A Review of the Empirical Literature on Conversational Agents and Future Research Directions

Completed Research Paper

Naim Zierau
University of St.Gallen
Müller-Friedberg-Str. 8,
CH-9000 St. Gallen
naim.zierau@unisg.ch

Edona Elshan
University of St.Gallen
Müller-Friederg-Str. 8,
CH-9000 St. Gallen
edona.elshan@unisg.ch

Camillo Visini
University of St.Gallen
Müller-Friedberg-Str. 8,
CH-9000 St. Gallen
camilloluca.visini@student.unisg.ch

Andreas Janson
University of St.Gallen
Müller-Friedberg-Str. 8,
CH-9000 St. Gallen
andreas.janson@unisg.ch

Abstract

The knowledge base related to user interaction with conversational agents (CAs) has grown dramatically but remains segregated. In this paper, we conduct a systematic literature review to investigate user interaction with CAs. We examined 107 papers published in outlets related to IS and HCI research. Then, we coded for design elements and user interaction outcomes, and isolated 7 significant determinants of these outcomes, as well as 42 themes with inconsistent evidence, providing grounds for future research. Building upon the insights from the analysis, we propose a research agenda to guide future research surrounding user interaction with CAs. Ultimately, we aim to contribute to the body of knowledge of IS and HCI in general and user interaction with CA in particular by indicating how developed a research field is regarding the number and content of the respective contributions. Furthermore, practitioners benefit from a structured overview related to CA design effects.

Keywords: Conversational Agent, Human-Computer Interaction, Literature Review, Research Agenda

Introduction

Recently, there has been a steep increase in the interest of researchers and practitioners into Conversational Agents (CA) (Pfeuffer et al., 2019). These agents assist by engaging with users via natural language (Janssen et al., 2020). Thereby, they herald an enormous potential for digital disruption both for company-based processes (i.e., through cognitive automation) and user-based processes (i.e., through humanization of human-computer interaction; Maedche et al., 2019). In regards to the latter, these agents represent a novel form of Information Systems (IS) that can be distinguished from other entities of IS by their high degree of interaction and intelligence (Maedche et al., 2019). These capabilities may fundamentally affect user perceptions and raise a number of theory- and design-related questions, most prominently revolving around an emergent conversation-based interaction paradigm (e.g., Clark et al. 2019). Hence, the transition to CAs amplifies several challenges and calls for future research (Pfeuffer et al., 2019).

Thus, in recent years a diverse body of empirical work has emerged on CAs in different disciplines most prominently in the IS and Human-Computer-Interaction (HCI) domains (Janssen et al., 2020). Thereby, researchers have studied CAs from a variety of user interaction outcomes (i.e., attitudes, perceptions, intentions, behavior). Moreover, they have explored the effects of a multitude of design elements offered by those interfaces on these interaction outcomes (e.g., Feine et al. 2019; Janssen et al. 2020). Thus, building upon extant theories and reflecting the complex nature of CAs, researchers to date have analyzed a growing number of relationships between independent and dependent variables (DV). However, the heterogeneity among variables investigated in these studies leads to difficulty in summarizing, analyzing, and evaluating findings succinctly from the overall body of empirical research on CAs. This leads to a fragmented literature base and sometimes contradictory research results as this paper shows.

Although initial literature reviews and meta-studies (i.e., taxonomic classifications on CAs; Janssen et al. 2020) emerged during the past years, the research is still scattered across different streams of research. For instance, existing literature reviews focus on specific sub-classes of CAs, such as pedagogical agents (Winkler & Söllner, 2018), or generally discuss literature on AI-based applications (Rzepka & Berger, 2018), while classification-based papers focus on the structural characteristics of CAs (Janssen et al., 2020). Thus, the scientific and practical knowledge that has grown dramatically in recent years as this review shows remains segregated. However, relatively young research fields such as the research on CAs need to arrive at an integrated conceptualization and synthesis of representative literature on which future research efforts can build on (Torraco, 2005). So far, to the best of our knowledge, such an integrated conceptualization does not exist, which results in research on CAs being terminologically fuzzy. Thus, we aim to encapsulate the rapidly developing empirical body of knowledge on CAs in a way that is concise, meaningful and provides value to researchers. Hence, this paper focuses on answering the following research question (RQ):

RQ: *What is the state of the art regarding user interactions with Conversational Agents?*

To answer this RQ, within the scope of this review, we examined 107 empirical publications on CAs in 20 seminal outlets related to IS and HCI research. We analyzed a multitude of findings across a selection of studies and aggregated the results from both quantitative and qualitative research. Based on the review, we extracted the most frequently studied constructs and developed four descriptive models pertaining to major user interaction outcomes. These models serve as a representation of the current state of the art on CA research, and the resulting gaps identify potential research avenues.

Conceptual Background

In recent years, scientific and industry interest in CAs has grown considerably (e.g., Feine et al., 2019; Pfeuffer et al., 2019). The basis for the new technology was laid in 1966 when Joseph Weizenbaum developed a computer program that communicated with humans via a text-based interface and which passed the touring test (Weizenbaum 1966). These text-based interfaces were later followed by the development of voice-based dialogue systems and embodied conversational agents in the 1980s (McTear et al., 2016). The increased interest in this system type results from a few overlapping trends. On the one hand, due to recent advances in AI, especially in regards to natural language processing, new generations of CAs emerged that can be used to automate an increasing number of tasks in areas such as health (Laumer et al., 2019), education (Winkler et al., 2019), and customer service (Qiu & Benbasat, 2009).

On the other hand, the conversational character of CAs enables new as well as potentially more convenient and personal ways to access content and services, thus ultimately enhancing user interactions with IS. Along with these developments, the scientific interest in how these interfaces affect user perceptions has also increased. Under the terms of *Intelligent Personal Assistant* (Hauswald et al., 2016)*, Smart Personal Assistant* (Knote et al., 2018), *Chatbot* (Følstad and Brandtzæg, 2017), and *Conversational Agent* (Feine et al., 2019), numerous studies have been undertaken in recent years. Here, we will highlight some key features of CAs as the overarching object of interest of this paper.

According to Maedche et al. (2019), CAs are distinguishable from other entities of IS based on their capabilities for *interaction* and *intelligence*. Regarding the first dimension, the ability to engage with users via natural language is formative to our understanding of CAs (Feine et al., 2020). Typically, CAs have relied on rigid behavioral patterns. Those agents could only respond to simple requests by matching user inputs against a set of stored patterns (McTear et al., 2016). However, novel forms of CAs can now process compound natural language and thereby respond to increasingly complex user requests (Knote et al., 2018). One example is Amazon`s Alexa, which supports users to carry out everyday tasks via an advanced voice interface, ultimately acting as their personal assistant (Benlian et al., 2019). These agents increasingly mimic human-to-human interaction (Feine et al., 2019; Purington et al., 2017) and enable a more convenient and natural way to interact with technology (Knote et al., 2019). In addition, modern CAs are now usually also characterized by an intelligent component (Maedche et al., 2019). This property makes CAs more adaptive to different users and given context situations. Thus, CAs are capable to *"learn"* using inputs such as environmental data and user preferences (Maedche et al., 2019). Building on a continuously growing data set, CAs can adapt and personalize their behavior over time, thus manifesting autonomous characteristics (Pfeuffer et al., 2019). We consider a broad variety of CAs incorporating also less advanced agents (i.e., rule-based or scripted CAs) to give a comprehensive overview of respective user interactions.

In sum, these capabilities may fundamentally affect user interactions with these systems and raise several questions related to the theoretical grounding and design elements of CAs. In this regard, it was shown that the human-like traits of CAs might trigger users to exhibit emotional, cognitive, or behavioral reactions that are reminiscent of human interactions (Krämer et al., 2005). Hence, researchers are increasingly relying on the *Computer Are Social Actors* (CASA) paradigm as their theoretical foundation to explain specific user interaction outcomes. Accordingly, humans perceive certain CA design elements (e.g., an avatar) that cause them to categorize a technical system as a relevant social entity (Nass et al., 1994). In this regard, design elements can be defined as the distinctive technical, contextual, and knowledge features that frame the CA (Janssen et al., 2020). Recently, the inventory of possible design elements both concerning verbal and non-verbal communication has been significantly increasing (Feine et al. 2019), which has enabled CA developers to address prevalent user concerns (i.e., lack of trust) and design increasingly convincing user interaction experiences (Pfeuffer et al., 2019). At the same time, a multitude of research emerged in different disciplines, most prominently in the IS and HCI domain, which empirically investigated the effect of specific CA design elements on various interaction outcomes. Thereby, most studies focused on single or few design elements or configurations and their effect on selected user perceptions, leading to a fragmented literature base, and sometimes contradictory research results.

