Why do Chatbots fail? A Critical Success Factors Analysis

Completed Research Paper

Antje Janssen
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany
janssen@iwi.uni-hannover.de

Lukas Grützner
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany
gruetzner@iwi.uni-hannover.de

Michael H. Breitner
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany
breitner@iwi.uni-hannover.de

Abstract

Chatbots gain more and more attention, both in research and in practice, and enter several application areas. While much research addresses technical or human-centered aspects, development, and adoption, little is known about Critical Success Factors (CSFs) and failure reasons of chatbots in practice. Design Science Research (DSR) oriented, we first analyze 103 real-world chatbots to examine the discontinuation rate of chatbots in 15 months. With a literature review and 20 expert interviews, we derive 12 specific CSFs and identify failure reasons which are evaluated in a focus group discussion with chatbot experts, afterwards. We explain chatbots' failure in practice, improve chatbot knowledge in Information Systems (IS) and Human Computer Interaction (HCI), and finally deduce recommendations and further research opportunities.

Keywords: Chatbot, conversational agent, failure reasons, critical success factors, design science research

Introduction

Due to technological advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) as well as increasing user acceptance, chatbots have gained tremendous popularity in research and practice over the last years (Adamopoulou and Moussiades 2020; Diederich et al. 2021). Practitioners see the chatbot market growing from $17.17 billion in 2020 to $102.29 billion in 2026, indicating the high relevance of the field (Mordor Intelligence 2021). This progress is also visible within an enormous increase of scientific publications about chatbots (Zierau et al. 2020; Adamopoulou and Moussiades 2020). Chatbots also known under the term conversational agents (Zierau et al. 2020) are mostly internet-based software systems that interact with humans within a simulated conversation to perform tasks (Brandtzaeg and Følstad 2018). These assistants are used to automate redundant processes in a wide variety of areas, such as education, health or customer support, e.g., to ensure 24/7 availability, to increase efficiency or to minimize customer support costs (Adamopoulou and Moussiades 2020; Janssen et al. 2020) and can be found on websites, social networks or apps (Janssen et al. 2020). Chatbots are seen as typical examples of HCI, as they are constantly changing due to further technological developments whereas the interaction with the user is

crucial for chatbot adoption and success (Adam et al. 2021). Although chatbots are not a new technology (Schumaker et al. 2007) and the general availability of technology and tools is increasing significantly (Galitsky 2019), from the end-user side, there still is a high failure rate of developed chatbots that are unable to understand user input, do not react (Seeger and Heinzl 2021; Brandtzaeg and Følstad 2018, Følstad et al. 2018; Filipczyk et al. 2016) or cannot longer be found on their previous used communication channels. These negative and frustrating experiences in interacting with chatbots are still seen as one of the key challenges in the deployment of chatbots (van der Goot et al. 2021; Følstad et al. 2018). Despite this, there has been minimal qualitative research to date exploring the exact reasons why organizations are taking their chatbots offline permanently or no longer maintaining them. Chatbot failure is annoying not only from the user's point of view, but also from the chatbot provider's perspective, who has put a lot of work, time and money into the development, as well as from a global perspective, as single negative chatbot experiences can damage the reputation of chatbots in general (van der Goot et al. 2021). Scientific literature considers various individual perspectives, such as single-case field reports from practitioners about challenges in deploying chatbots (Fiore et al. 2019), characteristics relevant to failure or success based on user surveys (Rodríguez Cardona et al. 2021; Diederich et al. 2021; Mozafari et al. 2021), and guidelines for improving chatbot elements (Brandtzaeg and Følstad 2018). Seeger and Heinzl (2021) proved that high failure rates among customer service chatbots may harm customer trust and stimulate negative word of mouth. These authors recommend using 'anthropomorphic design elements' to avoid these effects (Seeger and Heinzl 2021). An aggregation of all these technical, behavioral, and institutional aspects and perspectives on failure and success of chatbots is not yet available. A deeper understanding of the various factors from different stakeholders that impact chatbots' failures and success would increase the probability of a chatbot achieving success. For this reason, it is desirable to investigate the critical success factors (CSFs) affecting the development, deployment and usage of chatbots. We address these two research needs, which lead to our following research questions:

RQ1: What are reasons for chatbots' failure in practice?

RQ2: What are critical success factors for chatbots?

We use a DSR process model as a structural guide (Baskerville et al. 2018) to identify reasons for chatbot failure and the key factors that determine chatbot success. In four main steps (Doyle et al. 2019) we determine our research problem by analyzing 103 real-world chatbots and identifying reasons for chatbot failure from 154 academic papers and 20 expert interviews. We further develop and evaluate within a focus group discussion (FGD) twelve domain superior chatbot CSFs. We discuss our results and findings, and present implications, limitations and strategies for research and practice before concluding.

Research Design and Methodology

Design Science Research

To answer our research questions, we follow the DSR principles to address a design problem experienced by many service providers and service users, in particular to provide the foundations for the successful development and deployment of chatbots (Gregor and Hevner 2013). DSR is a problem-solving paradigm used to generate design knowledge and theoretical insights based on the formation of a theory-based artifact and/or the implementation of empirical design principles in form of constructs, methods, models, prototypes or design theories (Hevner et al. 2004; Baskerville et al. 2018; vom Brocke et al. 2020a). Within this study, the artifact in form of the CSFs can be categorized into the design theory. As structural guidance (Baskerville et al. 2018), we refer to a high level DSR process model. Doyle et al. (2019) compared multiple DSR process models, e.g., Peffers et al. (2007) six-step model, and identified four core steps. We adopt these generic process steps depicted by Doyle et al. (2019).

Following the DSR steps outlined by Doyle et al. (2019), the first step is to identify the problem and the necessity for research by analyzing 103 chatbots. The second step consists of designing and building our artifact, starting with the identification of a solution. We gain failure reasons and gather requirements for CSFs for chatbots based on the analysis of 154 scientific papers. We further conduct 20 expert interviews, to learn about their experiences regarding chatbots' failures and gather further requirements to verify and enhance our theory derived set of CSFs, which are then evaluated through an independent FGD in step three. In step four, reasons for failure of chatbots and CSFs are presented in our results section.

Step 1: Problem Identification by Chatbot Taxonomy Analysis

While the development, release, and deployment of chatbots is largely reported in scientific literature and practitioner reports, there are only isolated reports of failed chatbots, such as the Microsoft chatbot Tay (Brandtzaeg and Følstad 2018). From a user perspective, several researchers define chatbot failure within a human-to-chatbot interaction in the form of conversation irregularities, i.e., the chatbot's inability to perform tasks for the user through its inability to either interpret the intent or to provide proper responses (Filipczyk et al. 2016; Seeger and Heinzl 2021; Mozafari et al. 2021). From a more global perspective, we expand the chatbot failure definition by also including chatbots that have been discontinued by being taken offline.

To explore the extent to which chatbot discontinuation is an issue in practice, we revisited the sample of Janssen et al. (2020) who classified 103 real-world chatbots within a taxonomy development process according to Nickerson et al. (2013). The sample contains 103 chatbots from the widely used application domains customer service, daily life, e-commerce, finance and work & career (Janssen et al. 2020, Appendix pp. 8-10). In May 2019, with the aim of classifying a set of chatbots along several application domains, the authors first classified a set of 12 most popular chatbots on the database "botlist.co", before selecting 10% of the chatbots within each of 28 application areas on the chatbot database "chatbots.org" (Janssen et al. 2020). The sample includes chatbots from 35 different countries within six continents, like the USA, Belarus, Argentina, Nigeria, India and United Arab Emirates. We decided to reapply this dataset, because it enables us to identify chatbots from a broad spectrum of application domains that were taken offline within a relatively small time-horizon of 15 months, as well as allows us to utilize the design elements classification of the taxonomy to take first interpretations. Therefore, in September and October 2020, two of the authors revisited both all 103 chatbots to start a conversation and engage with them. Using the classification results, we calculated inter-coder reliability to measure the quality of agreement. This was done by applying Cohens' (1968) weighted kappa coefficient resulting in 0.882 which means almost perfect agreement for this value (Landis and Koch 1977). This leads us to the assumption that a bias caused by two coders can be precluded. In the context of an external user perspective, chatbots failed in our analysis if they were no longer available or did not react appropriately anymore in a human-to-chatbot dialogue. We focused on first insights engaging with chatbots instead of testing all primary tasks of a chatbot. If we could not find a chatbot via the original URL we used Google Search to search for the chatbot and company name.

