Artificial Intelligence in Organizations: Managing the Lifecycle of Conversational Agents

Cumulative dissertation with the aim of achieving a doctoral degree

at the Faculty of Mathematics, Informatics, and Natural Sciences

Department of Informatics

Universität Hamburg

submitted by

Tom Lewandowski

May, 2024

Hamburg, Germany

Date of disputation

26 . 08 . 2024

Evaluators

First Evaluator: Prof. Dr. Tilo Böhmann

Second Evaluator: Prof. Dr. Eva Bittner

Abstract

Motivation

Given the rapid advances of information technology, led by increasing dissemination and possibilities, many areas of science, society and the economy are being impacted upon. In this context, artificial intelligence (AI) has developed into a disruptive technology. Driven by the large availability of data, increased computing power, and algorithmic and technological advances, AI is a decisive factor in the digital transformation of organizations. As a result, companies are spending significant resources on AI-based systems and related applications, e.g., to improve customer engagement and remain competitive.

Conversational Agents (CAs), such as chatbots, are an emerging natural language-based application of AI, which provide organizations with an increasingly intelligent, social, learning, integrative, and cost-effective platform to support and automate the delivery of services and information. CAs offer new forms of scalability and availability (24/7), speed, and personalization. However, despite their possibilities, organizations are struggling to realize the full potential of CAs in real-world environments. AI-based CAs pose new challenges due to their unique characteristics. They are not sufficiently represented in current research, which leads to many open questions and opportunities regarding the management of CAs' lifecycle, especially in regard to different activities in the introduction, design, implementation, operation, and improvement phases.

From a broader, organization-wide perspective, the management of AI is an essential task in realizing the new value propositions. However, knowledge on how to structure AI transformations and effectively manage specific applications is limited, leading to a high failure rate of AI initiatives. There is a need for research to support organizations in systematically developing AI (implementation) competencies in order to master all the complexity of this new technology.

By switching viewpoints between organization-wide and system-specific perspectives, this dissertation contributes to the emerging field of (1) AI in organizations in general by providing guidance for navigating, managing, and (re)evaluating AI transformations, and (2) CAs in particular, by providing various forms of validated design knowledge for their lifecycle management. Together, these themes facilitate implications for a broader understanding of AI management in organizational settings.

Research Design

Anchored in the paradigm of Design Science Research (DSR), this dissertation follows a cumulative approach and reports on several iterations of a research project based on seven publications at the intersection of AI, CAs, and organizations. During this cumulative dissertation, different research methods were applied to the publications to support DSR efforts and to ensure research rigor. The applied methods were used to understand existing problems and provide a solid knowledge (data) base, to derive design knowledge in a structured way, and to support the construction and evaluation of artifacts to satisfy the project solution space. In this context, several structured literature reviews and qualitative data collection and analysis were conducted. In addition, structured methods were used to generate design principles and to evaluate the generated design knowledge, e.g., by implementing the Framework for Evaluation in Design Science Research (FEDS) paradigm.

Results

The central results of this thesis can be divided into two research perspectives. In each perspective, different conceptually and empirically validated design knowledge was developed. This design knowledge includes both prescriptive knowledge such as design principles and artifacts that are instantiated in the research project in the real-world.

The first part deals with a CA-specific perspective, in which primarily the management of the lifecycle of conversational agents and associated activities along the main phases were investigated. The specific AI-based information system CA was studied and insights were gained into the functionality and characteristics, design, implementation, development, and continuous improvement of these novel natural-language based information systems. The second perspective focuses on organization-wide AI transformations and examines the strategic design of AI projects across the organization. The findings offer concrete recommendations for action for practitioners and starting points for in-depth investigations for researchers. These insights help navigate, manage, and (re)evaluate AI strategies, programs, and initiatives. Both perspectives are mutually beneficial and have implications for each other: CA-specific findings help avoid pitfalls in organization-wide AI transformations, while the resulting knowledge about AI transformations is useful for understanding the fundamentals of CA management.

Contributions

This dissertation provides several theoretical and practical contributions at the intersection of AI and CA research in organizations. Knowledge and concrete evaluated and instantiated artifacts are delivered that contribute to both organization-wide AI transformation and the management of specific AI-based applications, such as CAs. In this context, from a CA-specific perspective and with the establishment of a dedicated CA research stream, the AI-based system CA is investigated and knowledge about the management of these novel systems is gained. Thus, this thesis contributes knowledge to the field of CAs by adopting an organizational perspective, thereby countering the trend of previous CA research, which has previously focused mainly on individual, behavioral, interactional, or technical design perspectives.

To this end, this thesis contributes a comprehensive conceptualization of the terminology of CAs and the relationships between their concepts and different characteristics. This leads to both potentials and complexities of CAs in organizations, which are subsequently examined and described in this dissertation. In doing so, the foundation has been laid for understanding the challenges and activities of the lifecycle management (LCM) of CAs, aggregated through various research activities from theory and practice. Based on this foundational and conceptual knowledge, this thesis charts a research agenda for CA management in organizations that reviews existing knowledge and identifies future research opportunities. Subsequently, this thesis contributes to the management of CAs in organizations by providing literature-based and empirically grounded design knowledge that prescribes the lifecycle of CAs and provides a system-wide and phase-based view of the technology. Various insights were presented along the phases of initiation, design, implementation, development, operation, and improvement. With regard to the latter phases, this dissertation has provided well-founded artifacts for the continuous evaluation and improvement of CAs. Examples include a design guideline for CAs, a quality criteria set for CAs, and a framework for structuring the evaluation of CAs and identifying areas for systematic improvement.

Finally, this dissertation offers an organization-wide perspective that focuses on AI transformation and takes a broader view of the key research questions. This thesis aims to shed light on how to approach AI transformations and provide concrete recommendations for action for practitioners and starting points for in-depth investigations for researchers. These insights help to navigate, manage, and (re)evaluate AI strategies, programs, and initiatives. In this context, a 3-D model to guide organization-wide AI transformation was developed and concrete recommendations for each dimension were presented. By switching between organization-wide and system-specific perspectives, this dissertation contributes to the emerging field of (1) AI in general by providing guidance for AI transformations, and (2) CAs in particular by providing various forms of validated design knowledge for their lifecycle management, which together facilitate implications for a broader understanding of AI management in organizational settings.

Limitations

The research conducted within this dissertation has limitations that provide opportunities for further investigation. The structured literature reviews (SLRs) conducted as part of this dissertation faced limitations resulting from the filtering steps used in terms of defining the scope of the review, including fixed search terms, database and literature selection, aggregation, and authors' judgment, which has potentially affected the depth of findings. However, structured methods were used to conduct the SLRs in a valid, transparent, and comprehensible manner.

In terms of qualitative data collection and analysis, the potential subjectivity of authors and the influence of experts' domain-specific experiences might influence the external validity of the research. Additionally, all interviews were conducted digitally due to the COVID-19 pandemic, which may have introduced unintended effects. However, many experts work for international companies in various industries, which has provided a range of experience and sufficient data saturation. With regard to the pandemic, the qualitative data analysis by various authors did not reveal any conspicuous side effects. Furthermore, the Design Science Research (DSR) framework used here also has its limitations, including its dependence on specific organizational contexts. These limitations suggest that more longitudinal studies and more diverse methodological approaches may be needed to further substantiate the results. Further studies could make use of even more varied evaluation methods, which could lead to more comprehensive and extended results based on an even broader context range.

Future Research

The findings of this cumulative dissertation indicate further promising starting points and directions for future research. The dissertation underscores the general need for more in-depth studies from an organizational, practical, or management-oriented perspective. Such studies are crucial for a better understanding of the identified specific characteristics of CAs, the challenges they pose for management, and the strategies that together can reduce the risks of early termination of their operation in organizations. Closely connected, more Design Science Research (DSR) oriented research and entrepreneurial approaches that produce evaluated socio-technical artifacts,

addressing CA lifecycle activities and providing robust design knowledge for both researchers and practitioners, are needed. In this context, there are still significant knowledge gaps in the field of CA lifecycle management, for example with regard to researching the detailed activities, roles, and collaborations necessary for effective CA integration, efficient operation and their continuous improvement. In this regard, further research that takes an interdisciplinary approach to CAs and unifies the previously fragmented research streams, disciplines, and application domains is recommended. The research agenda of this dissertation serves as a concise research roadmap for the management of CA in organizations and as a compass for researchers interested in deepening their knowledge in these areas.

Regarding recent technical advancements in the field of Natural Language Processing (NLP) and Machine Learning (ML) applications, the emergence of large language models (LLMs) in particular will lead to a more dynamic research landscape and therefore new enquiry opportunities, suggesting a paradigm shift in CA development and integration, fostering more natural, effective, and organizationally aligned conversational agents.

Kurzfassung

Motivation

Die zunehmende Verbreitung und die Möglichkeiten der Informationstechnologie schreiten schnell voran und wirken sich auf verschiedene Bereiche der Wissenschaft, Gesellschaft und Wirtschaft aus. In diesem Zusammenhang hat sich die Künstliche Intelligenz (KI) zu einer disruptiven Technologie für Unternehmen entwickelt. Angetrieben durch die große Menge an verfügbaren Daten, der steigenden Rechenleistung sowie algorithmische und technologische Fortschritte ist KI ein entscheidender Faktor für die digitale Transformation von Unternehmen. Infolgedessen investieren Unternehmen erhebliche Ressourcen in KI-basierte Systeme und Anwendungen, um z.B. den Kundenkontakt zu verbessern und wettbewerbsfähig zu bleiben.

Eine aufkommende, auf natürlicher Sprache basierende Anwendung von KI sind Conversational Agents (CAs) wie Chatbots, die Unternehmen eine zunehmend intelligente, soziale, lernende, integrative und kostengünstige Plattform zur Unterstützung und Automatisierung der Bereitstellung von Dienstleistungen und Informationen bieten. Sie erlauben neue Formen der Skalierbarkeit und Verfügbarkeit (24/7), Geschwindigkeit und Personalisierung. Trotz ihrer Möglichkeiten fällt es Unternehmen jedoch schwer, das volle Potenzial von CAs in der Praxis auszuschöpfen. KI-basierte CAs stellen Unternehmen aufgrund ihrer einzigartigen Eigenschaften vor neuen Herausforderungen. Sie sind in der aktuellen Forschung nicht ausreichend repräsentiert, was zu vielen offenen Fragen und Forschungsmöglichkeiten, z.B. in Bezug auf das Management ihres Lebenszyklus führt, insbesondere in Bezug auf unterschiedlichen Aktivitäten in den Phasen der Einführung, des Designs, der Implementierung, des Betriebs und der Verbesserung.

Aus einer breiteren, organisationsweiten Perspektive ist das Management von KI eine wesentliche Aufgabe, um die neuen Wertversprechen zu realisieren. Das Wissen darüber, wie KI-Transformationen strukturiert und spezifische KI-Anwendungen effektiv gemanagt werden können, ist jedoch begrenzt, was zu einer hohen Misserfolgsquote von KI-Initiativen führt. Es besteht Forschungsbedarf, um Unternehmen bei der systematischen Entwicklung von KI-(Umsetzungs-)Kompetenzen zu unterstützen, um die Komplexität von KI zu beherrschen.

Durch den Perspektivenwechsel zwischen organisationsweiten und systemspezifischen Ansätzen leistet diese Dissertation einen Beitrag zu dem aufkommenden Feld (1) der KI Transformation im Allgemeinen, indem sie Anleitungen für die Navigation, das Management und die (Neu-)

Bewertung von KI-Transformationen liefert, und (2) der CAs im Besonderen, indem sie verschiedene Formen von validiertem Designwissen für deren Lebenszyklusmanagement bereitstellt, die zusammengenommen Implikationen für ein breiteres Verständnis des KI-Managements im organisatorischen Umfeld ermöglichen.

Forschungsdesign

Die vorliegende Dissertation basiert auf dem Paradigma der Design Science Research (DSR). Sie verfolgt einen kumulativen Ansatz und berichtet über mehrere Iterationen eines Forschungsprojekts, das auf sieben Veröffentlichungen an der Schnittstelle von KI, CAs und Organisationen basiert. Während dieser kumulativen Dissertation wurden verschiedene Forschungsmethoden in den einbezogenen Publikationen angewandt, um die Forschung zu unterstützen und die Rigorosität der Forschung sicherzustellen. Die angewandten Methoden wurden eingesetzt, um existierende Probleme zu verstehen und eine solide Wissensbasis zu gewinnen, um Designwissen auf strukturierte Weise abzuleiten und um die Konstruktion und Evaluation von Artefakten zu unterstützen, die auf den Lösungsraum des Projekts abzielen. In diesem Zusammenhang wurden mehrere strukturierte Literaturrecherchen und qualitative Datenerhebungen und -analysen durchgeführt. Darüber hinaus wurden strukturierte Methoden zur Generierung von Design-Prinzipien und zur Evaluierung des generierten Design-Wissens eingesetzt, z.B. durch die Anwendung des Framework for Evaluation in Design Science Research (FEDS).

Forschungsergebnisse

Die zentralen Forschungsergebnisse dieser Arbeit lassen sich in zwei Forschungsperspektiven unterteilen, in denen unterschiedliches konzeptionelles und empirisch validiertes Gestaltungswissen erarbeitet wurde. Dieses Gestaltungswissen beinhaltet sowohl präskriptives Wissen in Form von z.B. Gestaltungsprinzipien als auch Artefakte, die im Forschungsprojekt in der realen Welt instanziiert wurden.

Der erste Teil befasst sich mit einer CA-spezifischen Perspektive, in der vor allem das Management des Lebenszyklus von Conversational Agents und die damit verbundenen Aktivitäten entlang der Hauptphasen untersucht wurden. Das spezifische KI-basierte Informationssystem CA wurde untersucht und es wurden Einblicke in die Funktionalität und Eigenschaften, das Design, die Implementierung, die Entwicklung und die kontinuierliche Verbesserung dieser neuartigen natürlichsprachlichen Systeme gewonnen. Die zweite Perspektive konzentriert sich auf organisationsweite KI-Transformationen und untersucht die strategische Gestaltung von KI-

Vorhaben in der gesamten Organisation. Die Ergebnisse liefern konkrete Handlungsempfehlungen für die Praxis und neue Ansatzpunkte für die Forschung. Diese Erkenntnisse helfen, KI-Strategien, -Programme und -Initiativen zu etablieren, zu steuern und (neu) zu bewerten. Beide Perspektiven ergänzen und beeinflussen sich gegenseitig: Die CA-spezifischen Erkenntnisse helfen, Fallstricke bei unternehmensweiten KI-Transformationen zu vermeiden, während das resultierende Wissen über KI-Transformationen nützlich ist, um die Grundlagen des CA-Managements zu verstehen.

Forschungsbeitrag

Diese Dissertation leistet mehrere theoretische und praktische Beiträge an der Schnittstelle der KI- und CA-Forschung in Organisationen. Es werden Erkenntnisse und konkrete evaluierte und instanziierte Artefakte bereitgestellt, die sowohl zur organisationsweiten KI-Transformation als auch zum Management spezifischer KI-basierter Anwendungen, wie CAs, beitragen. In diesem Zusammenhang wird aus einer CA-spezifischen Perspektive und mit der Etablierung eines dedizierten CA-Forschungsstroms das KI-basierte System CA untersucht und Wissen über das Management dieser neuartigen Systeme gewonnen. Damit trägt diese Arbeit Wissen zum Feld der CAs bei, indem eine organisationale Perspektive eingenommen wird und entgegnet damit dem Trend der bisherigen CA-Forschung, die sich zuvor vor allem auf individuelle, verhaltensbezogene, interaktionale oder technische Gestaltungsperspektiven konzentriert hat.

Hierbei stellt die Arbeit eine umfassenden Konzeptualisierung der Terminologie von CAs und der Zusammenhänge zwischen ihren Konzepten und unterschiedlichen Ausprägungen vor. Hieraus ergebenden sich sowohl Potenziale, als auch Komplexitäten von CAs in Organisationen, die anschließend in der Dissertation untersucht und beschrieben werden. Damit wurde der Grundstein für das Verständnis der Herausforderungen und Aktivitäten des Lifecycle Managements (LCM) im CA-Management gelegt, welches durch unterschiedliche Forschungsaktivitäten aus Theorie und Praxis aggregiert wurde. Aufbauend auf diesem Basiswissen wird in dieser Arbeit eine Forschungsagenda für das CA-Management in Organisationen aufgestellt, die das vorhandene Wissen verknüpft und zukünftige Forschungsmöglichkeiten aufzeigt. Im Anschluss, leistet diese Arbeit einen Beitrag zum Management von CAs in Organisationen, indem sie literaturbasiertes und empirisch fundiertes Designwissen bereitstellt, das den Lebenszyklus von CAs beschreibt und eine systemweite und phasenbasierte Sicht auf die Technologie vermittelt. Hierbei werden unterschiedliche Erkenntnisse zu den Phasen Planung, Design, Implementierung, Entwicklung, Betrieb und Verbesserung zusammengetragen und sowohl Herausforderungen als auch Gestaltungsmöglichkeiten vermittelt. Im Hinblick auf die letzteren Phasen hat die Dissertation fundierte Artefakte zur kontinuierlichen Evaluierung und Verbesserung von CAs geliefert.

Beispiele hierfür sind ein Gestaltungsleitfaden für CAs, ein Katalog anwendbarer Qualitätskriterien für CAs und ein Prozedurmodell zur Strukturierung der Bewertung von CAs und zur Ermittlung von Bereichen für systematische Verbesserungen.

Abrundend wird in dieser Dissertation eine unternehmensweite Perspektive eingenommen, die sich auf die KI-Transformation konzentriert und einen breiteren Blick auf unternehmensweite KI-Fragestellungen Fragen wirft. Diese Dissertation soll Aufschluss darüber geben, wie KI-Transformationen angegangen werden können, und konkrete Handlungsempfehlungen für Praktiker sowie Ansatzpunkte für vertiefte Untersuchungen für Forscher liefern. Diese Erkenntnisse helfen dabei, KI-Strategien, -Programme und -Initiativen zu planen, zu managen und (neu) zu bewerten. In diesem Zusammenhang wurde ein 3-D-Modell für die unternehmensweite KI-Transformation vorgestellt und konkrete Empfehlungen für jede Dimension präsentiert. Durch den Wechsel zwischen einer organisationsweiten und einer systemspezifischen Perspektive leistet diese Dissertation einen Beitrag zu den aufstrebenden Forschungsfeldern der (1) KI-Transformation im Allgemeinen, indem sie eine Orientierungshilfe für KI-Transformationen bietet, und (2) zu CAs im Besonderen, indem sie verschiedene Formen von validiertem Designwissen für ihr Lebenszyklusmanagement bereitstellt, was zusammengenommen Implikationen für ein breiteres Verständnis des KI-Managements im organisatorischen Umfeld ermöglicht.

Limitationen

Die im Rahmen dieser Dissertation durchgeführte Forschung weist Limitationen auf, die Möglichkeiten für weitere Untersuchungen eröffnen. Die strukturierten Literaturrecherchen (SLRs), die im Rahmen dieser Dissertation durchgeführt wurden, weisen Einschränkungen auf, die sich aus den filternden Schritten bei der Festlegung des Untersuchungsumfangs, den definierten Suchbegriffen, der Datenbank- und Literaturauswahl, der Aggregation und dem Urteil der Autoren ergeben und die Tiefe der Ergebnisse beeinflussen können. Dennoch wurden strukturierte Methoden verwendet, um die SLRs in einer validen, transparenten und nachvollziehbaren Weise durchzuführen. In Bezug auf die qualitative Datenerhebung und -analyse könnten die potenzielle Subjektivität der Autoren und der Einfluss der domänenspezifischen Erfahrungen der Experten die externe Validität der Forschung beeinflussen. Darüber hinaus wurden alle Interviews aufgrund der COVID-19-Pandemie digital durchgeführt, was zu unbeabsichtigten Nebeneffekten führen könnte. Viele der Experten arbeiten jedoch für internationale Unternehmen in unterschiedlichen Branchen, sodass ein breites Erfahrungsspektrum und eine ausreichende und vielseitige Datensättigung gegeben waren. In Bezug auf die Pandemie ergab die qualitative Datenanalyse durch verschiedene Autoren keine auffälligen Nebeneffekte.

Weiterhin hat der hier verwendete Design Science Research (DSR)-Rahmen auch seine Grenzen, einschließlich der Abhängigkeit von spezifischen organisatorischen Kontexten. Diese Einschränkungen deuten darauf hin, dass mehr Längsschnittstudien und vielfältigere methodische Ansätze erforderlich sein könnten, um die Resultate weiter zu untermauern. Zukünftige Studien könnten noch mehr unterschiedliche Evaluationsmethoden verwenden, was zu umfassenderen und breiter validierten Ergebnissen führen könnte, die auf zusätzlichen Organisationsumfeldern basieren.

Ausblick

Aus den Ergebnissen dieser kumulativen Dissertation lassen sich weitere vielversprechende Anknüpfungspunkte und -richtungen für weitere, zukünftige Forschung identifizieren. Die Dissertation unterstreicht den allgemeinen Bedarf an tiefergehenden Studien aus organisatorischen, praktischen beziehungsweise managementorientierten Blickwinkeln. Solche Studien sind entscheidend für ein besseres Verständnis der identifizierten spezifischen Charakteristika von CAs, der Herausforderungen, die sie für das Management darstellen, und der Strategien, mit denen gemeinsam die Risiken einer frühzeitigen Beendigung ihres Betriebs in Organisationen reduziert werden könnten.

Eng damit verbunden wird weitere Forschung empfohlen, die sich stärker an Design Science Research (DSR) bzw. unternehmerischen, pilotierenden und/oder realweltlichen (Forschungs-) Ansätzen orientiert und somit evaluierte sozio-technische Artefakte hervorbringt und sich dabei mit den Lebenszyklusaktivitäten von CAs befassen und fundiertes Gestaltungswissen für Forscher und Praktiker bereitstellen. In diesem Zusammenhang gibt es noch erhebliche Wissenslücken im Bereich des CA-Lebenszyklusmanagements, zum Beispiel in Bezug auf die Erforschung der detaillierten Aktivitäten, Rollen und Kooperationen, die für eine effektive CA-Integration, einen effizienten Betrieb und ihrer kontinuierlichen Verbesserung notwendig sind. In dieser Hinsicht werden weitere Untersuchungen empfohlen, die einen ähnlichen interdisziplinären Ansatz für CAs verfolgen und die bisher fragmentierten Forschungsrichtungen, Disziplinen und Anwendungsbereiche vereinen. Die Forschungsagenda dieser Dissertation dient als präziser Forschungsfahrplan für das Management von CAs in Organisationen und als Kompass für Forscher, die ihr Wissen in diesen Bereichen vertiefen möchten.

Weiterhin wird in Anbetracht der jüngsten technologischen Fortschritte im Bereich des Natural Language Processing (NLP)- und der Machine Learning (ML)-Anwendungen, insbesondere durch das Aufkommen von Large Language Models (LLMs), eine zunehmend dynamischere Forschungslandschaft erwartet. In Folge dessen entstehen neue Forschungsmöglichkeiten, die einen Paradigmenwechsel in der Entwicklung von CAs und ihrer Integration nahelegen und zu natürlicheren, effizienteren und an die Organisation angepassten Gesprächsagenten und verknüpften Konversationen führen werden.

Table of Contents

List of Figures

The publications' figures are not integrated in this list.

List of Tables

The publications' tables are not integrated in this list.

List of Abbreviations

AI	Artificial Intelligence
AIS	Association for Information Systems
API	Application Programming Interface
CA	Conversational Agent
CCP	Context-Content-Process Evaluation Framework
CS	Computer Science
DP	Design Principle
DSR	Design Science Research
FEDS	Framework for Evaluation in Design Science Research
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HCI	Human-Computer Interaction
IS	Information Systems
IT	Information Technology
ITSM	IT Service Management
LCM	Lifecycle Management[1]
LLM	Large Language Model
ML	Machine Learning
MR	Meta-Requirement
NLG	Natural Language Generation
NLP	Natural Language Processing

[1] Contrary to the publication by Alter et al. (2001), the word "lifecycle" is written together in this dissertation, which is the case in large parts of the IT service management (ITSM) literature.

