Assessing the Reusability of Design Principles in the Realm of Conversational Agents

Edona Elshan[1] [0000-0002-4737-2392], Christian Engel[1][0000-0001-7349-7933], Philipp Ebel[1] and Dominic Siemon[2] [0000-0002-2945-4167]

[1] University of St.Gallen, 9000 St.Gallen, Switzerland; {name.surname}@unisg.ch
[2] LUT University , 53850 Lappeenranta, Finnland; dominik.siemon@lut.fi

Abstract. Conversational Agents (CAs) provide the means to foster user experience design through seizing their interaction capability, knowledgeability, and human-like behavior. To support practice and academia in designing CAs, IS researchers have been creating design knowledge in the form of design principles (DPs) guided by the Design Science paradigm. However, scientific literature in this vein is dispersed and lacks an axis of cohesion and transferability to sustained practice usage. This raises the question of reusability of design principles in the realm of CAs. Therefore, in this study, we conduct a Systematic Literature Review to retrieve and assess design principles of existing design science papers dealing with CAs with regard to their reusability. Our findings indicate that the Design Science community, in our case in the domain of CAs, seems to face challenges in creating reusable design principles. We discuss this observation and provide avenues on how to move forward

Keywords: Conversational Agents, Design Science, Design Principles, Reusability Assessment.

1 Introduction

Conversational agents (CAs) are up-and-coming, fostering individualized interactions between users and companies due to their innovative properties, such as the various possibilities to interact, their knowledgeability, and their human-like behaviour. With the rise of end-user-oriented CAs, the access to these novel applications has been democratized. CAs can be regarded as software agents that are designed to aid users in performing various activities by interacting with users via natural language [1, 2]. These CAs aregradually evolving to become the dominant mode of delivering user experiences designed by miscellaneous service providers [3]. However, designing CAs in a manner that leads to high levels of user satisfaction still poses a challenging task for user experience designers due to the novelty of this class of systems and the ambiguity of design outcomes.

To counteract this, research has started to elaborate on corresponding design knowledge seizing Design Science Research (DSR) as a guiding paradigm to structure CA-oriented design endeavors [4]. Due to its suitability for constructing socio-technical artifacts, DSR has become widely used in the field of Information Systems [5]. The notion of design science is described as a process of originating novel arti-

facts in Information Systems. Thereby, a particular problem set is addressed and further assessed regarding the usefulness of addressing that specific set of problems in multiple steps. Outstanding DSR artifacts may result in impact through novelty and generalized theories of design, enabling practical application to real-world problems. Thus, DSR offers the opportunity of providing guidance to both practitioners and researchers within the development process and to ensure that the developed CA is serving its intended purpose.

However, the scientific literature is dispersed into different thematic axes and research areas [6, 7]. Furthermore, the scientific and practical knowledge about CAs has also grown in a dispersed manner, given a shortage of integrative perspectives to support CA development and design processes [8, 9]. This leads to challenges in the field of DSR, as emphasized by Iivari et al. [10, 11], as well as Cronholm and Goebel [12]. The authors stress the necessity for Design Science researchers to not neglect reusability of design principles (DPs), as design DPs "found applicable by practitioners and not useful in practice" pose a mismatch with the basic idea of DSR. Taking action on the raised concerns, Iivari et al.[11] recently presented a proposal for assessing existing DPs of DSR papers by systematically evaluating the extent of their reusability based on different criteria. This provides a suitable framework for appraising any DPs with respect to their reusability.

Against this backdrop, we aim to address the mentioned shortage of integrative perspectives on CA development and design processes; we assess DPs of existing research papers dealing with CAs by actively evaluating their reusability for practitioners. Thus, we formulated the following research question (RQ) that we aim to address in this paper, performing a two-step analysis.

RQ: To what extent are existing DPs in the realm of CAs reusable by other researchers and chatbot designers?

By answering this research question, we intend to contribute on the one hand to the DSR community by evaluating the usability of DSR artifacts in a specific context, on the other hand, we intend to contribute to the research field of CAs by shedding light on the reusability of DPs for academia as well as practice. We hope to foster and contribute to the ongoing discussion in the field of design science. Overall, this work shall present the first evaluation of DPs in a specific research area and should highlight opportunities for improvement and action for the future creation of prescriptive Design knowledge in IS. The remainder of this paper is structured as follows. First, we provide a brief overview of the current research on CAs and introduce our conceptual notion of DPs. In the next step, we describe the method regarding the systematic literature review as well as the reusability framework proposed by Iivari et al. [11]. Subsequently, we present our descriptive findings and insights into the current state of CA literature within design science research. Furthermore, we present our results regarding the reusability of prescriptive design knowledge, i.e., DPs. Based on this, we discuss our results, lay out the imminent limitations of our analysis, and provide areas for future research on advancing both the field of CAs and DSR. Finally, we close this paper with implications for practice and academia and some concluding remarks.