15 months after the initial screening by Janssen et al. (2020), 53 out of 103 chatbots proved to be failed. Some chatbots were undetectable on the websites (e.g., Annemiek), other chatbots were converted to live chats with human agents (e.g., Amanda), some websites were taken offline (e.g., Soa Seeks Check) or a chatbot did not respond anymore within the conversation (e.g., Jaquelina). Considering the application domains, it is noticeable that all e-learning chatbots (n=4) are no longer operational, as well as 69% of the finance chatbots (n=13) and 57% of the work & career chatbots (n=7). If we take a look on failure rate at the daily life (48%, n=48), e-commerce (44%, n=9) and customer service area (41%, n=22), a slight majority of chatbots is still running.

To get a first understanding of the reasons why chatbots failed, we used the chatbot design-elements taxonomy developed by Janssen et al. (2020) and their classification of chatbots to compare the two groups of operating (n=50) and discontinued (n=53) chatbots. The two groups are very similar in terms of e.g., collaboration orientation, duration of relation, or socio-emotional behavior. The greatest difference can be discovered in the dimension front-end user interface channel. While the group of failed chatbots consists of 40% social media chatbots and 28% website chatbots, the group of still operating chatbots contains 50% website chatbots and 28% social media chatbots. In addition, a distinction can be seen in the system architecture dimension, while the group of failed chatbots contains 72% reactive chatbots but also 28% proactive ones, the group of chatbots that are still operating contains only 14% proactive chatbots. Among the customer service chatbots, it is noticeable that the group of failed chatbots consists of 89% chatbots that did not offer additional human support, while 87% of chatbots are still operating that offer users the option to get in touch with a human. Differences also exist regarding the chatbot role within the dialogue, while 85% of the facilitator customer service chatbots still exist, 53% of the customer service chatbots that claimed to be experts failed.

Step 2.1: Solution Identification through Critical Success Factors

Considering that 53% of chatbots being no longer accessible after 15 months, we deduce that there is a high chatbot discontinuation rate. Based on the analysis of the distinctive characteristics of design elements, first clues for reasons, such as absence of additional human support, can be guessed why chatbots failed. The causality between the characteristics and the unavailability of a chatbot cannot be shown only analyzing the design-elements of chatbots externally. Our research goal is to explore the reasons why chatbots fail from a chatbot providers' and a scientific perspective. To manage chatbot failure risks, we identify chatbot CSFs on a global level that should be considered both from a scientific and a practical perspective.

CSFs can be seen as one of the oldest and widely researched subjects in IS research (Lee and Ahn 2008; Hawking and Sellitto 2010). We refer to the widely known and used CSF definition of Rockard (1979) and Bullen and Rockard (1981). According to these authors, CSFs are defined as "[…] the limited number of areas in which satisfactory results will ensure successful competitive performance for the individual, department or organization. CSFs are the few key areas where "things must go right" for the business to flourish and for the manager's goals to be attained" (Rockard 1979 p. 84-85; Bullen and Rockard 1981 p. 7). CSFs have been widely applied in various IS research fields, such as business process management (e.g., Trkman 2010), information technology (IT) projects (e.g., Iriate and Bayona 2020), agile analytic projects (e.g., Tsoy and Staples 2020) and software development (e.g., Ahimbisibwe et al. 2015). Trckman (2010) developed a theoretical framework based on theories of contingency, dynamic capabilities, and task technology fit to appropriately combine business environment and business processes. As part of a case study, 12 CSFs were identified, e.g., strategic alignment, and employee training and empowerment (Trckman 2010). Based on a literature review, Iriate and Bayona (2020) compiled and summarized the most cited and overlapping IT project CSFs, such as system quality, project management, and time. For agile analytic projects, Tsoy and Staples (2020) updated 25 attributes of potential CSFs previously identified by Chow and Cao (2008). Chow and Cao (2008) identified 12 CSFs with 25 associated attributes, e.g., ability to work in a team or a high customer involvement. From related agile project literature, Tsoy and Staples (2020) identified 10 additional attributes, e.g., for the CSF team capability, the new attribute of a sufficient team diversity corresponding to a high task complexity. Ahimbisibwe et al. (2015) identified 28 frequently cited CSFs for software development and implied that the importance of CSFs vary for agile and traditional projects. CSFs such as technological uncertainty or specification changes are much more frequently cited in an agile context while CSFs such as project planning, vision, and mission are more of concern in traditional projects (Ahimbisibwe et al. 2015).

Chatbots are mostly complex internet-based software systems (Brandtzaeg and Følstad 2018) and accordingly fall within the scope of IT projects (Karlsen et al. 2005). However, chatbots differ from other IS technologies because of their interaction and intelligence capabilities (Maedche et al. 2019). For example, the natural language interface of chatbots differs from other types of user interfaces. The design of chatbot interfaces, unlike graphical user interfaces for example, focuses not only on visual elements but also on communicative behaviors (Araujo 2018). Consequently, we decided not to compile existing CSFs from higher-level domains such as software development, but to identify specific CSFs for chatbots from scratch based on chatbot-related literature and expert interviews.

Before identifying CSFs, it has to be defined what is understood under the term success in the chatbot context. A human-chatbot interaction can be seen as successful if the intended task is completed appropriately by the chatbot (Seeger and Heinzl 2021). Feine et al. (2019) emphasize that a chatbot is from the user perspective successful when it is able to efficiently and satisfactorily perform longer conversations with a user while the user experiences the consumption-related fulfilment as enjoyable. We define success of chatbots from the organizational and main responsible managers' point of view by the fact that a chatbot is functioning and available, performs the tasks for which it is designed, and satisfies the users of a chatbot, whereas CSFs are the few essential requirements that must run perfectly to be successful with the chatbot (Rockard 1979; Bullen and Rockard 1981). Conversely, in our context, this means that a chatbot will fail if these key factors are not properly fulfilled. Consequently, we moved away from the previous definition that a chatbot fails when it is not responding appropriately or becomes unavailable within a human-to-chatbot dialogue to a broader definition with differentiated perceptions by not considering solely an external user perspective. In the following, using a concept-centric literature analysis and conducting expert interviews, we will first explore reasons for failure before identifying chatbot CSFs according to our definition.

Step 2.2: Gathering Knowledge from Scientific Literature

To derive requirements from theory, we conducted a systematic literature review following Webster and Watson (2002), Watson and Webster (2020) and vom Brocke et al. (2015). The review scope was structured in line with our research questions and focused on research and theories on failure and success of chatbots. We centered on the identification of CSFs and necessary requirements for chatbots. Reasons for chatbots' failure and CSFs for chatbots are consequently interconnected and considered together in our literature review, leading to the search string: ("chatbot" OR "chat bot" OR "conversational agent") AND ("unsuccess" OR "fail" OR "failure" OR "success" OR "success factor" OR "critical success factor"). From October to December 2020, five databases were searched to identify publications from established journals and conferences in IS, HCI and other relevant fields. After three screening phases in which we examined the *(total hits/relevant by title/relevant by abstract)* (i.e., ACM Digital Library (601/27/13), AIS Electronic Library (299/103/54), ScienceDirect (664/36/15), SpringerLink (2522/46/30), and Google Scholar (12000/68/38)), and deleted duplications, we gathered 126 papers. Using forward search (10), backward search (12), and similarity search (6), we completed our list, resulting in a total of 154 relevant papers.

Starting from this point, we explored reasons for chatbots' failure in practice. We summarized the various perspectives of failure, the current limitations, and shortcomings of chatbots, the various lessons learned, and important elements and guidelines for chatbot success into a comprehensive set of very basic CSFs for chatbots at the industry level. A concept matrix was built iteratively analyzing the 154 identified papers for general content and factors relevant to success or failure for chatbots. After approximately 100 papers, no new categories were discovered and existing ones were only slightly expanded, so we concluded our review as exhaustive. The papers were classified in 32 categories. Similar categories were then clustered, resulting in 10 potential CSFs for chatbots. The CSFs and associated categories from theory are shown in Table 2.

Step 2.3: Gathering Knowledge from Experts

To explore the underlying reasons for chatbot discontinuation within expert interviews, we first contacted the 53 companies who initially deployed the not operational or not accessible chatbots in our sample in November 2020. Unfortunately, it resulted in 53 negative or unanswered messages. We changed our approach and contacted individual chatbot experts directly. To identify those experts, renowned chatbot conferences, e.g., Chatbot Summit, were screened for keynote speakers. In addition, the LinkedIn network was used to search for experts based on their job descriptions. A total of 60 individuals were contacted via LinkedIn, which led to 20 semi-structured interviews with experts from five different countries (USA, Germany, the Netherlands, Switzerland and Israel). The chatbot experts come from different work domains, such as research, systems engineering, consulting, process ownership, or business leadership (Table 1). The organizations and companies affiliated with the experts are of different sizes, such as universities, startups, medium-sized companies, and large international corporations, which can be, e.g., determined by the number of employees. This focus allowed us to include a broad range of different company types. The organizations and companies analyze, develop, and distribute chatbots, provide chatbot-related services and infrastructure, and/or deploy chatbots.