NLU	Natural Language Understanding
RG	Research Goal
RQ	Research Question
SaaS	Software as a Service
SLR	Systematic Literature Review
TOE	Technology-Organization-Environment (Framework)
UI	User Interface
WSLC	Work System Life Cycle (Model)

The publications' abbreviations are not integrated in this list.

1 Introduction

The following chapter is divided into the motivation and problem statement, the presentation of the research questions, and the outline of the thesis.

1.1 Motivation and Problem Statement

The capabilities of information technology are growing at an unprecedented pace and increasingly exceed the cognitive abilities of human beings (Winkler et al., 2020). In this context, Artificial Intelligence (AI) has evolved as one of *"the most important and disruptive new technology for large organizations"* (Benbya et al., 2020, p. 1) and serves as a significant driver for the digital transformation of enterprises in the upcoming years (Subramaniam, 2023). Various AI-based systems that mimic and reproduce human skills and intelligence have emerged (Castillo et al., 2020), facilitated by the accessibility of vast data volumes, enhanced computational capabilities, and advances in learning algorithms (Rzepka & Berger, 2018). AI helps organizations turn data into value (Eitle et al., 2022), automate processes (Subramaniam, 2023), and innovate products, services, and business models (van Giffen et al., 2020). Presently, 37% of global organizations have embedded AI into their businesses and products (Jovanovic, 2023; Uba et al., 2023), indicating that AI *"is today a fixed point on the agenda of many companies aiming to leverage AI in their respective business context"* (Sagodi et al., 2022, p. 6851).

As a result, the management of AI in organizations is a central task in realizing the new value proposition with productive systems (van Giffen et al., 2020). More specifically, the professional management of AI-based systems signals the advent of a new era in the field of information technology (IT) management, representing a class of novel IT artifacts that require a new holistic understanding by managers on how to effectively communicate, lead, coordinate, and control them (Berente et al., 2021).

However, there needs to be more knowledge in structuring both (1) the organization-wide AI transformation and (2) the management of specific AI-based applications, presenting many activity fields and challenges for organizations. Helping organizations systematically develop AI (implementation) capabilities is still a scarce area of knowledge in research and practice (Uba et al., 2023). This is unsatisfactory given the challenges various organizations face in establishing organization-wide AI programs and initiatives and AI projects' generally high failure rate due to lack of guidance and inadequate capabilities coupled with exaggerated expectations (Uba et al., 2023). Today, AI adoption is still in its infancy for many organizations, with AI projects frequently remaining exploratory and experimental, and often failing to progress beyond the pre-production proof-of-concept stage (e.g., Schmelzer, 2022), while achieving little or no economic return (Benbya et al., 2020).

From a system-specific viewpoint, conversational agents (CAs) represent a particular and prominent case of AI that merits further investigation in the described tension area. Contemporary organizations are increasingly adopting CAs as intelligent and natural language-based interfaces to their digital services and information (Følstad et al., 2021; Gnewuch et al., 2018). Integrated into diverse front and back-end applications, such as websites or messaging applications (e.g., MS Teams), CAs support the progressive digitization and transformation of organizations by, for example, filtering or providing information or efficiently supporting employees in their daily work tasks (Zierau et al., 2020a). Other drivers for the expansion of CAs are the intuitive communication channel for users and the relatively low technical effort required for deployment, which are key factors driving their widespread adoption throughout organizations (Diederich et al., 2019b; Riikkinen et al., 2018; Xu et al., 2017).

According to a recent analysis, the global market for CAs, valued at $3.49 billion in 2021, is expected to grow to $22.9 billion by 2030, indicating their growing importance (Research and Markets, 2022). CAs offer new forms of personalization, speed, and (cost-)effectiveness (Lewandowski et al., 2023a), supporting users by retrieving, structuring, and analyzing information, or by augmenting and automating activities (Poser et al., 2022b). Regarding automation, Gartner projects that by 2026, CAs will automate one in ten agent interactions (Rimol, 2022). As a result, CAs are expected to deliver substantial economic value in existing and future applications, significantly impacting businesses and digital ecosystems (Seeger et al., 2021; Seiffer et al., 2021).

However, while researchers and practitioners are increasingly interested in the potentials and applications of CAs, e.g., in service encounters and workplace settings, as evidenced by new research studies (Janssen et al., 2020; Meyer von Wolff et al., 2019b), many CAs fall short of expectations or even fail (Gnewuch et al., 2017; Janssen et al., 2021b). Equipped with numerous novel characteristics, they offer many opportunities for organizations while also posing a variety of (design and management) challenges. CAs represent a novel subtype of AI-based information systems (IS), being natural language-based, social, and user-centric systems that interact with users in a dialog-like manner, while also being learning and intelligent systems with an unfinished and integrative character (see Section 2.2). Their successful adoption depends on organizational arrangements, including collaborative and continuous design, training, and development approaches that involve the efforts of IT, business, and service professionals (Lewandowski et al., 2021). In addition, CAs require novel human and data-centric management approaches (Lewandowski et al., 2021).

Despite the growing importance of CAs, research from an organizational and management-oriented perspective still needs to be conducted. Research in this context is essential for understanding the characteristics of CAs, the resulting management challenges, and activities to reduce the risk of failure and discontinuation in organizations. Nevertheless, only a few contributions investigate the organizational or management-oriented perspective of CAs (e.g., Corea et al., 2020; Diederich et al., 2019a; Essaied et al., 2020). Instead, the primary focus is on the individual (e.g., trust issues), conceptual (e.g., interaction design), or technical design aspects (e.g., NLP algorithms) (Diederich et al., 2019a; Janssen et al., 2020; Lu et al., 2020; Premathilake et al., 2021; Zierau et al., 2020a). Conversely, less is known regarding the management of CA applications in organizational contexts and studies investigating CA applications often ignore their long-term success (Lewandowski et al., 2021; Lewandowski et al., 2022a; Lewandowski et al., 2022b). Closely related, research on the strategic management of CA design, deployment, operation, and improvement is scarce.

Based on the outlined problems and knowledge gaps of these two research streams—the organization-wide AI transformation stream and the CA-specific lifecycle stream—this dissertation approaches the given problem domains by applying the DSR paradigm. In this context, the specific AI-based system CA is studied and knowledge about the management of these novel systems is gained. This results in a CA lifecycle and associated activities that allow implications for general AI management. In addition, an organization-wide perspective is applied through the investigation of AI transformation, which in turn allows implications for the management of CAs.

1.2 Research Questions

In the following paragraphs, the guiding research questions (RQs) are specified and briefly described.

> **RQ 1: What are the characteristics of CAs and what challenges do these characteristics pose for implementing CAs in organizations?**

Adopting a CA-specific perspective and establishing an appropriate research stream, RQ 1 aims to investigate the characteristics of this new technology. As many CAs fail in organizations due to insufficient knowledge of their specific concepts and the associated complexities, the question arises how CAs can be effectively integrated and managed. In recent years, several organizations have taken their CAs offline due to a lack of detailed knowledge which has resulted in an uncoordinated, dynamic, and highly exploratory development process (Janssen et al., 2021b). Therefore, this dissertation analyses and aggregates the various characteristics of CAs and derives the resulting LCM challenges from theory and practice to provide a structured conceptualization and fundamental understanding as a basis of this AI-based technology.

RQ 2: How to manage the lifecycle of CAs and specifically their improvement activities?

Given the general lack of information on the management of CAs in real-world contexts, and the specific shortage of guidance or knowledge on activities for the continuous management and improvement of this technology, RQ 2 addresses the lifecycle of CAs and corresponding activities. Therefore, the dissertation aims to develop methods for establishing and sustaining CAs in organizations. First authors already call for research on how organizations can most effectively implement/deploy (Janssen et al., 2020; Schuetzler et al., 2021), adopt (Essaied et al., 2020), and manage (Corea et al., 2020; Meyer von Wolff et al., 2021), and maintain CAs (Kvale et al., 2019) to prevent their failure and to sustain them. Thus, understanding the LCM of CAs can provide a structured, unified view of this complex IS. Implementing CA projects requires a multi-perspective design and development process (e.g., for the design of interaction and handover scenarios), which must be approached in a highly interdisciplinary and participatory manner (Lewandowski et al., 2021). The findings of this dissertation reveal that CAs fail due to various organizational and human-related challenges (see Section 5.4), necessitating a broader management perspective that encompasses a range of parallel activities, which can be guided by a system-wide and phased view of the technology, as examined in this dissertation.

RQ 3: How to drive organization-wide AI transformations?

RQ 3 expands to an organization-wide perspective, focusing on AI transformations and taking a broader view of the topics to guide organizations, which consequently impacts also CA management activities. Organizations are struggling to realize the potential of AI in general, and many projects fail in the early stages due to a lack of strategic guidance and established best practices for initiating AI transformation and driving organization-wide AI change. This dissertation aims to shed light on how to approach AI transformations and provide concrete recommendations for action for practitioners and starting points for in-depth investigations for researchers. These insights help navigate, manage, and (re)evaluate AI strategies, programs, and initiatives. In addition, they help specific technology or development teams (e.g., CA teams) lay the foundation for successful management of their AI technologies by emphasizing a structured approach to AI integration and the critical importance of alignment with organizational AI strategies and capabilities.

1.3 Outline of the Thesis

The structure of this dissertation is shown in **Table 1**. Following this introduction, Chapter 2 presents the theoretical foundations in the context of AI and CAs to provide a foundational knowledge base for the terms and concepts used in this dissertation and its related publications. Then, Chapter 3 describes the overall research design and the applied research methods to address the RQs. Thereafter, Chapter 4 details the publications included in this cumulative dissertation. Chapters 5 and 6 present and discuss the theoretical and practical contributions resulting from the conducted DSR project and its research activities. Chapter 7 focuses on the limitations of the chosen research design, while Chapter 8 examines avenues for further research in the context of this thesis. Finally, Chapters 9-15 contain the core publications of this dissertation, which have been published in different scientific outlets (e.g., conference proceedings or scientific journals).

Table 1. Thesis Outline

Wrapper	1. Introduction	2. Theoretical Foundations	3. Research Design	4. Publications
	5. Theoretical Contributions	6. Practical Contributions	7. Limitations	8. Implications for Further Research
Publications	9. Publication No. 1: Lewandowski et al. (2021): State-of-the-Art Analysis of Adopting AI-based Conversational Agents in Organizations: A Systematic Literature Review			
	10. Publication No. 2: Lewandowski et al. (2022a): Managing Artificial Intelligence Systems for Value Co-creation: The Case of Conversational Agents and Natural Language Assistants			
	11. Publication No. 3: Lewandowski et al. (2022b): Design Knowledge for the Lifecycle Management of Conversational Agents			
	12. Publication No. 4: Uba et al. (2023): The AI-based Transformation of Organizations: The 3D-Model for Guiding Enterprise-wide AI Change			
	13. Publication No. 5: Heuer et al. (2023): Rethinking Interaction with Conversational Agents: How to Create a Positive User Experience Utilizing Dialog Patterns			
	14. Publication No. 6: Lewandowski et al. (2023b): Leveraging the Potential of Conversational Agents: Quality Criteria for the Continuous Evaluation and Improvement			
	15. Publication No. 7: Lewandowski et al. (2023a): Enhancing Conversational Agents for Successful Operation: A Multi-perspective Evaluation Approach for Continuous Improvement			

2 Theoretical Foundations

The following chapter introduces the theoretical foundations to provide a fundamental knowledge base for the terms and concepts used throughout this dissertation and its included publications.[2] The section outlines the origin of the term AI and emphasizes its significance for organizations in Information Systems (IS) research. Further, this section introduces AI-based CAs, including their key developments, definition and demarcation, unique characteristics, and key preliminary work, highlighting their growing importance and research potential.

2.1 Artificial Intelligence in Organizations

Over the past few years, the landscape of AI has evolved profoundly, moving from a mere technical trend to an integral part of our daily lives (Maedche et al., 2019). AI-based systems are proliferating in various application domains and contributing to numerous innovations (L. Wang et al., 2020). The availability of massive amounts of data, increasing computational capabilities, and advances in learning algorithms (Rzepka & Berger, 2018) have led to various AI-based systems that mimic and complement human intelligence (e.g., Xiong et al., 2023). AI has far-reaching implications for organizations and society in general and, thus, for academic disciplines (Bawack et al., 2019). In organizational contexts, AI has emerged as a comprehensive collection of technologies that can learn and perceive data or objects, leading to widespread applications of AI technologies (Bawack et al., 2019). Consequently, there has been a significant increase in research attention to AI across all disciplines, with a particular focus on IS (Elshan et al., 2022b; Felderer & Ramler, 2021).

However, while interest in AI has grown considerably, there is no consensus in practice or academia on the exact meaning of the umbrella term "AI" (Alsheibani et al., 2019b; Nguyen et al., 2022). Since the term "AI" was first coined by Minsky and McCarthy in 1956 (McCarthy et al., 2006), AI has been investigated in numerous academic disciplines, from computer science (CS) to philosophy to futurology (Kühl et al., 2022) (see **Table 2**). Thereby, AI has undergone continuous redefinition and expansions across disciplines (Collins et al., 2021; Venkatesh, 2022). Similarly, terminology related to AI, such as machine learning (ML), and AI-based or intelligent systems is used inconsistently (Kühl et al., 2022). Typically, researchers think of AI as a broad concept that encompasses

[2] The theoretical foundations are based on the publications included in this dissertation. Thus, text components are included here that are similar or identical to components of the articles' foundation and background sections.

technologies that can mimic human behavior and learn to solve tasks usually performed by human intelligence (Castillo et al., 2020). In IS research, AI capabilities are typically studied in organizational environments where AI performs human-like decision making or problem solving (Benbya et al., 2021). An essential indicator of the intelligence exhibited by an AI-based system, historically relied upon, is the Turing test, which implies that for an AI to be deemed intelligent, it must perform a given task at least as proficiently as a human counterpart (Kühl et al., 2022). The attribute intelligence usually comprises perception, reasoning, understanding, and learning to engage with the environment (see also **Table 2**), solve problems, make or suggest decisions, and even exhibit creativity (Rai et al., 2019). AI-based systems are often described as algorithms that do not operate according to rules but instead use cognitive or conversational functions similar to the human brain and interface with large amounts of data at enormous scales and volumes (Johnson et al., 2021). In this context, AI technologies cover biometrics (e.g., computer vision), robotics, machine and deep learning, recommendation systems, and natural language processing (NLP) (Lewandowski et al., 2022a).

Table 2. Exemplary Definitions of Artificial Intelligence

Literature Stream	Definition
Information Systems Rai et al. (2019, p. 1)	*“The ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity.”*
Business Kaplan and Haenlein (2019, p. 15)	*“Defined as a system’s ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation […].”*
Service (Management) Huang and Rust (2018, pp. 155-156)	*“Manifested by machines that exhibit aspects of human intelligence (HI)[.][…] We distinguish four intelligences, in the order of their developmental history in AI. They are mechanical, analytical, intuitive, and empathetic.”*

Currently, organizations are intensifying their digital transformation endeavors through the integration of AI, leveraging its potential to automate tasks, optimize processes and services, and redefine their business models (Brynjolfsson & McAfee, 2017; Jöhnk et al., 2021). Thereby, AI heralds various potentials for organizations, including increased revenues, enhanced customer interactions, and improved business efficiencies, increasingly integrated as a crucial strategic, innovative, and thus IT transformational element in organizations to achieve competitive advantage (Alsheibani et al., 2020; Uba et al., 2023).

In particular, AI has had a lasting impact on many fields and organizations (L. Wang et al., 2020), especially in areas such as service encounters (Ostrom et al., 2019). In this context, customer service and its employees are undergoing a radical transformation, spurred by the adoption of ML, NLP, and related technologies for service-related tasks (Poser et al., 2022b). AI furnishes a new perspective to service contexts, essentially *"[…] to provide value in the internal and external service environments through flexible adaptation enabled by sensing, learning, decision-making and actions"* (Bock et al., 2020, p. 317). The ongoing adoption of AI is actively reshaping the customer service landscape, leading to the era of the service encounter 2.0 (Larivière et al., 2017) and unveiling entirely new opportunities for value co-creation (Bock et al., 2020). Customer service is gradually evolving from classic dyadic interactions between customers and the service provider to complex service systems consisting of configurations of resources, including people and technologies (Maglio et al., 2009), with technology increasingly coming to the forefront of service (Larivière et al., 2017). In this context, the *"[…] service interface is gradually evolving to become technology-dominant (e.g., Intelligent Assistants acting as service interface) rather than human-driven (i.e., service employee acting as service interface)"* (Larivière et al., 2017, p. 239). As AI-based assistants become more prevalent, organizations are capitalizing on their ability to increase the availability, accessibility, and personalization of service delivery while simultaneously reducing the cognitive load and overall workload of service employees through cognitive relief and automation (Poser et al., 2022b). This progressive substitution suggests a potential shift towards AI-based systems becoming the predominant actors in service interactions in the next few years (Xiao & Kumar, 2019), thereby enhancing the efficiency and cost-effectiveness of service delivery (Poser et al., 2022b; Xu et al., 2020).

2.2 AI-Based Conversational Agents

2.2.1 Key Developments in the Research on Conversational Agents

Although the technical possibilities differ from the current potentials in AI and data processing, the idea of communicating with computers has existed for several years. The nascent research on CAs has its roots in several fields of enquiry, some of which have seen decades of research and effort in industrial applications (Følstad et al., 2021). Since the 1960s, researchers have worked on text-based and later speech-based CAs to automate procedures and assist users with various tasks (Følstad et al., 2021). An early example was ELIZA, which allowed initial natural language-based interactions with a computer (Weizenbaum, 1966), a system that generated responses to text input, imitating a psychotherapist in a therapy session. However, early attempts at CAs were limited by technical constraints such as computational power and storage capacity, as well as simplistic non-learning algorithms. Consequently, they could not meet high expectations (Diederich et al., 2019a; Gnewuch et al., 2017). According to Dale (2016) and Klopfenstein et al. (2017), ELIZA and other previously developed CAs used simple, rule-based response mechanisms or relied on pattern-matching algorithms to generate responses.

Nevertheless, technological progress in recent decades has enabled the development of more sophisticated CAs that utilize novel AI, ML, and NLP algorithms and models (Berg, 2015; Gnewuch et al., 2017; Knijnenburg & Willemsen, 2016). In this context, the CA attempts to understand the user's intention behind the input prompt to provide an adequate response output. In particular, the techniques of supervised learning, unsupervised learning, and human-in-the-loop (where humans are involved in the training process) lead to increasingly better CAs (Radziwill & Benton, 2017; Wiethof & Bittner, 2021). As a result, they have gained widespread adoption and can now better address the needs of the general public and the mass market (Maedche et al., 2019).

Beyond technological advances, the sudden commercial interest in CAs is due to changes in the way people communicate: Messaging apps and smartphones are used by billions of people in their personal and professional lives, and messaging and voice interfaces are frictionless environments that allow for asynchronous conversations (Dale, 2016). Through the widespread surge in mobile internet and messaging platform adoption, users are now more inclined to natural language interactions, presenting promising business opportunities for organizations (Brandtzaeg & Følstad, 2017). In addition, the COVID-19 pandemic has accelerated these developments. Organizations are integrating CAs into numerous front and back-end systems, such as websites, corporate wikis, knowledge systems, and enterprise instant messengers (e.g., Microsoft Teams or Slack) (Stoeckli et al., 2019). This is also fueled by the increased ease of training and implementing CAs in general due to the abundance of open source code and NLP models, widely available development platforms, and implementation opportunities through cloud and software as a service (SaaS) options (Radziwill & Benton, 2017). Nowadays, various CAs with increasing capabilities and intelligence have been developed (Brandtzaeg & Følstad, 2017), which will continue to rise in the upcoming years in service and other settings.

CAs assist in the digitalization and automation of organizations by filtering information and providing efficient support for employees in daily tasks (Zierau et al., 2020a). Hence, with their scalability and 24/7 availability (Gnewuch et al., 2017; Xu et al., 2017), CAs can have a transformative impact on business operations by acting as a central service platform and first point of contact for customers, providing a convenient way to handle service requests more individually before human intervention (Zierau et al., 2020a), and reducing information overload for users (Xu et al., 2017). Further, they can assist employees in service encounters with cognitive relief by facilitating the performance of specific tasks (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). Customers are expected to resolve issues themselves via this novel User Interface (UI) before contacting customer service employees (Castillo et al., 2020). As a result, employees can focus on more complex, creative, and non-routine tasks. Consequently, CAs are used across various domains such as marketing, sales, health, entertainment, education, and other workplace applications (Diederich et al., 2019a, 2019b; Meyer von Wolff et al., 2019a). Customer service organizations, in particular, are investing significantly in CA technology (Gnewuch et al., 2017). With the help of CAs in the context of customer service, a shift towards convenient, automated, multi-lingual, and globally available support channels is already possible (Følstad et al., 2018a; Gnewuch et al., 2017). CAs play an active role in routine tasks that service employees have conventionally performed (Gnewuch et al., 2017; Herrera et al., 2019).

2.2.2 Definition of Conversational Agents

The widespread adoption of CAs has attracted considerable research interest, with a rapidly growing number of contributions. However, CA research has a strong interdisciplinary character and is fragmented into several research streams. Multiple perspectives and disciplines, including *"[…] informatics, management and marketing, media and communication science, linguistics and philosophy, psychology and sociology, engineering, design, and human-computer interaction"* (Følstad et al., 2021, p. 2916) are used to study CAs. This interdisciplinary research has introduced numerous designations such as chatbots (e.g., Dale, 2016), conversational UIs (e.g., Herrera et al., 2019), or dialog systems (e.g., McTear, 2021). In the service literature, CAs also partially overlap with the concept of service robots (e.g., Lu et al., 2020; Wirtz et al., 2021), leading to debates in the literature about their terminology and classifications (**Table 3** provides some exemplary definitions from renowned preliminary work on CAs). In this context, Gnewuch et al. (2017), for example, have divided these AI-based IS into two subclasses of communication modes: text-based CAs and speech-based CAs. The first class comprises *text-based CAs*, commonly known under synonyms such as chatbots or chatterbots (e.g., *ELIZA* or *Cleverbot*). In contrast, the second class embraces *speech-based* CAs as virtual or smart assistants (e.g., *Amazon Alexa* or *Apple's Siri*).

In this dissertation, the term "conversational agent" (CA) represents AI-based IS that are based *"on the idea that people interact with intelligent systems using natural language, just like engaging in a conversation with another human being"* (Gnewuch et al., 2017, p. 2). Thereby, CAs simulate a conversation by using NLP/NLU technologies to process and respond to language-based user input. In this context, the CA becomes an important conversational interaction partner and more critical to value creation by serving as a visible and customer-facing interface for large and integrated service systems (Wirtz et al., 2018). Thereby, they integrate *"[…] multiple (enterprise) data sources (like databases or applications) to automate tasks or assist users in their (work) activities"* (Meyer von Wolff et al., 2019a, p. 96). CAs can have different (human-like) representations and contexts of use, whereby they can serve different purposes, such as goal-oriented task completion, informational purposes, entertainment, and social chatter (Følstad et al., 2021). For example, instead of consulting a support hotline, an employee can intuitively and directly submit a support request to a CA via natural language. The CA serves as an instantaneous assistant and social actor by scanning diverse knowledge and data sources in the background and providing answers to requests.