This shortcoming could be addressed by an integrated analysis aggregating empirical insights on the diversity of CA design elements, which would increase our understanding of user behaviors and support us in identifying future research needs. Existing reviews on CAs either assume an overall perspective on AI-based technologies (e.g., Rzepka and Berger, 2018), which seems arguably too broad to draw meaningful conclusions from user interaction based on the specific characteristics of CAs. Alternatively, they focus on specific domains such as education (e.g., Winkler and Söllner 2018), which is too narrow to draw overall conclusions on user interactions based on the breadth of application contexts and possible design elements (Pfeuffer et al., 2019). Moreover, to the best of our knowledge, there exists no review that takes a distinct perspective on the empirical effects of CAs despite the accelerating and at the same time segregated manner in which practical and scientific knowledge in this area has been growing (Janssen et al., 2020). Hence, we address the lack of an integrative perspective by conducting a systematic analysis of the empirical literature on CAs, to identify validated findings and reveal relevant research needs.

Research Approach

Hereafter, we describe our research approach to review current empirical CA literature, which was informed by the methodological approach employed by Jeyaraj et al. (2006). To that end, we followed the steps of identifying, coding, validating, and analyzing quantitative and qualitative empirical findings.

Paper Selection Process

To identify relevant literature as the basis for the state-of-the-art analysis, we conducted a Systematic Literature Review (SLR) following Webster and Watson (2002) and vom Brocke et al. (2015). The overall scope of the conducted SLR can be defined along the dimensions of *process*, *source*, *coverage*, and *techniques* of the SLR (vom Brocke et al., 2015): Based on a *sequential search process* we searched relevant journals and conference proceedings from the field of IS and HCI literature as a *source*. Thereby, our literature search intends to reach a *representative coverage* of the design elements reported in the literature. Thus, to establish the basis for our analysis, we used a *comprehensive set of techniques* (i.e., keyword search, backward and forward search). To reach a high level of reproducibility and transparency of our research, we will describe the three single methodological steps that we undertook.

In the *first step*, we selected the search strings. Since we aimed to identify a wide range of literature on CAs, the search string is chosen to be rather broad. Based on recent publications, we identified different keywords researchers used to describe CAs. This resulted in the following search string: *("conversational agent" OR "chat bot" OR "chatbot" OR "dialogue system" OR "smart personal assistant" OR "smart assistant" OR "intelligent assistant")*. In the SLR, we used all variations of the keywords – singular, plural, hyphenated, or not hyphenated. In the *second step*, we selected the outlets. As our goal is to identify representative literature samples of different empirical research perspectives on user interaction with CAs, our search covers multiple journals and conference proceedings. We choose this approach since journal acceptance processes take substantially longer than conference proceedings to be processed, which would lead to neglecting some of the most relevant literature since CA research represents a young and nascent topic. For the selection of outlets, we identified two broad areas for deriving design elements of CAs – IS and HCI – as they cover a substantial share of literature on CAs.

Suitable journals and conference proceedings at the intersection of HCI and IS that provide an overview of high-quality and relevant research in the respective research fields were selected using the Basket of Eight and relevant IS journals and conferences based on the recommendations of the Special Interest Group on Human-Computer Interaction. Moreover, to safeguard the relevance of our results, we discussed our selection of journals and conference proceedings with two senior researchers from the field of interest, who were not involved in the writing process of the paper. Based on these inputs and their feedback, we selected 20 journals and proceedings for our keyword search, as seen in Table 1. Finally, in the *third step,* we identified suitable publications for our review. Searching in the title, abstract, and keywords of the papers, the outlet-based search reveals 383 hits. This number still contains literature not relevant to this paper. In an initial screening process, the identified papers are analyzed based on their abstracts. We only included papers that referred to any type of CAs and which provide empirical insights on user interaction with CAs. Papers dealing with this topic trivially or marginally, such as those generally dealing with technology acceptance of CAs, were removed from the sample. This resulted in a selection of 76 publications. Finally, we performed a forward and backward search to also capture papers not covered through the database search. Through screening the references and applying forward searches using Google Scholar, 31 articles were added to the list, including 24 papers from outlets that were not considered in the initial search, resulting in the final number of 107 relevant papers.

Table 1. Overview of searched journals and conference proceedings

Field	Outlets *(Hits/Relevant Publications/Additional Back- & Forward Searches in Brackets)*
IS	ACM Transactions on Information Systems *(16/0/0)*, Decision Sciences *(6/0/0)*, Decision Support Systems *(39/5/1)*, European Journal of Information Systems *(6/0/0)*, Information Systems Journal *(2/0/0)*, Information Systems Research *(6/0/0)*, Journal of Information Technology *(1/0/0)*, Journal of Management Information Systems *(21/1/0)*, Journal of Strategic Information Systems *(0/0/0)*, Journal of the Association for Information Systems *(1/0/0)*, Management Information Systems Quarterly *(0/0/0)*, Proceedings of the International Conference on Information Systems (ICIS) *(14/2/1). Proceedings of European Conference on Information Systems (3/0/2)*
HCI	ACM Transactions on Computer-Human Interaction *(17/6/0)*, Human-Computer Interaction *(12/3/0),* International Journal on Human-Computer Studies *(43/6/0)*, Journal of Computer-Mediated Communication *(6/0/0)*, Journal of the ACM *(1/0/0)*, User-Modelling and User-Adapted Interaction *(17/2/0)*, Proceedings of the Conference on Human Factors in Computing Systems *(172/51/3), Additional Back/Forward (24)*

Paper Analysis

The 107 relevant papers are analyzed from a concept-centric perspective using an abductive approach. Thereby, to aggregate the insights from identified studies, we developed a list of master codes and master code descriptions. This process was iterative and required multiple coding rounds of the identified papers by different researchers. Thereby, the iterative process started by two of the researchers to independently code a subset of 20 randomly chosen articles. For each of the 20 studies, we listed each dependent and independent variable (IV) as named by the author(s), which together form our initial list of "*author variables*" and "*author variable descriptions*".

Moreover, we captured contextual variables such as the application domain and task of the CA. Subsequently, we discussed how to combine variables across studies, which resulted in a list of "*master* *variables*" and "*master variable descriptions*". In some cases, both lists are identical, while other variables required further consideration. Next, we re-examined the initial subset set of 20 articles and mapped author variables to our master variables. During the next iteration, two researchers independently coded another subset of 20 articles. Thereby, we coded the dependent, independent, and structural variables and also mapped these variables to the growing list of master variables and descriptions. Afterward, these researchers discussed their independent findings. In case the respective findings differed, a third researcher, was involved in discussing the differences. For each iteration, we added new master variables and descriptions. As new variables emerged, we reviewed previously coded articles to determine if any needed to be refined based on the addition of new master variables. This process was concluded once all articles were coded.

Concurrent with the creation of a list of master variables and descriptions, we also coded for empirical relationships between an independent and a DV in each study. Thereby, following Lacity et al. (2010), we assigned four possible codes to the relationship between independent and DVs: "*+1*", "*-1*", "*0*" and "*M*". In this process, we coded *"+1"* for a positive relationship, *"-1"* for a negative relationship, and *"0"* for relationships that were studied but did not show any significant value in the empirical results. In quantitative studies, we used P < 0.10 as the requirement for a significant positive or negative relationship. In case the study was qualitative, we relied on the authors' argumentation signified by a robust theoretical anchoring, which we coded as *"M"*. All told, we coded 302 relationships between independent and DVs. Of these, 138 were positive and significant, 33 were negative and significant, 36 relationships mattered, and 95 relationships were not significant. The overrepresentation of positive effects on our DVs can either be explained by a strong focus on positive effects, or by the resistance of researchers to report negative effects regarding user interaction with CAs.

Results

We organized the findings into four sections. The first section examines descriptive statistics based on the meta-data of the underlying literature, the second section examines the DVs used in CA research, the third section examines the IVs used in CA research, and the fourth examines the relationships between independent and DVs.