To guide the interview process to some extent, we followed the recommendation of Myers and Newman (2007) and designed an interview guide. This guide was in advance also sent to the experts so that they could prepare for the interview. The content of the interview guide was based on the recommendations on the interview process by Bullen and Rockard (1981) and Caralli et al. (2004). The interview guide consisted of seven sections. First, the topic of chatbots' failure and the relevance CSFs was briefly introduced, followed by a description of the interviewee's tasks and goals related to chatbots. Then, an introduction to the method of identifying CSFs was given and the need for research was outlined. In the fourth step, the experts' personal experiences regarding the failure of chatbots was discussed. Starting on this, CSFs of chatbots were elaborated. These factors were then prioritized in the sixth step. Methods for measuring the CSFs were discussed in the final step. The 20 expert interviews took place in January and February 2021 and lasted between 23 minutes and 52 minutes, with an average of 35 minutes. All interviews were conducted via video chat or telephone and followed the interview process outlined in the previously described interview guideline. After completing the expert interviews, the interviews were denaturalized transcribed (Oliver et al. 2005) and coded within MAXQDA. The code system was based on the 32 categories previously identified in our literature review and the assigned potential CSFs and consisted therefore of two levels. The use of open coding (Wiesche et al. 2017) enabled us to iteratively expand and modify the initial list of coding tags

according to our findings, resulting in 40 categories clustered into 12 CSFs (Table 2). In addition, the personal experiences of the experts regarding the failure of chatbots were highlighted by a coding tag.

	ID	Job description	Company description	Number of employees
Interviews	Exp01	PhD candidate and researcher	University	-
	Exp02	PhD candidate and researcher	University	-
	Exp03	CEO and CPO	Consulting and chatbot development	1 - 10
	Exp04	System engineer	Consulting and chatbot development	1 - 10
	Exp05	Conversational interface designer	Consulting and chatbot development	11 - 50
	Exp06	CEO and company founder	Consulting, training, and certification	11 - 50
	Exp07	CEO and technical leader	Mortgage financing	11 - 50
	Exp08	CEO and CMO	Software	51 - 200
	Exp09	CEO and leader of development	Telecommunication	51 - 200
	Exp10	Chatbot product owner	Tourism	1000+
	Exp11	Business process owner chat	Tourism	1000+
	Exp12	Global head of AI	Software	1000+
	Exp13	Global product owner for chatbot and chat	Tourism	10000+
	Exp14	Chatbot product owner	Telecommunication	10000+
	Exp15	Head of e-commerce and sales innovation	Public transport	10000+
	Exp16	Sales engineer	Telecommunication	10000+
	Exp17	Product owner conversational AI	Banking	10000+
	Exp18	Leader digital assistance program	Chemical and pharmaceuticals	50000+
	Exp19	Strategy and innovation architect for AI	Hardware and software	100000+
	Exp20	Manager digital transformation and change	Automotive supplier	100000+
FGD	Exp21	PhD candidate and researcher	University	-
	Exp22	PhD candidate and researcher	University	-
	Exp23	PhD candidate and researcher	University	-
	Exp24	Post Doc and researcher	University	-
	Exp25	Chatbot, and AI consultant EMEA	Consulting and chatbot development	5000+

Table 1. Expert Profiles

Step 2.4: Extracting Failure Reasons in Practice & Developing an Initial Set of CSFs

We first extracted reasons for chatbots' failure in practice based on literature and expert interviews to answer RQ1. This led us to a set of six reasons, all of which have already been experienced by at least three experts. RQ2 was related to the CSFs of chatbots. The CSFs found in literature review in Step 2.2 and by expert interview conduction in Step 2.3 were synthesized into an initial set of CSFs. All ten potential CSFs previously identified in our literature review were confirmed by the 20 experts. Two additional CSFs were found, namely a chatbot development team and the factor developmental strategy. In addition, the categories of CSFs were adjusted by identifying new ones and expanding or deleting previous ones. The categories of the following CSFs, top management support, project resources, chatbot progress, chatbot design, user-centric use case, and technology and tool availability, were modified. In selecting the names to identify each cluster, i.e., the CSFs, an attempt was made to make the name descriptive enough for the reader to recognize the reference. The names chosen are more abstract than the concepts they represent. In some cases, the chosen category name was selected from the pool of concepts. In other cases, the chosen name was borrowed from terminology commonly used in literature (e.g., "Trust") (Følstad et al. 2018).

Step 3: Evaluation and Adjustment

To determine whether our artifact, i.e., the initial set of CSFs (Table 2), is comprehensible, understandable, and useful (Gregor and Hevner, 2013), we conducted an evaluation considering "what", "who" and "how" (Pries-Heje et al. 2008) which is a crucial step within DSR (vom Brocke et al. 2020b). Regarding "what", the object of evaluation, we decided to evaluate the design process as well as our design product (CSFs). Regarding the evaluation subject ("who"), we selected five individuals from three countries (Switzerland, Germany, Luxembourg) who were not previously involved in the development of the CSFs (Table 1). To test a broad applicability and understandability, we brought together people from research (Exp21, Exp22, Exp23, Exp24) and practice (Exp24, Exp25) with chatbot and/or CSF methodology experience. Concerning the "how", we decided to conduct a FGD to get as many different perspectives on the CSFs within a joint discussion as possible. In April 2021, the virtual FGD took place and lasted 92 minutes. In preparation, participants received a worksheet with our questions, an explanation of our research process and the CSFs

table with definitions. After all participants introduced themselves, the first two authors explained the problem statement, gave an overview of DSR, results from our literature review and the interviews on chatbots' failure, before presenting the CSFs. Questions of process understanding were clarified and the definition of chatbot failure and the analysis of 103 chatbots were discussed. Overall, our DSR to identify reasons for chatbots' failure and CSFs appeared to be comprehensible and coherent. The reasons identified in practice were confirmed by the participants. Exp25 emphasized that for already developed chatbots, it is often necessary to tease out the real reasons for failure, as companies do not like to talk about what they have done wrong. Exp25 noted that the reasons for chatbot failure identified within this study are very well differentiated from each other, which the expert notes is very helpful, and reports that in reality it is usually a combination of several reasons. Exp24 remarked with regard to the interviewed experts that there may be significant geographical and cultural differences, e.g., due to legal regulations, and brings up an example from Belarus, where it is common to change the cell phone tariff with two messages to a chatbot.

The CSFs were perceived by participants as comprehensible and extensive. Exp 25: "I would like to see companies sometimes simply take three steps at the beginning and think about: what I actually want to do in the project and what do I have to consider? And I could imagine the CSFs list being very helpful for this in practice, especially for companies that don't yet have so much experience." But it was noted that the CSFs will continue to evolve. Exp23 claimed: "If the same study were conducted 10 years from now, I guarantee we would see different success factors." Exp22 described that while the CSFs are understandable, it would be helpful to have a brief description text for each of the twelve CSFs. We used these results of the evaluation to improve the CSFs. While most categories were perceived as understandable and clearly delineated, we renamed the category "adjustability and extensibility" to "simple editing and extensibility of design elements". We also merged the categories "access to database" and "connection to backend systems" which was addressed by Exp21.

Results and Findings

Chatbots' Failures Reasons in Practice

To answer RQ1, we present reasons for failure in practice extracted from scientific literature and expert interviews. Our findings in Step 2.2 indicate that scientific literature provides many single-case field reports from practitioners about challenges in creating chatbots, elements relevant to failure or success based on user responses, and guidelines for improving these chatbot elements. But to date, there has been little qualitative research examining the exact reasons why organizations take their chatbots offline permanently or stop maintaining them. In a brief attempt, Brandtzaeg and Følstad (2018) cite two real-world examples of failure. Ikea's chatbot Anna failed because it could not balance robotic and human aspects, resulting in customers abusing the chatbot (Brandtzaeg and Følstad 2018). Microsoft's chatbot Tay, in turn, gave unethical responses shortly after its release and was taken offline as a result (Brandtzaeg and Følstad 2018; Zemčík 2020). Chatbots often fail in part because they do not meet user expectations and companies tend to focus on business-centric use cases rather than user-centric use cases (Zamora 2017; Grudin and Jacques 2019). Based on user interviews and questionnaires, user trust and privacy concerns as well as perceived ease of use and usefulness have also been identified as reasons that can lead to chatbots' failures (Rodríguez Cardona et al. 2021; Følstad et al. 2018; Mozafari et al. 2021).