Table 3. Exemplary Definitions of Conversational Agents

Weizenbaum (1966, p. 36)	*"[…] is a program which makes natural language conversation with a computer possible."*
McTear et al. (2016, p. 1)	*"[…] enable people to interact with smart devices using spoken language in a natural way—just like engaging in a conversation with a person."*
Gnewuch et al. (2017, pp. 2-3)	*"CAs build on the idea that people interact with intelligent systems using natural language, just like engaging in a conversation with another human being [.][…] Text-based CAs are often referred to as chatbots or natural dialogue systems, which can be interacted with using text messages."*
Bittner et al. (2019, p. 284)	*"CAs are "computer programs that interact with humans using natural languages" and their goal is to simulate human conversation. […] a computer program that can be considered as an assistant for users."*
Janssen et al. (2020, p. 212)	*"Chatbots are conversational agents (CA) that enable users to access data and services as well as exchange information by simulating a human conversation. This conversation is conducted in form of a natural language dialogue about a common topic. The […] conversation resembles a human-to-human conversation in that the chatbot responds to the input and keeps the conversation going by analyzing single words, phrases and sentence constructions."*
Meyer von Wolff et al. (2019a, p. 96)	*"[…] is an application system that provides a natural language user interface for the human-computer-integration. It usually uses artificial intelligence and integrates multiple (enterprise) data sources (like databases or applications) to automate tasks or assist users in their (work) activities."*

Følstad et al. (2021, pp. 2-3)	*"Chatbots are conversational agents providing access to information and services through interaction in everyday language. […][This] encompasses conversational agents for goal-oriented task completion, informational purposes, entertainment, and social chatter."*
Schuetzler et al. (2021, pp. 3-4)	*"[…] provide a text-based natural language interface that gives users a way to interact with a system. […] A well-designed chatbot can save time, guide users through simple steps to accomplish a task, gather information and interact in a much more personal way than other options such as web forms, search engines or apps."*
Diederich et al. (2022, pp. 3-4)	*"CAs are based on the idea of interacting with users through natural language as in human-to-human conversations. […] These agents take on different forms that are distinguished by communication mode, embodiment, and the context in which they are used."*

2.2.3 Characteristics, Components, and Concepts of Conversational Agents

While the development of a CA has become much more accessible, the underlying system is complex by nature (Maroengsit et al., 2019). Depending on the CA model and its overall architecture, CAs exhibit various novel characteristics, components, and concepts that influence their management throughout the lifecycle. Beyond the described possibilities and applications, the design, integration, evaluation, and improvement of these AI-based systems pose new challenges to organizations that must be considered in order to prevent the failure and discontinuation of CAs (Janssen et al., 2021b; Lewandowski et al., 2021; Meyer von Wolff et al., 2021). The specificities identified in this dissertation, briefly presented in the following paragraphs, impact traditional working methods, demanding more flexible strategies for CA teams. The implementation of CA projects requires a multidimensional design and development process. Aspects such as (1) the natural language and social interface and, in this context, the intelligent and user-centered interaction and representation of CAs, (2) the technical AI/NLP components, (3) the connection to existing knowledge bases, and the integration into existing service landscapes and systems are subject to continuous scrutiny and improvement. Due to CAs (4) unfinished, limited and learning character, this requires an interactive, interdisciplinary and participatory mindset, complemented by innovative change management, monitoring, and evaluation principles (for an overview of the CA architecture and the overall design fields, see **Figure 1**). The following sections briefly present the individual aspects according to their characteristics, conceivable components, and concepts.

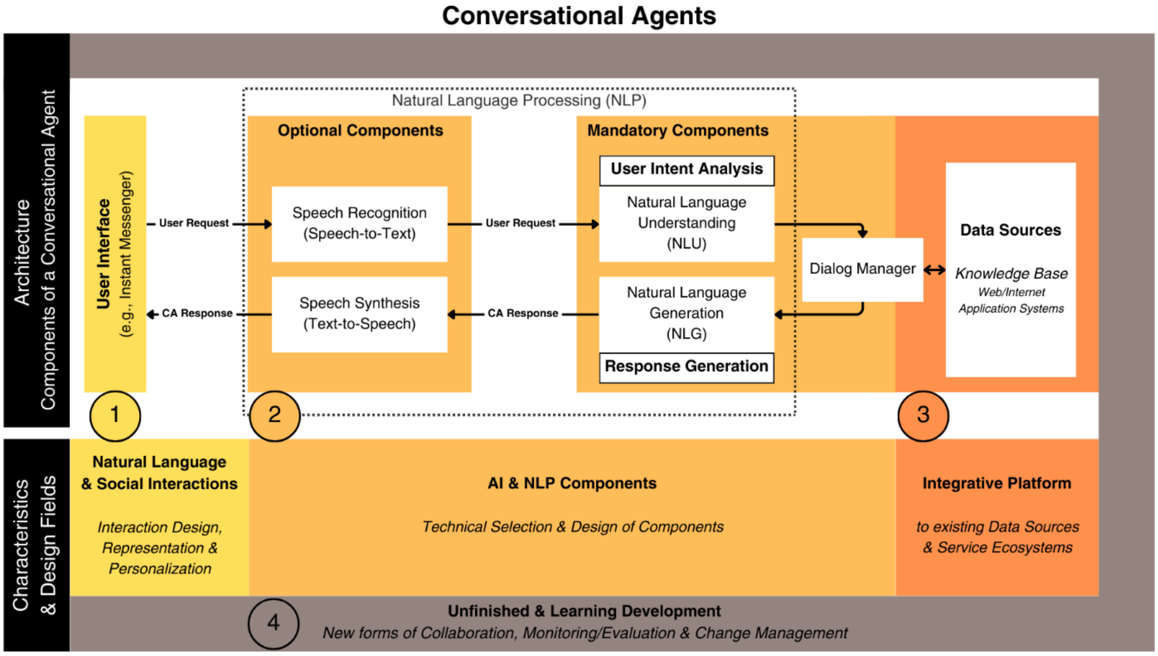

Figure 1: General Conversational Agent Architecture, Components and Design Fields (based on Adamopoulou & Moussiades, 2020; Bohus & Rudnicky, 2009; Meyer von Wolff et al., 2019a)

Natural language and social interactions (Design Field 1 in Figure 1): CAs provide a natural language-based user interface to enterprise applications and resources (see also Section 2.2.2). Employees or customers interact with the system as a new sociotechnical application class (Maedche et al., 2019). *"[…] The assistant's knowledgeability and human-like behavior, often summarized as artificial intelligence"* (Knote et al. 2019, p. 2025) bear a great potential to assist, solve, or automate tasks intuitively.

Therefore, a defining characteristic of CAs is their *sociability and often task-oriented character*. CAs are augmented by a user-centric and intelligent component that extends the IT landscape of organizations (Stoeckli et al., 2019). They emerge as social actors with increasing intelligence, autonomy, and personality. They interact with employees in various new use cases and ways, learn from collaboration, and gradually make independent decisions (Seeber et al., 2020). Users can establish relationships with CAs as "Teammates" (Bittner et al., 2019), impacting traditional service delivery by enabling new individualized and convenient sociotechnical interactions (Klaus & Zaichkowsky, 2020). This evolution requires human-like, user-centered, and socially interactive design (Lewandowski et al., 2022a). Consequently, IT teams are approaching the design of CAs differently from systems with a classical graphical user interface, e.g., endowing them with social features, names, avatars, and communicative behaviors to attract users' attention and simulate natural conversation (McTear et al., 2016). In addition to the CA representation, dialog flows and trees must also be designed, which confront CA teams with new tasks. Nevertheless, improving the user experience of CAs remains a challenge due to the lack of a comprehensive overview to determine whether they are well-designed and valuable, and due to the lack of widely applied approaches to evaluate and improve them, as outlined in the interdisciplinary chatbot agenda of Følstad et al. (2021).

AI and NLP components (Design Field 2 in Figure 1): Second, understanding and processing human language is integral to modern CAs to provide realistic and intuitive communication. Thus, NLP, natural language understanding (NLU), and natural language generation (NLG) algorithms are central to the design, development and implementation of CAs. At a high level, NLP is a subfield of AI that deals with the processing and analysis of natural language, including technologies such as text mining, speech recognition, and machine translation, to enable computer programs to understand and generate human language. NLU and NLG are both subcomponents of NLP (see **Figure 1**). This leads to an architecture that opens up new technological scope and, therefore, new design components and options for a CA team (see **Figure 1**). In the literature, CAs are described as a transformative technology because users can communicate conveniently, intuitively, and more naturally through a conversational human interface rather than a Graphical User Interface (GUI) (Dale, 2016; Diederich et al., 2020; McTear, 2020; McTear et al., 2016).

Core components of the CA architecture: As shown in **Figure 1**, the main difference between text and speech-based CAs is an optional speech recognition (speech-to-text) and text-to-speech conversion component, depending on whether the input is text or voice. After the optional conversational component processes the input ("the audio signal from the user is captured and passed through a speech recognition module"; Bohus & Rudnicky, 2009, p. 333), the different CAs work in a relatively similar way (Adamopoulou & Moussiades, 2020; Berg, 2013; Bohus & Rudnicky, 2009; Meyer von Wolff et al., 2019a): An NLU component analyzes the text input and extracts the main patterns for the dialog manager, who then identifies the user's intent and performs an action. This action could be retrieving, refining, and structuring knowledge from a knowledge base, editing a record in a database, or performing an action in an application system. Finally, the response generation creates appropriate responses and uses NLG to create a written or spoken natural language response sent back to the user interface. In cases where the user needs feedback or the dialog manager needs more information to complete the task, NLG generates a text response/another question for the user back to the conversational user interface. The dialog manager component is vital in storing and updating a conversation context, including detected intents, entities, or missing information, and asking the user for clarification.

Intents and entities as central CA constructs: Intent refers to the user's intention expressed in an utterance, such as a question or a statement, which ultimately leads to action after analysis and recognition (Kohne et al., 2020). Multiple utterances can be combined into an intent or trigger an intent. For example, the intent "Forgot password" can be triggered by diverse utterances, such as "My login from the mail program no longer works" or "I have forgotten the password for my mailbox," conveying the same meaning (Example adapted and translated from Kohne et al., 2020, pp. 44-45). In this regard, entities provide additional context to the request. For example, in the request "My login for my Microsoft Outlook no longer works," the specific entity is Microsoft Outlook as an application. Entities act as keywords that help the CA or, respectively, the NLU component understand the subject or meaning of an utterance to ensure better accuracy and, thus, a better user experience (Khan & Das, 2018).

Building the CA architecture, selecting CA frameworks (such as RASA.ai, SAP CAI, Google Dialogflow, or Microsoft LUIS), (re-)training the NLP component, and creating, clustering, and managing various intents and associated utterances pose significant challenges, especially for large-scale CAs that need to handle large numbers of user requests effectively. The development and maintenance of CAs is an ongoing process, as frameworks, intents, and entities need to be continuously maintained.

Integrative platform (Design Field 3 in Figure 1): Third, CAs represent integrative platforms, which require a new form of service design and consideration of (technical) integration in ecosystems. As Meyer von Wolff et al. (2019a, p. 96) describe, CAs integrate *"[…] multiple (enterprise) data sources (like databases or applications) to automate tasks or assist users in their (work) activities."*

The role of CAs in customer service and the need for service design: Compared to conventional service delivery in customer service settings, which consists of a dyadic interaction between a customer and a service provider (representing the face of the organization), CAs will progressively represent the predominant customer-facing part (platform) of an extensive and integrated service system (Ostrom et al., 2019; Wirtz et al., 2018). Customer service encounters represent the prevalent channel used in service-oriented business models (Gnewuch et al., 2017; Ostrom et al., 2019) to provide information, advice, and assistance between providers and customers (Janssen et al., 2021a). In doing so, customer service has a vital function in generating income and revenue, as well as representing an organization and its products by ensuring that customers are satisfied with their business (Cui et al., 2017). To evaluate the performance of a customer service provider, service quality is a crucial factor (Gronroos, 1988; Johnston, 1995), which can be defined as the result of a comparison between service expectations and what is perceived to be received (Parasuraman et al., 1985). From the provider's perspective, a significant challenge for customer service is to enhance efficiency and reduce resources without compromising service quality (Frei, 2006; Gnewuch et al., 2017). In fact, customer service is often a resource-intensive activity within an organization (Cui et al., 2017). Currently, many service requests are handled manually, resulting in a high error rate and a failing to meet user expectations due to the time-consuming process (Xu et al., 2017). From the perspective of service employees, technological advances and the growth of information are reshaping the work of knowledge workers (Semmann et al., 2018). Traditionally, customer service employees spend a significant amount of time answering questions via phone or through messaging applications, often with repetitive questions from a variety of customers (Cui et al., 2017). Pervasive challenges include a high volume and complexity of requests, and rising customer expectations for service quality (Corea et al., 2020; Hu et al., 2018).

In this context, processing and responding to incoming inquiries has become more complex and customers expect ever-faster response times (Xu et al., 2017). As a result, service employees face high-stress situations that ultimately lead to poor service quality (Semmann et al., 2018). In this context, CAs emerged and are envisioned to play a key role in customer service (Zierau et al., 2020b), promising *"[…] to create a fast, convenient, and cost-effective channel for communicating with customers"* (Gnewuch et al., 2017, p. 1). However, integrating CAs into existing service systems requires thoughtful consideration. From a service (eco)systems perspective, CAs require new service design approaches emphasizing close collaboration between domain experts and developers to align CAs with business needs. Key aspects, such as handoffs from the CA to a service agent, require different configurations and designs. For example, designing a handoff (e.g., Poser et al., 2022a; Poser et al., 2021) between a CA and an agent requires expertise in developing technical solutions and innovating business processes.

Integration into frontends and solid interaction with a knowledge base in the backend: From an architectural perspective, solid integration is needed to connect the CA to different data sources (backend), such as existing application systems and knowledge bases via Application Programming Interfaces (APIs), to enable tasks such as retrieving information or creating a support ticket, and to facilitate communication such as sending emails to responsible employees (Meyer von Wolff et al., 2021; Sousa et al., 2019). In addition, this integration should cover enterprise messengers and self-service platforms (Meyer von Wolff et al., 2020b) to ensure a seamless user experience (Pereira & Díaz, 2018). Their constant adaptability and demand for real-time interaction require more flexible service strategies. Ongoing training and monitoring become essential to ensure the quality of conversations and content (de Lacerda & Aguiar, 2019; Janssen et al., 2021b). Moreover, research examining the integration of these new social actors into existing business processes, services, actors, stakeholder structures, and enterprise workflows is scarce.

Unfinished and learning development (Design Field 4 in Figure 1): Another unique characteristic of AI-based CAs is their intelligence and ability to learn and improve via naturalistic interactions (see also **Design Field 1**). Therefore, CAs can be classified as *learning and intelligent systems,* subject to continuous development and the introduction of, so far, unsolved challenges (Lewandowski et al., 2021; Zierau et al., 2020a). Initially characterized by limited capabilities, also referred to as *unfinished IS* in related publications in this dissertation (e.g., Lewandowski et al., 2021; Lewandowski et al., 2022b), the learning progress of CAs depends on the application domain and the commitment of the actors to train these systems. Further, the learning progress of CAs is highly contextual and depends on usage (Clark et al., 2019; Zierau et al., 2020c). Initially, CAs have a small number of stored intents and knowledge (Design Field 2). Consequently, they can only handle light and simple initial tasks with low cognitive and emotional complexity (Wirtz et al., 2021), while expectations from managers, employees, and customers are high. However, CAs can continuously improve as they access more data and connect to different sources and systems in the IT and service landscape (see also Design Field 3) (Castillo et al., 2020; Xiao & Kumar, 2019). Over time, CAs profit from a scaling effect, allowing them to make more recommendations and decisions, and take actions with minimal or no human intervention (Xiao & Kumar, 2019). However, before this state can be achieved, a new understanding and engagement of all actors involved in the service is required. On the one hand, ambition is needed to participate in a continuous improvement process (Stieglitz et al., 2018), and on the other hand, customers are skeptical about the use of CAs (e.g., due to initially limited capabilities) while service employees can develop negative attitudes towards CAs (e.g., due to loss of autonomy or job insecurity). Since customers have nearly similar expectations regarding service delivery, e.g., regarding the service levels (Castillo et al., 2020), one question is how to manage CA limitations right from the start. Service failures caused by CAs could reduce service quality, resulting in a loss of customer resources and a shift from collaborative and interactive value co-creation to value co-destruction.

Despite being characterized in various articles as increasingly *proactive* (e.g., Cui et al., 2020; Janssen et al., 2020; Meyer von Wolff et al., 2019a)*, autonomous* (Premathilake et al., 2021; Xiao & Kumar, 2019; Zierau et al., 2020a), *dynamic* (Manseau, 2020; Meske et al., 2020)*, self-learning* (e.g., Meyer von Wolff et al., 2019a), and *adaptive* systems (Zierau et al., 2020a) aware of their environment utilizing contextual knowledge, CAs raise numerous questions for organizations.

Overall, as CAs are often even unfinished and in a learning state, there is a need for innovative approaches to understanding their implementation, integration, change, evaluation, and improvement throughout their lifecycle. This is imperative due to their limited functionality at the outset, which requires diverse collaborative, interdisciplinary design and management activities (Janssen et al., 2021b; Lewandowski et al., 2023b; Meyer von Wolff et al., 2021).

2.2.4 Preliminary Work on Conversational Agents and Established Research Streams

Conversational agents have emerged as an important interdisciplinary research topic in recent years, and several practical and scientific challenges and questions remain to be addressed (Følstad et al., 2021). As described in the research agenda of Følstad et al. (2021), research results on CAs have been published in various (top) conferences and journals in different scientific communities (e.g., Diederich et al., 2022; Gnewuch et al., 2023; Maedche et al., 2019).

Despite the growing interest in CAs in fields such as HCI, CS, and IS (Gnewuch et al., 2017), the research remains fragmented across different disciplines, application areas, and communities. In addition, the findings often lack integration (Følstad et al., 2021; Lu et al., 2020; Zierau et al., 2020a). Current research explores possible application areas (e.g., Laumer et al., 2019a; Meyer von Wolff et al., 2020a), CA application goals (e.g., Brandtzaeg & Følstad, 2017), and models for human-CA collaboration scenarios and democratization of CAs (Følstad et al., 2021). Furthermore, existing research on CAs has predominantly focused on technical (design) aspects, individual and behavioral issues, conceptual aspects, and social, ethical, and privacy issues, as outlined below.

More specifically, previous research has investigated **technical (design) aspects,** such as framework and platform selection (Diederich et al., 2019a, 2019b; Følstad et al., 2021), and specific perspectives, including continuous technical adaptations (e.g., NLP algorithm retraining (Meyer von Wolff et al., 2022)), knowledge base adjustments (Janssen et al., 2021b; Jonke & Volkwein, 2018), and improvement of individual CA functionalities and dialog flow based on previous mistakes identified by chatlogs (e.g., Kvale et al., 2019). These efforts aim to underpin CAs with appropriate technologies, frameworks, and algorithms (Følstad et al., 2021), and improve the NLP and understanding capabilities (e.g., Dahl, 2013) to prevent CA failures in conversations.

Scholars have also investigated user attitudes toward CAs including motivation and behavioral implications such as user trust (e.g., Brandtzaeg & Følstad, 2017; Go & Sundar, 2019; Seeger et al., 2017; Zierau et al., 2020b). In this context, CA research focuses on **individual and behavioral issues**, primarily focusing on user-level aspects, examining perceived human likeness, social support, enjoyment, and affordance theory (Lee & Choi, 2017; Stoeckli et al., 2019; Zierau et al., 2020b), or within the broader context of IS acceptance theories, such as the "Technology Adoption Model" (TAM, e.g., Pillai and Sivathanu (2020)), "Adoption Use and Impact Framework" (AUI, e.g., Bawack et al. (2019)), and "Unified Theory of Acceptance and Use of Technology" (UTAUT, e.g., Laumer et al. (2019b)). Various studies explore how and why people interact differently with nonhuman counterparts, comparing human-to-human and human-to-CA interactions, identifying psychological mechanisms (Gnewuch et al., 2023).

Moreover, prior studies have also concentrated on **CA representation and conceptual aspects**. Contributions related to the design and evaluation of CAs are beginning to emerge. According to Følstad et al. (2021), there is a rapidly growing body of work on CA interaction design (e.g., Ashktorab et al., 2019), CA personalization (e.g., Laban & Araujo, 2020; Shumanov & Johnson, 2021), use of interaction elements (e.g., Jain et al., 2018), social cues (e.g., Feine et al., 2019a; Seeger et al., 2021), and capability representation. Furthermore, researchers have explored the **social, ethical, and privacy challenges** associated with the use of CAs (e.g., Ischen et al., 2020; Ruane et al., 2019; Wambsganss et al., 2021). The broader context of AI and also CAs has received significant attention from policymakers and regulators, sparking discussions about ethics, privacy management, and trust (Chung et al., 2017; Følstad et al., 2021). Concerns about the ethics of AI, including its disruptive nature, potential impact on the labor market, and misuse by malicious actors, have led to reflections on accountability and bias (Følstad et al., 2021).

In practice, various conversational agents have emerged over time. Since 2011, speech-based CAs have gained importance as virtual or intelligent assistants and digital companions (e.g., Amazon Alexa or Apple's Siri). Termed the *"year of the chatbot"* (Dale, 2016, p. 811), in 2016 text-based CAs gained importance and were increasingly available in various contexts. Well-known public examples are IKEA's Anna, Microsoft's Tay, and Starbucks' Barista Bot (Brandtzaeg & Følstad, 2018; Diederich et al., 2022). With the introduction of enterprise messengers such as Slack or Microsoft Teams, CAs are also becoming increasingly important in internal and external organizational contexts, such as in workplace applications or customer service scenarios (Meyer von Wolff et al., 2020b; Stoeckli et al., 2019).

Despite their hype in research and practice, many organizations still fail to seize CAs' potential because they lack knowledge regarding the management of CA applications in organizational contexts (Corea et al., 2020; Meyer von Wolff et al., 2021) and studies investigating CA applications often ignore their long-term success (Corea et al., 2020; Rodríguez Cardona et al., 2019). Closely related to this, research regarding the strategic management of CAs' introduction, operation, and improvement is scarce (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). However, successfully introducing and managing CAs depends on straightforward operation and maintenance processes and diligences (Kvale et al., 2019). In this context, existing literature lacks a dedicated organizational-level perspective but instead takes a specific conceptual or technical perspective, often based on laboratory settings (Diederich et al., 2019a; Laumer et al., 2019a; Meyer von Wolff et al., 2020a). Guidance in integrating CAs into existing organizational processes, governance structures, and work routines, and understanding how their adoption differs from other AI-based and conventional IS is limited (Lewandowski et al., 2021). First authors call for research on how organizations can most effectively initiate, implement/deploy (Janssen et al., 2020; Schuetzler et al., 2021), adopt (Essaied et al., 2020), and manage (Corea et al., 2020; Meyer von Wolff et al., 2021) and maintain CAs (Kvale et al., 2019). An understanding of CAs' lifecycle management (LCM) can provide a structured, unified view of this dynamic and novel IS, identifying requirements to ensure reliable, consistent, and cost-effective handling of planned and unplanned changes based on previous issues. First authors already call for a *"[…] switch from chatbot design research to rather an organizational or management view […][, since] organizational and individual issues have the highest influence […]"* (Meyer von Wolff et al., 2021, pp. 12-13) and for *"practice-based requirements[, which] can provide insights that may not have been captured in scientific literature"* (Corea et al., 2020, p. 5827).

Furthermore, the SLRs at the beginning of this dissertation underscore the need for more design science-oriented research or entrepreneurial approaches (Peffers et al., 2007; Ries, 2011) to pilot AI-based CAs in sociotechnical environments (Briggs et al., 2019) and real-world settings throughout their lifecycle. Many CA projects fail because CAs emerge from lab settings, and the problem to be solved is imprecise or isolated from real processes and natural organizational contexts (Lewandowski et al., 2021; Lewandowski et al., 2022a). In a real-world environment, CA teams are confronted by rapid changes and high dynamics, in which it is generally impossible to predict how users will interact and what information will be retrieved long-term (Janssen et al., 2021b). CAs have gained significant research attention, encompassing specific conceptual or usability-related aspects to technical design. However, detailed theoretical and practical knowledge is lacking for the management, including questions regarding the initiation, design, integration, operation, and continuous improvement process of CAs throughout their lifecycle (Lewandowski et al., 2022b; Meyer von Wolff et al., 2021).