Section 1: Descriptive findings

Figure 1. shows that the number of identified publications has been steeply growing during the last years. The youngest paper is from 2020 and the oldest paper from 1996. The majority of papers have been published within the last four years, which supports our initial assumption that CAs represent an emerging research field. This is also reflected by the fact that most papers are from conference proceedings, which gives testament to the relative youth of the field. Moreover, it is worth noting that a multitude of investigated papers is from the HCI discipline, while publications in IS outlets are only recently directing attention towards CAs. The first contributions are rather explorative, incorporating a multitude of investigated variables, while recent papers are more specific concerning the theoretical lenses applied and the effects investigated. Further, the examined contributions included empirical data from various application contexts and data sources.

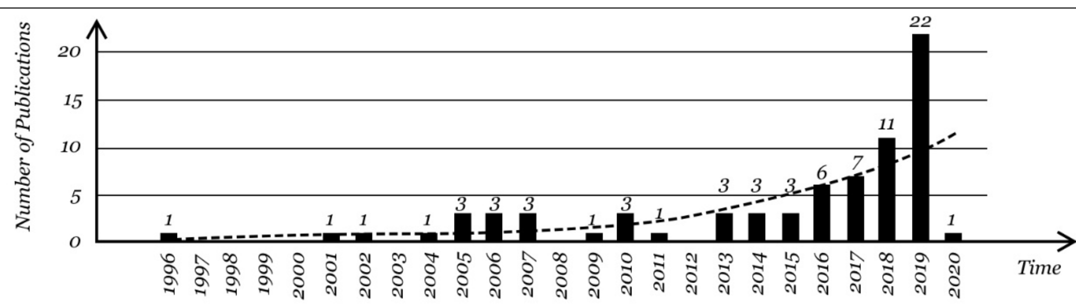

Figure 1. Number of Publications over Time (n=74, included in Analysis of Findings)

Section 2: Findings on Dependent Variables

The publications at hand adopted a wide dispersion of DVs. We identified 390 DVs. We categorized these 390 DV into five broad categories: DVs that investigated perceptual and attitudinal outcomes. Moreover, one category summarizes behavioral intentions and outcomes, which represents a broad umbrella term based on a scarcity of findings in this regard (25 findings.). Thus, the five major variables that evolved from our review are *rapport, social presence*, *trust*, and *utility* as well as *behavioral intention and outcomes*.

Rapport (31.3%). Researchers have generally examined a plethora of outcomes related to the quality of the social bond between the user and the CA, which is also referred to as *rapport* (Pecune et al., 2018). A third of all studied outcome variables can be assigned to this category. Prominently studied variables in this category are the *likeability* of the CA (e.g., Chin and Yi 2019; Miehle et al. 2018) the degree of *involvement* or *engagement* experienced by users (e.g., Van Es et al. 2002; Vugt et al. 2008), and the *perceived closeness* (Bickmore & Picard, 2005; SeoYoung Lee & Choi, 2017).

Social Presence (16.4%). Another important outcome category represents *social presence*, which does not seem surprising since many researchers work on recreating human-CA interactions that are experienced as human-like. In this regard, *social presence* can be defined as "*the extent to which other beings in the world appear to exist and react to the user*" (Heeter, 1992). Within this category, researchers focused on *perceived humanness* (e.g., Candello et al. 2017) and *social presence* (e.g., Cho 2019) as the two main outcome variables.

Trust (15.2%). Additionally, a major perceptual outcome category is reflected by user *trust*. As many researchers cite a lack of trust as one of the central adoption barriers for AI-based technologies, this sentiment has also been important to trust researchers in regards to CAs making *trust* as one the main variables being in the focus of CA research (e.g., Kang and Wei 2018). Trust is usually defined as an expectation that another entity "will perform a particular action important to the trustor [i.e., user], irrespective of the ability to monitor or control that other party [i.e., CA]" (R. C. Mayer et al., 1995, p. 712). However, authors also investigated trust-related concepts such as *credibility* (e.g., Cowell and Stanney 2005) or *privacy* perceptions (e.g., Benlian et al. 2019), which we incorporated in this section.

Utility (31.1%). A multitude of authors investigated productivity-related perceptions, which we summarized under the category of *utility*. Thereby, prior researchers have looked at *usefulness* (e.g., Qiu and Benbasat 2010), *ease of use (*e.g., Van Es et al. 2002*)*, the *quality of interaction* (e.g., Ashktorab et al. 2019) *satisfaction* (e.g., Chaves and Gerosa 2018), and *helpfulness* (e.g., Berry et al. 2005).

Behavioral Intention and Outcomes (6.0%). Finally, few authors also investigated behavioral intentions and outcomes, whereby actual outcomes were in the minority. In regards to behavioral intention, prior authors have looked at the *intention to use* (e.g., Wuenderlich & Paluch, 2017) and *willingness to interact* (e.g., Candello et al., 2017). Behavioral outcomes incorporated findings related to interaction behavior (e.g., length of interaction; Strait et al., 2015) or behavior change (e.g., exercise behavior; Bickmore & Picard, 2005).

Section 3: Findings on Independent Variables

We identified 390 IVs used in CA research. To facilitate the discussion of this high quantity of IVs, we categorized them into five broader categories. Thereby, our allocation into aggregated dimensions is based on a taxonomic classification of social cues of CAs introduced by Feine et al. (2019), which we extended based on our coding by selected categories, as some design elements did not fit these categories (i.e., *interaction*). Each category is briefly discussed below.

The category ***Auditory (3.3 %)*** refers to all design elements that can be perceived via the sense of hearing except the words itself (Burgoon et al., 2013). Within this dimension, those elements have been analyzed that relate to *voice qualities* representing permanent and adjustable characteristics of the voice. In total, these cues have been investigated 10 times. For example, Yu et al. (2019) have studied the impact of the voice's gender (female vs. male) on different perceptual outcomes. Although this category also hypothetically includes nonlinguistic vocals and sounds, there were no studies in our sample addressing these elements.

Within the category ***Interaction (14.5 %)***, we summarize all design elements that refer to the underlying structural representation of the interaction both in regards to its communication mode and its turn-taking mechanism. Overall, researchers have studied often the choice of *interaction mode*. Moreover, researchers studied the influence of preset answers, which reflects the *degree of freedom* employed in the conversation. The former category has been studied 34 times, whereby most researchers compared chat and voice interfaces (Soomin Kim et al., 2019). In comparison, the latter category was studied less but was found equally influential for user perceptions (Diederich et al. 2019).

Design elements in the category ***Invisible (10.6 %)*** refer to all CA characteristics that cannot be perceived by the sense of hearing or seeing (Knapp et al., 2013). These elements are often referred to as "the silent language"(Hall, 1990). This category can be divided into four subcategories: *Chronemics* refers to the role of timing in conversation and is reflected in studies that focus on the design of conversation flows (e.g. Winkler et al. 2019), system response times (e.g., Gnewuch et al. 2018) or the role of synchronicity (e.g., Park and Sundar 2015), which in total have been studied 11 times. *Intelligence* refers to elements that express the cleverness of the agent, which was exemplarily studied by Xu et al. (2017). The other two categories, *personality* (i.e., personality traits) and *haptics* (i.e., tactile sensations), were comparatively less frequent in the research field. Among personality, e.g., Cafaro et al. (2013) examined the personality traits.

The category ***Visual (34.3 %)*** refers to all nonverbal design elements that are not invisible, and can visually be perceived except words itself (Leathers & Eaves, 2015). This category can be distinguished into five subcategories, which have been broadly studied (104 times). The most prominently researched variable is *agent appearance* (46 times). The embodiment of the agent has gained much attention in our sample (e.g., McBreen and Jack 2001; Nunamaker et al. 2011). Another studied aspect of the agent's appearance was gender (e.g., Pfeuffer et al. 2019). Furthermore, *kinesics*, which refers to body movements, such as demeanors (e.g., Krämer et al. 2013) and gaze patterns (e.g., Van Es et al. 2002) were addressed in total 25 times. *Computer-mediated communication* (CMC) refers to visual elements that augment or modify written texts and were examined 23 times. Here, the effects of using emojis (e.g., Park and Sundar 2015), typos (e.g., Westerman et al. 2019) or videos and images (Huber et al., 2018) were researched. The other two categories, *entrainment* (i.e., the adjustment of visual elements to the user) and *proxemics* (i.e., the role of distance in communication), have not yet been studied in detail (10 times for both).