2 Conceptual Background

2.1 Conversational Agents

AI-based CAs assist users by interacting with them using natural language [13]. CAs can respond to user input, adapt their responses, and build up a dialogue with them, similar to human-human interaction. CAs are distinguished from other intelligent system entities by their interaction and intelligence capabilities [13]. Mainstream conversational intelligent systems follow strict behavioral patterns. Those agents could only match user inputs against stored patterns [1]. However, CAs can now process compound natural language and thus respond to more complex user requests [14]. They can also adapt their responses to the user's workflows, knowledge state, and dialogue routines. CAs are thought to improve quality in various personal and professional tasks. They are expected to increase worker productivity by adapting to their tasks and routines [15]. They are intended to improve user comfort and well-being in private settings.

Considering the widespread of CAs, the industry anticipates high user adoption [16]. However, the opposite has proven to be true [17]. The high level of contextualization required to provide a flawless user experience makes designing a new CA a difficult task. However, design elements are the distinctive technical, conceptual, and knowledge features that frame a CA [18]. Even though a large body of research, mostly in the IS domain, investigated the design of CAs in various contexts, there are no general CA design guidelines, only high-level suggestions and domain-specific advice. As a result, many CAs confuse, frustrate, and even annoy users [19]. Thus, an integrated analysis aggregating design science insight on the diversity of CA design knowledge could help us better understand CA design and identify future research needs. Moreover, despite the rapid yet segregated growth of practical and scientific knowledge in this area, we are unaware of any review or evaluation that focuses on CAs design knowledge. Thus, we address the lack of an integrative perspective by systematically analyzing the DS literature on CAs to identify design knowledge, assess reusability, and identify research needs.

2.2 DPs as Generalizable Design Knowledge

Hevner et al. [20] consider DSR as a proactive paradigm when it comes to IT as DSR enables entities to address relevant problem sets in IT through the conception and assessment of novel artifacts. Therefore, when it comes to technology, DSR appears to be more hands-on and agile as this kind of research is in the continuous process of new artifact creation and utility assessment of corresponding addressed problems. Baskerville et al. [21] show that DSR requires the creation of artifacts and the formulation of their design. Thereby, DSR artifacts show a considerable impact on the dimensions of theory as well as practice. DSR is widely seen as an opportunity to respond to calls for academics to commit themselves to work that has a greater resonance outside the scientific community [22].

Taking up on this, scholars have been exploring the codification of knowledge in the field of DSR (e.g., [23]). The purpose of creating design knowledge in DSR projects is homogenous in its overarching goal of creating utility but heterogeneous in

how this is realized in as many different configurations as possible for creating and evaluating design knowledge in IS exist [24]. Taking this into consideration, Sturm and Sunyaev [25] observed that most knowledge is used by practitioners for designing artifacts, i.e., CAs. A form of representation of design knowledge can be found in DPs. Following Gregor and Hevner [5], DPs can be described as "generalized knowledge contribution in DSR". Moreover, DPs are comparable to Gregor and Jones [26] component of Level 3. The idea for the term of DP was shaped by the research of Gregor and Hevner [5] and Sein et al.[27]. It is noteworthy that Gregor and Hevner [5] have a slightly different perspective on DPs than Sein et al. [27]. While Sein et al. [27] are of the opinion that DPs should be theory-ingrained as a part of an action design research approach, Gregor and Hevner [5] argue that DPs emerge from DSR efforts.

Early on, Sein et al. [27] argued that DPs are how prescriptive knowledge should be expressed within design science research. Nevertheless, it should not be neglected that, in addition to DPs, other forms of presentation can be used to record design knowledge. Thus, design knowledge can also appear in patterns or requirements (e.g., [28]). Nevertheless, within the DSR community, it is widely shared that DPs are seen as the right way to formalize DSR efforts and abstract the research findings to allow knowledge accumulation within the community [29]. This accumulation of knowledge is reflected in how DPs describe how other instances belonging to the same class of systems can be created and designed [27]. According to Chandra et al. [30], design principles have two audiences. On the one hand, they add to a corpus of knowledge [31] regarding the design of various types of IT artifacts [32, 33]. On the other hand, they are also meant to provide practitioners with meaningful insights that can be used to develop new versions of related artifacts.