3 Research Design

The following chapter describes the research design to address the RQs in this thesis. Anchored in the paradigm of Design Science Research (DSR), introduced in Section 3.1, this chapter explains the principles, practices, and guidelines essential to understanding DSR and the methodological rationale of the paradigm. Next, Section 3.2 discusses the derived research strategy based on the corresponding paradigm and explains how DSR and its guidelines were operationalized in research that led to a series of publications using Hevner's (2007) three-cycle view model. In this context, the environment - the INSTANT research project - is also presented. Finally, Section 3.3 presents the applied methods used to achieve the research objectives in the operationalized paradigm and explains how they contributed to the dissertation (e.g., to ensure relevance and rigor during artifact creations and evaluations).

3.1 Research Paradigm

This dissertation applies the DSR paradigm in order to answer the RQs. *"Design science research has been practiced in Computer Science, Software Engineering and Information Systems for decades"* (Iivari, 2007, p. 39). In recent years, DSR has *"staked its rightful ground as an important and legitimate Information Systems (IS) research paradigm"* (Gregor & Hevner, 2013, p. 337), and has been widely employed in a variety of top IS journals and conferences (Baskerville et al., 2018; Prat et al., 2015). With roots in the sciences of the artificial (Nunamaker et al., 1990; Simon, 1996), DSR is concerned with the creation and evaluation of novel and innovative artifacts to address real-world problems (Simon, 1996) often in natural settings such as organizational contexts, in order to improve the (problem) environment (Hevner, 2007; Hevner et al., 2004). In doing so, DSR aims to generate prescriptive knowledge in the form of artifacts, such as software, methods, models, or concepts, with a problem-and-solution orientation, i.e., it aims to develop practical solutions to address problem spaces (vom Brocke et al., 2020). DSR represents a valuable approach to bridging the gap between practice (relevance) and theory (rigor) by providing actionable solutions and thus, research contributions to an observed and understood research problem (Peffers et al., 2007).

Generally, DSR *"seems to be more a research paradigm"* (Baskerville, 2008, p. 442) than, for example, a methodology since methods are usually thought to involve predefined processes or steps. In contrast, DSR follows boundaries but leaves room for creativity and adaptability of procedures (Baskerville, 2008). In this context, numerous research contributions have been made in recent years that have presented different and partially not entirely clear-cut principles, objectives, practices, and guidelines for DSR (e.g., Hevner, 2007; Hevner et al., 2004; Kuechler & Vaishnavi, 2012; Peffers et al., 2007). This dissertation is based on Hevner's (2004) seven DSR guidelines, presented below. Current research contributions extend these fundamental guidelines, which are presented as an extension in the following, as the instructions were used during the publications in this dissertation.

Guideline 1 - Design as an Artifact: Hevner et al. (2004, p. 83) emphasize in the DSR guidelines that the result of every DSR project is the creation of a *"viable artifact in the form of a construct, a model, a method, or an instantiation."* The resulting artifact must be effectively described to ensure its implementation and application in the problem domain (Hevner et al., 2004). The most essential and central task of a DSR endeavor is to create artifacts, where a verifiable contribution is facilitated not only by rigor in the artifact development process but also in the evaluation by documenting the design process and its outcomes (Peffers et al., 2007).

Guideline 2 - Problem Relevance: *"The objective of design-science research is to develop technology-based solutions to important and relevant business problems"* (Hevner et al., 2004, p. 83). As Hevner (2007, p. 89) stated: *"Good design science research often begins by identifying and representing opportunities and problems in an actual application environment."* Through an understanding and delineation of the problem, which includes elements such as domain, stakeholders, time and space, the act of *"positioning a DSR project in the problem space"* (vom Brocke et al., 2020, p. 7) is achieved, which serves to establish the situational context and research objectives of the DSR project. In this context, DSR is a problem-solving paradigm that continuously shifts perspectives between design/construction and evaluation activities in order to effectively develop (technology-based) solutions (called artifacts) to previously derived relevant and unsolved (real-world) problems, thereby contributing new knowledge to the body of scientific evidence (Hevner & Chatterjee, 2010). The artifacts designed are useful and fundamental in comprehending the previously derived (business) problem (Hevner & Chatterjee, 2010).

Guideline 3 - Design Evaluation: In the context of DSR, design and evaluation activities follow cycles, whereby the evaluation provides essential feedback on the artifact regarding utility, quality, and efficacy (Hevner et al., 2004). In a sense, DSR involves *"learning through the act of building"* (Kuechler & Vaishnavi, 2008, p. 489), which typically results in multiple iterations of the DSR process (vom Brocke et al., 2020) involving multiple stakeholders (e.g., researchers, developers, practitioners, and others; Morana et al., 2018). Here, a DSR process approaches the final artifact through the design phase (e.g., from simplified conceptualizations to holistic instantiations of the problem, through inherently iterative and incremental activities), which satisfies the requirements and constraints of the problem meant to be solved (Hevner et al., 2004). Therefore, a rigorous DSR project follows evaluation strategies to generate feedback (Venable et al., 2016), consisting of evaluation episodes using different evaluation methods and criteria, applied in this dissertation (see Section 3.3.4). In summary, research in DSR aims to create artifacts that can be objectified and evaluated to confirm their effectiveness and relevance.

Guideline 4 - Research Contributions: Hevner's (2004, p. 83) fourth guideline states that *"effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies."* While in Hevner's (2004) DSR approach, the constructed artifact (e.g., constructs, models, methods, or instantiations; March & Smith, 1995) represents the central contribution, other scholars have added in recent years that DSR contributions range from very novel artifacts to rigorous theory development, while noting the impact of the technology on users, organizations, and society (Baskerville et al., 2018). In conducting DSR, this cumulative thesis and its included publications follow the DSR knowledge contribution types of Gregor and Hevner (2013), which divide the contributions into three levels. These range from level 1—the implementation or instantiation of a software artifact—to level 2, which describes operational knowledge, such as models or principles, to abstract, mature knowledge at level 3, which includes, for example, design theories (Walls et al., 1992). Compared to other research paradigms, DSR represents a form of science that gains knowledge through the creation and evaluation of artifacts rather than only through empirical observation and is, therefore, suitable for dissertation work in natural, real-world contexts (see **Section 3.2.1** – INSTANT project).

Guideline 5 - Research Rigor (and Relevance): According to Hevner et al. (2004, p. 83), *"Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact."* Thereby, rigor describes how research is conducted and, in this context, the use of established methods for the individual DSR activities, such as the collection of initial research data, its analysis, the formulation of requirements for the artifact, and the construction and evaluation of the resulting artifact (Hevner et al., 2004). Further, *"[…] rigor is derived from the effective use of the knowledge base"* (Hevner et al., 2004, p. 88). Therefore, DSR projects build on extant knowledge such as scientific foundations (theories and models), experiences, and expertise to ensure the innovation of the contribution (e.g., the artifact) (vom Brocke et al., 2020). This approach safeguards against routine designs (Hevner et al., 2004). In DSR projects, clear contributions build on or reuse existing knowledge (vom Brocke et al., 2020). In addition, these projects incorporate validation checks that strengthen the credibility and substantiate the contribution (vom Brocke et al., 2020).

However, as elaborated in the forthcoming research (e.g., Baskerville et al., 2018; Hevner, 2007), DSR pursues a dual synthesis as the IT artifact is a vehicle for research and practice impact (Baskerville et al., 2018). In this context, DSR projects aim to produce research results that are relevant and applicable in practice in order to contribute directly to solving real-world problems. Therefore, DSR inputs requirements from the contextual environment into the research (Hevner, 2007). This contextual input helps define problems and requirements accurately, and establishes acceptance criteria for the resulting artifact. However, as Hevner (2007, p. 97) points out, *"practical utility alone does not define good design science research. It is the synergy between relevance and rigor and the contributions along both the relevance cycle and the rigor cycle that define good design science research."*

Guideline 6 - Design as a Search Process: *"The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment"* (Hevner et al., 2004, p. 83). Here, the design science process describes an inherent, iterative search process for an effective solution—the artifact—to address realistic/real-world IS problems (Hevner et al., 2004). Simon (1996) characterizes the design process as a generate and test cycle. Creating the artifact requires a high degree of innovation and creativity, whereby the process depends on the environment and the initial problem (Hevner et al., 2004).

Guideline 7 - Communication of Research: As Hevner et al. (2004, p. 90) stated in the DSR guidelines, *"Design-science research must be presented both to technology-oriented as well as management-oriented audiences. Technology-oriented audiences need sufficient detail to enable the described artifact to be constructed (implemented) and used within an appropriate organizational context. This enables practitioners to take advantage of the benefits offered by the artifact and it enables researchers to build a cumulative knowledge base for further extension and evaluation."*

"Communicate the problem and its importance, the artifact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences such as practicing professionals" (Peffers et al., 2007, p. 56) is a mandatory activity. DSR is fundamentally based on rigorous methods and emphasizes high transparency throughout the DSR process. The transparency of each step and decision (e.g., problem definition, requirements formulation, or

evaluation strategy) requires the most objective documentation and communication of research and results to researchers and other relevant audiences, such as practitioners (Hevner et al., 2004). DSR projects are usually longitudinal projects (Baskerville et al., 2018). Different contributions are published at different points along the research journey, and different types of established methods are used to derive intermediate results (Baskerville et al., 2018).

Next, Section 3.2 explains the specific DSR research strategy and refers to the DSR Guidelines. Afterward, Section 3.3 provides a detailed description of the methods.

3.2 Research Strategy

In IS research, the DSR paradigm is well established (Gregor & Hevner, 2013), providing a substantial body of knowledge and encompassing various approaches, methods, and frameworks to guide researchers in conducting DSR projects (Morana et al., 2018). Hevner's (2007) three-cycle view model, accompanied by seven guidelines (Hevner et al., 2004), stands out as one of the most renowned frameworks in this context. Widely applied in various studies, this framework has been recognized by the IS community in recent years for its rigor and relevance (Prat et al., 2015). For this dissertation, the three-cycle view model and the seven guidelines are consistent with the research objectives. The dissertation aims to produce artifacts that address real-world problems in natural organizational contexts and enable the design/development and continuous improvement of CAs for organizations.

As emphasized by Baskerville et al. (2018, p. 369), *"DSR projects are typically longitudinal streams of research. Varied contributions will appear at different points along the research stream. Researchers must identify the appropriate times to present and publish the research contributions in terms of the continually evolving artifacts and design theories."* In this dissertation, Hevner's (2007) three-cycle view model is operationalized as a flexible and adaptable framework, instrumental in navigating the research and organizing the longitude publication process to achieve the desired research outcomes. Utilizing Hevner's (2007) three-cycle view model contributes to the structured and effective operationalization of DSR, facilitating the synthesis of meaningful insights from the extensive research endeavors undertaken in this dissertation.

The model consists of three iterative cycles that can be repeated several times in a project and are often closely interwoven. The relevance cycle facilitates the identification of real-world problems and input requirements from the contextual environment, such as an organization which enables the creation of artifacts (Hevner, 2007) (see also Guidelines 2 and 5 in **Section 3.1**). The relevance cycle helps identify and understand the application domain, including existing structures, people, systems, challenges, and opportunities, to derive the requirements or criteria a potential solution must meet (Hevner, 2007). The rigor cycle, conversely, ensures that the research builds on existing knowledge and that the solution is differentiated from previous findings, thus making a clear contribution (Hevner, 2007) (see also Guidelines 1, 4 and 5 in **Section 3.1**). The foundation for this is the extensive knowledge base, which includes existing knowledge from other publications, such as scientific foundations, experience and expertise, other (meta) artifacts, and models that inform the research project (Hevner, 2007).

According to Hevner (2007), DSR projects have a twofold contribution: first, to expand the knowledge base and return the collected knowledge in an abstract form (goal of the rigor cycle), and second, to find a valid solution (e.g., new and innovative artifacts) for the identified application domain and thus improve the environment (goal of the relevance cycle). To ensure this, the design cycle is at the center of both cycles, representing *"the heart of any design science research project"* (Hevner, 2007, p. 90), linking the two cycles and iterating between the construction of an artifact, its evaluation and improvement, based on subsequent feedback, to create rigorous and relevant artifacts (see also Guideline 1, 3 and 6 in **Section 3.1**). The cycle generates alternatives until a satisfactory design is achieved (Simon, 1996).

These three DSR cycles are operationalized in research that led to a series of publications to address the RQs of this dissertation and to communicate the research (see also Guideline 7 in **Section 3.1**), as outlined in **Figure 2**, while the scholarly standards and established rigorous rules and guidelines of DSR are followed (e.g., Baskerville et al., 2018; Hevner et al., 2004). The real-world environment is outlined in Section 3.2.1, describing the conducted research project in which the relevant data for the publications were generated and which were impacted by the results of this dissertation. Sections 3.2.2, 3.2.3 and 3.2.4 describe the three cycles and detailed activities performed. The applied methods within the cycles are described in detail in Section 3.3.

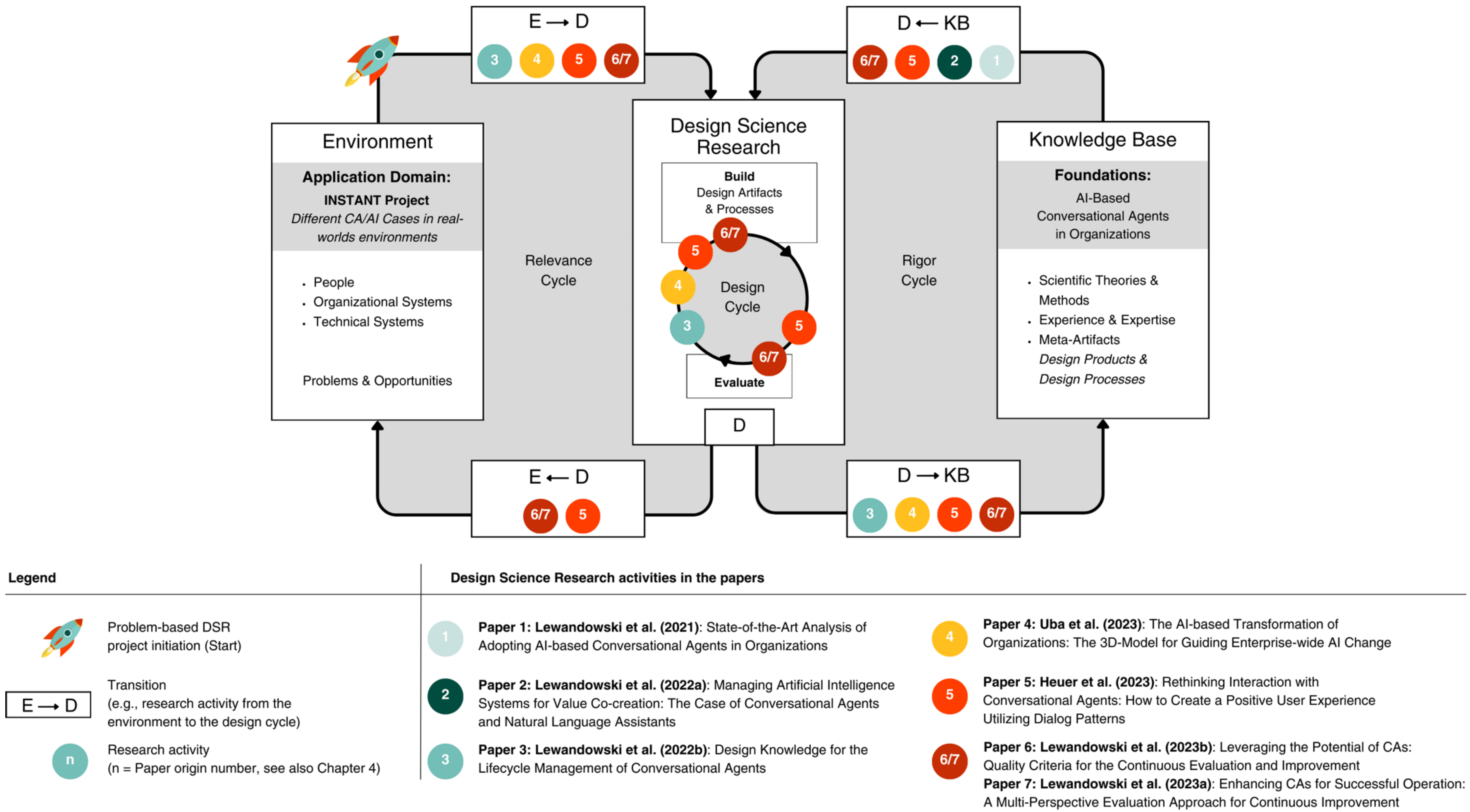

Figure 2: DSR Research Activities in this Dissertation (according to the DSR research cycles from Hevner, 2007)

3.2.1 The Environment: The INSTANT Research Project

The research project "Intelligent Collaboration of Humans and Language-Based Assistants" (INSTANT) dealt with the development, implementation, use, and improvement of CAs based on AI solutions for customer service (Böhmann et al., 2023).[3] The goal was to implement CAs, such as chatbots, in the daily work of customer service agents. These AI-based assistants facilitate the execution of specific tasks and provide decision support by retrieving, refining, structuring, and analyzing work-relevant information (Semmann et al., 2018). Accordingly, one of the research project's goals was to integrate human actors and intelligent assistants meaningfully and to provide practical design knowledge for managing AI-based IS and their long-term use in customer service organizations (Semmann et al., 2018). The German Federal Ministry of Education and Research (BMBF) funded the research project, which lasted three years. The publications included in this dissertation were part of the research project and were developed with the different project partners in the natural organizational environment.

As described by Hevner (2007, pp. 88-89), *"Design science research is motivated by the desire to improve the environment by the introduction of new and innovative artifacts,"* with the application domain that initiates the research and is later improved consisting of *"the people, organizational systems, and technical systems that interact to work toward a goal"* (see also Guidelines 1, 2 and 5 in **Section 3.1**). In this dissertation, the INSTANT project provided different application contexts, which are summarized as "the environment" in this DSR study.

In this context, the INSTANT research project was structured in seven work packages involving six partners. In addition to three departments of the University of Hamburg with different areas of research and expertise (from interaction design to technical implementation in the field of language technology), there was a fixed collaboration with three corporate partners as well as an additional, complementary collaboration in the work packages with other organizations that use AI-based CAs in different application contexts (see **Figure 3**).

The project followed a practice-oriented approach based on real-world piloting and different types of field experiments. During the project, all partners worked closely together in Hamburg, Munich and in virtual space. An iterative, agile approach with development cycles in the INSTANT laboratory and the transfer of these findings to three parallel pilot streams was a central component

[3] For more detailed information on the research project, see also: https://instant.informatik.uni-hamburg.de/ or for the 3x3 of chatbots: https://www.inf.uni-hamburg.de/de/inst/ab/itmc/material/instant

of the project in order to develop empirically validated findings (see also Guideline 3 in **Section 3.1**). Both the results from the INSTANT laboratory and, above all, the experiences from the fields of CA application were combined to form the basis for the more extensive development and piloting cycles in the corporate partners' work systems. In total, three CAs were accompanied in their lifecycle and with the different partners across companies and universities, were collectively designed, implemented, evaluated and further refined throughout the three parallel pilot streams.

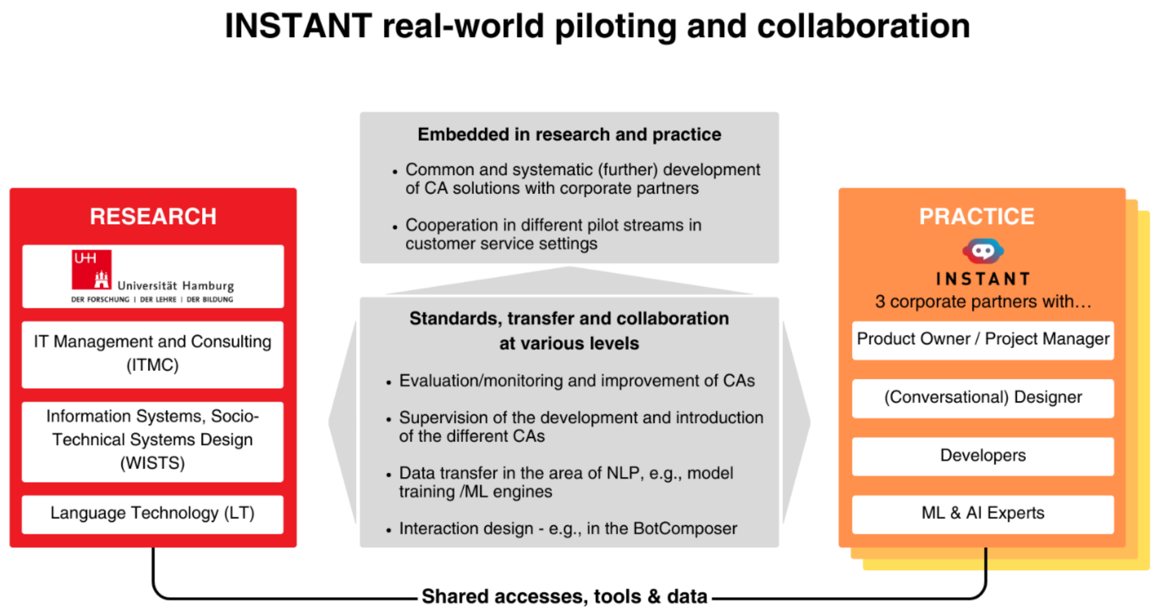

Figure 3: The INSTANT Setup in regard to Real-world Piloting and joint Collaboration

The research and corporate partners worked closely together in real-world laboratories (labs) to realize an appreciative and effective work design for AI-based chatbots in customer service. Different CA scenarios were analyzed, tested, and evaluated in common field experiments. The research project brings together an interdisciplinary team of collaboration, language technology, service researchers, and corporate partners that serve as a proving ground (Semmann et al., 2018). Specifically, we interviewed an interdisciplinary team of corporate partners consisting of users, product owners, (conversational) designers, AI experts, CA developers, and project managers with differing backgrounds and experience to gain a comprehensive view of the application environment at the beginning of the project. Later, the real-world labs allowed the analysis, testing, and evaluation of different combinations of the service triangle (see **Figure 4**) in different contexts of interaction work. Further, qualitative and quantitative survey procedures and iterative improvement processes were carried out in the different pilot streams and phases of experimentation, as a "design as a search process" with the three organizations (see also Guideline 6 in **Section 3.1**). This included the analysis of conversation samples, user feedback, and KPIs to continuously develop the CAs in a user-centered and multi-perspective way in co-creation with the corporate partners.

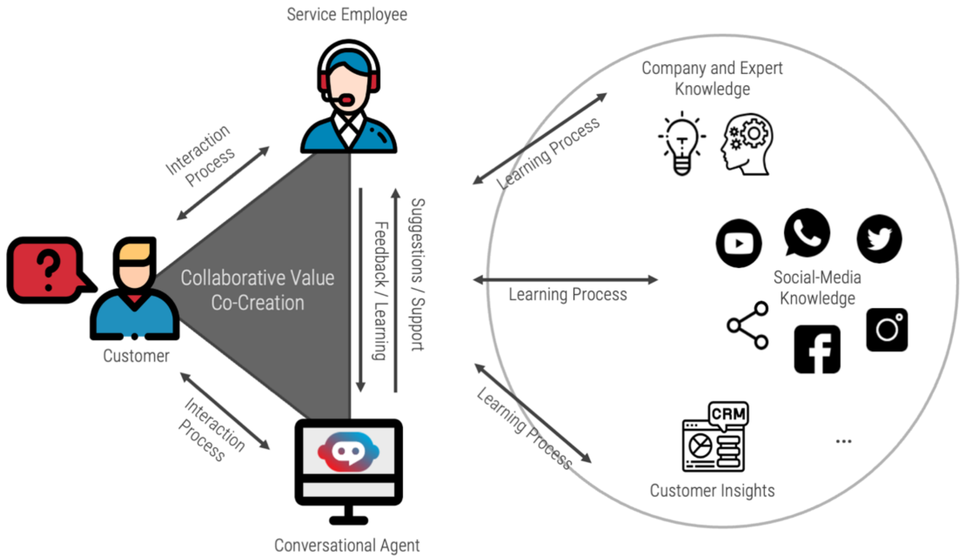

Figure 4: The Service Triangle: Joint Value (co-)creation through Customers, AI and Employees (Böhmann et al., 2023)

The practice-oriented research in a real-world environment allowed the solutions and findings to be directly integrated into the organizations' structures, such as work processes, thus achieving a sustainable and application-oriented impact. The results from the real-world labs were directly incorporated into the design of methodology for introducing and consolidating CAs in organizational service systems. Special attention was paid to the positive evolution of work processes and tasks through need-based support and relief of employees.