The category ***Verbal (37.3 %)*** refers to all CA elements that can be expressed by words either written or spoken (Antaki, 2008). Within this dimension, the *conversation style*, which refers to how something is being communicated, was the most researched IV (53 times). For instance, Mayer et al. (2006) studied the effects of relational strategies. The aspect of *content* captures all elements that relate to the literal meaning of a message and was researched a total of 22 times. For example, Akahori et al. (2019) have looked at the effects of self-disclosure. Similar attention has been given to *adaptivity* (22 times), which refers to the verbal adaptation of the CA to the users. Within this category, researchers have studied the use of contextual information (e.g., Vtyurina et al. 2017), user content (e.g., Schuetzler et al. 2014), or the absence of adaptivity (e.g., Engelhardt et al. 2017).

Section 4: Findings on the Relationship between Independent and Dependent Variables

In this section, we summarize major findings concerning the 49 relationships we coded between 16 IV and four DV (perceptual and attitudinal outcomes) as well as the regarding behavioral intention and outcomes (that we discuss more generally based on a scarcity of coded relationships). At this detailed level, the frequency with which findings were replicated across studies was minimal and did not provide a very coherent or comprehensive picture of CA research. Hence, to study these relationships in a way that is concise, and helpful to researchers, we moved to a higher unit of analysis by reporting the 277 findings using our four categories of DV and the five categories of IV. Although precision is reduced in aggregating to the broader categories of DVs, we gain a better overall understanding of the determinants of perceptual and attitudinal outcomes of CAs. Thereby, we also aim to investigate the consistency of the empirical evidence.

For this reason, we indicated the consistency of the empirical evidence. Therefore, we used *(!!)* if the evidence was consistent in more than 80 percent of the reviewed relationships. Accordingly, if the relationship provided less robust results, we used a *(!)* evidence that reflected a consistency of 60 to 80 percent of the reviewed relationships. To indicate that a relationship did not meet the set thresholds, we indicated it with a *(0)*. We then applied the same logic to relationships if the evidence originated from a qualitative study and was argumentative. Therefore, if the relationship was consistent in more than 80 percent of the reviewed relationships, we indicated it with a *(MM)*. Accordingly, we applied the same logic and used an *(M)* if the relationship was less robust. For instance, if a relationship was studied six times, and one represented an insignificant relationship, and the remaining evidence showed a significant relationship, we categorized this example with a *(!!)*.

Further, we included the experimental variables and marked them. Experimental variables refer to variables that have been studied less than five times. Those variables, as well as the inconclusive findings, guide in identifying avenues for further investigation and discussing perspectives for a research agenda.

Independent Variables on Rapport

Figure 2 provides an overview of the relationships between *rapport* and the five IVs. In this model, 92 findings are synthesized and depicted based on consistency within the subgroups of IVs, providing an answer to the question: “Which determinants of *rapport* were reported by past empirical research on CAs?”.

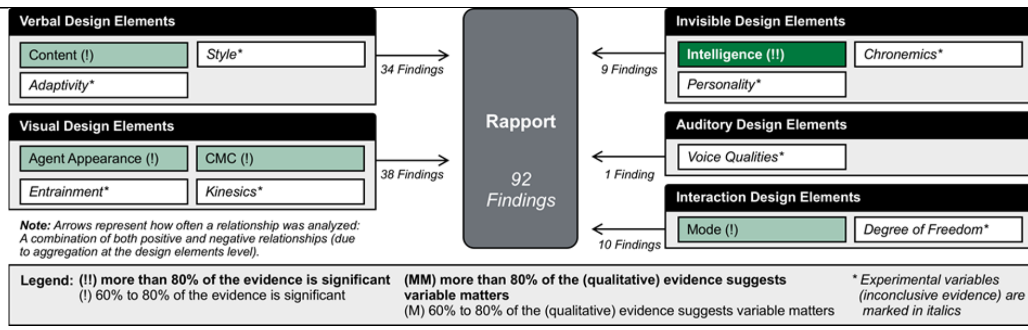

Figure 2. CA Outcomes regarding Rapport

In this model, we identified 1 variable showing consistency of 80% or more, as well as 4 variables showing consistent results in 60% of all instances. The other 8 variables were marked as experimental variables due to having less than 3 observed samples (4 variables) or showing inconsistent findings (4 variables).

Past research offers evidence that ***verbal design elements*** are an antecedent of *rapport* towards the CA. Researchers investigated the variables *style* (20 observations (OBS), e.g., Clark et al. 2019), *adaptivity* (10 OBS, e.g., Lee et al. 2019), and *content* (4 OBS, e.g., Clark et al. 2019). In this regard, content and entrainment were identified as having a significant impact. For instance, it was shown that eliciting similar interests (Clark et al., 2019) and the degree of matching or coordination in word counts of the CA and the user positively influences rapport-building (Pecune et al., 2018). Further, our findings indicate that ***visual design elements*** are determinants of *rapport*. Researchers have studied the variables *agent appearance* (19 OBS, e.g., Sproull et al. 1996), *kinesics* (12 OBS, e.g., Krämer et al. 2013), *entrainment* (4 OBS, e.g., Qiu and Benbasat 2010), and *CMC* (3 OBS, e.g., Westerman et al. 2019). Thereby, *agent appearance* and *CMC* were found to be significant. For instance, enriching the CA’s message by the way of typos and capitalization uncovered a significant influence on the social attractiveness of the CA (Westerman et al., 2019). Moreover, including typos and capitalization as manifestations of CMC increased the social attractiveness of the CA (Westerman et al., 2019). In our model, also variables related to ***invisible design elements*** were found to be significant and consistent determinants of *rapport*. Researchers have inquired into the variables *intelligence* (6 OBS, e.g., Schuetzler et al. 2019), *chronemics* (2 OBS, e.g., Winkler et al. 2019), and *personality* (1 OBS, Cafaro et al. 2013).

In contrast, past research was found to have directed only limited attention to the influence of ***auditory design elements*** on *rapport* between the user and CA. Only one variable (i.e., voice pitch) was studied (Yu et al., 2019), indicating no conclusive evidence. Additionally, our findings indicate some evidence regarding the influence of ***interaction design elements*** as determinants of *rapport* between the user and CA. Researchers studied two variables, *mode* (9 OBS, e.g., Miehle et al. 2018), and *degree of freedom* (1 OBS, Jeong et al. 2019). The influence of *mode* was found to be significant. For instance, employing voice-based interfaces increased users’ self-disclosure towards the CA (Yu et al., 2019).

To summarize our findings on the DV *rapport:* The most significant and consistent evidence regarding determinants of this outcome dimension was found to be related to the group of variables coded as *intelligence (invisible)*. Other consistent findings were found regarding the variables categorized as *agent appearance* and *CMC (*both *visual), mode (interaction),* as well as *content* (*verbal).*

Independent Variables on Social Presence

Our findings regarding the outcome dimension of *social presence* are outlined in Figure 3. In this model, 48 findings are synthesized and depicted based on their consistency within the subgroups of the five IVs, providing an answer to the question: "Which determinants of *social presence* were reported by past empirical research on CAs?".

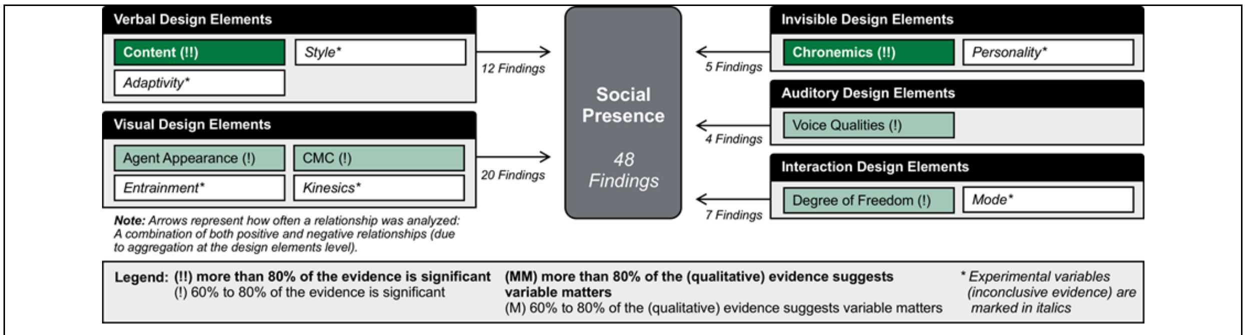

Figure 3. CA Outcomes regarding Social Presence

In this model, we found 2 variables showing consistency of 80% or more, as well as 4 variables showing consistency of 60% or more. The 7 remaining variables were categorized as experimental variables due to having less than 3 observed samples (4 variables) or showing inconsistent findings (3 variables).