Recently, Purao et al. [23] discussed that it is naive to assume that practitioners simply take the DPs designed by scholarly researchers and apply them in their context. Other researchers have already addressed the underlying problem, namely that DPs reported in research articles currently find little application in practice (e.g., [11, 22]). Gregor et al. [22] assume that one of the main factors causing this is the formulation of DPs. Their study notes that there are many discrepancies and inconsistencies in the literature when it is a matter of how DPs should be formulated. In this context, they propose a formulation scheme that should help researchers to formulate DPs better. This awareness of the issue is also being shared through research by others. Lukyanenko and Jeffrey [34]conclude that design knowledge should be formulated in "clear, accessible and unambiguous language". In connection with the use of language to describe the design knowledge, Gregor et al. [22] note that hardly any attention is currently paid to the "people’s aspect" within DPs. These perceptions are also common to the framework's authors for reusability evaluation. Iivari et al. [11] claim that DPs should make explicit who the target community is, which should reuse this prescriptive form of knowledge.

3 Research Approach

3.1 Paper Selection Process

Our procedure for this paper contains two main steps: gathering and analyzing relevant literature, which is conducted as follows. First, to evaluate existing DSR research papers on CAs using the suggested proposal by Iivari et al. [11], we reviewed current work in design science research, particularly on CAs. To identify relevant literature as the basis for the systematic analysis, we conducted a systematic literature review (SLR) following Webster and Watson [35] and vom Brocke et al. [36]. The overall scope of the conducted SLR can be defined along the dimensions of process, source, coverage, and techniques of the SLR [36]. We used a comprehensive set of techniques to establish our data set and thereby the basis for the reusability evaluation (i.e., keyword search, backward search, and forward search).

Selection of search string. To identify a wide range of literature on CAs, the search string is chosen to be rather broad. Based on recent literature reviews (e.g., [6, 7]), we identified different keywords researchers used to describe CAs. This resulted in the following search string:

"*conversational agent" OR "chatbot" OR "chat bot" OR "interactive agent" OR "talkbot" OR "virtual assistant" OR "artificial intelligence assistant" OR "smart personal assistant" AND "design science"*

Selection of databases. For the literature search process in the first phase, we needed to identify relevant papers for our research paper to test the reusability of DPs. Research in the context of CAs has recently gained a lot of attention; therefore, we have chosen to conduct our literature review database-based. To ensure a high quality of the papers we analyze and evaluate, we relied on highly-ranked databases [37]. We conducted our search mainly using the following databases: AIS Electronic Library, IEEE Xplore Digital Library, ScienceDirect, ProQuest, and EBSCOhost. Eventually, we conducted a final search query using Google Scholar.

Selection of papers. By searching in the title, abstract, and keywords of the papers, the database-based search reveals 1032 hits. The identified papers are analyzed based on their abstracts in an initial screening process. We only included papers that referred to any type of CAs and presented any kind of design knowledge (DPs, guidelines, or design decisions). This first screening resulted in a set of 350 papers. In a subsequent step, we excluded paper that did not conduct design science research, did not report their design knowledge as well as papers that did not present design knowledge in our chosen context of CAs. Finally, the forward and backward search was carried out. Through screening the references and applying forward as well as backward searches using GoogleScholar, 21 articles were added to the set. Thus, resulting in the final number of 86 papers. After reviewing the search hits, reading through their abstract, and conducting a full-text search, we reached a total of 35 research papers that we can analyze regarding the reusability of their DPs subsequently. We excluded many of the articles since they did not formulate DPs but instead reported the design knowledge differently.

3.2 Paper Analysis

In the next step, we proceeded with our analysis of the selected research papers referring to the proposal for minimum reusability evaluation of DPs by Iivari et al. [11]. We thoroughly read this paper to understand the rationales of the proposal for evaluating DPs according to their reusability and the suggested procedure. In this context, we used the framework for reusability evaluation to analyze our selected set of academic literature thoroughly and comprehensively. Therefore, we carefully read the selection of DSR papers on CAs once again, screening for DPs, elements, or artifacts and depicting them in a structured spreadsheet. Subsequently, we evaluated every single DP, element or artifact suggested in science by assessing them concerning the specific criteria, accessibility, importance, novelty and insightfulness, actability and guidance, and effectiveness suggested by the authors. Based on the applicability check of IS research from Rosemann and Vessey [38], Iivari et al. [11] propose a reusability evaluation for DPs as artificats of DSR. It is important to note that the criteria do not have measurable scales or classifications; they merely indicate whether the criterion is applicable. Therefore, we decided to rate each of the DPs on a 5-point-Liker scale and assess whether their reusability would be low or high.
Furthermore, the framework requires that particular order of the criteria is followed, and therefore, the reusability framework cannot be called "flat" [10]. In this sense, the DPs are not considered reusable if one of the criteria is answered with no. We briefly discuss the individual criteria and describe their content in the following.