Based on the experts' personal experiences narrated within the interviews, we present six generic reasons for the failure of chatbots in practice, identified in Step 2.3. Each of these reasons is based on the experts' experience with an actual failed chatbot. Hypothetical reasons for chatbots' failures were not incorporated.

Not enough resources: Six experts have come across the lack of resources as a reason for the failure of a chatbot. A person primarily responsible for a chatbot leaves the company, not enough money is made available for a sufficient technical infrastructure, or a third party providing relevant services spontaneously breaks away. One interview partner expressed firmly: "Okay, look, that doesn't make sense like that. We're going to redo the planning and we're going to put everything back to square one, because the effort […] is significantly greater than what we want to achieve, and our core competence is simply elsewhere […]. One aspect, why chatbot projects fail is that the effort behind it is underestimated and […] too few resources are provided" (Exp16).

No business case: The second reason for the failure of chatbots in practice encountered by six experts is the missing of a business case. The potential value creation for the company is significantly less than the effort involved. A chatbot project is set up without properly evaluating the benefits or cost. One expert described the situation as followed: "So SMS based chatbots can be pretty costly as far as your return on investment. So when you're sending out several text messages a day to people [...] you might be spending [...] a couple of dollars a day, whereas you might not be getting any return, which is one of the experiences I had developing a covid-19 chatbot, so that one, we had to take offline after a few months" (Exp01).

Wrong use case: Five experts witnessed how a chatbot failed because of a wrong use case. A chatbot was deployed for a task where the basic chatbot technology did not match the required task. One interviewee described: "Basically, the idea was that you could get a complete construction financing via a chatbot [...]. But if the process doesn't fit what you want to map with a chatbot [...] to process a complete construction financing via the chatbot, then we fail. That simply doesn't fit together. It is simply very difficult to map a deeper complexity only via a chatbot" (Exp07).

Law regulations, data security, and liability concerns: The fourth identified reason for the actual failure of chatbots in practice, mentioned by five experts, is legal regulations and privacy and liability concerns of the organizations using a chatbot. One participant described the problematic situation encountered as below: "We actually developed a chatbot for a major bank at the end of last year [...] it was about reporting, how to report business numbers [...] and it actually failed because the EU has terminated the Privacy Shield agreement with the USA and as a result there is now legal uncertainty, in a sense" (Exp05).

Ignorance of user expectation and bad conversation design: Four experts stated that they have witnessed chatbots fail due to ignored user requirements and poor conversation design. An interviewee described how chatbots have failed because of a disregard for good conversational design: "So usually, companies reach out to us when their chatbots are not performing the way they ought to do [...]. They focus too much on the technology [...] on knowledge management, or [...] on the business process. So what you see a lot of times is that they [...] sort of create a flow chart of what the business process looks like. And then they pretty much add some words to it [...] that doesn't work, because people don't talk the way your business is organized" (Exp06).

Poor content: Last identified reason for chatbots' failures in practice is poor content. Three participants experienced that a chatbot failed because the content requested by users was wrong, not complete, or not up-to-date. Exp10 said that the lack of relevant data, regarding to current conversational topics of users, leads to failure. While another expert stated his experience regarding incomplete data as followed: "So the chatbot itself, from the technology, it's possible that you tap into external sources [...]. You need a lot of sources [...]. So, the questions that came in were of course all over the place. One wanted to know whether Corona could be dangerous during pregnancy. The other wanted to know something, whether one can still play soccer [...] that has blown up the information pool of the Robert-Koch-Institute" (Exp07).

Final Critical Success Factors for Chatbots

To answer RQ2, our final cross-domain set consists of 12 CSFs for chatbots (Table 2). 10 CSFs were identified based on an extensive literature review of 154 papers. These CSFs were confirmed by the 20 expert interviews, and an additional two CSFs were found. Subsequently, the set was evaluated and adjusted by five more experts. In addition, the 12 CSFs were divided into three variables describing the interrelation of the CSFs.

Technology and tool availability: This CSF addresses the limitations of available underlying technology and infrastructure for chatbot applications which was addressed by 73 scientific papers. Five experts mentioned that the availability of basic chatbot technology and related tools is important. Exp01 asserts that the increase in the availability of software tools for conversational design, conversational AI, tools for testing chatbot frameworks, and project management tools in recent years have led to chatbot projects being more successful. The development process of chatbot technology started back in the last century, but it is still ongoing (Exp19). Chatbots have evolved from single-line, text-based chat systems that initially supported human-to-human conversations to modern, complex, knowledge-based models that began to emulate dialogue systems using natural language understanding and dialogue management (Galitsky 2019; Kepuska and Bohouta 2018). Exp06 noted that chatbot technology still has many limitations, and Exp18 described it as a "pretty immature technology" and that "things are always kind of a bit buggy".

CSF	Associated category from coding	P*	Example authors	E*
Technology availability	Technology & tool availability	70	Galitsky 2019; Schumaker et al. 2007	5
User centric use cases	Adequate use case	26	Rodríguez Cardona et al. 2019; Zamora 2017	16
	User requirements	17	Følstad & Brandtzaeg 2020; Meyer von Wolff et al. 2020	13
	Acceptance to change operation methods	-	-	1
	General chatbot technology acceptance	17	Weber & Ludwig 2020; Mesbah & Pumplun 2020	6
Chatbot promotion	Communicating the intention to introduce/use a chatbot	1	Aoki 2020	8
Chatbot design	Data security	26	Lai et al. 2018; Følstad et al. 2018	7
	Technical design elements	68	Janssen et al. 2020; Yuan et al. 2019	15
	Conversational design elements	73	Kvale et al. 2019; Gnewuch et al. 2020	13
	Design elements' simple editing and extensibility	1	Koetter et al. 2019	4
	Databases & backend systems accessibility	27	Kruse et al. 2019; Johannsen et al. 2020	8
	Word sensitivity	8	Yu et al. 2016; Canhoto & Clear 2020	2
	Level of intent & content understanding	37	Følstad & Brandtzaeg 2020; AbuShawar & Atwell 2016	7
	Technical robustness & chatbot efficiency	2	Nguyen & Sidorova 2017; Weber & Ludwig 2020	3
Chatbot progress	Testing & training	17	Johannsen et al. 2018; Vijayaraghavan & Cooper 2020	8
	Continuous monitoring, updating and improvement	27	Jonke & Volkwein 2018; Brandtzaeg & Følstad 2018	13
	Chatbot self-development	15	Zemčík 2020; Hancock et al. 2019	-
Top management support	Changing company structure and workflows	-	-	2
	Manage top management expectations in short & long term	-	-	12
	Top management support	7	Benbya et al. 2020; Pumplun et al. 2019	4
Project resources	Transparent cost management	-	-	2
	External resources	-	-	5
	Human resources	10	Galitsky 2019; Kruse et al. 2019	8
	Technical resources	16	Desouza et al. 2020; Winkler & Roos 2019	10
Developmental strategy	Highly dynamic long-term process (instead of a classic project)	-	-	7
	Multidisciplinary process (not a pure IT and engineering based and driven)	-	-	2
	Start small, go big (quick wins)	-	-	3
Chatbot developing team	Team composition	-	-	6
	Team building	-	-	2
	Clear definition of used success and performance metrics to evaluate chatbot	-	-	2
	Content management core team	-	-	2
Usefulness	User expectation	67	Følstad & Brandtzaeg 2020; Weber & Ludwig 2020	5
	User understanding of chatbot capabilities	17	Følstad et al. 2018; Aoki 2020	4
	Perceived usefulness (based, e.g., on TAM)	47	Wuenderlich & Paluch 2017; Følstad & Skjuve 2019	13
Usability	Unexperienced user guidance	3	Weber & Ludwig 2020; Piccolo et al. 2018	5
	Seamless chatbot integration in customer journeys	1	Kuligowska 2015	2
	Ease of use (based, e.g., on TAM)	36	Rodríguez Cardona et al. 2021; Rese et al. 2020	4
Trust	Trust in chatbot and operating company	8	Følstad et al. 2018; Sanny et al. 2020	2
	Trust in chatbot technology	49	Fiore et al. 2019; Nordheim et al. 2019	2
	Privacy concerns	30	Rodríguez Cardona et al. 2021; Kim et al. 2020	5

V* = Variable classification (Exo* = Exogenous, Endo* = Endogenous, Mod* = Moderator)
P* = Number of papers, E* = Number of experts

Table 2. Critical Success Factors for Chatbots

User-centric use cases: A suitable use case for chatbot deployment is a valid business case with reasonable scope that adds value to both the business (Exp18) and the customer (Exp14, Exp02). Similarly, Rodríguez Cardona et al. (2021) and Zamora (2017) recommend a user-centric approach, where different requirements of potential users should be considered throughout the development phase to ensure that value-added chatbot elements are prioritized. Exp03, Exp07, and Exp12 note that it must also be a logical use case suitable for the use of chatbot technology. In addition, legal frameworks such as privacy regulations and ethical considerations that distinguish between acceptable and unacceptable practices must also be

considered (Rodríguez Cardona et al. 2019). The general acceptance of potential users to use a chatbot represents another characteristic. Two interviewees elaborated that some people do not want to communicate with a chatbot or are afraid that this technology leads to the loss of their job and therefore have a negative stance on it (Exp14, Exp20).