Past research offers some evidence that ***verbal design elements*** of conversational agents are determinants of *social presence*. Researchers have investigated the variables *style* (6 OBS, e.g., Bickmore and Schulman 2007), *content* (3 OBS, Kobori et al. 2016), and *adaptivity* (3 OBS, Schuetzler et al. 2014). Only *content* was found to be significant for *social presence*, i.e. small-talk utterances increased the perception of the liveliness of the agent (Kobori et al., 2016). Moreover, in our sample, we found considerable evidence of ***visual design elements*** being determinants of *social presence*. Researchers investigated the variables *CMC* (8 OBS, e.g., Candello et al. 2017), *agent appearance* (7 OBS, e.g., Lee et al. 2019), *kinesics* (3 OBS, Van Es et al. 2002), and *entrainment* (2 OBS, Qiu and Benbasat 2010). For instance, a CA with a humanoid embodiment was found to be perceived as significantly higher in *social presence* as compared to a CA with no embodiment features. Additionally, our study uncovered considerable evidence suggesting that ***invisible design elements*** are determinants of *social presence*. Researchers have investigated the variables *chronemics* (3 OBS, e.g., Gnewuch et al. 2018) and *personality* (2 OBS, e.g., Liao et al. 2018). For instance, dynamic delays in system response time, compared to near-instant responses, were observed to invoke higher perceptions of *social presence* and naturalness of the interaction (Gnewuch et al., 2018). Further, previous research on ***auditory design elements*** identified consistent and significant evidence on *social presence* (4 OBS, Qiu and Benbasat 2009). *Voice qualities* was found to be a significant determinant of *social presence*. For instance, low pitch contour and high flanging increments were found to significantly affect perceptions of humanness (Muralidharan et al., 2014). In addition, past research studying ***interaction design elements*** on *social presence* identified consistent and significant evidence. Researchers have studied the variables *mode* (4 OBS, Cho 2019) and *degree of freedom* (3 OBS, Diederich et al. 2019). Thereby, only variables related to the *degree of freedom* produced significant evidence. For example, pre-defined answer options were found to negatively affect perceptions of humanness (Diederich et al., 2019).

Summarizing our findings on the DV *social presence,* the most significant and consistent evidence regarding its determinants was found to be related to the two groups of variables coded as *content* (*verbal*), and *chronemics (invisible).* Other consistent and significant evidence was found regarding the variables *agent appearance* and *CMC (*both *visual), voice qualities (auditory),* and *degree of freedom (interaction).*

Independent Variables on Trust

Figure 4 provides an overview of the relationships between *trust* and the five IVs. In this model, 45 findings are synthesized and depicted based on their consistency within the subgroups of the IVs, providing an answer to the question: "Which determinants of *trust* were reported by past empirical research on CAs?".

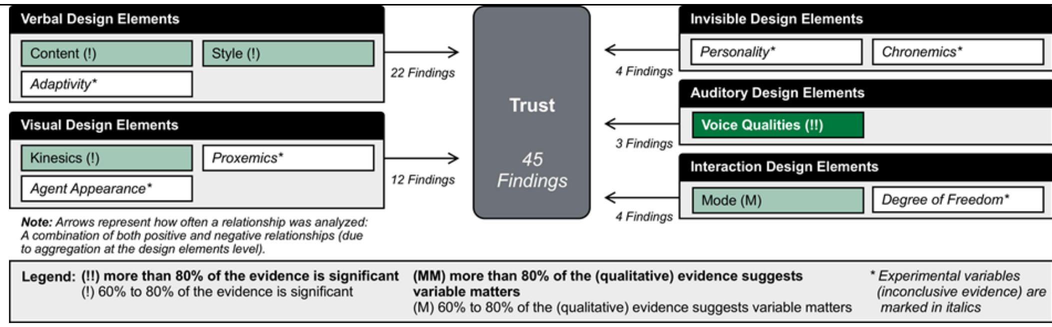

Figure 4. CA Outcomes regarding Trust

In this model, we found one variable showing consistency of 80% or more, as well as three variables showing consistency of 60% or more. Another six variables were categorized as experimental variables, due to having less than 3 observed samples (4 variables) or showing inconsistent findings (2 variables). The one remaining variable was categorized as argumentative.

Past research offers some evidence that ***verbal design elements*** are determinants of *trust*. Researchers have focused their investigation on the variables *style* (9 OBS, e.g., Kang and Wei 2018), *adaptivity* (8 OBS, Engelhardt et al. 2017), and *content* (5 OBS, e.g., Benlian et al. 2019). The variable *style* produced nearly consistent results and was marked six times in a significant relationship. In contrast, the connection between adaptivity and *trust* showed mixed results. The variable content has, in turn, shown promising results. In this case, three findings were significant, and two more were investigated qualitatively. Contrary, past research offers little evidence that ***visual design elements*** of conversational agents are determinants of *trust*. Prior research has investigated the variables *agent appearance* (8 OBS, e.g., Nunamaker et al. 2011), *kinesics* (3 OBS, e.g., Elkins and Derrick 2013), and *proxemics* (1 OBS, Benlian et al. 2019). Similarly, prior research has yet not been able to provide evidence that ***invisible design elements*** of conversational agents are determinants of *trust*. Nevertheless, past research has investigated the variables *personality* (3 OBS, e.g., Nordheim et al. 2019) and *chronemics* (1 OBS, Benlian et al. 2019). Concerning the variables personality and chronemics and their influence on trust, no conclusive findings have been elaborated. Additionally, there is little evidence that ***auditory design elements*** of conversational agents are determinants of *trust*. Nevertheless, past research has focused on *voice qualities* (3 OBS, e.g., Muralidharan et al. 2014) and found strong and consistent results. Moreover, prior investigations have not offered any evidence that ***interaction design elements*** are determinants of *trust*. However, the results obtained have a qualitative character.

Summarizing our findings on the DV *trust,* the most significant and consistent evidence regarding determinants of this outcome dimension was found to be related to the *auditory* design element of *voice qualities*. Additional evidence concerned the two groups of variables coded as *style and content* (*verbal*), *mode* (*interaction*), and *kinesics (visual)*. Other variables have not yet been able to show significant evidence in relation to *trust*. However, there are some promising avenues for future research.

Independent Variables on Utility

Our findings regarding the outcome dimension of *utility* are outlined in Figure 5. In this model, 92 findings are synthesized and depicted based on their consistency within the subgroups of the IVs, providing an answer to the question: "Which determinants of *utility* were reported by past empirical research on CAs?". We found three variables showing consistency of 80% or more, as well as three variables showing consistency of 60% or more. Another seven variables were categorized as experimental variables, due to having less than 3 observed samples (4 variables) or showing inconsistent findings (3 variables).

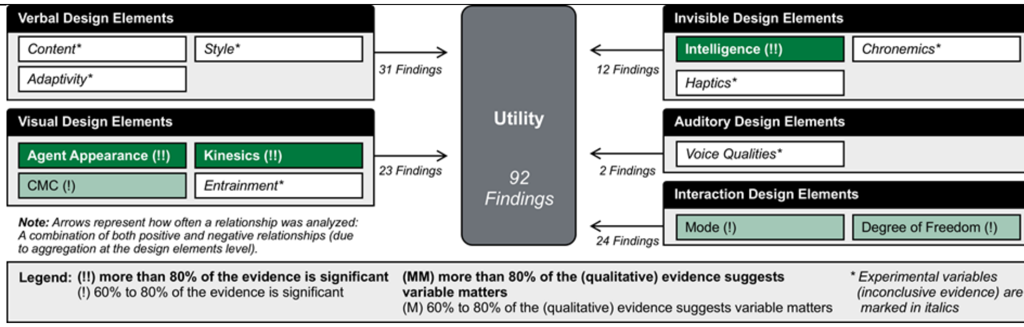

Figure 5. CA Outcomes regarding Utility

Past research offers no evidence that ***verbal design elements*** are determinants of *utility*. Researchers investigated the variables *style* (13 OBS, e.g., Kim et al. 2019), *content* (8 OBS, e.g., Kobori et al. 2016), and *adaptivity* (10 OBS, Engelhardt et al. 2017). The variables *style*, *adaptivity*, and *content* of the conversation have not yet been able to show evidence that the relationship to *utility* is significant. In contrast, we found considerable evidence of ***visual design elements*** being determinants of *utility*. Prior research has focused on the variables *CMC* (10 OBS, Westerman et al. 2019), *kinesics* (6 OBS, Cowell and Stanney 2005), *agent appearance* (5 OBS, e.g., McBreen and Jack 2001), and entrainment (2 OBS, Koulouri et al. 2016). For example, the embodiment of the CA with facial expressions was perceived as more useful, and the users seemed to be more satisfied than the faceless CA (Sproull et al. 1996).

