS, McKendrick R et al (2016) Almost human: anthropomorphism increases trust resilience in cognitive agents. J Exp Psychol Appl 22:331–349

Diederich S, Brendel AB, Kolbe LM (2019a) On conversational agents in information systems research: analyzing the past to guide future work. In: Proceedings of the international conference on Wirtschaftsinformatik, pp 1550–1564

Diederich S, Brendel AB, Lichtenberg S, Kolbe LM (2019b) Design for fast request fulfillment or natural interaction? Insights from an online experiment with a conversational agent. In: Proceedings of the European conference on information systems (ECIS). Stockholm

Diederich S, Janßen-Müller M, Brendel AB, Morana S (2019c) Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: insights from an experiment with a conversational agent. In: Proceedings of the international conference on information systems (ICIS). Munich

Diederich S, Lichtenberg S, Brendel AB, Trang S (2019d) Promoting sustainable mobility beliefs with persuasive and anthropomorphic design: insights from an experiment with a conversational agent. In: Proceedings of the international conference on information systems (ICIS). Munich

Dzindolet MT, Peterson SA, Pomranky RA et al (2003) The role of trust in automation reliance. Int J Hum Comput Stud 58:697–718

Elson JS, Derrick DC, Ligon GS (2018) Examining trust and reliance in collaborations between humans and automated agents. In: Proceedings of the Hawaii international conference on system sciences (HICSS). Waikoloa Village, pp 430–439

Feine J, Gnewuch U, Morana S, Maedche A (2019) A taxonomy of social cues for conversational agents. Int J Hum Comput Stud 132:138–161

Følstad A, Brandtzæg PB (2017) Chatbots and the new world of HCI. Interactions 24:38–42

Gefen D, Straub D (1997) Gender differences in the perception and use of e-mail: an extension to the technology acceptance model. Manag Inf Syst Q 21:389–400

Gefen D, Straub D (2003) Managing user trust in B2C e-services. e-Service J 2:7–24

Gefen D, Straub D (2005) A practical guide to factorial validity using PLS-graph: tutorial and annotated example. Commun Assoc Inf Syst 16(5):91–109

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). Seoul

Gnewuch U, Morana S, Adam MTP, Maedche A (2018) Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. In: Proceedings of the European conference on information systems (ECIS). Portsmouth

Go E, Sundar SS (2019) Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput Hum Behav 97:304–316

Gong L (2008) How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. Comput Hum Behav 24:1494–1509

Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. Manag Inf Syst Q 37:337–355

Gregor S, Jones D (2007) The anatomy of a design theory. J Assoc Inf Syst 8:312–334

Gregory RW, Muntermann J (2014) Heuristic theorizing: proactively generating design theories. Inf Syst Res 25:639–653

Hevner AR (2007) A three cycle view of design science research. Scand J Inf Syst 19:87–92

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. Manag Inf Syst Q 28:75–105

Holtgraves T, Han TL (2007) A procedure for studying online conversational processing using a chat bot. Behav Res Methods 39:156–163

Iivari J (2015) Distinguishing and contrasting two strategies for design science research. Eur J Inf Syst 24:107–115

Junglas I, Goel L, Abraham C, Ives B (2013) The social component of information systems—how sociability contributes to technology acceptance. J Assoc Inf Syst 14:585–616

Knijnenburg BP, Willemsen MC (2016) Inferring capabilities of intelligent agents from their external traits. ACM Trans Interact Intell Syst 6:1–25

Komiak SYX, Benbasat I (2006) The effects of personalization and familiarity on trust and adoption of recommendation agents. Manag Inf Syst Q 30:941–960

Koufaris M (2002) Applying the technology acceptance model and flow theory to online consumer behavior. J Assoc Inf Syst 13:205–223

Kuechler W, Vaishnavi V (2008) Theory development in design science research: anatomy of a research project. In: Proc Third Int Conf Des Sci Res Inf Syst Technol May, vol 7-9, pp 1–15

Lee SY, Choi J (2017) Enhancing user experience with conversational agent for movie recommendation: effects of self-disclosure and reciprocity. Int J Hum Comput Stud 103:95–105

Liao QV, Hussain MM, Chandar P et al (2018) All Work and no Play? Conversations with a question-and-answer chatbot in the wild. In: Proceedings of the ACM CHI conference on human factors in computing systems. Montréal

Luger E, Sellen A (2016) "Like having a really bad PA": the gulf between user expectation and experience of conversational agents. In: Proceedings of the ACM CHI conference on human factors in computing systems. San José, pp 5286–5297

MacDorman KF (2006) Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley. In: Proceedings of the ICCS/CogSci-2006 long symposium: toward social mechanisms of android science. Lawrence Erlbaum, Vancouver

MacDorman KF, Green RD, Ho CC, Koch CT (2009) Too real for comfort? Uncanny responses to computer generated faces. Comput Hum Behav 25:695–710

Maedche A, Morana S, Schacht S et al (2016) Advanced user assistance systems. Bus Inf Syst Eng 58:367–370

Maedche A, Legner C, Benlian A et al (2019) AI-based digital assistants. Bus Inf Syst Eng 61(4):535–544

Mayer RE, Johnson WL, Shaw E, Sandhu S (2006) Constructing computer-based tutors that are socially sensitive: politeness in educational software. Int J Hum Comput Stud 64:36–42

McQuiggan SW, Lester JC (2007) Modeling and evaluating empathy in embodied companion agents. Int J Hum Comput Stud 65:348–360