The first criterion is accessibility, intended to determine whether the DP has been formulated in such a manner that a member of the target community can grasp it with ease and without much effort [10]. Roseman and Vessey [38] also state that accessible representation is a representation that has adapted tone, style structure, and semantics to the target audience. If this is not the case, practitioners may not understand the DPs or their consequences, and they may not find acceptance and application outside the scientific literature. A possible workaround to enhance comprehensibility and usability is for researchers to formulate a practitioner-oriented version of their DPs [11]. Although this suggestion seems to be reasonable, it stands in contradiction to the actual purpose of DSR [5]. The authors argue that the importance is assessed in terms of the severity and relevance of the real problems they ultimately propose to overcome [11]. In a similar vein, importance is interpreted by Roseman and Vessey [38]. Namely, research is to be classified as being of importance if it "meets the needs of practice by addressing a real-world problem in a timely manner, and in such a way that it can act as the starting point for providing an eventual solution" (p.3). Having a look at the novelty and insightfulness criterion, it is evident that the novelty is typically assessed by fellow researchers solely and seldom by practitioners [11]. Thereby, the authors propose that practitioners should evaluate whether they perceive that the DPs have any kind of impact in a real-world context and do not only display knowledge that they already know (confirmatory). An extension of the novelty evaluation to include practitioners seems to make sense to the extent that research sometimes lags behind practice. Consequently, it is possible that something that is described as "new and innovative" in research may not be innovative in practice. This phenomenon was

observed, for example, in the development of digital maturity assessment models and can currently be observed with artificial intelligence maturity assessment models that were developed instead by consulting firms rather than scholars. The fourth criterion considers the actability and guidance of DPs. In the understanding of the authors' reusability framework, actability refers to the assumption that the DPs "can be acted and carried out in practice, i.e., under the control of the practitioners in question" [11].

Further, they suggest that DPs shall be "realistic to be carried out". It is important to note that Chandra Kruse et al. [29] have highlighted that tacit knowledge is compromised within DPs. In addition, they argue that no set of DPs is adequate to design the proposed instance without further guidance. Iivari et al. [11] take up this point and add that through appropriate guidance, this problem shall be addressed without being too restrictive. Lastly, the framework proposes to assess the effectiveness of the DPs. The authors refer "to effects or consequences of reusing the DPs in the adopting unit" (p.12), even though DPs might have effects at different levels as well. Having these criteria in mind, we were challenged to change perspectives and neutrally evaluate the DPs from a practitioner's point of view and assess whether the criteria are fulfilled or not. To facilitate the evaluation process, we bore in mind the presented example by Iivari et al. [11] and continuously referred to the more detailed elaboration of the defined criteria to be able to evaluate to what extent the design artifacts fulfill the criteria for reusability.

4 Results

We organized the findings into two sections. The first section examines descriptive statistics based on the meta-data of found literature. The second section examines the reusability of the proposed DPs in the context of CAs.

In total, we have analyzed 35 publications in the context of conversational agents that apply Design Science methodologies in order to generate design knowledge. The youngest paper is from 2021, and the oldest paper from 2005. Although at that time, the concept of DPs had not yet been introduced to the DSR, it was possible to identify approaches of guidelines in the paper which go in the direction of DPs. Therefore, this paper was kept in the dataset. Most papers have been published within the last three years, which supports our initial assumption that CAs indeed represent an emerging research area in DSR. This argument is underpinned by the fact that most papers are from conference proceedings, which gives testament to the relative youth of the field. In addition to distribution along the timeline also a distribution along application domains is visible within our results. It appears that there exists no application domain, that appears not suitable for design science research. This is evident from the many different contexts and application areas in which CAs are used. Further, it is noteworthy that a multitude of investigated studies is conducted in the HCI discipline, while publications in IS conferences and outlets are only recently picking up. Next, we provide more specific insights concerning the reusability of DPs in the context of CAs. In this sense, the following table summarizes the results along with the criteria of the reusability evaluation framework. Due to space limitations, the table includes only five papers. However, the analysis and evaluation of the whole data set identified 35

papers suggesting DPs in any context of CAs as their contribution to research and practice. It is also worth noting that the sample of papers selected in the following table is drawn entirely at random. In addition, we do not claim the actual quality of the papers but only examine their DPs.