Users seemed to be more satisfied when the CA had a controlled but normal gaze pattern than when it had a randomized gaze pattern (Van Es et al., 2002). Furthermore, past research offers evidence that ***invisible design elements*** are determinants of *utility*. Researchers have focused on the variables *intelligence* (6 OBS, e.g., Xu et al. 2017), *chronemics* (5 OBS, Chaves and Gerosa 2018), and *haptics* (1 OBS, Kim et al. 2018). For instance, the perceived usefulness was higher, when the CA was empowered by deep learning, than when it did not (Xu et al., 2017). On the contrary, we found no evidence that ***auditory design elements*** are determinants of *utility*. Only the influence of voice qualities has been investigated, but no significant evidence was found (Tian et al. 2017). *Voice qualities*, i.e., the distinctive characteristics between acted and natural speech, did not affect how well the CA recognized the users' emotions.

Additionally, past research offers some evidence that ***interaction design elements*** are determinants of *utility*. Prior researchers have found significant evidence when looking at the variable *mode* (15 OBS, Miehle et al. 2018). Concerning the *degree of freedom* (9 OBS, Mu & Sarkar 2019), no conclusive findings have been elaborated. For example, Akahori et al. (2019) were able to show that the main effects of the number of agents had a significant influence on understandability.

Summarizing the DV *utility,* the most significant and consistent evidence regarding determinants of this outcome dimension was related to the three groups of variables coded as *intelligence (invisible),* and *agent appearance* as well as *kinesics* (both *visual*). Other consistent and significant evidence was found regarding the variables *CMC (visual),* and *degree of freedom* as well as *mode (*both *interaction).*

Independent Variables on Behavioral Intention and Outcomes

Findings on the dimension of *behavioral intention and outcomes* are scarce (25 findings). Moreover, this category represents a broad category incorporating different intentions and behaviors, which is why we did not create an independent model that indicates the consistency within the subgroups of the IVs. In sum, past research mostly looked at visual (11 OBS; e.g., Strait et al. 2015) and verbal (10 OBS, e.g., Kobori et al. 2016) design elements. Only a minority of findings looked at interaction (2 OBS, e.g., Miehle et al. 2018) and invisible (2 OBS, e.g., Cafaro et al. 2013) design elements. No study in our sample investigated the effect of auditory design elements on behavioral intentions or outcomes. Summarizing our findings, we did not record any consistent findings within this subgroup. However, we find that while visual and verbal design elements are starting to be investigated, others like auditory features have not yet attracted any research.

Discussion and Development of a Research Agenda

In the next paragraphs, we discuss the findings based on our research space. This is possible as both quantitative and qualitative assessments have been purposefully aggregated. The research agenda is positioned in the research space that we have spanned by this review linking important attitudinal outcomes of CAs to categories of design elements. Furthermore, we relate our results to the overall HCI framework. Table 1 provides a summary of the progress in various fields of research. The Harvey Balls indicate how developed a research field is regarding the number and content of the respective contributions. Fields where no research had been found, were classified as very low. Fields, in which only a few items were found or the contributions represent only a first attempt to research (i.e., only qualitatively), were classified as low. Fields with some contributions or mixed results were classified as moderate. Fields with several contributions and consistent findings were classified as high. Fields that have already been fully covered (very high) were not found. In connection with the descriptive and thematic findings, this assessment provides the opportunity to assume the need for further research. In each thematic discussion, research opportunities are localized and characterized. Thus, we follow the framework of Müller-Bloch and Kranz (2015). Also, practitioners get a structured overview of existing knowledge in the field of CA design.

Table 2. Summary of Research Progress

Design Elements	Attitudinal and Perceptional Outcomes			
	Rapport	**Social Presence**	**Trust**	**Utility**
Interaction	◔	◔	◔	◑
Visual	◑	◕	◔	◕
Verbal	◑	◑	◑	◑
Auditory	◔	◔	◕	◔
Invisible	◕	◔	◑	◑

Legend: Low ◔, Moderate ◑, High ◕

The publications regarding *rapport* provide insights into design elements that contribute to forming a social bond with the agent. In total, we have found 92 relationships involving *rapport* within 40 unique studies, which testifies to the high research interest in this aspect of user interaction with the agent. Thereby, several relationships showed consistent and robust evidence of a positive effect on *rapport*. Notably, the CAs intelligence seems to afford a positive user evaluation of relationship to the agent (e.g., Xu et al. 2017). Moreover, we found consistent evidence that *agent appearance* and *CMC* (both visual) and *content* (verbal) may be positively related to *rapport*. However, there are still a few inconsistent findings, as shown in this review, which prompt further research. Potentially fruitful future research directions are particularly evident regarding the influence of the following design elements:

- **Mode (Interaction):** Understanding how *rapport* emerges between the user and CA in the context of different interaction modes has been the focus of several studies: In our review, we extracted 9 findings related to the influence of the interaction mode on *rapport*. For instance, D'Mello and colleagues (2010) investigated how the mode of interaction (text-based vs voice-based) influenced dynamics between students and a CA tutor, but the limited sample of the study moderates the explanatory power. Thus, we argue that exploring CA interaction mode and its influence on *rapport* represents a potentially insightful direction for future research.
- **Style (Verbal):** Across different studies, 20 findings were concerned with investigating the influence of the CA's verbal style on *rapport* dynamics between the user and the CA. However, from an aggregate perspective, the results were mixed. Six findings reported inconclusive evidence regarding any relationship. Three studies argued a positive relationship but did not provide any quantitative evidence. Therefore, we identify *rapport* between the user and CA in the context of different CA verbal styles as a potentially worthwhile avenue for future research.

Regarding *social presence,* we identified 48 findings within 25 unique studies. Thereby, especially those design elements that represent social cues of CAs not surprisingly showed a strong and consistent effect on *social presence*. Typically, CA visual design is the primary research area when considering *social presence*. Though, the findings related to appearance are not as consistent as one would expect, especially since agent appearance is a well-researched area when considering early research on other agents (Nowak & Biocca, 2003). Thus, future research should incorporate a more nuanced and configurational view on these design elements, since we suspect that the interrelationship could be key for understanding *social presence* with CAs. Hence, we propose the following prospects for future research:

- **Agent Appearance (Visual):** Future research should provide a more nuanced understanding of what appearance elements positively influence social presence and to what degree. Arguing from an "uncanny valley" perspective (Mori et al., 2012), specific elements of agent appearance could be related more consistently to leveraging *social presence* perceptions than others, especially anthropomorphic design elements (Pfeuffer et al., 2019).
- **Configurational View on Social Presence:** Connected with the first avenue for future research, we suggest that agent appearance from a visual perspective as well as other elements leveraging *social presence* should not be treated in isolation and carefully considered in a configurational view. For instance, with other aspects such as the personality of the agent, which has been, so far, neglected in research. For example, Amazon's Alexa has a very minimalistic appearance but rather a high degree of *social presence* through other elements fostering *social presence* (Purington et al., 2017).