McTear M, Callejas Z, Griol D (2016) The conversational interface: talking to smart devices. Springer, Basel

Morana S, Friemel C, Gnewuch U et al (2017) Interaktion mit smarten Systemen – Aktueller Stand und zukünftige Entwicklungen im Bereich der Nutzerassistenz. Wirtschaftsinformatik & Management 5:42–51

Mori M (1970) The Uncanny Valley. Energy

Mori M, MacDorman KF, Kageki N (2012) The Uncanny valley. IEEE Robot Autom Mag 19:98–100

Mosier KL, Skitka LJ (1996) Human decision makers and automated decision aids: made for each other? In: Automation and human performance: theory and applications. Routledge, pp 201–220

Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56:81–103

Nass C, Steuer J, Tauber ER (1994) Computers are social actors. In: Proceedings of the ACM CHI conference on human factors in computing systems. Boston, p 204

Nunamaker JF, Derrick DC, Elkins AC et al (2011) Embodied conversational agent-based kiosk for automated interviewing. J Manag Inf Syst 28:17–48

Peffers K, Tuunanen T, Niehaves B (2018) Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. Eur J Inf Syst 27:129–139

Pfeuffer N, Benlian A, Gimpel H, Hinz O (2019) Anthropomorphic information systems. Bus Inf Syst Eng 1–16

Powers A, Kiesler S (2006) The advisor robot: tracing people's mental model from a robot's physical attributes. In: Proceedings of the 2006 ACM conference on human–robot interaction. Salt Lake City

Qiu L, Benbasat I (2010) A study of demographic embodiments of product recommendation agents in electronic commerce. Int J Hum Comput Stud 68:669–688

Reeves B, Nass C (1996) The media equation: how people treat computers, television and new media like real people and places. Cambridge University Press, Cambridge

Saffarizadeh K, Boodraj M, Alashoor TM (2017) Conversational assistants: investigating privacy concerns, trust, and self-disclosure. In: Proceedings of the international conference on information systems (ICIS). Seoul

Schroeder J, Schroeder M (2018) Trusting in machines: how mode of interaction affects willingness to share personal information with machines. In: Proceedings of the Hawaii international conference on system sciences (HICSS). Waikoloa Village, pp 472–480

Schuetzler RM, Grimes GM, Giboney JS, Buckman J (2014) Facilitating natural conversational agent interactions: lessons from a deception experiment. In: Proceedings of the international conference on information systems (ICIS). Auckland

Schuetzler RM, Grimes GM, Giboney JS (2018) An investigation of conversational agent relevance, presence, and engagement. In: Proceedings of the Americas conference on information systems (AMCIS). New Orleans

Seeger A-M, Pfeiffer J, Heinzl A (2017) When do we need a human? Anthropomorphic design and trustworthiness of conversational agents. In: Special interest group on human–computer interaction. Seoul

Seeger A-M, Pfeiffer J, Heinzl A (2018) Designing anthropomorphic conversational agents: development and empirical evaluation of a design framework. In: Proceedings of the international conference on information systems (ICIS). San Francisco

Seidel S, Chandra Kruse L, Székely N et al (2017) Design principles for sensemaking support systems in environmental sustainability transformations. Eur J Inf Syst 27:221–247

Seymour M, Riemer K, Kay J (2018) Actors, avatars and agents: potentials and implications of natural face technology for the creation of realistic visual presence. J Assoc Inf Syst 19:953–981

Stock RM (2018) Can service robots hamper customer anger and aggression after a service failure? In: Proceedings of the international conference on information systems (ICIS). San Francisco

Stock RM, Merkle M (2018) Customer responses to robotic innovative behavior cues during the service encounter. In: Proceedings of the international conference on information systems (ICIS). San Francisco

Tinwell A, Sloan RJS (2014) Children's perception of uncanny human-like virtual characters. Comput Hum Behav 36:286–296

Toxtli C, Monroy-Hernández A, Cranshaw J (2018) Understanding chatbot-mediated task management. In: Proceedings of the ACM CHI conference on human factors in computing systems. Montréal

Urbach N, Ahlemann F (2010) Structural equation modeling in information systems research using partial least squares. J Inf Technol Theory Appl 11:5–40

Vaccaro K, Agarwalla T, Shivakumar S, Kumar R (2018) Designing the future of personal fashion experiences online. In: Proceedings of the ACM CHI conference on human factors in computing systems. Montréal

Venable J, Pries-Heje J, Baskerville R (2016) FEDS: a framework for evaluation in design science research. Eur J Inf Syst 25:77–89

Verhagen T, van Nes J, Feldberg F, van Dolen W (2014) Virtual customer service agents: using social presence and personalization to shape online service encounters. J Comput Commun 19:529–545

Walls JG, Widmeyer GR, El Sawy OA (1992) Building an information system design theory for vigilant EIS. Inf Syst Res 3(1):36–59

Wang N, Johnson WL, Mayer RE et al (2008) The politeness effect: pedagogical agents and learning outcomes. Int J Hum Comput Stud 66:98–112

Wiese E, Weis PP (2019) It matters to me if you are human – examining categorical perception in human and nonhuman agents. Int J Hum Comput Stud 133:1–12

Wünderlich NV, Paluch S (2017) A nice and friendly chat with a bot: user perceptions of AI-based service agents. In: Proceedings of the international conference on information systems (ICIS). Seoul

Xu A, Liu Z, Guo Y et al (2017) A new chatbot for customer service on social media. In: Proceedings of the ACM CHI conference on human factors in computing systems. Denver, pp 3506–3510