The results of Table 1 indicate that almost all papers, with Lechler et al. [39] as an exception, somehow evaluated their DPs. The authors of the four papers demonstrated how the principles were implemented in a first mockup [40] or showed how the DPs were translated into design decisions and then implemented in prototypes [41] or even executed in the real world [42]. Other evaluation methods used were, for example, user experience evaluations [43], qualitative experiments [44], focus group discussion [45], or interviews [46].

Table 1. Evaluation of DPs of Expository Papers.

		Wambsganss et. al [40]	Winkler & Roos [41]	Meier et al. [47]	Lechler et al. [39]	Gnewuch et al. [48]
Design Princples		6 DPs for CA for course evaluations	11 DPs for CA for online educational context	4 DPs for CAs in health awareness context	6 DPs for CAs for feedback exchange	4 DPs for CAs in customer service
Evaluation Method		Online questionnaire	Focus Group Discussion	Experiment	N/A	Field Study for prototype
	Accessibil-	Mostly	Most of the times not clear	Sometimes not clear what is addressed	Formulation quite abstract.	Mainly accessible.
	Importance	Overall addresses a real-world problem, DP not.	DP, not referring to a significant real-world problem.	Overall important real-world problem, DPs not referring this.	Address mostly an important real- world problem.	Most of the DP address a real-world problem.
	Novelty and Insight-	DPs are somehow insightful, not novel.	DPs might be insightful, yet not very novel.	DPs are insightful for the context, but not very novel.	Some DPs are insightful, yet not all of them novel.	Novel, regard to state of research at that time.
	Actability and Guidance	Rather high actability & guidance.	Most probably actable and can be carried out in practice	Most probably actable and can be carried out in practice.	Actability relatively moderate.	Guidance within DPs appropriate.
	Effective- ness	Positive, online survey with students	Positive, experiment	Positive,: questionnaire	Not sure if positive	Might affect adopting unit positively
Final Evaluation		Moderate Reusability	Rather low Reusability	Moderate Reusability	Rather low Reusability	Moderate Reusability

Although there are many different approaches to how DPs have been evaluated, they are rarely evaluated with the help of the target community, which should use the DPs one day. Instead, most of the evaluations are conducted with the end-user, although many focus on the artifact that the DP is reflected in rather than the DP itself. For

example, Wambsganss et al. [40] formulated six DPs, which they shaped by applying the DSR approach of Hevner [49]. The DPs were composed of literature research and interviews with end-users. Thus, requirements for a CA were derived from the interviews. However, a look at these requirements shows that not all of them are aimed at concrete aspects of design. To evaluate the DP, the authors of this paper have created a mockup of a CA. For the design of this CA, they first transferred the DPs into a set of 12 design features. Unfortunately, the reader is not told how this step is done. Therefore, it is difficult to understand if the DPs could have been converted into other design features. Although design features are then used for the mockup, the authors describe the goal of the evaluation as follows: "The evaluation serves to verify if the DPs are of value to the lecturers and students and to identify change requests and additional DPs. This would lead one to conclude that the DPs were evaluated with the later designer of the artifact (the lecturer), but later in the process, it is explained that the evaluation was done with the help of 28 students. In their next step, they tested the CA in a real-world setting with 12 students and one lecturer. However, even this evaluation did not directly aim to evaluate the DP. In this "proof-of-usefulness" evaluation, the authors tested the CA they designed against a conventional survey tool.

In their study, Winkler et al. [41] proposed a set of 11 principles for the design of a CA as a learning tutor. For the evaluation of their DPs, they conducted a proof-of-concept evaluation in the form of a focus group discussion. With this, they aimed to check the validity of the requirements as well as the translation into DPs. Similar to the evaluation conducted by Wambsganss et al. [40] also, these authors did not involve real practitioners in their evaluation but rather students as end-users of the designed artifact. This issue can also be observed in other papers.