The contributions regarding *trust* provide insights into design elements that form *trust* in the agent. In total, we have found 45 relationships involving *trust* within 23 unique studies. Thereby, several relationships showed strong evidence, which testifies to the relative maturity of this research stream. Especially, voice qualities seem to be strongly associated with trust-formation (e.g., Muralidharan et al. 2014). Moreover, we found consistent evidence that *content* and *style* (verbal), *kinesics* (visual), and *mode* (interaction) may be positively related to trust. However, these are results that should be corroborated by further studies and replicated in different contexts. Further research opportunities are especially evident regarding the points of:

- **Personality (Invisible).** The reasoning of Nordheim et al. (2019) indicates that attributing personality to CAs shall positively affect perceptions of trustworthiness. However, the papers in our sample show inconclusive evidence regarding the effects of agent personality. Hence, we propose that future research further explores the attribution of personality dimensions (i.e., Big Five) on CAs and their effects on *trust*.
- **Agent Appearance (Visual).** In general, agent appearance has been explored relatively often regarding user *trust* (e.g., Nunamaker et al., 2011). However, the findings only herald mixed results. It seems that agent appearance is vital in some contexts and others not. Hence, there is a need to explore the effect of appearance design from a more nuanced perspective taking into account the respective context or task.

In the analyzed literature, manifold insights on how design elements contribute to creating *utility* for the user have been made. In this research stream, we identified 92 findings involving *utility* within 41 unique studies. Especially those characteristics afford *utility* that are concerned with the accessibility or functionality aspect of HCI. Thus, we found strong evidence for the effects of degree of freedom (interaction), intelligence (invisible), agent appearance, and kinesics (visual). Moreover, we found consistent evidence *adaptivity* (verbal), *CMC* (visual), and *mode* (interaction) may be positively related to *utility*; however, these are results that should be corroborated by further studies and replicated in different contexts. In general, the results regarding *utility* perceptions are already quite profound. However, we see merit, especially in the following research directions:

- **Adaptivity (Verbal):** Based on theoretical reasoning and qualitative data, several researchers highlighted that adaptivity might afford high potential for creating *utility* as users expect personalized content. However, other researchers argue that standardized content may contribute to ease of use (Chin & Yi, 2019). Hence, we propose to investigate the effect of CA adaptivity on *utility* in different contexts.
- **Mode (Interaction)**: Researching the *utility* of interaction with CA is undoubtedly one of the most common research side effects of research in the field of CA. However, very few have been concerned with merely investigating the influence of the chosen interaction mode on *utility*.

Furthermore, we only found 25 outcomes related to intentions and behavioral outcomes. This represents another vital research gap as the discipline is in dire need to investigate the influence of all aspects of CAs and their features both on actual user behavior and on evoked behavior changes in real-life settings in various domains.

Finally, we have identified some overarching research opportunities based on the overall positioning of the reviewed literature. According to Li and Zhang (2005), HCI is concerned with the interaction between an IS and a user. This interaction is shaped by the characteristics of the system, the user, and the task and context. Interaction results can include perceptions, attitudes, intentions, and behaviors (Li & Zhang, 2005). In the scope of reviewing the literature, we found that task context (i.e., support, assistance, coaching function) has been rarely implicated in the research design. Additionally, user characteristics have been hardly considered in the research model beyond being a control variable. Here, we see an important research gap as both task and user characteristics may dramatically influence the effect of CA design on user perceptions and behaviors.

Limitations

Despite us following established guidelines and attempting to rigorously analyze the identified empirical literature on user interaction with CAs, this SLR has several limitations that should be considered. First, the scope of this SLR is not exhaustive. Despite due diligence, the scope of the SLR might not be fully exhaustive, and our search strategy may have missed relevant publications. Nevertheless, we aimed at capturing a broad spectrum of research on user interactions with CAs by employing a journal as well as a proceedings-based search. Second, the indicated relationships between the design elements and user outcomes are based on our interpretation of prior empirical research. Furthermore, the number of findings ultimately coded and included in our dataset was limited. Thus, it is not our intention to suggest any kind of causality between the design elements and user outcomes. By employing the method introduced by Jeyaraj et al. (2006), it was our objective to elucidate the variables studied, and offer a conceptual structuring of the empirical findings on design elements and their influence on outcomes. Fourth, bias within the results was visible, which consisted of a strong overrepresentation of positive effects. Finally, the resulting research agenda imposes further limitations. As such, the resulting and presented research agenda cannot be regarded as complete. Additional research questions and streams might be formulated for each of the agenda’s parts.

Conclusion

The holistic evaluation of the empirical academic literature regarding user interaction with CAs is crucial to uncover potential research fields and gaps for shaping future empirical CA research. For this purpose, we conducted a systematic literature review to study which design elements had a significant influence on design outcomes. Following Jeyaraj et al. (2006), we identified, coded, validated, and analyzed quantitative and qualitative empirical findings on user interaction with CAs. We, therefore, analyzed the 107 identified research papers and systematically identified existing knowledge as well as future research needs. Despite the already impressive growth of existing CA research, the field continues to rapidly evolve. Hence, based on a systematic examination of relationships between major IVs and DVs, we have suggested areas of future research representing understudied or inconclusive yet promising variables and interactions.

References

Akahori, W., Miyake, A., Sugiyama, H., Watanabe, M., & Minami, H. (2019). *Paired Conversational Agents for Easy-to-Understand Instruction*. In *CHI EA '19*.

Antaki, C. (2008). Discourse analysis and conversation analysis. *The SAGE Handbook of Social Research Methods*, 431–446.

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). *Resilient chatbots: Repair strategy preferences for conversational breakdowns*. In *CHI '19*.

Benlian, A., Klumpe, J., & Hinz, O. (2019). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*.

Berry, D. C., Butler, L. T., & De Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies*, *63*(3), 304–327.

Bickmore, T., & Schulman, D. (2007). *Practical approaches to comforting users with relational agents*. 2291–2296. In *CHI EA '17*.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293–327.

Burgoon, J. K., Guerrero, L. A., & Manusov, V. (2013). *Nonverbal signals*. Cengage Learning.

Cafaro, A., Vilhjálmsson, H. H., Bickmore, T. W., Heylen, D., & Schulman, D. (2013). *First impressions in user-agent encounters: The impact of an agent's nonverbal behavior on users' relational decisions*. In *AAMAS 2013*.

Candello, H., Pinhanez, C., & Figueiredo, F. (2017). *Typefaces and the perception of humanness in natural language chatbots*. 3476–3487. In *CHI '17*.

Chaves, A. P., & Gerosa, M. A. (2018). *Single or Multiple Conversational Agents? An Interactional Coherence Comparison*. 1–13. In *CHI '18*.

Chin, H., & Yi, M. Y. (2019). *Should an agent be ignoring it? A study of verbal abuse types and conversational agents' response styles*. In *CHI EA '19*.

Cho, E. (2019). *Hey Google, Can I Ask You Something in Private?* In *CHI '19*.

Clark, L., Munteanu, C., Wade, V., Cowan, B. R., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., & Murad, C. (2019). What Makes a Good Conversation? *In CHI '19*.

Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies*, *62*(2), 281–306.

Diederich, S., Brendel, A. B., Lichtenberg, S., & Kolbe, L. (2019). *Design for fast request fulfillment or natural interaction? insights from an experiment with a conversational agent*. In *ECIS 2019*.

D'Mello, S. K., Graesser, A., & King, B. (2010). Toward spoken human–computer tutorial dialogues. *Human–Computer Interaction*, *25*(4), 289–323.

Elkins, A. C., & Derrick, D. C. (2013). The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, *22*(5), 897–913.

Engelhardt, S., Hansson, E., & Leite, I. (2017). *Better Faulty than Sorry: Investigating Social Recovery Strategies to Minimize the Impact of Failure in Human-Robot Interaction*. 19–27. In *WCIHAI 2017*.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*.

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). *Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction*. In *ECIS 2018*.

Hall, E. (1990). *The silent language (Anchor Book ed.)*.

Hauswald, J., Laurenzano, M. A., Zhang, Y., Yang, H., Kang, Y., Li, C., Rovinski, A., Khurana, A., Dreslinski, R. G., & Mudge, T. (2016). Designing future warehouse-scale computers for Sirius, an end-to-end voice and vision personal assistant. *ACM Transactions on Computer Systems (TOCS)*, *34*(1), 1–32.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, *1*(2), 262–271.

Huber, B., McDuff, D., Brockett, C., Galley, M., & Dolan, B. (2018). *Emotional dialogue generation using image-grounded language models*. 1–12. In *CHI '18*.