The research project concluded with the final result of the CA lifecycle, which included the co-created knowledge, such as lessons learned, recommended best practices, and actionable recommendations regarding the effectiveness of different chatbot applications (Böhmann et al., 2023). A "3x3" logic ("plan, do, learn") structures the research results, communicated by the end of the project (Böhmann et al., 2023) according to Guideline 7 (see **Section 3.1**). In this context, practical insights were created to help organizations successfully plan, introduce, operate, and improve AI-based CAs in order to relieve and support employees in customer service. The interrelationships and dependencies in the service triangle (see **Figure 4**) were considered so that joint value creation was possible while ensuring service quality, and meaningful and value-oriented work simultaneously.

3.2.2 Rigor Cycle

Initiated by the identified problems and opportunities in the environment, particularly in the context of the INSTANT project (see **Section 3.2.1**), several data collections from the knowledge base (KB) were conducted (shown as KB → D in Figure 2: DSR Research Activities in this Dissertation). Following Hevner's (2007, p. 90) guidance that *"the rigor cycle provides past knowledge to the research project to ensure its innovation"*, researchers must extensively explore and reference the knowledge base to confirm that the designs produced contribute meaningfully to the research (see also Guideline 4 and 5 in **Section 3.1**).

This dissertation draws on the rich knowledge base of IS with links into HCI, CS, and service research communities. Initially, insights were gained through two systematic literature reviews (SLRs) at the beginning of the DSR project (Lewandowski et al., 2021; Lewandowski et al., 2022a), aiming to establish a systematic and aggregated understanding of AI-based CAs (e.g., their novel characteristics) and their organizational challenges throughout their lifecycle (for more details, see also **3.3.1**). These activities laid the groundwork and entry point for subsequent research and design activities, as in Lewandowski et al. (2022b). Based on this starting point, further structured literature reviews were conducted in the publications by Heuer et al. (2023), Lewandowski et al. (2023b), and Lewandowski et al. (2023a), in order to create a knowledge base for the design ("to inform design activities") and evaluation (see KB → D in **Figure 2**).

Throughout this process, the rigor cycle was iteratively traversed, extracting existing knowledge multiple times and contributing new knowledge back to the academic community (additions to the knowledge base, see D → KB in **Figure 2** through publication and communication processes, see also Guideline 7 in **Section 3.1**). The systematic incorporation of existing knowledge through the rigor cycle aimed to prevent the creation of routine designs (Hevner et al., 2004). Thereby, the contributions (Heuer et al., 2023; Lewandowski et al., 2022b; Lewandowski et al., 2023a; Lewandowski et al., 2023b; Uba et al., 2023) extend the knowledge base by providing additions in the form of new design artifacts (see also Guideline 1 in **Section 3.1**). Examples include the CA lifecycle and its associated design principles (DPs) (Lewandowski et al., 2022b), the design guideline for CAs (Heuer et al., 2023), or the quality criteria set for the design, evaluation, and improvement of CAs (Lewandowski et al., 2023a).

3.2.3 Relevance Cycle

The relevance cycle initiates DSR activities within a specific application context, encompassing people, organizational systems, and technical systems (Hevner, 2007). By closely observing this context, the cycle facilitates the identification of problems and opportunities (see Guideline 2 in **Section 3.1**), serving as the foundational trigger for commencing a DSR project to improve the real-world environment (Hevner, 2007; Peffers et al., 2007; vom Brocke et al., 2020). In this dissertation, the INSTANT project provided different application contexts, consolidated in the following under the term "environment" in this DSR study (see **Section 3.2.1**). The INSTANT project and the problems it encountered posed by the novel AI-based systems identified at the outset were the starting point for the DSR project (see **Figure 2**). Subsequently, the relevance cycle was iteratively traversed and supplemented with extant knowledge from the rigor cycle to guide design activities (see also Guideline 1 and 5 in **Section 3.1**).

Publications, such as Lewandowski et al. (2022b), Uba et al. (2023), Heuer et al. (2023), Lewandowski et al. (2023b), and Lewandowski et al. (2023a) extracted knowledge (shown as E → D in Figure 2: DSR Research Activities in this Dissertation) from the environment along different phases of the CA lifecycle. These phases encompassed aspects regarding, e.g., the project initiation, design and operation, as well as the monitoring and improvement of such systems. In this context, qualitative data was systematically collected, analyzed, and interpreted throughout the publications to understand problems and opportunities to formulate requirements and DPs (for more details, see **Section 3.3.2**). For instance, Lewandowski et al. (2022b) built upon factors and issues extracted from the knowledge base through SLRs in the publications of Lewandowski et al. (2021) and Lewandowski et al. (2022a), and conducted an interview study and evaluation, as part of the rigor cycle. Thus, the knowledge of the rigor cycle (see KB → D in **Figure 2**) and relevance cycle (E → D) informed the design activity to formulate requirements and, later on, prescriptive design knowledge of the lifecycle (Lewandowski et al., 2022b).

Beyond extracting practical knowledge, new artifacts were generated and evaluated with diverse stakeholders, influencing the real-world environment (see D → E in **Figure 2**). These artifacts were not merely conceptual but concretely instantiated and applied, enhancing the research environment at the corporate partners involved in the INSTANT research project. One illustrative example is the publication Lewandowski et al. (2023a), which represents an example of several iterations through the relevance cycle. In the publication, quality criteria were instantiated and improved based on feedback from the environment from field testing, and finally incorporated into the application context. In the rigor cycle context, the publications of Heuer et al. (2023), Lewandowski et al. (2023b), and Lewandowski et al. (2023a) are concrete examples of impacting the real-world environment by concrete instantiations and artifact-based evolutions (see D → E in Figure 2). Thereby, various aspects of the configurated AI-based systems or technologies were examined and assessed to demonstrate their utility, quality, efficacy, and overall impact (Hevner et al., 2004), using the Framework for Evaluation in Design Science Research (FEDS) (see **Section 3.3.4** and Guideline 3 and 6 in **Section 3.1**).

3.2.4 Design Cycle

The *"design cycle is where the hard work of design science research is done"* (Hevner, 2007, p. 91). Influenced by the rigor and relevance cycle, the design cycle consists of activities related to the construction of an artifact, its evaluation, and subsequent feedback to further improve the design (Hevner, 2007). In this DSR core, the final artifact will be approached through iterative design and evaluation activities, progressing from simplified conceptualizations to comprehensive instantiations of the problem (see also Guideline 6 in **Section 3.1**). This progression ensures that the resulting artifact meets the specified requirements and constraints of the problem intended to be solved (Hevner et al., 2004), for example, through multiple interactions and exchanges with stakeholders from the problem domain.

In the context of this cumulative dissertation, numerous design and evaluation activities were undertaken to contribute to design knowledge within the design cycle, as illustrated in **Figure 2** (see D). The publications by Lewandowski et al. (2022b), Uba et al. (2023), Heuer et al. (2023), Lewandowski et al. (2023b), and Lewandowski et al. (2023a) involved the development of artifacts, often in multiple stages. Heuer et al. (2023), Lewandowski et al. (2023b) and Lewandowski et al. (2023a) instantiated and evaluated these artifacts in real-world scenarios, significantly impacting the environment and contributing design knowledge at various levels.

Following the classification of contribution types in the DSR of Gregor and Hevner (2013), the dissertation contributes with the publication of Lewandowski et al. (2022b) to level two by creating nascent design theory knowledge as operational and prescriptive design knowledge in the form of DPs prescribing the lifecycle of AI-based CAs. This artifact includes key activities and meta-requirements (MRs) derived from practical insights and the knowledge bases (for more information on DPs, see **Section 3.3.3**). Further, the publication of Heuer et al. (2023) contributes to level two by creating and evaluating operational CA design guidelines. Additionally, it contributes to level one by offering user-centered CA solution alternatives based on the design guidelines to address identified interaction problems in the real-world environment.

With the steps in Lewandowski et al. (2023b) and Lewandowski et al. (2023a), the dissertation further contributes to level two by creating an operational artifact in the form of a set of quality criteria, including a procedure model (design knowledge). Moreover, this dissertation contributes to level one through various artifact instantiations, applying the quality criteria in real-world contexts to enhance existing CAs. The publications' authors derived and generated prescriptive knowledge from the descriptive knowledge extracted and evaluated from the knowledge base (Drechsler & Hevner, 2018). This knowledge further served as a normative blueprint for practitioners and as a starting point for further research.

To ensure the viability of the outcomes of the DSR activities, different evaluation steps were conducted throughout this dissertation following the Framework for Evaluation in Design Science (FEDS) (Venable et al., 2016). These steps involved defining appropriate evaluation strategies, structuring evaluation episodes, and adapting to specific DSR project settings (see also **Section 3.3.4** and Guideline 3 and 6 in **Section 3.1**). For example, Lewandowski et al. (2023a) evaluated the initial literature-based quality criteria set through semi-structured interviews with experts from the real-world environment to expand the set in a second design cycle (formative ex-ante approach to evaluate the quality criteria set). Afterward, the publication's authors revised the set. They conducted a summative naturalistic ex-post evaluation of the quality criteria set by supervising its case-based instantiation in the affected IT organization to improve the CA in a structured and normative way by emphasizing its usefulness and relevance. Another example is the publication by Heuer et al. (2023), which created a prototype and a design guideline, evaluated formative and naturalistic by (1) conducting a focus group and between-subjects design to compare the existing CA with the optimized version and (2) evaluated summative ex-post the final design guideline via a structured survey. Overall, the evaluations consist of naturalistic evaluations with situated implementation and improvements of existing artifacts (CAs) in the applications, whereas more abstract knowledge is generated, such as design guidelines, quality criteria sets for the construction of AI-based solutions, or the procedure model as a blueprint for CA evaluation, which in turn were evaluated with the experts in the respective application contexts.

3.3 Research Methods

During this cumulative dissertation, different research methods were applied in the included publications to support the DSR endeavors and ensure the research rigor (Hevner et al., 2004) while answering the RQs. The applied methods were used to understand existing problems and provide a solid knowledge (data) base (e.g., through literature reviews, see **Section 3.3.1**, or interviews, see **Section 3.3.2**), to derive design knowledge in a structured way (see **Section 3.3.3**), or to support the construction and evaluation of artifacts (see **Section 3.3.4**) to meet the project solution space. The following subsections briefly introduce and describe the methods applied in this dissertation.

3.3.1 Literature Review

This cumulative dissertation consists of comprehensive, structured literature reviews (SLRs) underpinning each publication (see also **Section 3.2.2** – Rigor Cycle). In this context, SLRs are an essential foundation for initiating relevant and rigorous academic research because (1) advances in knowledge must build on existing work, and (2) in order to contribute new knowledge, scholars must know where the frontier of knowledge is positioned (Xiao & Watson, 2019). An accurate literature review can facilitate theory development, analyze, summarize, or synthesize the current state-of-the-art to target open areas of research (e.g., by creating novel IT artifacts, while justifying their novelty, vom Brocke et al., 2015) and identify areas in need of investigation (Paré et al., 2015; Rowley & Slack, 2004; Webster & Watson, 2002). When conducted appropriately, an SLR can (1) provide *"powerful information sources for researchers"* (Paré et al., 2015, p. 191) (e.g., as a stand-alone SLR) and thus a solid starting point for further research efforts, or (2) *"serve as background for an empirical study […,] commonly used as justification for decisions made in research design, provide theoretical context, or identify a gap in the literature the study intends to fill"* (Xiao & Watson, 2019, p. 94). Thus, an SLR can play a central role in preparing an empirical or DSR-oriented study to assess the validity and quality of prior work from the outset to facilitate substantial contributions (e.g., well-grounded artifacts, Hevner, 2007).

Over the past few years, IS researchers have developed several approaches aimed at improving the validity, reliability, traceability, and replicability of SLRs (e.g., Paré et al., 2015; vom Brocke et al., 2009; Webster & Watson, 2002; Xiao & Watson, 2019). These approaches provide systematic guidance on various aspects, including the overall structure, the procedural methodology, and the execution of individual steps within the SLR process. In the publications, the authors followed the well-established procedural instructions of vom Brocke et al. (2009), vom Brocke et al. (2015), and

Xiao and Watson (2019), while applying different research methodologies such as the taxonomy of Cooper (1988), the search term list according to Brink (2013), or the concept matrix (Webster & Watson, 2002).

Table 4 provides an overview of the SLRs conducted and the methodologies used in their respective sub-stages. In addition, it provides a detailed description of each SLR, highlighting the interrelationships between the different SLRs and their objectives and contributions. Thereby, this dissertation builds on both: stand-alone SLRs, e.g., to capture the current state of the art and discuss research gaps (e.g., Lewandowski et al., 2021), and process-embedded SLRs, e.g., in DSR studies serving as background SLR as part of the rigor cycle of DSR projects (e.g., Lewandowski et al., 2023a) (see also KB → D in **Figure 2**: DSR Research Activities in this Dissertation). The literature review process and inclusion and exclusion criteria are outlined in detail in the research methods sections of the publications.

Table 4. Applied SLRs in this Dissertation

Publication	Description and Goal	References
Lewandowski et al. (2021)	Stand-alone SLR in five steps following vom Brocke et al. (2009) to obtain a first structured overview of managing CAs and to chart an agenda of relevant management and adoption factors by searching the CS, IS and HCI literature.	(Brink, 2013; Cooper, 1988; vom Brocke et al., 2009; Webster & Watson, 2002)
Lewandowski et al. (2022a)	Another SLR was conducted following the methodology outlined by vom Brocke et al. (2009) within the realm of service literature journals and databases. The purpose of this SLR was to underpin the foundation of the papers as part of the rigor cycle in the context of a DSR project, including aspects such as motivation, research background, and central findings (recommendations), structured according to the multilevel framework of Grotherr et al. (2018).	(Brink, 2013; Cooper, 1988; Rowley & Slack, 2004; vom Brocke et al., 2009)
Lewandowski et al. (2022b)	Building on the preliminary SLR of Lewandowski et al. (2021), the SLR serves as a foundation for uncovering several issues from the emerging CA literature that impact the adoption and management of CAs as opposed to general AI-based and traditional IS applications.	(Brink, 2013; Cooper, 1988; vom Brocke et al., 2009; Webster & Watson, 2002)
Lewandowski et al., 2023a; Lewandowski et al., 2023b	In these papers, an SLR was conducted as part of the rigor cycle in the context of a DSR project (Hevner, 2007) to derive the initial criteria for evaluating CA quality, which formed the literary basis for the quality criteria set. The publication's	(Brink, 2013; Cooper, 1988; Rowley & Slack, 2004; vom Brocke et al., 2009; Xiao & Watson, 2019)

	authors followed the five-step process of vom Brocke et al. (2009) and applied different methodologies in the sub-steps.	
Heuer et al. (2023)	A structured literature review was conducted in accordance with vom Brocke et al. (2009) to identify practical interaction problems between CAs and users in order to extend or improve the problems previously identified through a qualitative-empirical study.	(vom Brocke et al., 2009)

3.3.2 Qualitative Data Collection and Analysis

A major strength of IS research is its diversity of research methods (Venkatesh et al., 2013; Wilde & Hess, 2006), with an increase in qualitative research involving data collection and analysis in recent years (Myers, 2019). Qualitative research is valuable because it allows researchers to understand the context in which decisions and actions occur (Myers, 2019). Context is often necessary for a meaningful explanation of human behavior, and engaging with people through qualitative methods is the most effective way to gain insight into these contexts (Myers, 2019). Typical qualitative methods in IS to obtain appropriate data include case studies (e.g., Wilde & Hess, 2007; Yin, 2003), questionnaires (e.g., Kaplan & Maxwell, 2005; Myers, 2019), interviews (e.g., Gläser & Laudel, 2009; Meuser & Nagel, 2009a; Myers, 2019), and focus groups (e.g., Morgan, 1996; Tremblay et al., 2010), as well as document and material collection (e.g., Creswell & Creswell, 2017; Mayring, 2014; Morgan, 2022), all of which were used and their results analyzed during this cumulative dissertation.

Additionally, qualitative data collection helps evaluate IS artifacts in the environment (Kaplan & Maxwell, 2005; Venable et al., 2016).

Because qualitative research involves the systematic, in-depth study of individuals in natural settings, the use of open-ended interviews to understand their experiences and perspectives on specific issues is one of the most commonly used methods (Kaplan & Maxwell, 2005). This also formed the primary basis in this dissertation for understanding the research subject in its "rich" social, cultural and environmental context (Myers, 2019, pp. 9-10) (see also **Section 3.2.3** Relevance Cycle). In doing so, *"the qualitative interview is a powerful research tool. It is an excellent means of gathering data, and has been used extensively in IS research"* (Myers & Newman, 2007, p. 23). It represents a situation where IS researchers conduct research with real people in real organizations (Myers & Newman, 2007), possessing real problems requiring a solution (e.g., IS artifacts, Peffers et al., 2007). In particular, qualitative interviews facilitate the capture of interviewees' perspectives in order to understand deeply contextualized, nuanced, and authentic accounts of participants' external and internal worlds, experiences, and interpretations of situations (Schultze & Avital, 2011).

Throughout the dissertation process, a series of qualitative interviews was conducted with appropriate individuals at differing levels of experience and across various organizations, following the instructions of Gläser and Laudel (2009), Myers and Newman (2007), Myers (2019), and Meuser and Nagel (2009a). In preparation for the different interview studies, various semi-structured interview guides were developed based on a preliminary theoretical reasoning stage according to the process of Gläser and Laudel (2009) with (1) background information, preparation and introductory questions, (2) core questions on the respective research topic, and (3) closing questions and conclusions (Myers, 2019). This (1) ensured a systematic approach and comparable data (Meuser & Nagel, 2009a) and (2) the consideration of the nascent state of the literature identified with previous SLRs to ensure rigor and relevance (Hevner, 2007). However, because the interviews were semi-structured, the authors of the publications followed a prepared but non-binding interview guide that allowed (1) room for improvisation, as there was no complete script (Myers & Newman, 2007), and (2) flexible deviations based on the interview subjects and their behavior (Myers, 2019). The interview guides were continuously updated based on the results of the previous interviews, the organizational contexts (in the INSTANT research project – see **Section 3.2.1**), and the objectives of the consecutive research studies.

Interviews were performed in the publications of Lewandowski et al. (2022b), Uba et al. (2023), Lewandowski et al. (2023a), Heuer et al. (2023), and Lewandowski et al. (2023b) to acquire a data foundation from the environment for the construction of various artifacts in the context of the relevance cycles (Hevner, 2007) (illustrated as E → D in **Figure 2**: DSR research activities in this dissertation). In addition, interviews were conducted in the publications of Lewandowski et al. (2023a), Heuer et al. (2023), and Lewandowski et al. (2023b) to evaluate the artifacts. Furthermore, the authors of the publications collected qualitative data through a series of focus groups (e.g., Heuer et al., 2023; Uba et al., 2023).

All interviews were conducted via video conferencing systems, such as Zoom or Microsoft Teams, with audio recordings made for subsequent transcription and following data analysis. For the data extraction and analysis, the thesis adhered to the established instructions of Mayring (2014), Corbin and Strauss (1990), and Rädiker and Kuckartz (2019). Qualitative content analysis was performed using *MAXQDA software* to sort, code, synthesize, summarize, and interpret the qualitative data of this dissertation (Rädiker & Kuckartz, 2019). In doing so, the data analysis was not limited to the interview data as a source, as we often coded the transcribed protocols with additional company materials (e.g., slides and internal documents from the INSTANT research project), as in Uba et al. (2023), to allow for a broad contextual analysis (Mayring, 2014). Following an intercoder reliability check, the researchers (authors of the publications) continuously compared and adjusted an initial set of codes (and category systems) to ensure the validity of the results in the distinct studies (Mayring, 2014). For example, the authors of the publications used open, axial, and selective coding to explore aspects of interest, find relationships, and ultimately identify aspects that were explicitly related to each research objective. Throughout the coding process, the authors of the publications engaged in ongoing discussions to validate and maintain the consistency of the separate codes/category system and interpretations of the material, thereby enhancing the validity of research findings. In this process, the *segment matrix in MAXQDA software*, which is very similar in structure to the Webster and Watson (2002) concept matrix, was used as a helpful framework.

3.3.3 Derivation of Design Principles

"Prescriptive research occupies an indispensable position in the repertoire of the information systems (IS) discipline" (Chandra et al., 2015, p. 4039). In this context, prescriptive (design) knowledge derived from DSR activities serves as a fundamental bridge between theory and practice, enabling IS researchers to translate theory into practical (organizational) application contexts and vice versa (Chandra et al., 2015). Consequently, it is essential to contribute beyond concrete instantiations applicable in a limited (organizational) application context toward addressing abstract problem classes in their entirety (Chandra et al., 2015; Sein et al., 2011). In this respect, DPs are an essential category of prescriptive design knowledge (Cronholm & Göbel, 2019; van Aken, 2004) prior to the design and instantiation of concrete artifact instances in DSR projects, since rigorously formulated DPs can organize the construction of IS artifacts from a higher "meta-level" and, thus, help and improve, for example, IS development, application, and management processes (Cronholm & Göbel, 2018; Gregor, 2002; Gregor et al., 2020; Möller et al., 2020). Thereby, DPs aim not only to codify knowledge from a single research project and thus preserve it in prescriptive statements but also to enable its reuse to transfer the (abstract) knowledge to other problems, scenarios, or projects that are subject to similar boundary conditions (Möller et al., 2020; Wache et al., 2022).

A DP can be characterized as a *"fundamental rule […] [derived from] extensive experience and/or empirical evidence, which provides design process guidance to increase the chance of reaching a successful solution"* (Fu et al., 2015, p. 2). In doing so, it *"capture[s] the knowledge gained about* *the process of building solutions for a given domain, and encompass[es] knowledge about creating other instances that belong to this class"* (Sein et al., 2011, p. 45), and can serve as a blueprint or architecture prescribing the design of an IS artifact (Cronholm & Göbel, 2019; Gregor & Jones, 2007). In this dissertation, the guidance of Gregor et al. (2020) and Möller et al. (2020) on deriving DPs was adopted (e.g., in Lewandowski et al., 2022b) to direct the description of abstract propositions and enable their validated design. First, the development taxonomy of Möller et al. (2020) was applied, including the first six process steps for DP development. Second, DPs were based on the formulation template of Gregor et al. (2020) to ensure a precise goal, context, and mechanism grounded in its derivation by the relationships between DP elements (Gregor et al., 2020).

3.3.4 Artefact Evaluation: FEDS

The *"evaluation of design artefacts and design theories is a key activity in Design Science Research (DSR), as it provides feedback for further development and (if done correctly) assures the rigour of the research"* (Venable et al., 2016, p. 77). Within the IS discipline, evaluation refers to the (systematic) examination and assessment of various aspects of systems or technologies to demonstrate their utility, quality, efficacy, and overall impact (Hevner et al., 2004). There are several methods and approaches for conducting effective artifact evaluations (Stockdale & Standing, 2006). To support and guide the artifact evaluation process in this dissertation, the publications' authors employed the Context-Content-Process (CCP) Evaluation Framework (Stockdale & Standing, 2006) and the FEDS Framework for Evaluation in Design Science (Venable et al., 2016) to define appropriate evaluation strategies and structure the different evaluation episodes, e.g., to match the specific DSR project settings.

On the one hand, the CCP evaluation framework by Stockdale and Standing (2006) serves as a valuable tool to comprehensively prepare evaluations by considering *what* is being evaluated, *who* is involved in the evaluation, *when* the evaluation will take place, and *how* it will unfold, while fostering a profound understanding of the underlying objectives (the "*why*") with a holistic consideration of the broader context, content, and the dynamics of both the internal and external environments (Stockdale & Standing, 2006; Symons, 1991). This framework provides a high-level structure that prepares a detailed breakdown of sub-constructs or elements for evaluation (Stockdale & Standing, 2006). Building on this, the FEDS Framework by Venable et al. (2016) was used, which proposes a detailed four-step process for selecting a holistic and structured evaluation approach within a DSR project. This process begins with predefined goals (step 1) and the context or object of the evaluation, leading to the selection of an appropriate evaluation strategy or strategies (step 2). Factors influencing this choice include, for example, the context in which the artifact will be used (purely technical vs. problem- and social-centered with a real user focus), the design risks posed by the context, and the cost of the evaluation episodes.