Interestingly Meier et al. [47] discuss critical findings regarding their DPs after their evaluation with end-users and experts. However, in their key findings, only the user's perception and usefulness are addressed. No further indications of the applicability of the DPs are given. However, this conflicts with the intended goal of their DPs, which the authors of this paper state as: "the presented DPs contribute to the information systems discipline by providing important guidance in designing successful CAs for practical challenges […]".

We think that, especially with regard to the comprehensibility of the DPs, it would be useful to have this tested by a potential user (e.g., CA designer) of the DPs. Further, our analysis reveals that most of the DPs are not accessible without further explanations. More specific, many of the DPs contain very specific terms, which are not understandable for inexperienced potential users. Tavanapour et al. [50], for instance, describe in their paper that the CA should resemble a social actor. However, they do not elaborate on what a social actor is or what it looks like. As a result, it is not clear how this DP should be instantiated. This may ultimately lead to the DPs not being applied for this very particular reason. Regarding the importance, it can be stated that most of the papers contain a problem that is important for the real world or the question came from a real-world problem. We think that this is since in many DSR studies requirements are also derived from the practitioners or users and therefore are incorporated into the DPs. However, it must be pointed out that on the level of the DPs, the real-world problem does not always have priority. Here the question arises, whether it

is sufficient that the set of DPs and their instance address a real-world problem or whether each individual DP should represent this circumstance.

With regards to the criterion of novelty and insightfulness, it must be noted that this is a difficult-to-evaluate criterion. An assessment of the novelty of the DPs presupposes that the reader is well versed in the field of CAs as well as in the specific context. Nevertheless, we have tried to evaluate whether the papers are novel or not through the seniority of the papers and their proposed DPs. However, this criterion cannot be verified merely by a temporal comparison, such as whether a DP formulated in 2021 takes up a new aspect compared to the previous DPs. In order to be able to carry out a comprehensive evaluation of novelty and insightfulness, the respective context must also be considered. It may be that certain DPs have a new meaning in the context of CAs, but this design knowledge has already been applied to other instances of artifacts. The question now is whether the transfer to a new class of systems is sufficiently innovative and new, or whether more than this "incremental" innovation is needed. Altogether, the results shown here based on these five research contributions can be transferred to the entire data set. However, there is hardly a single paper in the entire data set that presents flawless DPs. This is problematic from several points of view. First of all, this could mean that research, such as the development of taxonomies based on these DPs, also contains design elements that are not understandable to practitioners. In a subsequent logical step, the question arises of whom these DPs should be formulated. Our analysis has shown that the DPs are usually not formulated in an understandable way for practitioners.

Nevertheless, not only practitioners fall back on these DPs. It seems to be a bit worrying that other researchers are not able to transfer these DPs to their context. Consequently, it would be possible that research could lapse into creating DPs for each specific use case.

5 Discussion and Concluding Remarks

Our goal was to address the shortage of integrative perspectives on CA development and design processes. Thus, we assessed DPs of existing research papers dealing with CAs by evaluating their reusability using the framework proposed by Iivari et al. [11]. This shall contribute to research and practice in a three-fold manner. First, reviewing the reusability of DPs in the field of CAs, led to the insight that in DSR, the community is still faced with a low level of reusability of DPs. To assess reusability for practice, we left our scientific perspective and viewed DPs through a practice lens. Our analysis reveals that in many cases, the evaluation of the first criterion, i.e., accessibility, led to negative assessment as the DPs were not formulated in a way facilitating to be easily understood by non-academic readers. This finding is also in line with prior insights [22, 23], which have already noted that the formulation of DPs needs to be rethought to facilitate sustainable use in practice. Consequently, we encourage future research to position and formulate DPs guided by the notion established by Gregor et al.[22]. This bears the potential to foster transparency and clarity when communicating and conveying design knowledge. We also anticipate that this will provide more guidance when instantiating respective DPs into concrete artifacts.

Regarding transferability of design knowledge, the majority of analyzed papers faced challenges in using DPs to bridge the gap between conceptual scientific knowledge and its deployment in practice. In this sense, we also think that other approaches are needed to evaluate DPs in the context of design efforts. In many research papers, the DPs were not evaluated according to their usefulness and usability of the designed artifacts, although, in many research papers, this was the intended purpose of the evaluation. By using the reusability evaluation framework in a specific context, we show its applicability in another domain of IS research. This indicates a certain level of generalizability of the framework, which raises the question of whether the reusability evaluation framework could benefit the review process of research articles to account for the utility of the developed Design Science artifacts. Therefore, future research opportunities could be building up on Ivari et al. [11] by further detailing the single analysis steps to reach higher levels of intersubjectivity and comparability between individual evaluations. A next step could be formulating concrete questions that need to be answered in the individual criteria and thus to move in the direction of a standardized assessment. We believe that this be very useful for a review process. The DP formulation scheme [22] offers a promising foundation for enhancing comparability when assessing the reusability of DPs. This could be facilitated by combining and integrating both approaches in a structured method. However, also the framework for reusability must also be studied critically. While numerous criteria are available, evaluating DPs using these criteria is not always straightforward and leaves a great deal of space for interpretation.