Janssen, A., Passlick, J., Cordona, D. R., & Breitner, M. H. (2020). Virtual Assistance in Any Context: A Taxonomy of Desgin Elements for Domain-Specific Chatbots. *Business & Information Systems Engineering*.

Jeong, Y., Lee, J., & Kang, Y. (2019). *Exploring Effects of Conversational Fillers on User Perception of Conversational Agents*. 1–6. In *CHI EA '19*.

Jeyaraj, A., Rottman, J. W., & Lacity, M. C. (2006). A review of the predictors, linkages, and biases in IT innovation adoption research. *Journal of Information Technology*, *21*(1), 1–23.

Kang, J., & Wei, L. (2018). *"Give Me the Support I Want!" The Effect of Matching an Embodied Conversational Agent's Social Support to Users' Social Support Needs in Fostering Positive User-Agent Interaction*. 106–113. In *HAI 2018*.

Kim, Seyeong, Row, Y., & Nam, T.-J. (2018). *Thermal Interaction with a Voice-based Intelligent Agent*. In *CHI EA '18 2018*.

Kim, Soomin, Lee, J., & Gweon, G. (2019). *Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality*. In *CHI '19*.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Knote, R., Janson, A., Eigenbrod, L., & Söllner, M. (2018). *The what and how of smart personal assistants: Principles and application domains for IS research*. In *MKWI18*.

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). *Classifying Smart Personal Assistants: An Empirical Cluster Analysis*. In *Proceedings of the 52nd HICSS*.

Kobori, T., Nakano, M., & Nakamura, T. (2016). *Small talk improves user impressions of interview dialogue systems*. In *SIGDIAL 2016*.

Koulouri, T., Lauria, S., & Macredie, R. D. (2016). Do (and say) as I say: Linguistic adaptation in human–computer dialogs. *Human–Computer Interaction*, *31*(1), 59–95.

Krämer, N. C., Iurgel, I., & Bente, G. (2005). *Emotion & motivation in embodied conversational agents*. In *AISB 2005*.

Krämer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you—The effects of a virtual agent‘s smile on users' evaluation and behavior. *International Journal of Human-Computer Studies*, *71*(3), 335–349.

Lacity, M. C., Khan, S., Yan, A., & Willcocks, L. P. (2010). A review of the IT outsourcing empirical literature and future research directions. *Journal of Information Technology*, *25*(4), 395–433.

Laumer, S., Maier, C., & Gubler, F. T. (2019). *Chatbot Acceptance in Healthcare: Explaining User Adoption of Conversational Agents for Disease Diagnosis*. In *ECIS 2019*.

Leathers, D. G., & Eaves, M. (2015). *Successful nonverbal communication: Principles and applications*. Routledge.

Lee, M., Ackermans, S., van As, N., Chang, H., Lucas, E., & IJsselsteijn, W. (2019). *Caring for Vincent: A Chatbot for Self-Compassion*. In *CHI '19*.

Lee, SeoYoung, & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, *103*, 95–105.

Lee, Sunok, Kim, S., & Lee, S. (2019). *" What does your Agent look like?" A Drawing Study to Understand Users' Perceived Persona of Conversational Agent*. In *CHI EA '19*.

Li, N. L., & Zhang, P. (2005). The intellectual development of human-computer interaction research: A critical assessment of the MIS literature (1990-2002). *JAIS*, *6*(11), 9.

Liao, Q. V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M. P., Wang, D., Muller, M., Shami, N. S., & Geyer, W. (2018). *All Work and No Play?* In *CHI '18*.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering*, *61*(4), 535–544.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, *20*(3), 709–734.

Mayer, R. E., Johnson, W. L., Shaw, E., & Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human-Computer Studies*, *64*(1), 36–42.

McBreen, H. M., & Jack, M. A. (2001). Evaluating humanoid synthetic agents in e-retail applications. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *31*(5), 394–405.

McTear, M., Callejas, Z., & Griol, D. (2016). Conversational interfaces: Devices, wearables, virtual agents, and robots. In *The Conversational Interface* (pp. 283–308).

Miehle, J., Minker, W., & Ultes, S. (2018). *Exploring the impact of elaborateness and indirectness on user satisfaction in a spoken dialogue system*. In *UMAP '18*.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100.

Mu, J., & Sarkar, A. (2019). *Do We Need Natural Language? Exploring Restricted Language Interfaces for Complex Domains*. 1–6. In *CHI EA '19*.

Müller-Bloch, C., & Kranz, J. (2015). *A framework for rigorously identifying research gaps in qualitative literature reviews*. In *ICIS 2015*.

Muralidharan, L., de Visser, E. J., & Parasuraman, R. (2014). The effects of pitch contour and flanging on trust in speaking cognitive agents. 2167–2172). In *CHI EA '14*.

Narducci, F., Basile, P., de Gemmis, M., Lops, P., & Semeraro, G. (2019). An investigation on the user interaction modes of conversational recommender systems for the music domain. *User Modeling and User-Adapted Interaction*, 1–34.

Nass, C., Steuer, J., & Tauber, E. R. (1994). *Computers are social actors*. In *CHI '94*.

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. *Interacting with Computers*, *31*(3), 317–335.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *12*(5), 481–494.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems*, *28*(1), 17–48.

Park, E. K., & Sundar, S. S. (2015). Can synchronicity and visual modality enhance social presence in mobile messaging? *Computers in Human Behavior*, *45*, 121–128.

Pecune, F., Chen, J., Matsuyama, Y., & Cassell, J. (2018). *Field trial analysis of socially aware robot assistant*. 1241–1249. International Conference on Autonomous Agents and MultiAgent Systems.

Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic Information Systems. *Business & Information Systems Engineering*, *61*(4), 523–533.

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). *Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo*. In *CHI EA '17*.

Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems*, *25*(4), 145–182.

Qiu, L., & Benbasat, I. (2010). A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer Studies*, *68*(10), 669–688.

Rzepka, C., & Berger, B. (2018). *User interaction with AI-enabled systems: A systematic review of IS research*. In *ICIS 2018*.

Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2019). The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*, *97*, 250–259.

Schuetzler, R. M., Grimes, M., Giboney, J. S., & Buckman, J. (2014). *Facilitating natural conversational agent interactions: Lessons from a deception experiment*. In *ICIS 2014*.

Shamekhi, A., Czerwinski, M., Mark, G., Novotny, M., & Bennett, G. A. (2016). *An exploratory study toward the preferred conversational style for compatible virtual agents*. In *IVA 2016*.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, *11*(2), 97–124.

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. In *CHI '15*.

Tian, L., Moore, J. D., & Lai, C. (2017). *Recognizing emotions in spoken dialogue with acoustic and lexical cues*. In *ISIIA 2017*.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, *4*(3), 356–367.

Van Es, I., Heylen, D., van Dijk, B., & Nijholt, A. (2002). *Gaze behavior of talking faces makes a difference*. 734–735. In *CHI EA '02*.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, *37*, 205–224.

Vtyurina, A., Savenkov, D., Agichtein, E., & Clarke, C. L. (2017). *Exploring conversational search with humans, assistants, and wizards*. In *CHI EA '17*.

Vugt, H. C. V., Bailenson, J. N., Hoorn, J. F., & Konijn, E. A. (2008). Effects of facial similarity on user responses to embodied agents. *ACM TOCHI*, *17*(2), 1–27.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), 13–23.

Weizenbaum, J. (1966). ELIZA---a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I believe in a thing called bot: Perceptions of the humanness of "chatbots." *Communication Studies*, *70*(3), 295–312.

Winkler, R., & Söllner, M. (2018). *Unleashing the potential of chatbots in education: A state-of-the-art analysis*. In *AOM 2018*.

Winkler, R., Söllner, M., Neuweiler, M. L., Rossini, F. C., & Leimeister, J. M. (2019). *Alexa, can you help us solve this problem? How conversations with smart personal assistant tutors increase task group outcomes*. In *CHI EA '19*.

Wuenderlich, N. V., & Paluch, S. (2017). *A nice and friendly chat with a bot: User perceptions of AI-based service agents*. In *ICIS 2017*.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). *A new chatbot for customer service on social media*. 3506–3510. In *CHI '17*.

Yu, Q., Nguyen, T., Prakkamakul, S., & Salehi, N. (2019). *" I Almost Fell in Love with a Machine" Speaking with Computers Affects Self-disclosure*. In *CHI EA '19*.