Chatbot promotion: Establishing a clear communication and integration of a chatbot is important (Exp08, Exp15). While eight experts emphasized the relevance of chatbot promotion, this was mentioned only once in literature (Aoki 2020). Users need to be aware of and become familiar with the chatbot (Exp05, Exp17). Similar, Aoki (2020) highlights about the use of chatbots in public institutions, that communicating the intention to utilize a chatbot is an inexpensive step to raise user's awareness and trust regarding the bot. Exp20 firmly stated that "of course you have to actively promote it".

Chatbot design: This CSF describes all the elements and capabilities of a chatbot that need to be considered during its development. Technical design elements like the intelligence quotient or an avatar (Janssen et al. 2020) have to be considered next to conversation relevant elements, e.g., a structured conversation flow, which must be as natural as possible regarding dialogue flow and formulations (Kvale et al. 2019; Koetter et al. 2019). The development of a knowledge base, access to relevant databases and the connection of different backend systems increases the capabilities of a chatbot and ensures a higher usefulness to the user (Exp09, Exp14, Exp15, Exp20). Exp05 and Exp20 said that the option to escalate to a human if the conversation stagnates must be considered to handle damage control and prevent the conversation from getting stuck (Weber and Ludwig 2020). Data protection measurements must be clarified in advance (Exp11, Exp14) and sufficient IT security must be guaranteed (Exp16, Exp20). Exp13 described the consequences of disregarding this aspect as followed "If something goes wrong with it, you don't want to be in charge of the project". In addition, Exp12 and Exp19 note that the overall framework of a chatbot must be easily adjustable without a lot of coding to implement dynamic requirements.

Chatbot progress: Chatbot progress describes the initial training of a chatbot, testing functionality and content, as well as refining design elements and updating the knowledge base during live deployment. To prevent a chatbot from failing during its use, it must be tested and validated (Vijayaraghavan and Cooper 2020; Ruane et al. 2018). Exp18 and Exp19 recommend iteratively testing the content and language model of a chatbot changing test groups to test different content and different ways of asking questions. Many experts emphasize that a chatbot must be continuously monitored and improved based on the data collected and user feedback. Exp12 and Exp18 point out that it is generally impossible to predict how users will interact with a chatbot and what information will be requested long-term, so a good monitoring and updating process is needed. "There's a shift in terms of what people are asking over time. And then you need to be on top of that case" (Exp12). These statements confirm the recommendations of Brandtzaeg and Følstad (2018), and Jonke and Volkwein (2018) to maintain and enhance the content of a chatbot. The category of self-development of a chatbot identified in our literature review (e.g., Zemčík 2020) was not confirmed by the experts and was dropped. The unsupervised evolvement of chatbot systems is not desirable and, in the worst case, leads not only to linguistic hostility but also to psychological or physical damage, e.g., through the incorrect instruction of treatment methods in areas such as healthcare (Exp01, Exp02).

Top management support: This CSF describes the needed support of projects and changes of business structures within an organization by the organization's upper management. Exp05 mentions that a chatbot needs a defined status in the company's communication strategy. Exp11 and Exp13 further add that a chatbot needs to be integrated into the corporate structure, which requires an adaptation of the workflows. These changes have to be backed up by the top management to deal with conflicts of interest (Pumplun et al. 2019). This also includes a shift in authority. Different project members and their tasks must be equally valued and given voice by the top management (Exp06). New job roles that challenge existing structures will emerge and chatbot developing teams will gain more influence on internal decision-making processes (Benbya et al. 2020). In addition to the findings from our literature review, the experts underline that top management and the chatbot project team must be in frequent contact with each other (Exp08, Exp10, Exp17). Varying short- and long-term expectations as well as different understanding of success criteria for chatbots must be synchronized and the added value offered by a chatbot clearly communicated (Exp12).

Project resources: This CSF specifies the resources and skills required to develop, implement, and maintain a chatbot, forcing companies to adjust their technical, human, and financial resources (Kruse et al. 2019). Exp11 and Exp14 mention that sufficient human resources on a full-time basis must be allocated. People

with different skill sets such as system engineers, AI trainers, conversational designers or copyrighters are required (Exp06). These developers require in-depth knowledge of a broad range, including areas such as AI and SQL databases (Galitsky 2019; Kruse et al. 2019). In cases where the core competence of one's own organization does not lie in the development of chatbots, it is recommended to outsource some of the tasks and work together with specialists (Exp15, Exp16). Exp15 and Exp18 specify that the partners must be economically stable and that they must also exist in a few years, as a chatbot usually is operated for many years. Since a chatbot requires many different resources over a long period of time and can lead to ambiguity and resentment among stakeholders, proper and comprehensive cost management must be upheld (Exp11, Exp18).

Developmental strategy: Seven experts mentioned that a basic understanding of the characteristics of a chatbot must be created in the organization while this CSF has not received any attention in scientific literature so far. Chatbots are not projects that are finished at a specific point, but an evolving process (Exp06, Exp13, Exp19). The process of a chatbot project does not correspond to a waterfall-like or agile method process with fixed target dates, where a chatbot can be launched as a finished product at the end (Exp19). A chatbot is supported by a "highly agile project" and is developed in a longer term with many individual tests and constantly changing test groups (Exp18) and should rather be treated "like an entire eco system" (Exp06). Exp01 and Exp06 further add that this is not a simple engineering project, but a balancing act between engineers, conversational designers, copyrights, and other professionals involved. Initiating a chatbot with a small scope helps to get qualitative user feedback as quickly as possible allowing to easier developing and scaling a chatbot (Exp08, Exp13, Exp19).

Chatbot developing team: The second CSF solely identified by the expert interviews consists of four categories. The responsible development team must be composed of different work domains with different responsibilities, since varying expertise is needed from areas such as systems engineering, user experience design, psychology, or copyrighting (Exp01, Exp06, Exp19). To ensure a seamless development process, these experts with completely different skill sets need to be synthesized and get to work as a cohesive team (Exp02, Exp06). Exp06 and Exp14 point out that due to the differences in skills and perspectives, it is important to have common clarity on the definitions of the evaluation metrics used. In the long term, there must be a core team responsible exclusively for content management (Exp11, Exp18).

Usefulness: This CSF comprises categories dealing with the ability of a chatbot to efficiently perform the tasks desired by the user, considering the user's preferred workflow. Exp08 summarized this factor as follows, "The most important thing is of course the subject, does the chatbot satisfy the customer's need in the end?!". A chatbot must also be able to provide almost completely coverage of what the user expects (Exp12). Similarly, Weber and Ludwig (2020) and Wuenderlich and Paluch (2017) outline in their studies that a chatbot must be able to perform the tasks expected of it faster and more accurately than the user himself or another human counterpart, e.g., call centers, could. Exp07, Exp11 and Exp12 further asserted that users are required to understand how a chatbot works.

Usability: Usability refers to the current use of a chatbot and is composed of the three identified categories usability, user guidance, and seamless integration of a chatbot into the customer journey. Users ought to perceive the chatbot as an easy-to-use and smoothly functioning system (Rodríguez Cardona et al. 2021; Rese et al. 2020). Exp12 describes this as, "Obviously, the interface needs to be intuitive". Exp05 said that the structure of the dialogue flow and its comprehensibility are important, too. A chatbot must be located at the point in the customer journey where the customer needs or expects it (Exp12). In line with Weber and Ludwig (2020) and Piccolo et al. (2018), Exp09 and Exp11 commented that inexperienced users must be initially guided to avoid getting lost in the interaction process, e.g., by a short introduction.