Additionally, it should be noted that a DP is only reusable if it meets all requirements; if one of the criteria, for example, novelty or insightfulness, is not met, the DP is classified as not reusable. We would like to emphasize that it is feasible for a DP to be reusable even if the novelty value is not exceptionally high. As a result, a weighting of the criteria or a revision to the minimal standards for designating a DP as reusable should be explored in the future.

However, our paper does not come without limitations. First, the scope of this SLR cannot claim to be exhaustive. However, we intended to reach a representative coverage of Design Science literature in the domain of CAs by applying a rigorous research method for searching and analyzing the papers. Second, the indicated reusability and by this the assessment along the criteria of accessibility, importance, novelty and insightfulness, actability and guidance, and effectiveness of the DPs are based on our interpretation of the reported design knowledge in the studies. Thus, a certain residual level of subjectivity remains as this process involves individual human judgement. However, by assessing the DPs independently, we aimed to mitigate this issue. Future studies could also ask CAs designers independently to assess the DPs to prevent subjectivity.

References

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

2. Pfeuffer N, Benlian A, Gimpel H, Hinz O (2019) Anthropomorphic information systems. Business & Information Systems Engineering 61:523–533

3. McLean G, Osei-Frimpong K (2019) Hey Alexa… examine the variables influencing the use of artificial intelligent in-home voice assistants. Computers in Human Behavior 99:28–37

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

5. Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. MIS quarterly 337–355

6. Diederich S, Brendel AB, Morana S, Kolbe L (2022) On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research

7. Elshan E, Zierau N, Engel C, et al (2022) Understanding the Design Elements Affecting User Acceptance of Intelligent Agents: Past, Present and Future. Information Systems Frontiers 1–32

8. Brandtzaeg PB, Følstad A (2017) Why people use chatbots. In: International conference on internet science. Springer, pp 377–392

9. Jain M, Kumar P, Kota R, Patel SN (2018) Evaluating and informing the design of chatbots. In: Proceedings of the 2018 Designing Interactive Systems Conference. pp 895–906

10. Iivari J, Hansen MRP, Haj-Bolouri A (2018) A Framework for Light Reusability Evaluation of Design Principles in Design Science Research. In: 13th International Conference on Design Science Research and Information Systems and Technology: Designing for a Digital and Globalized World

11. Iivari J, Hansen MRP, Haj-Bolouri A (2020) A proposal for minimum reusability evaluation of design principles. European Journal of Information Systems 0:1–18. https://doi.org/10.1080/0960085X.2020.1793697

12. Cronholm S, Göbel H (2018) Guidelines Supporting the Formulation of Design Principles. In: 29th Australasian Conference on Information Systems (ACIS), Sydney, December 3-5, 2018

13. Maedche A, Morana S, Schacht S, et al (2016) Advanced User Assistance Systems. Business & Information Systems Engineering 58:367–370. https://doi.org/10.1007/s12599-016-0444-2

14. Knote R, Janson A, Söllner M, Leimeister JM (2020) Value co-creation in smart services: a functional affordances perspective on smart personal assistants. Journal of the Association for Information Systems

15. Jalaliniya S, Pederson T (2015) Designing wearable personal assistants for surgeons: An egocentric approach. IEEE Pervasive Computing 14:22–31

16. Nordheim CB, Følstad A, Bjørkli CA (2019) An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. Interacting with Computers 31:317–335

17. Zierau N, Elshan E, Visini C, Janson A (2020) A Review of the Empirical Literature on Conversational Agents and Future Research Directions

18. Janssen A, Passlcik D, Rodrıguez Cardona D (2020) Virtual Assistance in Any Context - A Taxonomy of Design Elements for. Business & Information Systems Engineering 62:

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

20. Hevner A, Prat N, Comyn-Wattiau I, Akoka J (2018) A pragmatic approach for identifying and managing design science research goals and evaluation criteria