The next step is to decide which specific aspects of the artifact (including its features, goals, and requirements) will be evaluated and which criteria/properties to evaluate as a baseline (step 3) before the evaluation episodes are planned and designed in detail (step 4). In this context, essentially, Venable et al. (2016) distinguish between two core dimensions: the functional purpose (formative vs. summative evaluation) and the evaluation paradigm (naturalistic vs. artificial evaluation).

Throughout this cumulative dissertation, a series of evaluation episodes were conducted in the publication projects, adhering to various predefined evaluation strategies to ensure a well-structured and accurate evaluation of the artifacts developed. For example, in Lewandowski et al. (2023b), the publications' authors conducted two naturalistic evaluation episodes between the design cycles (Hevner, 2007) to validate and expand the derived CA quality criteria set. In addition, the evaluations were performed to assess whether the set of CA quality criteria could help organizations improve their CAs in a structured and normative way by emphasizing their usefulness and relevance (Hevner, 2007). To this end, a formative ex-ante approach was initially employed to evaluate the initial literature-based quality criteria set through interviews with questions about all quality criteria (Lewandowski et al., 2023a) to ask experts from an IT organization with professional experience in CA projects and external researchers about the utility, quality, efficacy and overall impact of the quality criteria set (Hevner et al., 2004). These formative evaluations provided the empirical basis for actionable improvements throughout the DSR cycles (Venable et al., 2016). Finally, a summative naturalistic ex-post evaluation of the refined quality criteria set was carried out by supervising its case-based instantiation within an IT organization using the FEDS framework. To achieve this, the publications' authors created mockups using Figma (2022) and performed A/B tests, a methodology inspired by Young (2014), to compare the novel criteria-based mockups with the current state version of an implemented CA (for more details, see Lewandowski et al., 2023a).

Another set of exemplary evaluation episodes in this work was conducted in the publication by Heuer et al. (2023). First, a formative naturalistic ex-ante evaluation (Venable et al., 2016) was conducted, in which prototypes were evaluated by a focus group (Morgan, 1996) using a between-subjects design (Charness et al., 2012) and the system usability score (Bangor et al., 2009). This evaluation involved the application of a previously established design guideline and the use of derived scenarios to compare the prototypes effectively. In addition, a summative ex-post evaluation was executed through a structured survey with pre-defined statements to rate the quality of the resulting artifact using a Likert scale (Likert, 1932).

In sum, these naturalistic evaluation episodes should serve to explore the effects and performance of the derived artifacts in their intended environment (real people, real systems, and real environments) to ensure the validity and quality of the derived knowledge outcomes regarding the effectiveness of the artifact in real use, *"as the artifact improves in quality and risks become low enough for real use by real users"* (Venable et al., 2016, p. 81).

4 Publications

The following chapter details the publications included in this cumulative dissertation.

4.1 Overview of the Included Publications

This cumulative dissertation includes seven peer-reviewed publications (see **Sections 9 to 15**) created during this dissertation and published in various reputable outlets to address the RQs in Section 1.2. The publications are divided into one journal article (Lewandowski et al., 2023a), five articles in conference proceedings (Heuer et al., 2023; Lewandowski et al., 2021; Lewandowski et al., 2022b; Lewandowski et al., 2023b; Uba et al., 2023), and one chapter in the Palgrave Handbook of Service Management (Lewandowski et al., 2022a). **Table 5** provides an overview of all included publications in chronological order by publication date. The publications have been reformatted to provide a uniform appearance in the context of this dissertation. (see **Sections 9 to 15**). All publications were published in their respective outlets at the time of submission of this dissertation.

Publication No. 1, Lewandowski et al. (2021) was invited to submit an extended version of the publication to the Pacific Asia Journal of the Association for Information Systems (PAJAIS).

Publication No. 4, Uba et al. (2023) won the Best Paper Award at the Hawaii International Conference on System Sciences (HICSS) in the Organizational Systems and Technology Track.

In addition, publication No. 6, Lewandowski et al. (2023b), was nominated as the best paper at the Hawaii International Conference on System Sciences (HICSS) and invited by the editor-in-chief to submit an extended version of the HICSS paper to the journal Electronic Markets (EM), resulting in publication No. 7 (Lewandowski et al., 2023a).

Table 5. List of Included Publications

No.	Publication	Section
1	*Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021)* **State-of-the-Art Analysis of Adopting AI-based Conversational Agents in Organizations: A Systematic Literature Review** Proceedings of the Pacific Asia Conference on Information Systems (PACIS), Dubai (UAE), A Virtual AIS Conference.	9
2	*Lewandowski, T., Grotherr, C., & Böhmann, T. (2022a)* **Managing Artificial Intelligence Systems for Value Co-creation: The Case of Conversational Agents and Natural Language Assistants** In: Edvardsson, B., and Tronvoll B. (Eds.), The Palgrave Handbook of Service Management (pp. 945-966). Springer International Publishing, Cham.	10
3	*Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022b)* **Design Knowledge for the Lifecycle Management of Conversational Agents** Proceedings of the International Conference on Wirtschaftsinformatik (WI), Nürnberg (Germany), A Virtual Conference.	11
4	*Uba, C., Lewandowski, T. & Böhmann, T. (2023)* **The AI-based Transformation of Organizations: The 3D-Model for Guiding Enterprise-wide AI Change** Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS), Hawaii (USA).	12
5	*Heuer, M., Lewandowski, T., Weglewski, J., Mayer, T., Kubicek, M., Lembke, P., Ortgiese, S., & Böhmann, T. (2023)* **Rethinking Interaction with Conversational Agents: How to Create a Positive User Experience Utilizing Dialog Patterns** Proceedings of the HCI International 2023 (HCII), Copenhagen (Denmark).	13
6	*Lewandowski, T., Poser, M., Kučević, E., Heuer, M., Hellmich, J., Rayhklin, M., Blum, S., & Böhmann, T. (2023)* **Leveraging the Potential of Conversational Agents: Quality Criteria for the Continuous Evaluation and Improvement** Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS), Hawaii (USA).	14
7	*Lewandowski, T., Kučević, E., Leible, S., Poser M., and Böhmann, T. (2023)* **Enhancing Conversational Agents for Successful Operation: A Multi-Perspective Evaluation Approach for Continuous Improvement** Journal: Electronic Markets (EM)	15

4.2 Descriptions

The following tables provide information about the included publications, such as the author's name, title of the publication, year and outlet of publication, ranking of the publication, type of publication, research objective and question of the publication, methodology used, research contribution, and co-author contributions for each publication. In cases where Tom Lewandowski, the author of this dissertation, is not the first author, his involvement in the publication and the findings derived from the publication are presented.

Section 9

Table 6. Summary of Publication No. 1: Lewandowski et al. (2021)

Citation	Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-Art Analysis of Adopting AI-based Conversational Agents in Organizations: A Systematic Literature Review. Proceedings of the Pacific Asia Conference on Information Systems (PACIS), Dubai (UAE), A Virtual AIS Conference.		
Link / DOI:	*https://aisel.aisnet.org/pacis2021/167*		
Ranking	VHB Publication Media Rating 2024: **IS Conference Proceedings: C**		
	VHB-JQ 3: **C**	WKWI: **B**	CORE2018: **A**
Type	*Conference:* Completed research paper		
Track	IS Implementation and Adoption		
Methodology	Systematic literature review (SLR)		
Research question	Which factors need to be taken into account to adopt an AI-based CA in an organization in contrast to other (AI-based) IS?		
Research contribution	Driven by the growing importance of conversational agents in organizational contexts and the concurrent lack of knowledge on how to adopt and manage this novel and innovative type of AI-based IS, the authors of this publication conducted a systematic literature review to explore the research topic. The research contributed by providing an initial and organized overview of the IS "conversational agent," highlighting its increasing levels of intelligence, personality, and autonomy, including a systemization of its specific characteristics as a social actor with a distinct natural language-based representation and features such as a learning, unfinished, and interactive character. Building on this examination, the publication aimed to provide a first structured overview of the strategic management of these systems. The paper presented organizational, technical, and environmental factors that need to be considered to initiate and sustain CA projects in organizations. Finally, it identified open questions regarding their management, governance, and design, and formulated an agenda for future research opportunities.		
Co-authors' contribution	Jasmin Delling, Christian Grotherr and Tilo Böhmann are co-authors of this publication. Jasmin Delling assisted in the SLR and in the preparation of the results. Christian Grotherr provided overall feedback and edited, structured and proofread the publication. Tilo Böhmann provided feedback on the discussion section.		

Section 10

Table 7. Summary of Publication No. 2: Lewandowski et al. (2022a)

Citation	Lewandowski, T., Grotherr, C., & Böhmann, T. (2022). Managing Artificial Intelligence Systems for Value Co-creation: The Case of Conversational Agents and Natural Language Assistants. In: Edvardsson, B., and Tronvoll B. (eds.), The Palgrave Handbook of Service Management (pp. 945-966). Springer International Publishing, Cham.		
Link / DOI:	*https://doi.org/10.1007/978-3-030-91828-6_45*		
Ranking	VHB Publication Media Rating 2024: -		
	VHB-JQ 3: -	WKWI: -	CORE2018: -
Type	*Chapter in the Handbook of Service Management*		
Methodology	Design Science Research (DSR), Systematic literature review (SLR)		
Research contribution	Organizations are increasingly using CAs as a form of AI in service encounters to support and automate activities due to their many benefits, including cost-effectiveness and scalability. Despite these benefits, organizations need assistance in realizing the full potential of CAs in real-world service settings. This research contributes by analyzing the emerging service literature on CAs and offering insights from a design science research (DSR) project focused on implementing CAs in a service setting. First, the publication identifies challenges associated with the design, implementation, and operation of CAs in service systems. Second, through the lens of a multilevel framework for service systems, the publication presents insights into how CAs can be designed, integrated into contemporary service systems, and managed for value co-creation. This multilevel perspective enhances our understanding of the design challenges of implementing such AI-based systems in service contexts. The case of CAs illustrates that service system design must facilitate learning cycles at the individual micro-level and the institutional macro-level to succeed in increasingly dynamic environments. In particular, the publication contributes key aspects for facilitating engagement with these platforms, including technical design, interaction design, and service (integration) design. Institutional design requirements are also addressed, including recommendations for data governance, privacy, and security.		
Co-authors' contribution	Christian Grotherr and Tilo Böhmann are co-authors of this publication. Christian Grotherr assisted in organizing the results of the SLR and the findings of the research project along the multi-level framework. Christian Grotherr and Tilo Böhmann provided overall feedback and edited the publication.		

Section 11

Table 8. Summary of Publication No. 3: Lewandowski et al. (2022b)

Citation	Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022). Design Knowledge for the Lifecycle Management of Conversational Agents. Proceedings of the International Conference on Wirtschaftsinformatik (WI), Nürnberg (Germany), A Virtual Conference.		
Link / DOI:	*https://aisel.aisnet.org/wi2022/ai/ai/3/*		
Ranking	VHB Publication Media Rating 2024: **IS Conference Proceedings: B**		
	VHB-JQ 3: **C**	WKWI: **A**	CORE2018: **C**
Type	*Conference:* Completed research paper		
Track	Design, Management & Impact of AI-based Systems		
Methodology	Systematic literature review (SLR), interview study (qualitative data collection and analysis), DP development		
Research question	How to manage the lifecycle of conversational agents?		
Research contribution	AI-based conversational agents (CAs) pose new organizational challenges. However, current research tends to overlook their organizational implications, leaving many questions unanswered about how to effectively manage their deployment, operation, and improvement. To address this gap, this publication contributes design knowledge that considers explicitly the organizational perspective of CAs. Therefore, this publication conducted a SLR and qualitative interview study to identify and analyze issues and challenges with CAs. Subsequently, MRs were developed, and DPs were derived. This publication adds valuable insights to the evolving field of CAs, which has predominantly focused on individual, behavioral, interactional, or technical design aspects, neglecting organizational contexts and related management issues. In doing so, the publication identifies critical phases, activities, and requirements for the effective management of CAs, emphasizing that their successful implementation relies on well-structured organizational arrangements. These arrangements include collaborative efforts, and continuous training and development approaches involving IT, business, and service professionals.		
Co-authors' contribution	Marvin Heuer, Pascal Vogel and Tilo Böhmann are co-authors of this publication. Marvin Heuer contributed to the data analysis and interpretation of the interview study (including the derivation of issues and MRs) and to the formulation of the research background and results sections (Sections 11.2.2 and 11.4). Pascal Vogel and Tilo Böhmann provided overall feedback, and edited and proofread the publication.		

Section 12

Table 9. Summary of Publication No. 4: Uba et al. (2023)

<table>
<tr><td>Citation</td><td colspan="3">Uba, C., Lewandowski, T. & Böhmann, T. (2023). The AI-based Transformation of Organizations: The 3D-Model for Guiding Enterprise-wide AI Change. Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS), Hawaii (USA).</td></tr>
<tr><td>Link / DOI:</td><td colspan="3">https://hdl.handle.net/10125/103377</td></tr>
<tr><td rowspan="2">Ranking</td><td colspan="3">VHB Publication Media Rating 2024: IS Conference Proceedings: B</td></tr>
<tr><td>VHB-JQ 3: C</td><td>WKWI: B</td><td>CORE2018: A</td></tr>
<tr><td></td><td colspan="3">Best Paper Award

</td></tr>
<tr><td>Type</td><td colspan="3">Conference: Completed research paper</td></tr>
<tr><td>Track</td><td colspan="3">Practice-Based Information Systems</td></tr>
<tr><td>Methodology</td><td colspan="3">Multi-case interview study (qualitative data collection and analysis)</td></tr>
<tr><td>Research question</td><td colspan="3">What are the key activities for driving enterprise-wide AI change and capabilities?</td></tr>
<tr><td>Research contribution</td><td colspan="3">Organizations struggle to unlock the potential of AI, as many AI projects fail in the early stages due to a lack of guidance and best practices for initiating AI transformation in organizations and driving enterprise-wide AI change. This publication contributes by shedding light on how to approach AI adoption and transformation, and the challenges organizations face. Based on insights from eleven organizations of different sizes and industries, the publication introduces four transformation types characterized by different AI transformation stages and journeys. In addition, a 3-dimensional model was developed to guide enterprise-wide AI transformation and provide concrete recommendations for each dimension. The findings help navigate, manage, and (re)evaluate AI strategies for enterprise-wide transformation.</td></tr>
<tr><td>Co-authors' contribution</td><td colspan="3">Chikaodi Uba and Tilo Böhmann are co-authors of this publication.</td></tr>
</table>

Co-authors' contribution	The following activities were performed in collaboration with Tilo Böhmann: Preparation and execution of the multi-case interview study and focus groups as part of the INSTANT research project. Preliminary analysis of the results, which served as a data foundation for the development of the publication idea. The following activities were conducted in collaboration with Chikaodi Uba: Coding, data analysis and interpretation of interviews, focus groups and materials. Chikaodi Uba developed the idea, framework and model for the publication. Tilo Böhmann assisted with the conceptualization of the publication and the derivation of the model, while Tom Lewandowski provided feedback on the model in joint sessions. Tom Lewandowski is the author of sections 12.1 Introduction and 12.2 Conceptual Background. Chikaodi Uba is the author of the remaining sections 12.3 Research Design, 12.4 Types & Selected Example Cases, 12.5 Recommendations for Action, and 12.6 Concluding Remarks. Tom Lewandowski provided overall feedback and edited the sections. In addition, Tilo Böhmann gave overall feedback, and edited and proofread the publication.

Section 13

Table 10. Summary of Publication No. 5: Heuer et al. (2023)

Citation	Heuer, M., Lewandowski, T., Weglewski, J., Mayer, T., Kubicek, M., Lembke, P., Ortgiese, S., & Böhmann, T. (2023). Rethinking Interaction with Conversational Agents: How to Create a Positive User Experience Utilizing Dialog Patterns. Proceedings of the HCI International 2023 (HCII), Copenhagen (Denmark). In: Lecture Notes in Computer Science, vol 14033, pp. 283–301. Springer, Cham.		
Link / DOI:	*https://doi.org/10.1007/978-3-031-35708-4_22*		
Ranking	VHB Publication Media Rating 2024: **IS Conference Proceedings: B**		
	VHB-JQ 3: **C**	WKWI: **B**	CORE2018: -
Type	*Conference:* Completed research paper		
Track	Design, User Experience, and Usability		
Methodology	Action Design Research (ADR), Systematic literature review (SLR), interview study (qualitative data collection and analysis), survey, evaluation		
Research question	How can a user's CA experience be improved by analyzing, understanding, and optimizing interaction problems?		
Research contribution	The goal of CA design is to facilitate long-lasting interactions with users and resolve their requests. However, in the process of analyzing and maintaining current AI-based CAs, user satisfaction is often low because the CA lacks understanding and provides unsatisfactory solutions to users. This publication contributes by examining the reasons for CA failure from a user perspective and implementing an ADR project to improve real-world CAs in the long term. In this context, the publication presents key interaction problems (findability, welcome message, dialog control, and fallback problems) based on a monitoring process and (chatlog) analysis. In addition, users were interviewed to determine their expectations and requirements for a successful CA interaction. On this basis, CA prototypes were developed and evaluated. This study contributed with a central design guideline that can be used in CA development for its design, evaluation and improvement process.		
Co-authors' contribution	Marvin Heuer, Joffrey Weglewski, Tom Mayer, Max Kubicek, Patrick Lembke, Simon Ortgiese and Tilo Böhmann are co-authors of this publication. The following activities were performed in collaboration with Joffrey Weglewski, Tom Mayer, Max Kubicek, Patrick Lembke and Simon Ortgiese: Preparation and execution of the SLR, interview study and evaluation as part of the INSTANT research project. Preliminary analysis of the results and development of the idea for the publication. The following activities were realized in cooperation with Marvin Heuer: Final coding, data analysis and interpretation of interviews, focus groups and materials. Tom Lewandowski is the author of sections 13.1 Introduction and 13.2 Conceptual Background. Marvin Heuer provided overall feedback and edited the sections. Marvin Heuer is the author of the remaining sections, 13.3 Methodology, 13.4 Results, 13.5 Discussion and 13.6 Conclusion. Tom Lewandowski provided feedback and edited the sections. Tilo Böhmann gave valuable feedback on the discussion.		

Section 14

Table 11. Summary of Publication No. 6: Lewandowski et al. (2023b)

Citation	Lewandowski, T., Poser, M., Kučević, E., Heuer, M., Hellmich, J., Rayhklin, M., Blum, S., & Böhmann, T. (2023). Leveraging the Potential of Conversational Agents: Quality Criteria for the Continuous Evaluation and Improvement. Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS), Hawaii (USA).			
Link / DOI:	*https://hdl.handle.net/10125/103055*			
Ranking	VHB Publication Media Rating 2024: **IS Conference Proceedings: B**			
	VHB-JQ 3: **C**	WKWI: **B**	CORE2018: **A**	**Best Paper Nominee**
Type	*Conference:* Completed research paper			
Track	Artificial Intelligence-based Assistants			
Methodology	DSR, Systematic literature review (SLR), interview study and focus groups (qualitative data collection and analysis), evaluation (prototypes/mockups)			
Research question	What are relevant criteria for continuously evaluating the quality of CAs, and how can they be applied?			
Research contribution	Despite the hype surrounding CAs in research and practice, organizations fail to sustain these communication tools in their operations. This struggle arises from a lack of knowledge on how to effectively evaluate and enhance the quality of CAs throughout their lifecycle. This publication contributes by conducting a multi-step design science research (DSR) project that aggregates insights from the literature, supplemented by real-world experience, to derive a systematized and synthesized set of CA quality criteria. First, the publication contributes the set of criteria that serves organizations as an overview of relevant aspects to evaluate and improve the quality of CAs as part of their operations. In combination with the application of the prototype method, the instantiation of the criteria set can pave the way to systematically evaluate and improve CAs by comparing different versions. Second, in this context, the paper presents an initial procedure model related to the instantiation of the criteria set, which serves as a blueprint for applying the criteria set. The procedure model allows for structuring the evaluation of CAs and discovering areas for systematic improvement.			
Co-authors' contribution	Mathis Poser, Emir Kučević, Marvin Heuer, Jannis Hellmich, Michael Raykhlin, Stefan Blum, and Tilo Böhmann are co-authors of this publication. The following activities were performed in collaboration with Mathis Poser, Jannis Hellmich, Michael Raykhlin, and Stefan Blum: Preparation and execution of the SLR, interview study and evaluation as part of a common research project. Preliminary analysis of the results and development of the idea for the publication. The following activities were performed in collaboration with Emir Kučević: Emir Kučević assisted in the data analysis and interpretation of the DSR study (including the conceptualization of the quality criteria set and the procedure model) and in the formulation of section 14.5 Case-Based Instantiation. Further, Mathis Poser, Emir Kučević, Marvin Heuer, and Tilo Böhmann provided overall feedback and edited the publication.			

Section 15

Table 12. Summary of Publication No. 7: Lewandowski et al. (2023a)

Citation	Lewandowski, T., Kučević, E., Leible, S., Poser M., and Böhmann, T. (2023). Enhancing Conversational Agents for Successful Operation: A Multi-Perspective Evaluation Approach for Continuous Improvement. Electronic Markets 33, 39.		
Link / DOI:	*https://doi.org/10.1007/s12525-023-00662-3*		
Ranking	VHB Publication Media Rating 2024: **IS Journals: B**		
	VHB-JQ 3: **B**	CORE2020: **A**	WKWI: **A**
	Impact Factor (Clarivate Analytics Master Journal List 2022)[4]: **8.5**		
Type	*Journal Article:* Research Paper		
Issue	AI-based Assistants and Platforms		
Methodology	DSR, Systematic literature review (SLR), interview study, evaluation, focus groups		
Research question	What are relevant criteria for continuously evaluating the quality of CAs and how can they be applied?		
Research contribution	Many enterprises fail to realize the full potential of CAs because they lack knowledge of how to evaluate and improve the quality of CAs to sustain them in organizational operations. This publication aims to fill this knowledge gap by building on the design science research project and data from Publication No. 6: Lewandowski et al. (2023b), aggregating the insights from the literature and practice, and extending and re-evaluating both contributed artifacts. This publication aims to systematize the continuous evaluation and improvement of CAs to counteract failure in organizational environments. To operate a CA successfully, measurements or criteria for orientation are needed to adapt CAs to user needs. Consequently, this paper makes a twofold contribution: (1) presenting a set of pertinent criteria for evaluating CA quality and (2) introducing an extended procedure model as an integral part of instantiating the quality criteria set within an IT organization, and prescribing its application and evaluation activities. The criteria set and the procedure model define a cyclical criteria-based evaluation process that can be triggered by different impulses and provide an actionable approach for practice.		
Co-authors' contribution	Emir Kučević, Stephan Leible, Mathis Poser, and Tilo Böhmann are co-authors of this publication. Emir Kučević contributed to the formulation of Section 15.5 Case-Based Instantiation and Section 15.6 Discussion. Stephan Leible contributed to the formulation of sections 15.3 Research approach, 15.4 Quality Criteria Set and 15.6 Discussion. In addition, both authors provided feedback on all sections of the journal publication and assisted with editing and proofreading. Mathis Poser and Tilo Böhmann provided feedback on the discussion section.		

[4] More detailed information on the journal ranking can be found at: *https://www.electronicmarkets.org/about-em/ranking/*

5 Theoretical Contributions

5.1 Overview

Chapter 5 presents the theoretical contributions of this dissertation, which are the result of the conducted DSR project and its research activities (see **Figure 2**). The presentation of the contributions is based on **Figure 5**. Following a CA-specific perspective and thus the establishment of the CA research stream, Section 5.2 presents the conceptualization of the terminology of CAs and the interrelation of their concepts and characteristics. The resulting potentials and complexities of CAs are then described in Section 5.3, laying the groundwork for understanding challenges and activities in organizational CA management, aggregated in Section 5.4. Section 5.5 contributes a research agenda for CA management in organizations, reviewing existing knowledge and highlighting future research directions. Section 5.6 introduces the CA lifecycle, providing a system-wide and phased-based view of the technology. Sections 5.7 and 5.8 contribute artifacts for the operation phase to evaluate and improve CAs continuously. Finally, Section 5.9 expands to an organization-wide perspective, discussing and guiding the AI transformation and taking a broader view of the topics that also impact CA management.