Trust: End users trust towards general chatbot technology regarding reliability and goodwill of the chatbot itself must be given (Fiore et al. 2019; Kim et al. 2020). Exp02 states that the confidence of the user in a carefree use of a chatbot, i.e., trusting that a chatbot technology will work as intended, that it will not harm the user, is critical for success, especially in application areas such as healthcare. Accordingly, a good reputation of the brand or the associated organization is important, as trust in a chatbot depends on the users' previously established trust in the service provider (Følstad et al. 2018; Sanny et al. 2020). Exp02 commented it as follows, "There has to be a basic trust in a chatbot and its background". Users trust in guaranteed data and privacy protection must be assured, since low perceived risks, e.g., not having to provide sensitive information, increases trust (Nordheim et al. 2019; Rodríguez Cardona et al. 2021). Exp20 described this as users want to be sure that their data will not be misused.

Discussion, Implications, Limitations, and Outlook

To investigate chatbot failure and identify CSFs for chatbots, we analyzed 103 real-world chatbots, performed an extensive literature review, and conducted 20 expert interviews within our DSR based process. The analysis of a chatbot user was indirectly possible through the analysis of user-centered HCI and IS literature (e.g., chatbot technology acceptance (e.g., Mesbah and Pumplun 2020) or user expectation (e.g., van der Goot et al. 2021)) which allowed us to get an all-encompassing overview of chatbot CSFs.

We have examined chatbot failure from two different angles, from the chatbot user perspective and from the chatbot provider perspective, which required us to focus on different chatbot definitions. In analyzing 103 real-world chatbots, we defined chatbots as having failed externally if they no longer answered us, could no longer be found, or were transformed into a live chat. Since we unsuccessfully contacted all 53 discontinued chatbot providers, we cannot make definitive statements whether some of these chatbots may have been successful and were taken offline for other reasons. But our analysis gives a general insight into the high discontinuation rate within the global chatbot market. Further, within the sample, it is difficult to assess the extent to which further chatbots failed from the company's point of view, e.g., because of too few hits, too few leads (Janssen et al. 2021), or from the user's point of view, e.g., due to not executing the desired request (e.g., successful request for travel reimbursement (Exp24)). Failure is a sensitive topic that people do not like to talk about, which was a barrier in the search for interview partners. This could also have been a reason why no company or chatbot developer of the 53 no longer functioning chatbots of the dataset of Janssen et al. (2020) wanted to talk to us about the reasons. But our twenty interviewees confirmed that failure is very relevant, even if it still for them receives very little attention in literature. Conducting expert interviews (Table 1), we address this research need describing six reasons why chatbots failed in practice from the chatbot provider side which is a novel perspective on failure compared to other articles focusing only on the chatbot user perspective (e.g., Filipczyk et al. 2016; Seeger and Heinzl 2021; Mozafari et al. 2021). As confirmed by the experts within our FGD, reasons for chatbots' failures and our CSFs can help practitioners to continuously manage risks. As discussed in the evaluation, multiple reasons of failure often coincide (Exp25). In further research, it is worth to investigate in a survey which combinations often occur and to what extent geographical, cultural and application domain-specific differences exist.

From our literature review and the expert interviews, we identified a total of 12 CSFs and 40 categories that are relevant to the success of a chatbot (Table 2). In terms of the number of papers addressing the items, it is noticeable that the focus of scientific research is in line with expert opinions on what is important for the success of chatbots in many areas. 73 papers focus on elements of conversational design, named by 13 experts. Similarly, 68 papers focus on technical design elements, which were identified as critical to success by 15 of the 20 respondents. In contrast to scientific research, there is a much higher focus on the promotion of a chatbot and the correct design of use cases in practice, e.g., CSF chatbot promotion and the associated category are mentioned by 8 experts but only by one paper. In addition, we were able to identify two CSFs in practice, respectively the chatbot development team and the development strategy, that were not mentioned in the literature, which could also reveal research opportunities.

The majority of identified CSFs from experts and scientific research in the chatbot area show parallels to other HCI and IS research areas, but some CSFs are chatbot specific. Exp24 and Exp25 highlight that CSFs such as top management support or project resources are known in other IS research areas. The CSF top management support has been identified in the context of IT projects (Ahimbisibwe et al. 2015) and business systems implementation (Shapouri and Najjar 2020). Similarly, the importance of continuous improvement of systems described by the CSF chatbot progress has been observed in the context of business process management. Continuous improvement is critical to ensure long-term benefits (Trckman 2010). Within a DevOps context (van Belzen et al. 2019), the CSFs technology and tool availability as well as the need for adequate resources, e.g., technical infrastructure, have also been identified. Taking a closer look at the CSFs mentioned exclusively in the interviews, e.g., team composition, it is noticeable that several categories relate to project management issues which were mentioned in other HCI and IS topics, such as for agile project management (Ahimbisibwe et al. 2015; Tsoy and Staples 2020) and ERP implementation (Ahmad and Cuenca 2013). However, categories of the CSF chatbot design, such as technical design elements, e.g., additional support for escalating to a human when the conversation stagnates (Janssen et al. 2020), and conversational design elements, e.g., additional anthropomorphism features (Gnewuch et al. 2020), also highlight differences from previous CSFs from other IS and HCI research areas. It becomes

apparent that in practice not only chatbot specific aspects, such as the design elements are important for success but also project related aspects, such as top management support and the development team. Developing and deploying chatbots is a highly complex interplay of factors, where a broader view is necessary. We see a great added value in the all-encompassing consideration of CSFs in chatbot deployment since all factors that one has to consider for successful chatbots are listed. In recent years, the discussion of how much chatbots differ from other technologies has arisen (e.g., Zierau et al. 2020). Here, our list of CSFs can be a starting point for comparing chatbots with other technologies based on CSFs. But, even if the individual aspects may not differ, the interrelation and interaction of CSFs is crucial for chatbots' success which should be further examined in future studies. Therefore, e.g., Hilberts' (2005) framework could help to structure the identified CSFs into endogenous, exogenous and moderator variables and to identify interrelations between CSFs.

Our focus was to get a chatbot application domain superior overview of failure reasons and CSFs, which is why we emphasized on broad insights within the real-world chatbot analysis, interview partner selection, and within our literature analysis. In our chatbot analysis, we discovered that many customer service chatbots that did not offer additional human support no longer exist, which belongs to the technical design elements category mentioned in practice and literature. Whether these specific chatbots really failed because of this reason is speculation. But the current application areas, such as education (e.g., Winkler and Roos 2019), health (e.g., Kim et al. 2020) or customer service (e.g., Følstad et al. 2018) differ greatly from each other which is why the category importance for chatbot success may also vary across the application domains. It is worth to examine the application domains in isolation and break down from our CSFs list crucial application domain-specific CSFs. As diverse as chatbots are considering design elements (Janssen et al. 2020), chatbot implementation processes are equally diverse (Exp23, Exp25). We give an all-encompassing overview of the reasons for failure and chatbot CSFs, since in practice a chatbot is often developed and deployed in diverse countries (Exp17). Also, from the degree of how much of a chatbot is developed and maintained internally (e.g., by company employees) or externally (e.g., by consultancies and chatbot developing firms), in practice, all combinations exist which should be further examined in future.

While our literature review, chatbot sample analysis from 35 countries and search for interviewees were on a global scope, Exp24 in the evaluation mentioned that we may have a certain geographic and cultural bias as we interviewed experts from Europe, Middle East and USA. Since there are also geographically and culturally varying degrees of use and market penetration of chatbots (Diederich et al. 2021), it makes sense to also involve interview partners from other regions, such as Asia and Africa (Exp24, Exp25). Regarding the aforementioned real experienced reasons of failure, five experts mentioned law regulations and data security aspects that depend very much on the government. In this context, it is worth to examine the reasons for failure and the CSFs based on, e.g., geographical borders or across cultural dimensions. To identify these cultural factors within a survey, a large number of diverse participants is needed to have any significance. By following the four core DSR steps outlined by Doyle et al. (2019), based e.g., on Peffers et al. (2007), we focused on developing and evaluating CSFs. Future research could address how to put these CSFs into practice as part of a case study to avoid chatbot failure.

Conclusion

While we found that chatbots' failures have been addressed very little in scientific literature, through an analysis of 103 chatbots that we revisited after 15 months, we were able to identify that there is a high discontinuation rate in the chatbot field and captured experienced reasons of chatbots' failures through 20 qualitative expert interviews. To manage the chatbot failure risks, our research goal was to identify fundamental CSFs for chatbots based on our extensive literature review as well as 25 experts within interviews and a FGD for evaluation. For further researchers and practitioners being aware of these reasons of failure as well as CSFs of chatbots, we contribute to managing risks while deploying and maintaining chatbots. In addition, our CSFs analysis shows some deviations between chatbot-related topics addressed in practice and research which can be used to identify further research needs. The list of reasons for chatbots' failures and the CSFs represents the current state of research and technology. As the field is evolving rapidly in terms of e.g., AI, it is necessary to repeat this analysis and update the failure and CSFs listings.