21. Baskerville R, Baiyere A, Gregor S, et al (2018) Design science research contributions: Finding a balance between artifact and theory. Journal of the Association for Information Systems 19:3

22. Gregor S, Kruse LC, Seidel S (2020) The Anatomy of a Design Principle

23. Purao S, Chandra Kruse L, Maedche A (2020) The Origins of Design Principles: Where do… they all come from?

24. Engel C, Leicht N, Ebel P (2019) The Imprint of Design Science in Information Systems Research: An Empirical Analysis of the AIS Senior Scholars' Basket

25. Sturm B, Sunyaev A (2019) A Good Beginning Makes a Good Ending: Incipient Sources of Knowledge in Design Science Research

26. Gregor S, Jones D (2007) The anatomy of a design theory. Association for Information Systems

27. Sein MK, Henfridsson O, Purao S, et al (2011) Action design research. MIS quarterly 37–56

28. Knote R, Baraki H, Söllner M, et al (2016) From requirement to design patterns for ubiquitous computing applications. In: Proceedings of the 21st European Conference on Pattern Languages of Programs. pp 1–11

29. Kruse LC, Seidel S, Purao S (2016) Making use of design principles. In: International Conference on Design Science Research in Information System and Technology. Springer, pp 37–51

30. Chandra Kruse L, Purao S, Seidel S (Forthcoming.) How Designers Use Design Principles: Design Behaviors and Application Modes. Journal of the Association for Information Systems

31. Wieringa R (2009) Design science as nested problem solving. pp 1–12

32. Baskerville R, Pries-Heje J (2010) Explanatory Design Theory. Bus Inf Syst Eng 2:271–282. https://doi.org/10.1007/s12599-010-0118-4

33. Fischer C, Winter R, Wortmann F (2010) Design theory. Business & Information Systems Engineering 2:387–390

34. Lukyanenko R, Parsons J (2020) Design Theory Indeterminacy: What is it, how can it be reduced, and why did the polar bear drown? Journal of the Association for Information Systems

35. Webster J, Watson RT (2002) Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly xiii–xxiii

36. Vom Brocke J, Simons A, Riemer K, et al (2015) Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. CAIS 37:9

37. Levy Y, Ellis TJ (2006) A systems approach to conduct an effective literature review in support of information systems research. Informing Science 9:

38. Rosemann M, Vessey I (2008) Toward improving the relevance of information systems research to practice: the role of applicability checks. Mis Quarterly 1–22

39. Lechler R, Stöckli E, Rietsche R, Uebernickel F (2019) Looking beneath the tip of the iceberg: The two-sided nature of Chatbots and their roles for digital feedback exchange

40. Wambsganss T, Winkler R, Schmid PS, Söllner M (2020) Designing a Conversational Agent as a Formative Course Evaluation Tool. In: 15th International Conference on Wirtschaftsinformatik

41. Winkler R, Roos J (2019) Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors

42. Gnewuch U, Morana S, Heckmann C, Maedche A (2018) Designing conversational agents for energy feedback. In: International Conference on Design Science Research in Information Systems and Technology. Springer, pp 18–33

43. Chen Z, Lu Y, Nieminen MP, Lucero A (2020) Creating a Chatbot for and with Migrants: Chatbot Personality Drives Co-Design Activities. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. pp 219–230

44. Ahmad R, Siemon D, Fernau D, Robra-Bissantz S (2020) Introducing" Raffi": A Personality Adaptive Conversational Agent. In: PACIS. p 28

45. Elshan E, Ebel P (2020) Let's Team Up: Designing Conversational Agents as Teammates

46. Strohmann T, Fischer S, Siemon D, et al (2018) Virtual Moderation Assistance: Creating Design Guidelines for Virtual Assistants Supporting Creative Workshops. In: PACIS. p 80

47. Meier P, Beinke JH, Fitte C, et al (2019) FeelFit-Design and Evaluation of a Conversational Agent to Enhance Health Awareness. In: ICIS

48. Gnewuch U, Morana S, Maedche A (2017) Towards designing cooperative and social conversational agents for customer service

49. Hevner AR (2007) A three cycle view of design science research. Scandinavian journal of information systems 19:4

50. Tavanapour N, Poser M, Bittner EAC (2019) SUPPORTING THE IDEA GENERATION PROCESS IN CITIZEN PARTICIPATION - TOWARD AN INTERACTIVE SYSTEM WITH A CONVERSATIONAL AGENT AS FACILITATOR. 18