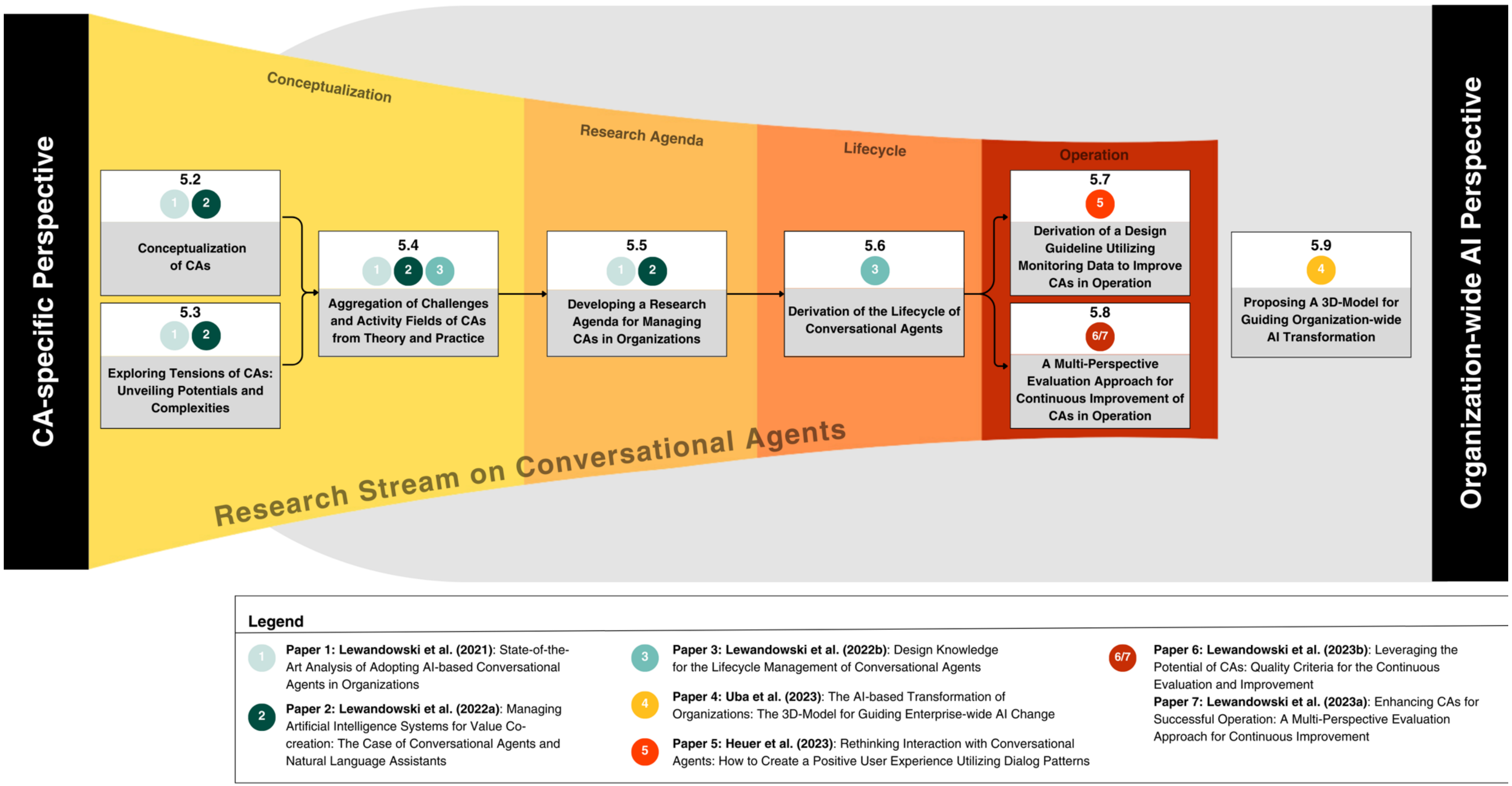

Figure 5: Overview of the Research Contributions along the DSR Research Activities

5.2 Conceptualization of Conversational Agents

5.2.1 Conceptualization of Conversational Agents to Understand their Resulting Design and Management Challenges

The publications by Lewandowski et al. (2021) and Lewandowski et al. (2022a) offer a first comprehensive overview of CAs in organizational settings, providing a conceptualization based on two separate SLRs. As outlined, CAs require new management approaches throughout their lifecycle due to their specificities (see **Section 2.2.3** for a detailed elaboration of their conceptualization). Depending on the CA model and its overall architecture, CAs exhibit various novel characteristics, components, and concepts—identified and organized in this thesis—resulting in various design fields (see also **Figure 1**). Despite the growing interest in CAs in research domains such as HCI, CS, IS, and Service Science, research in this area remains fragmented into different streams, with results often being separated (Lewandowski et al., 2021). The strong interdisciplinary nature of CA research has led to debates in the literature about the terminology and classifications of CAs (Følstad et al., 2021).

Moreover, despite the significant attention CAs have received in research and practice, many organizations struggle to realize the full potential of these communication tools due to a limited understanding of the technology and its characteristics, leading to challenges in design and management in particular. In recent years, several organizations have taken their CAs offline due to a lack of CA knowledge, resulting in an uncoordinated, dynamic, and highly exploratory development process (Janssen et al., 2021b). Structuring their design fields along a clear conceptualization can enable organizations and CA teams to reflect and design CAs in a more multidimensional way, rather than focusing on isolated aspects.

5.2.2 Concepts and Characteristics of Conversational Agents: An Aggregated Overview

Motivated by these research deficiencies, the studies conducted in this dissertation (Lewandowski et al., 2021; Lewandowski et al., 2022a) contribute to a fundamental understanding and comprehensive perspective of the technology by providing an aggregated overview of the analyzed and previously segmented literature with a systematized conceptualization of the term "conversational agent". For a more detailed presentation, see **Section 2.2.3**. The following is a brief overview of CA design fields, including the central concepts and characteristics.

CAs are characterized as **natural language, social, and user-centric systems that require (interaction) design and evaluation (Design Field 1 in Figure 1).** As stated in Lewandowski et al. (2023a, p. 4), a *"unique characteristic of CAs is their **sociability**. As social IS, they are capable of interacting with users via natural language, representing a new sociotechnical application class. These AI-based systems impact traditional service delivery and enable new individualized and convenient sociotechnical interactions, requiring humanlike, user-centered, and socially interactive IS design."*

Further, CAs are a class of **AI- or, specifically, NLP-based systems that interact with users through natural language in a dialog-oriented manner (Design Field 2 in Figure 1)**, leading to a distinct technical architecture and components. Understanding and processing human language is an integral part of modern CAs in order to provide realistic and intuitive communication. Therefore, technologies such as NLP, NLU, and NLG are crucial for CAs and require new expertise, e.g., in technology selection and development. In general, CAs require a new way of understanding because they need to be designed, trained, evaluated, and improved differently depending on ongoing (software) development.

In addition, CAs represent integrative platforms, requiring integration into several front and backends, **which require a new form of service design and consideration of (technical) integration into ecosystems (Design Field 3 in Figure 1)**. CAs integrate different knowledge bases as scalable, cost-effective and integrative platforms for organizations to support employees by retrieving, structuring, and analyzing information in work processes and tasks (see **Figure 1**). In doing so, they gain access to increasing amounts of data and are connected to different sources and systems in the IT and service landscape (Castillo et al., 2020; Xiao & Kumar, 2019).

Finally, CAs have a new level of intelligence and the ability to learn and improve through naturalistic interactions. *"As such, they can be classified as a form of **learning and intelligent IS**"* (Lewandowski et al., 2023a, p. 4). This results in **Design Field 4 in Figure 1**, which conceptualizes CAs as unfinished and learning systems that require new approaches to (change) management and closely-connected collaboration. CAs often start with limited capabilities and their learning progress depends on the application domain and the commitment of the actors to train these systems. In this context, the lifecycle of CAs is highly human-dependent and requires joint continuous development, training, evaluation, and monitoring activities between IT departments and affected business units. Accordingly, the learning progress of CAs is highly contextual and thus dependent on the actual application. This endeavor is complicated by rapid changes and high dynamics, where it is generally impossible to predict how users will interact and what information will be retrieved in the long-term (Janssen et al., 2021b).

5.3 Exploring Tensions of Conversational Agents: Unveiling the Potentials and Complexities

The foundational synthesis of the two SLRs' research findings highlights the growing importance of CAs in organizational contexts and academic research (Lewandowski et al., 2021; Lewandowski et al., 2022a). As extensively detailed in this dissertation and its included publications, CAs are recognized for their diverse application scenarios and (often perceived as disruptive) potentials. However, despite their potentials, these publications have also revealed that several CAs fail early in real-world contexts, where the dissertation has contributed to an understanding of the inherent tension between their potentials and the complexities surrounding their implementation. Organizations often have insufficient knowledge, false and often exaggerated expectations, or lack of acceptance, (employee/user) readiness, and skills when using CAs. To address this issue, this dissertation examines the reasons for failure, **aggregating potentials and applications**, but also **complexities and challenges**, to provide a baseline understanding for further investigations, as documented in the publications of Lewandowski et al. (2021), Lewandowski et al. (2022a), and (Lewandowski et al., 2022b), which systematically collect and organize these findings.

5.3.1 The Potentials of Conversational Agents: Applications and Opportunities

Currently, CAs are particularly used in interactive domains such as customer service and support, marketing, sales and entertainment, teaching and education, and in various workplace **applications** (Diederich et al., 2019a, 2019b; Meyer von Wolff et al., 2019a). Apart from professional work contexts, they have also gained widespread popularity in various private application domains (Meyer von Wolff et al., 2020a). Being integrated into various front and back-end applications, such as websites or messaging applications (e.g., MS Teams), CAs support the ongoing digitization and transformation of organizations by, for example, filtering or providing information or efficiently supporting employees in their daily work tasks (Zierau et al., 2020a). Further drivers for the proliferation of CAs include the intuitive communication channel for users and the low technical burden for organizations to deploy a CA (Diederich et al., 2019b; Riikkinen et al., 2018; Xu et al., 2017).

Regarding their **potentials**, CAs promote a new form of flexibility, quality, speed, and personalization of the customer relationship (Lewandowski et al., 2021). In addition, CAs are portrayed as a highly scalable and cost-effective solution that saves money by replacing manual tasks performed by previously required employees (Oracle, 2016; Wirtz et al., 2021; Wirtz et al., 2018). A transformation to an innovative, convenient, automated, self-learning, 24/7 communication channel available to employees or customers worldwide and in multiple languages is feasible (Brandtzaeg & Følstad, 2017; Følstad et al., 2018a; Gnewuch et al., 2017). In this context, CAs also offer shorter resolution times and high availability (Waizenegger et al., 2020). Consequently, CAs are expected to deliver significant economic value in existing and future applications (Seeger et al., 2021; Seiffer et al., 2021).

Particularly in the context of customer service, CAs are a popular research topic (Feine et al., 2019a; Zierau et al., 2020a) for their ability to enhance service efficiency, experience, and quality (Zierau et al., 2020b). Unlike traditional service provision characterized by dyadic interactions between a customer and a service provider (representing the "face" of the organization), CAs will progressively represent the prevailing customer-facing part of an extensive and integrated service system (Ostrom et al., 2019; Wirtz et al., 2018). These agents will shift service encounters from human-centric to technology-dominated (Castillo et al., 2020), influencing conventional service offerings and facilitating personalized interaction channels for customers (Klaus & Zaichkowsky, 2020; Zierau et al., 2020b), making them highly valuable to organizations.

From the perspective of service employees, CAs have the potential to support, augment, or automate human-centered tasks by providing solution strategies, decision support, and problem-solving capabilities (e.g., by retrieving and structuring information). Pervasive challenges include a high volume and complexity of requests and increasing customer expectations for service quality (Corea et al., 2020; Hu et al., 2018). As a result, service employees are confronted with high-stress situations that ultimately reduce service quality (Semmann et al., 2018). In this regard, CAs are promising for reducing the occupational stress of service and interaction with employees amidst increasing demands and information overload (Semmann et al., 2018). From the customer's perspective, CAs also appear as novel actors in the foreground of various customer support settings, promoting a new form of speed and personalization of customer relationships. Consequently, CAs can appear as new service actors between providers and customers, enabling novel value co-creation scenarios.

5.3.2 Complexities in the Implementation of Conversational Agents: Understanding the Tension between Potentials and their Challenges

Despite the growing interest in the massive potentials of CAs in organizational contexts—as evidenced by new research studies—many CAs fail to meet expectations. Introducing CAs in organizational environments does not always have a positive impact as the technology is still error-prone and fails in interactions (Gnewuch et al., 2017; Riquel et al., 2021). As outlined with the systematic conceptualization of CA in Section 5.2, CAs are complex systems with numerous new characteristics and components that require multi-perspective design approaches, which in turn lead to challenges and new activities for organizations and their CA teams (see **Section 5.4**).

Lewandowski et al. (2021) and Lewandowski et al. (2022a) have revealed and explored this field of tension between the potentials of CAs and their high complexity, which has been the fundamental motivation for further research. One reason for the limited success of CAs is the tension between the high expectations and tremendous interest resulting from the numerous potential options described in Section 5.3.1, which often leads to premature deployment driven by unrealistic expectations and management pressure, compounded by a lack of understanding of the CA development process and quality standards. This hasty deployment and insufficient knowledge often leads to non-use, dissent, or complete failure, as highlighted by, for example, Janssen et al. (2021b) and Lewandowski et al. (2022b). Unsatisfactory CA design and limited capabilities can result in a frustrating user experience that triggers resistance and a loss of trust in the CA, further hindering its successful adoption in real-world organizational environments (Weiler et al., 2022). Attributable to inadequate CAs, employees have developed negative feelings towards CAs and their providers in recent years (Diederich et al., 2020; Feine et al., 2019b; Schuetzler et al., 2021).

The failure of CAs is not only disappointing for employees but also for vendors who invest significant resources in their development (Janssen et al., 2021b; van der Goot et al., 2021). However, in this context, there remains a gap in a detailed understanding of the challenges and reasons for the failure of CAs in practice (Janssen et al., 2021b), addressed in the following in this dissertation.

5.4 Aggregation of Challenges and Activity Fields of Conversational Agents from Theory and Practice

Motivated by (1) the high failure rate of CAs (see **Section 5.3.2**), (2) the lack of a comprehensive understanding of the deeper reasons for failure, and (3) the general lack of knowledge about the management of CA technology, including its influencing challenges, required activity fields and organizational implications, this dissertation has pursued an integration of insights from theory and practice through a multidisciplinary and multistep approach. Starting with a systematization of the CA technology and its characteristics (as outlined in **Section 5.2**) and the elaboration of the problem space and the areas of tension (see **Section 5.3**), the publications Lewandowski et al. (2021), Lewandowski et al. (2022a), and Lewandowski et al. (2022b) further contributed with a first aggregation of challenges and activity fields (see **Table 13** for an overview).

5.4.1 Challenges of Adopting Conversational Agents: A Strategic and Factor-Based Perspective

In this context, the study by Lewandowski et al. (2021) revealed that the current body of research knowledge is highly fragmented, with findings scattered across different disciplines. Existing research often neglects the long-term adoption and success of CA applications, focusing instead on individual applications, designs, and technical features rather than taking an organizational or management perspective that could benefit different stakeholders within organizations. Current research explores possible applications (e.g., Laumer et al., 2019a; Meyer von Wolff et al., 2020a), objectives of CA applications (e.g., Brandtzaeg & Følstad, 2017), the challenges in using them (e.g., Gnewuch et al., 2017), how to design CAs for higher user-acceptance (e.g., Bittner et al., 2019; Gnewuch et al., 2018), or how to improve the NLP capabilities (e.g., Dahl, 2013). Conversely, it is unclear what the requirements are for using CAs in organizations and what perspectives must be considered in their implementation and management. Similarly, there are no frameworks to guide the organizational adoption, integration into processes, and ongoing development of CAs (Corea et al., 2020; Essaied et al., 2020).

To address this research gap, Lewandowski et al. (2021) conducted an SLR in the context of the IS, HCI, and CS literature domains to examine and aggregate the state-of-the-art of CA adoption in organizational contexts and the associated management challenges. This work contributes to the study of CAs by providing researchers and practitioners with a first and structured overview of CA management, offering an integrated perspective on the technological, organizational, and environmental factors critical to initiating and sustaining CA projects in organizations. Building on this examination, the publication aims to provide organizations with insights to mitigate CA failures. The publication discusses these systems from diverse perspectives, supporting strategic preparation.

From a theoretical perspective, this study organizes its identified findings using a systematized set of factors along the Technology-Organization-Environment (TOE) dimensions proposed by Depietro et al. (1990). The TOE framework serves as an organization-level theory for framing the constraints and opportunities that influence the adoption of IT innovations within enterprises. TOE has been widely applied in other technology domains, such as cloud computing and knowledge management systems (Pumplun et al., 2019). In Lewandowski et al. (2021), the framework is applied to provide insights into technical and organizational factors that influence the adoption of CAs, such as changes in the communication structure of employees, their skills and forms of collaboration, or the topic of data management activities. The framework also considers environmental factors, such as regulation or constraining technology adoption, such as privacy-related factors in the context of CAs. The choice of the TOE as a conceptual framework allows for the coding and structuring of the identified contributions along its dimensions. In addition, by identifying and aggregating foundational factors, this contribution highlights contexts where conventional knowledge about IS adoption may not apply. Furthermore, the paper extends the TOE framework with new factors that take into account the unique characteristics, components, and resulting design/management fields of CAs (see **Section 5.2**).

In this regard, Lewandowski et al. (2021) present an **organizational-strategic perspective** and discuss (1) fundamental challenges in forms of factors, such as the need for a long-term strategy, vision, and roadmap, (2) **technical factors**, such as the need for data management to ensure the availability and quality of (dialog-capable) training data and the basis for dialog design, and (3) **environmental factors**, such as privacy and ethical concerns. In addition, (4) **organizational factors** are discussed. The publication emphasizes that the introduction of CAs is *"much more than a classic software introduction"* (Lewandowski et al., 2021, p. 11). CAs bring additional hurdles that cannot be solved in a purely top-down, large-scale, project-oriented manner. In recent years, CAs have gained negative attitudes among employees due to limited language understanding and skill levels. In addition, employees have ethical and privacy concerns, including heightened perceptions of privacy risks, leading to non-use of CAs. In the early stages of adoption, concerns often outweigh the benefits of CAs, creating a barrier as CAs can only be improved with extensive and ongoing training.

While traditional IS adoption is influenced by factors such as top management support or resource allocation (Depietro et al., 1990), Lewandowski et al. (2021) emphasize the need to shift the focus for CA adoption to a broader perspective that is more people-centric, data-centric, and ecosystem-integrated. Successful adoption of CAs requires a great deal of persistence and active, long-term agreements that involve collaborative and continuous development approaches between IT departments and affected business units. Employees must understand that the CA will be a limited system for an extended period of time. The success of the adoption depends on the cooperation and acceptance of the employees' role as continuous "knowledge integrators" in various activities.

5.4.2 Challenges and Activity Fields of Conversational Agents in Service Environments: A Multi-level Framework Perspective for Service Systems Design

Lewandowski et al. (2022a) conducted a complementary SLR and analysis of the emerging literature in the service domain and provided insights from a DSR project on the implementation of CAs in a service setting to identify challenges in the design, implementation, and operation of CAs in service systems (see also INSTANT project, **Section 3.2.1**). The findings of Lewandowski et al. (2022a) are structured using Grotherr et al.'s (2018) multi-level framework for service system design as conceptual frame. This framework builds on the tiered understanding of value co-creation and actor engagement posited by Storbacka et al. (2016), connecting micro-level engagement activities to macro-level phenomena such as value co-creation and the corresponding institutional arrangements. Grotherr et al.'s (2018) framework allowed for a comprehensive examination of technology design, work/service design, and institutional design in the context of CAs. Thereby, Lewandowski et al. (2022a) contribute to the field of service management and AI by providing a structured approach for integrating CAs into service systems, and harnessing the potential of AI for service innovation and value co-creation, by revealing activity fields.

Through the lens of the multi-level service systems framework, this paper presents insights into how CAs can be designed and managed for value co-creation. The paper aims to broaden the perspective of CA design beyond the currently dominant technological perspective by applying a service systems perspective. The conducted research contributes to the understanding of novel interaction processes between customers, employees, and conversational agents. It highlights practical challenges in designing and deploying these AI-based systems to support service operations effectively. The findings are based on the implementation of CAs in real-world service environments, following the findings of the INSTANT project. Thereby, the engagement platform exemplifies the CA, and the required activities are applied according to the *Engagement* and *Institutional Design* of the multi-level framework.

Engagement Design captures the ongoing design, development and enhancement of socio-technical components such as engagement platforms. This includes the activities necessary to enable the evolution of the CA from its initial value proposition. The publication presents and discusses challenges and activities in the context of technical, interaction, and service design (see **Table 13** for a broad overview of the challenges and activity fields). For example, *Interaction Design* is essential for CAs as they emerge as new front-line social actors, exhibiting humanlike characteristics, interacting with customers and solving problems. Many customers will develop relationships with CAs as they take on more and more tasks as their capabilities evolve. For example, a customer conversation (such as a complaint) is characterized by small talk between the customer and the service agent, where the agent can respond to the customer's feelings (such as frustration). As the conversation with the customer is increasingly carried out by the agents, who learn from the collaboration and increasingly make their own decisions, the service literature recommends a cooperative and anthropomorphic interaction design to promote actor engagement. In addition, a strong competitive advantage can be achieved through good speech understanding and dialog design, as well as a comfortable and natural customer experience.

Further, *Service Design* is a highly relevant activity field for enabling actor engagement. Customer service is often a standardized practice in traditional customer support service systems, with pre-defined processes, roles, and task responsibilities, including resource acquisition and hand-offs. Customers are often clearly guided step-by-step through a predetermined process. As the CA now performs simple tasks, it must be smoothly integrated into the concurrent service process operation. In this context, the paper discusses different types of integration, including recovery and handover strategies.

The case of CAs shows that service system design needs to facilitate learning cycles at the individual micro-level and institutional macro-level to succeed in increasingly dynamic environments. Changes in actors' practices and institutions must be integrated to realize value. Moreover, changes in one service system's institutions must be integrated and aligned with other institutions into a broader service ecosystem context (Vargo & Lusch, 2016). This perspective is precious for the transformation of extant service systems with AI. In this context, the central role of CAs for the service frontline is described (including shaped value propositions and business models). Introducing CAs in the service frontline transforms the interaction touchpoints with the customer and, thus, the entire customer journey and the value proposition, leading to CA-shaped business models that require new management approaches and market-competitive service designs. In this context, CAs change the value creation process and value proposition by (1) representing a new customer channel and (2) establishing new forms of customer relationships, leading to new forms of revenue streams and reduced costs in the long run.

Essentially, Lewandowski et al. (2022a) complement the findings of Lewandowski et al. (2021) by offering a broader examination of the challenges of designing and managing CAs, focusing on the service design perspective. The multi-level framework highlights the interdependencies between the (re)design of CA technologies, the corresponding work processes, service interactions and customer touchpoints, and, in this context, aspects of engagement design and institutional arrangements that shape the beneficial design and use of AI. The publication contributes to research on AI in service science and guides researchers and practitioners in designing service innovations in the context of CAs.