References

AbuShawar, B., & Atwell, E. 2016. "Usefulness, Localizability, Humanness, and Language-benefit: Additional Evaluation Criteria for Natural Language Dialogue Systems," *International Journal of Speech Technology* (19:2), pp. 373-383.

Adam, M. T. P., Gregor, S., Hevner, A., & Morana, S. 2021. "Design Science Research Modes in Human-Computer Interaction Projects," *AIS Transactions on Human-Computer Interaction* (13:1), pp. 1-11.

Adamopoulou, E., & Moussiades, L. 2020. "An Overview of Chatbot Technology". In *Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations*.

Ahimbisibwe, A., Cavana, R. Y., & Daellenbach, U. 2015. "A Contingency Fit Model of Critical Success Factors for Software Development Projects: A Comparison of Agile and Traditional Plan-based Methodologies," *Journal of Enterprise Information Management* (28:1), pp. 7-33.

Ahmad, M. M., & Cuenca, R. P. 2013. "Critical Success Factors for ERP Implementation in SMEs," *Robotics and Computer-integrated Manufacturing* (29:3), pp. 104-111.

Aoki, N. 2020. "An Experimental Study of Public Trust in AI Chatbots in the Public Sector," *Government Information Quarterly* (37:4), 101490, pp. 1-10.

Araujo, T. 2018. "Living Up to the Chatbot Hype: The Influence of Anthropomorphic Design Cues and Communicative Agency Framing on Conversational Agent and Company Perceptions," *Computers in Human Behavior* (85), pp. 183-189.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. 2018. "Design Science Research Contributions: Finding a Balance Between Artifact and Theory," *Journal of the Association for Information Systems* (19:5), pp. 358-376.

Benbya, H., Davenport, T. H., & Pachidi, S. 2020. "Artificial Intelligence in Organizations: Current State and Future Opportunities," *MIS Quarterly Executive* (19:4), pp. ix-xxi.

Brandtzaeg, P. B., & Følstad, A. 2018. "Chatbots: Changing User Needs and Motivations," *Interactions* (25:5), pp. 38-43.

Bullen, C. V., & Rockart, J. F. 1981. "A Primer on Critical Success Factors," *Center for Information Systems Research Working Paper (69)*, pp. 1-64.

Canhoto, A. I., & Clear, F. 2020. "Artificial Intelligence and Machine Learning as Business Tools: A Framework for Diagnosing Value Destruction Potential," *Business Horizons* (63:2), pp. 183-193.

Caralli, R. A., Stevens, J. F., Willke, B. J., & Wilson, W. R. 2004. *The Critical Success Factor Method: Establishing a Foundation for Enterprise Security Management,* (No. CMU/SEI-2004-TR-010). Carnegie-Mellon Univ Pittsburgh PA Software Engineering Inst.

Chow, T., & Cao, D. B. 2008. "A Survey Study of Critical Success Factors in Agile Software Projects," *Journal of Systems and Software* (81:6), pp. 961-971.

Cohen, J. 1968. "Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit," *Psychological Bulletin* (70:4), pp. 213-220.

Desouza, K. C., Dawson, G. S., & Chenok, D. 2020. "Designing, Developing, and Deploying Artificial Intelligence Systems: Lessons from and for the Public Sector," *Business Horizons* (63:2), pp. 205-213.

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. 2021. "On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research", *Journal of the Association for Information Systems,* Online first.

Doyle, C., Luczak-Roesch, M., & Mittal, A. 2019. "We need the Open Artefact: Design Science as a Pathway to Open Science in Information Systems Research," In *Proceedings of the International Conference on Design Science Research in Information Systems and Technology*.

Feine, J., Morana, S., & Gnewuch, U. 2019. "Measuring Service Encounter Satisfaction with Customer Service Chatbots using Sentiment Analysis," In *Proceedings of the International Conference on Wirtschaftsinformatik*.

Fiore, D., Baldauf, M., & Thiel, C. 2019. "Forgot your Password Again?" Acceptance and User Experience of a Chatbot for In-company IT Support," In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*.

Filipczyk, B., Gołuchowski, J., Paliszkiewicz, J., & Janas, A. 2016. "Success and Failure in Improvement of Knowledge Delivery to Customers using Chatbot — Result of a Case Study in a Polish SME," In *Successes and Failures of Knowledge Management*, pp. 175-189.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. 2018. "What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study," In *Proceedings of the International Conference on Internet Science*.

Følstad, A., & Skjuve, M. 2019. "Chatbots for Customer Service: User Experience and Motivation," In *Proceedings of the International Conference on Conversational User Interfaces*.

Følstad, A., & Brandtzaeg, P. B. 2020 "Users' Experiences with Chatbots: Findings from a Questionnaire Study," *Quality and User Experience* (5:1), pp. 1-14.

Galitsky, B. 2019. "Chatbot Components and Architectures," In: Galitsky, B. (eds.). *Developing Enterprise Chatbots*. Springer, Cham, pp. 13-51.

Gnewuch, U., Yu, M., & Maedche, A. 2020. "The Effect of Perceived Similarity in Dominance on Customer Self-disclosure to Chatbots in Conversational Commerce," In *Proceedings of the European Conference on Information Systems*.

Gregor, S., & Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

Grudin, J., & Jacques, R. 2019. "Chatbots, Humbots, and the Quest for Artificial General Intelligence," In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

Hancock, B., Bordes, A., Mazare, P. E., & Weston, J. 2019. "Learning from Dialogue After Deployment: Feed Yourself, Chatbot!," *arXiv:1901.05415*.

Hawking, P., & Sellitto, C. 2010. "Business Intelligence (BI) Critical Success Factors," In *Proceedings of Australasian Conference on Information Systems*.

Hilbert, A. 2005. "Critical Success Factors for Data Mining Projects," In: Baier, D. et al. (eds.). *Data Analysis and Decision Support,* Springer, Berlin, Heidelberg. pp. 231-240.

Iriarte, C., & Bayona, S. 2020. "IT Projects Success Factors: A Literature Review," *International Journal of Information Systems and Project Management* (8:2), pp. 49-78.

Janssen, A., Passlick, J., Rodríguez Cardona, D., & Breitner, M. H. 2020. "Virtual Assistance in Any Context: A Taxonomy of Design Elements for Domain-Specific Chatbots," *Business & Information Systems Engineering* (62:3), pp. 211-225.

Janssen, A., Rodríguez Cardona, D., & Breitner, M. H. 2021. "More than FAQ! Chatbot Taxonomy for Business-to-Business Customer Services," In: Følstad A. et al. (eds.). *Chatbot Research and Design*. Springer, Cham, Lecture Notes in Computer Science, 12604, pp. 175-189.

Johannsen, F., Leist, S., Konadl, D., & Basche, M. 2018. "Comparison of Commercial Chatbot Solutions for Supporting Customer Interaction," In *Proceedings of the European Conference on Information Systems*.

Johannsen, F., Schaller, D., & Klus, M. F. 2020. "Value Propositions of Chatbots to Support Innovation Management Processes," *Information Systems and e-Business Management*, pp. 205-246.

Jonke, A. W., & Volkwein, J. B. 2018. "From Tweet to Chatbot – Content Management as a Core Competency for the Digital Evolution," In: Linnhoff-Popien, C. et al. (eds.). *Digital Marketplaces Unleashed,* Springer, Berlin, Heidelberg pp. 275-285.

Karlsen, J. T., Andersen, J., Birkely, L. S., & Ødegård, E. 2005. "What Characterizes Successful IT Projects," *International Journal of Information Technology & Decision Making* (4:4), pp. 525-540.

Kepuska, V., & Bohouta, G. 2018. "Next-generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," In *Proceedings of the IEEE Annual Computing and Communication Workshop and Conference*.

Kim, J., Park, S., Robert, L. P. 2020. "Bridging the Health Disparity of African Americans through Conversational Agents", *Digital Government: Research and Practice* (2:1), pp. 4:1-4:7.

Koetter, F., Blohm, M., Drawehn, J., Kochanowski, M., Goetzer, J., Graziotin, D., & Wagner, S. 2019. "Conversational Agents for Insurance Companies: From Theory to Practice," In *Proceedings of the International Conference on Agents and Artificial Intelligence*.

Kruse, L., Wunderlich, N., & Beck, R. 2019. "Artificial Intelligence for the Financial Services Industry: What Challenges Organizations to Succeed," In *Proceedings of the Hawaii International Conference on System Sciences*.

Kuligowska, K. 2015. "Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents," *Professionals Center for Business Research*, pp. 1-16.

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. 2019. "Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues," In *Proceedings of the International Workshop on Chatbot Research and Design*.

Lai, S. T., Leu, F. Y., & Lin, J. W. 2018. "A Banking Chatbot Security Control Procedure for Protecting User Data Security and Privacy," In *Proceedings of the International Conference on Broadband and Wireless Computing, Communication and Applications*.

Landis, J. R., & Koch, G. G. 1977. "An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers," *Biometrics* (33:2), pp. 363-374.

Lee, S., & Ahn, H. 2008. "Assessment of Process Improvement from Organizational Change," *Information & Management (*45:5), pp. 270-280.

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), pp. 535-544.

Mesbah, N., & Pumplun, L. 2020. ""Hello, I'm here to help you"-Medical Care where it is Needed Most: Seniors' Acceptance of Health Chatbots," In *Proceedings of the European Conference on Information Systems*.

Meyer von Wolff, R., Hobert, S., Masuch, K., & Schumann, M. 2020. "Chatbots at Digital Workplaces–A Grounded-Theory Approach for Surveying Application Areas and Objectives," *Pacific Asia Journal of the Association for Information Systems* (12:2), pp. 64-102.

Mordor Intelligence. 2021. "Conversational Systems Market – Growth, Trends, Covid-19, Impact and Forecasts (2021-2026)," https://www.mordorintelligence.com/industry-reports/chatbot-market, (accessed 2021/05/03).

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. 2021. "Trust me, I'm a Bot – Repercussions of Chatbot Disclosure in Different Service Frontline Settings," *Journal of Service Management*. Online first.

Myers, M. D., & Newman, M. 2007. "The Qualitative Interview in IS Research: Examining the Craft," *Information and organization* (17:1), pp. 2-26.

Nguyen, Q. N., & Sidorova, A. 2017. "AI Capabilities and User Experiences: A Comparative Study of User Reviews for Assistant and Non-assistant Mobile Apps," In *Proceedings of the American Conference on Information Systems*.

Nickerson, R. C., Varshney, U., & Muntermann, J. 2013. "A Method for Taxonomy Development and its Application in Information Systems," *European Journal of Information Systems* (22:3), pp. 336-359.

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), pp. 317-335.

Oliver, D. G., Serovich, J. M., & Mason, T. L. 2005. "Constraints and opportunities with interview transcription: Towards reflection in qualitative research," *Social Forces*, (84:2), pp. 1273-1289.

Peffers, P., Tuunanen, T., Rothenberger, M.A., & Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, (24:3), pp. 45-77.

Piccolo, L. S., Roberts, S., Iosif, A., & Alani, H. 2018. "Designing Chatbots for Crises: A Case Study Contrasting Potential and Reality," In *Proceedings of the International BCS Human Computer Interaction Conference*.

Pries-Heje, J., Baskerville, R., & Venable, J. R. 2008. "Strategies for Design Science Research Evaluation," In *Proceedings of the European Conference on Information Systems*.

Pumplun, L., Tauchert, C., & Heidt, M. 2019. "A New Organizational Chassis for Artificial Intelligence-Exploring Organizational Readiness Factors," In *Proceedings of the European Conference on Information Systems*.

Rese, A., Ganster, L., & Baier, D. 2020. "Chatbots in Retailers' Customer Communication: How to Measure their Acceptance?", *Journal of Retailing and Consumer Services* (56), 102176, pp. 1-14.

Rockart, J. F. (1979). "Chief Executives Define their own Data Needs," Harvard Business Review, (57:2), pp. 81-93.

Rodríguez Cardona, D., Werth, O., Schönborn, S., & Breitner, M. H. 2019. "A Mixed Methods Analysis of the Adoption and Diffusion of Chatbot Technology in the German Insurance Sector," In *Proceedings of the American Conference on Information Systems*.

Rodríguez Cardona, D., Janssen, A., Guhr, N., Breitner, M. H., & Milde, J. 2021. "A Matter of Trust? Examination of Chatbot Usage in Insurance Business," In *Proceedings of the Hawaii International Conference on System Sciences*.

Ruane, E., Faure, T., Smith, R., Bean, D., Carson-Berndsen, J., & Ventresque, A. 2018. "Botest: a Framework to Test the Quality of Conversational Agents Using Divergent Input Examples," In *Proceedings of the International Conference on Intelligent User Interfaces Companion*.

Sanny, L., Susastra, A., Roberts, C., & Yusramdaleni, R. 2020. "The Analysis of Customer Satisfaction Factors which Influence Chatbot Acceptance in Indonesia," *Management Science Letters* (10:6), pp. 1225-1232.

Schumaker, R. P., Ginsburg, M., Chen, H., & Liu, Y. 2007. "An Evaluation of the Chat and Knowledge Delivery Components of a Low-level Dialog System: The AZ-ALICE Experiment," *Decision Support Systems* (42:4), pp. 2236-2246.

Seeger, A.-M., & Heinzl, A. 2021. "Chatbots often Fail! Can Anthropomorphic Design Mitigate Trust Loss in Conversational Agents for Customer Service?," In *Proceedings of the European Conference on Information Systems*.

Shapouri, F., & Najjar, L. 2020. "Critical Success Factors in Implementing Business Intelligence Systems," In *Proceedings of the Americas Conference on Information Systems*.

Trkman, P. 2010. "The Critical Success Factors of Business Process Management," *International Journal of Information Management* (30:2), pp. 125-134.

Tsoy, M., & Staples, D. S. 2020. "Exploring Critical Success Factors in Agile Analytics Projects," In *Proceedings of the Hawaii International Conference on System Sciences*.

Van Belzen, M., Trienekens, J., & Kusters, R. 2019. "Critical Success Factors of Continuous Practices in a DevOps Context," In *Proceedings of the International Conference on Information Systems Development*.

van der Goot M. J., Hafkamp L. & Dankfort Z. 2021. "Customer Service Chatbots: A Qualitative Interview Study into the Communication Journey of Customers," In: Følstad A. et al. (eds) *Chatbot Research and Design* Springer, Cham, Lecture Notes in Computer Science, 12604, pp.190-204.

Vijayaraghavan, V., & Cooper, J. B. 2020. "Algorithm Inspection for Chatbot Performance Evaluation," *Procedia Computer Science* (171), pp. 2267-2274.

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* (37:1), pp. 205-224.

vom Brocke, J., Hevner, A., & Maedche, A. 2020a. "Introduction to Design Science Research," In: Hevner, A. & Chatterjee, S. (eds.). *Design Science Research Cases*, Springer, Cham, pp. 1-13.

vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. 2020b. "Special Issue Editorial–Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space," *Journal of the Association for Information Systems* (21:3), pp. 520-544.

Watson, R. T., & Webster, J. 2020. "Analysing the Past to Prepare for the Future: Writing a Literature Review a Roadmap for Release 2.0," *Journal of Decision Systems* (29:3) pp. 129-147.

Weber, P., & Ludwig, T. 2020. "(Non-) Interacting with Conversational Agents: Perceptions and Motivations of Using Chatbots and Voice Assistants," In *Proceedings of the Conference on Mensch und Compute*r.

Webster, J., & Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. xiii-xxiii.

Wiesche, M., Jurisch, M. C., Yetton, P. W., & Krcmar, H. 2017. "Grounded Theory Methodology in Information Systems Research," *MIS Quarterly* (41:3), pp. 685-701.

Winkler, R., & Roos, J. 2019. "Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors," In *Proceedings of the International Conference on Information Systems*.

Wuenderlich, N. V., & Paluch, S. 2017. "A Nice and Friendly Chat with a Bot: User Perceptions of AI-based Service Agents," In *Proceedings of the International Conference on Information Systems*.

Yu, Z., Xu, Z., Black, A. W., & Rudnicky, A. 2016. "Chatbot Evaluation and Database Expansion via Crowdsourcing," In *Proceedings of the Chatbot Workshop of LREC*.

Yuan, L., Dennis, A., & Riemer, K. 2019. "Crossing the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments," In *Proceedings of the Hawaii International Conference on System Sciences*.

Zamora, J. 2017. "I'm Sorry, Dave, I'm Afraid I can't do that: Chatbot Perception and Expectations," In *Proceedings of the International Conference on Human Agent Interaction*.

Zemčík, T. 2020. "Failure of Chatbot Tay was Evil, Ugliness and Uselessness in its Nature or do we judge it through Cognitive Shortcuts and Biases?" *AI & Society* (36), pp. 361-367.

Zierau, N., Elshan, E., Visini, C., & Janson, A. 2020. "A Review of the Empirical Literature on Conversational Agents and Future Research Directions," In *Proceedings of the International Conference on Information Systems*.