5.4.3 Challenges and Activity Fields of the Management of Conversational Agents: A Lifecycle Perspective

However, the two studies described in Section 5.4.1 and 5.4.2 were based on an initial understanding of the challenges, aggregated from the SLRs conducted. They were focused primarily on CA initiation, organizational aspects, and initial design features and integration within service systems. Later investigations, in contrast, emphasized the need for a continuous, lifecycle-oriented approach to CAs. While the identified challenges and design areas are relevant to the initiation and planning of CA projects, the core efforts arise in their ongoing use (e.g., to evaluate and increase CAs' overall quality to sustain them in organizational operations; Lewandowski et al., 2023a). In this context, Lewandowski et al. (2022b) point out that CAs, as dialog-based, social, integrative, and learning systems, have a highly human-dependent lifecycle, with continuous development, monitoring, and evaluation activities between IT departments and affected business units.

Building upon the results of both SLRs, Lewandowski et al. (2022b) conducted an empirical interview study to identify management challenges and activity fields along the lifecycle of CAs (see **Table 13**). Closely related to this, Lewandowski et al. (2022b) found that research on the strategic management of CA initiation, development and training, implementation, operation, and improvement is scarce. The understanding of the LCM of CAs can contribute by providing a structured, unified perspective on this dynamic and novel IS, facilitating activities with the integration of resources to ensure reliable, consistent, and cost-effective management of planned and unplanned changes based on past experiences (Alter, 2013) (for more information about the CA lifecycle, see **Section 5.6**).

Lewandowski et al. (2022b) identified 13 core challenges (C) along the LCM of CAs. Regarding the **initiation and planning phase**, one challenge is to address a *committed long-term vision and roadmap for CAs (C1)*, which can be attributed to the failure to address a clear, valuable, and scalable business problem. This circumstance can lead to insufficient resources and a lack of commitment across all levels. CA development often runs parallel with day-to-day business operations, and thus, the massive effort required is underestimated. Closely related, organizations in general and CA teams, in particular, have *insufficient knowledge, wrong expectations (C2)*, and a lack of acceptance of CA, e.g. due to their new characteristics. Establishing interdisciplinary CA teams (e.g., with conversational designers and ML experts) on the one hand and upskilling employees on the other hand, can support CA development. *Incorrect expectations may underestimate the preparation effort in terms of maturity (e.g., quality of data, technology preparation, dialog design, functionality, C3),* and CA may thus go live too early (e.g., driven by management pressure), leading to non-use of the CA and sometimes to a permanent dissent. In addition, the application of CAs often underestimates *environmental issues,* and in this context, the involvement of potentially *inhibiting stakeholders in the organizations (C4)*. The lack of involvement and the underestimation of inhibiting parties (e.g., the data protection department or the works council) can lead to the non-use or termination of the CA project before it fully arrives in the organization.

In the **development and training phase**, integration issues, data management, conversation design, and training, as well as communication issues become central, requiring several competencies in CA teams. First, CA development requires the *integration of the CA into relevant technical systems (C5)* (Gnewuch et al., 2017; Meyer von Wolff et al., 2020a) and the handling of data from various systems to create a seamless orchestration point (Corea et al., 2020). On the technical side, a challenge is that CAs are developed in isolation from technical structures (e.g. from existing (business) architectures and (frontend/backend) systems, data sources) and/or a modernization of the IT architecture is not considered (e.g. provision of APIs). CAs have been developed in isolation, e.g., in an innovation project, without system thinking. Closely connected, on the business side, the *integration of CAs into already existing workflows and business processes (C6)* is neglected and CAs are developed in isolation from current processes (e.g., feedback cycles and handovers). However, on a cultural level, there is often a lack of understanding of these many new activities and the need for *continuous collaboration, feedback, and communication (C7)*. There is often a lack of responsibility, roles, capacity, and freedom to ensure that underestimated development efforts get underway. The further development of a CA requires the continuous involvement of company stakeholders from different areas (e.g., works council) as well as the creation of new roles/freedoms to ensure development efforts (e.g., data, sampling, analysis, training, intent management, and monitoring). Strongly related is the undervaluation of the *competencies required for CA development in a team (C8)*. Organizations must pay more attention to the required developer expertise and the development of new skill sets (e.g., trainers, conversation designers, and modelers), since ignorance can result in possible lock-in effects on CA (platform) vendors and their frameworks. As social systems, personalization and conversation design skills must be considered at an early stage in CA projects. The appearance must be defined, and the integration into work processes must be addressed by designing conversion processes for regular operation and unsatisfactory processes in error-handling scenarios. Another challenge is that the CA application does not involve *data management activities (C9)*. CA training depends on accessing and preparing many (often heterogeneous, unstructured) data sources that are difficult to integrate and process into high-quality data sets for training activities. Several authors emphasize data availability, preparation, timeliness, and NLP conformity (Meyer von Wolff et al., 2020b; Zierau et al., 2020a).

During the **integration and change phase**, *planning domain expert involvement (C10)* is crucial to ensure long-term success. Because developers often do not have the business process and domain expertise (e.g., concrete knowledge of use cases, conversations, and processes), it is critical to involve domain experts continuously. This can include, for example, establishing long-term use case responsibility. This responsibility can be fully or partially delegated to business departments or other product managers, who ensure that the use case is functioning, well-designed, and kept up to date. As unfinished IS, CAs require continuous training and are dependent on the provision of knowledge (Lewandowski et al., 2021). In addition, the precise integration into existing service contexts and the associated change must be planned. Lewandowski et al. (2022) recommend a step-by-step go-live, in which the maturity of the CA is increased successively, and the user group is expanded in small steps so that a direct failure of the project (e.g., due to still very limited CA capabilities) is avoided.

In the **operation and monitoring phase**, *continuous training and maintenance (C11)* are necessary to prevent CAs from becoming outdated. Many CAs fail because they do not receive continuous further development and training. However, the conversation flows, functionalities, data, knowledge, and technical components should be constantly updated, analyzed, trained, and feedback collected to ensure utility and overall quality of the CA. In this context, one challenge

(C12) describes the CA application not *having a continuous monitoring for demonstrating behavior* (e.g., chatlog analysis) of the CA to the supported domains (e.g., metrics/dashboards). An overarching evaluation strategy can help analyze and improve CA quality from multiple perspectives (Lewandowski et al., 2023a). In general, many CAs fail due to a lack of evaluation and, in this context, often because *the organizations generally have a poor feedback and improvement and communication culture (C13)*, which is needed for the continuous development of a CA, as different knowledge is needed at different stages of development.

Article:	Lewandowski et al. (2021)				Lewandowski et al. (2022a)		Lewandowski et al. (2022b)			
Conceptual Frame:	Technological-Organizational-Environmental (TOE)				Multi-level Framework for Service System Design		Work System Life Cycle (WSLC) Model			
Research Method:	Structured literature review (SLR)				Structured literature review		Both SLRs (Article 1 & 2) & Interview study			
Challenges and the resulting design and activity fields:	Foundation	Technical	Organizat-ional	Environ-mental	Engagement	Institutional	Initiation	Development & Implementation & Training	Change	Operation, Monitoring & Improvement
Strategy and preliminary considerations • Long-term vision and roadmap (business problem) • Management commitment and resources • Key stakeholder identification and involvement • Expectations or understanding of the IS	X						X		(X)	
Initiation and planning • Interdisciplinary CA Development Teams (Team Setup & Expertise) • Make or Buy • Appropriate use cases	X		X				X	(X)		(X)
Value proposition, customer journey, and business model						X	X			
Organizational readiness and employee engagement • CA-related employee training (upskilling to address concerns) • Integration with CA training processes	X		X			X	X		(X)	
Culture of collaboration and continuous improvement • Limited systems that require ongoing stakeholder involvement and continuous improvement • Incorporating the diverse knowledge of domain experts • Use case responsibilities • Culture of continuous feedback and communication			X		X	X	X	X		(X)
Data Management • Availability and Quality of (Training) Data • Conversational data sets to train the NLP component		X				X		X		
(Technical) integration into and modernization of the IT landscape		X						X		
Service Design: (Organizational) Integration into Service Ecosystems • Work/Business Processes & Handover • Service Operations • Co-creation scenarios					X			X	X	
Development: The developed CA solution and related quality/maturity • Technical capabilities (e.g., selection of AI/NLP components) • Interaction and conversational design • Functionalities and knowledge	(X)	X			X			X		X
Ethics & System Transparency				X		X	(X)		X	
Legal Issues: Privacy & Security				X		X	(X)		X	
Operation: Continuous evaluation/monitoring, development, and training • Maintenance and enhancement of CA's interaction design, technical components, functionalities, and data/knowledge base • Behavior demonstration (data-driven monitoring): Continuous collection and analysis of feedback, chat logs and other monitoring data						X				X

Table 13. CA Challenges and the Resulting Activity Fields

5.5 Developing a Research Agenda for Managing Conversational Agents in Organizations

In addition to the aggregation and systematization of current research on CAs by the two SLRs, resulting in the conceptualization of CAs (see **Section 5.2**) and an overview of their potentials and complexities (see **Section 5.3**), and their challenges and activities along the CA lifecycle (see **Section 5.4**), Lewandowski et al. (2021) and Lewandowski et al. (2022a) contribute a research agenda for the management of CAs in organizations, thereby identifying, accumulating, and discussing existing CA knowledge. These publications also highlight overlooked issues related to CAs' LCM and envision future research opportunities. Selected gaps are further addressed in subsequent publications by Lewandowski et al. (2022b), Heuer et al. (2023), Lewandowski et al. (2023b), and Lewandowski et al. (2023a).

(1) Investigation of the Organizational Perspective for the Management of CA-Related Activities: Despite the growing importance of CAs, research that focuses on an organizational and management-oriented perspective remains scarce. However, research in this context is essential for understanding the characteristics of CAs, the resulting management challenges, and activities to reduce the risk of failure and discontinuation in organizations. Nevertheless, only a few contributions investigate the organizational or management-oriented perspective of CAs (e.g., Corea et al., 2020; Diederich et al., 2019a; Essaied et al., 2020). Instead, many contributions investigate CAs only from specific isolated perspectives, such as individual (e.g., trust issues), conceptual (e.g., interaction design), or technical design aspects (e.g., NLP algorithms) (Diederich et al., 2019a; Janssen et al., 2020; Lu et al., 2020; Premathilake et al., 2021; Zierau et al., 2020a). However, various research contributions highlight that CAs do not meet expectations and often disappear after their instantiation in concrete service environments or real organizational contexts (e.g., Diederich et al., 2022; Gnewuch et al., 2017; Janssen et al., 2021b). Lewandowski et al. (2022b) revealed that CAs often fail due to organizational and employee-dependent issues in their lifecycle. First authors already call for a *"switch from CA design research to [...][a] management view [...][, since] organizational and individual issues have the highest influence"* (Meyer von Wolff et al., 2021, pp. 12-13) and for *"practice-based requirements[, which] can provide insights that may not have been captured in scientific literature"* (Corea et al., 2020, p. 5827).

CA management comes with many novel activities and design fields that other AI applications (e.g., image recognition) do not have. Some of the challenges in the AI literature (e.g., long-term management support or data quality) (Jöhnk et al., 2021; Pumplun et al., 2019) are consistent with

CA management. However, CAs as dialog-based, social, learning, and integrative IS have a highly human-dependent lifecycle and depend on new collaborations and several new organizational-dependent activities.

(2) Towards a Multi-Perspective Understanding of CAs—The Need for Interdisciplinary Research: Extensive research has been conducted on CAs (Cui et al., 2017; Zierau et al., 2020b). However, as described in the previous section, much of the literature approaches CAs within specific research domains such as HCI, CS, IS, and service science, while focusing on isolated (design) aspects (Diederich et al., 2019a; Janssen et al., 2020). In doing so, current findings are often relatively fragmented across disciplines and application domains and thus lack a coherent axis of transferability for sustained practical use (Elshan et al., 2022a; Følstad et al., 2021; Li & Suh, 2022). In this context, experts in the field call for more collaboration and integration in research on CAs (e.g., Følstad et al., 2021), and encourage further research on issues related to their management, which require systematic and evaluated multi-perspective approaches, measures, models, or frameworks to guide researchers and practitioners towards the long-term success of CAs. In this context, interdisciplinary research and resulting multi-perspective artifacts can be helpful in understanding the complex nature of CA management.

(3) DSR-oriented and real-world research for CAs: Most of the literature discusses CAs from a conceptual or technical perspective, often in isolated settings and a specific research domain (Diederich et al., 2019a; Janssen et al., 2020). As many CA projects are unsuccessful because CAs emerge from laboratories and the problem to be solved is imprecise or isolated from real-world processes (Corea et al., 2020), there is a need for more design science-oriented research or entrepreneurial approaches (Peffers et al., 2007; Ries, 2011) to pilot AI-based CAs in socio-technical settings (Briggs et al., 2019). Broad and in-depth research and evaluated design knowledge are needed to support researchers and practitioners. Therefore, Lewandowski et al. (2021) and Lewandowski et al. (2022a) recommend more research conducted in real-world organizational environments.

(4) Examining the Lifecycle and Activities of CAs: In general, less is known about the management of CA applications in organizational, real-world contexts, and studies investigating CA applications often ignore their long-term success (Corea et al., 2020; Rodríguez Cardona et al., 2019). Closely related, research on the strategic management of CA deployment, operation, and improvement is scarce (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). However, the successful introduction and management of CAs depends on clear operational and maintenance processes and diligences (Kvale et al., 2019). Their successful adoption depends on organizational arrangements, including collaborative and continuous training and development approaches involving the efforts of IT, business, and service professionals (Lewandowski et al., 2021). Guidance on integrating CAs into existing organizational processes, governance structures, and workflows, and on how their adoption and management differs from other AI-based and conventional IS, is limited (Lewandowski et al., 2021). Essaied et al. (2020) describe how studies have conducted initial investigations of organizational adoption of AI without focusing on a specific type of AI solution.

First authors call for research on how organizations can most effectively implement/deploy (Janssen et al., 2020; Schuetzler et al., 2021), adopt (Essaied et al., 2020), and manage (Corea et al., 2020; Meyer von Wolff et al., 2021), and maintain CAs (Kvale et al., 2019) to prevent their failure and to sustain them. While existing studies reveal initial issues and factors that influence the successful adoption of AI-based systems (e.g., Kruse et al., 2019; Pumplun et al., 2019) and CAs (e.g., Corea et al., 2020; Meyer von Wolff et al., 2021; Schuetzler et al., 2021), research does not yet provide procedural guidance regarding the continuous management (and improvement) of CAs across their lifecycle. Thereby, an understanding of CAs' LCM can provide a structured, unified view of this dynamic and novel IS and link resources in order to ensure a reliable, consistent, and cost-effective handling of planned and unplanned changes based on previous issues (Alter, 2013).

(5) Research to address CA foundations: Initiation, Implementation and Change: CA research is needed that belongs to its initiation and implementation phases (Alter, 2013), in addition to pure development to ensure organizational and customer readiness, to facilitate the business problem-CA fit, and to achieve user acceptance of the CA rollout (e.g., with sufficient CA maturity and without alienating users). However, little is known about how to prepare for the go-live of CAs and how to manage the overall change. Lewandowski et al. (2022a) recommend a step-by-step go-live, in which the maturity of the CA is increased successively, and the user group is expanded in small steps so that a direct failure of the project (e.g., due to still very limited CA capabilities) is avoided. However, there is a need for further research to underpin this knowledge with concrete recommendations.

(6) Continuous Evaluation and Improvement Mechanisms to ensure CA Long-term Success: As identified in the SLRs, many publications address the design of CAs. However, few address the overall maturity of CAs. Further, CAs are only marginally or not continuously evaluated to ensure their improvement, successful operation, and overall progress and success in organizations (Janssen et al., 2021b; Meyer von Wolff et al., 2021). Future studies should investigate and establish measurement tools and maturity criteria for different stages of CA development to ensure effective

LCM. As shown, CAs require ongoing design, development, and improvement efforts. While the identified challenges and design fields are relevant to initiating and planning CA projects, the core efforts arise in their continued application (e.g., to evaluate and increase CAs' overall quality to sustain them in organizational operations; Lewandowski et al., 2023a). In this context, it is essential for researchers and practitioners to develop, for example, a set of maturity or quality criteria that address a broader range of requirements for the long-term success of CAs. Systematizing continuous evaluation and improvement processes can mitigate CA failures in organizational environments. In addition, new strategies are needed to engage employees in continuous development and ensure sustainable success. Studies should explore new activities, roles, and collaborations that are essential in this context.

5.6 Derivation of the Lifecycle of Conversational Agents

As discussed in the previous sections, existing research on CAs often focuses on specific, and isolated perspectives (Lewandowski et al., 2022b). However, the contributions in this dissertation have shown that CAs fail due to several organizational and human-related challenges (see **Section 5.4**), which require a broader management perspective that encompasses a range of parallel and interdisciplinary activities that need to be examined in real-world environments.

In this context, Lewandowski et al. (2022b) contribute to the management of CAs in organizations by providing literature-based and empirically grounded design knowledge that prescribes the lifecycle of CAs and sets a foundation for further research activities in this dissertation. The research is in line with previous CA contributions (Corea et al., 2020; Janssen et al., 2021b; Meyer von Wolff et al., 2021, 2022), which emphasize that although some key issues in conventional IS management are also present in the CA lifecycle, CAs require new perspectives due to their specific characteristics. Therefore, Lewandowski et al. (2022b) identify and analyze individual challenges (see **Section 5.4.3**), building on the findings of Lewandowski et al. (2021) and Lewandowski et al. (2022a), extended by insights from an empirical study, to formulate meta-requirements (MRs).

Subsequently, these MRs served as the basis for developing prescriptive and supportive design knowledge in the form of DPs under consideration of the Work System Life Cycle (WSLC) of Alter (2013) as the guiding conceptual framework. The WSLC includes the phases of initiation, development, implementation, and operation/maintenance (Alter, 2001). The authors of the publication included in this dissertation have built their study on this model because it encompasses most existing LCM models for IS, processes, and projects (Alter, 2001) and, within its iterative and adaptive framework, provides a broad view of an IS lifecycle in organizations. Further, the WSLC provides an analytical and design framework for the incremental management of CAs as a novel form of AI-based IS in organizations since their management raises many issues and there are no approaches to guide practitioners on how to manage this class of IS (Lewandowski et al., 2021). Further, CAs need an integrated, collaborative, socio-technical, and interdisciplinary view (Lewandowski et al., 2021) instead of a *"system-as-technical artifact perspective"* (Alter, 2013, p. 74), a viewpoint that is also fostered in the WSLC model (Alter, 2013). The DPs, formulated following the methodology of Gregor et al. (2020) and Möller et al. (2020) (see **Section 3.3.3**), facilitate an understanding of the LCM of CAs, including a structured and unified view of this emerging and dynamic phenomenon.

Based on the 13 challenges and 9 MRs briefly outlined in Section 5.4.3 (depicted in **Figure 6** as issues, here unified under "challenges"), 7 prescriptive DPs were derived to guide and manage the CA lifecycle through the phases of **planning and initiation**, **development and training**, and **integration and change** as well as **operation,** including **evaluation, monitoring and improvement.**

The **planning and initiation phase** encompasses strategically preparing for CA deployment, ensuring readiness, commitment, and long-term engagement with the new technology. This phase lays the groundwork for CA development and promotes the long-term continuation of CA in organizations. Therefore, this phase includes all activities and aspects required to launch the CA project. Successfully implementing CA projects requires managers' understanding and support for CA technology. This will ensure the integration of relevant stakeholders and the provision of the necessary resources over the long term, especially if there are limited CA capabilities. For example, it is important to determine whether an organization has the initial requirements for a CA application. The CA must address an apparent, scalable business problem and vision to ensure that the CA is more than a proof of concept. Pre-evaluating use cases and developing a roadmap are critical steps to ensure economic justification and organizational feasibility, ensuring ongoing commitment and resource availability (**DP$_1$**).

Further, formulating a roadmap helps organize use cases/features, build a CA team, and establish baseline expectations for development time and CA functionality to ensure the project receives adequate effort. In addition, a CA application needs to establish a collaborative and continuous development culture right from the initiation. For example, CA training requires new roles and interdisciplinary team structures for tasks such as preparing NLP-ready datasets, managing intents, and creating conversation/interaction designs while also considering organizational implications and maintaining communication with domain experts. Various new and multi-perspective roles are needed, combining design, technical, service-oriented, and other specialized skills (**DP$_2$**). Furthermore, early identification of potential stakeholders and affected parties is also critical for securing long-term commitments and aligning with business and regulatory requirements.

The **development and training phase** focuses on implementing the CA project. Organizations, or more specifically, CA teams, need to establish conversational/interaction design, data management, use case-oriented, and integration-oriented activities with respect to business processes and related service systems. In this context, **DP$_3$** emphasizes the establishment of data access and management capabilities to create NLP-ready datasets for CA development and training. As a conversational user interface, a CA provides intuitive access to different use cases, making data quality critical to user acceptance and engagement. The use cases offered by the CA can, therefore, only be as good as the data and knowledge elements in the background and the formulations (e.g., for intents and utterances) in the foreground, as well as the accuracy of the language models trained to interpret user intentions and provide solutions.

Closely related is **DP$_4$**, prescribing development and training activities, which comprises a continuous (software) development process in which numerous tasks are necessary to identify and define functions, design and increase the dialog quality (including the general CA representation), and ensure that information and technology are up-to-date. This includes selecting and integrating appropriate technologies (e.g., chatbot technology mix or monitoring tools), and ensuring technical and process integration (e.g., connection to service systems and workflows). In addition, the appropriate process or product owners should be involved, and, if necessary, clear lines of responsibility should be established for the specialist department. These often have the most knowledge about the respective use case and the product, system, or process. In this process, members of the respective departments could, for example, provide the relevant data (e.g., via interfaces) and content (e.g., conversation-based) for the use case. As a "learning and living" IS, a CA is partly defined by the selection of use cases and the scope of functionalities. These influence the perceived usefulness and capability of the CA and, thus, ultimately, its use. The selection of initial, meaningful use cases in the planning phase and the determination of use case ownership in the development phase are critical steps in preparing further use cases and implementation activities.

Regarding the **integration and change phase, DP$_5$** prerequisites the integration of the CA into technical and organizational structures (**MR$_4$**). Work integration is necessary to effectively and efficiently guide the user through the process and to contact a service agent when necessary. If the CA cannot respond to a request, fallback options are required to avoid conversation loops. In the event of a CA failure, the conversation flow should allow the CA to hand off user requests to customer service agents. On the one hand, these so-called “handovers” can be realized by agents’ seamless takeover of chat interaction in real time. On the other hand, handovers can also be forwarded to customer service as a summary of the user’s request (e.g., in the form of a ticket) and processed by agents with a time delay and resolved by contacting the user again. For these handovers, integration points must be considered when designing the conversation architecture and dialog flows. In summary, a CA application must consider integration into existing work and service systems, as well as the IT infrastructure.

Complementing this, **DP_6** addresses the human-centric aspects of the CA lifecycle, emphasizing CA-related education and user preparation to meet pre-rollout expectations (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). A communication strategy, expectation, and stakeholder management at different levels are required to enable a low “introduction threshold.” A successive CA rollout with gradual approval of small user groups, in which features are improved to avoid limited maturity (e.g., dialog design and NLP behavior) is recommended. The integration into a broad application landscape (e.g., common messenger frontends) can help enable rapid acceptance/adoption by employees (for more details, see **Section 5.4.3**). In general, a high level of maturity (functional, technical, and interactional) should ensure long-term commitment. However, this depends on the level of involvement of the stakeholders in the development and training phase.

Finally, the **evaluation/monitoring and improvement phase**, covered by **DP_7**, ensures that the CA behavior is evaluated from diverse perspectives to maintain performance (“keep the CA working”). This safeguards that the content is up to date and assists in identifying improvement opportunities to enhance the overall quality of the CA. The CA application requires establishing ongoing monitoring activities, including new skills and roles, to uncover the actual CA behavior and, thus improvement potentials. Data-driven monitoring and improvement deals with preparing and procuring specific data points for assessing and evaluating the behavior and technical capability of the CA. A compilation and selection of appropriate data points from chat logs, user ratings, qualitative feedback (e.g., interview data, focus groups, A/B testing), NLP metrics (NLP language comprehension and intent scores), and KPIs is required to learn from interactions and inform continuous improvement efforts.

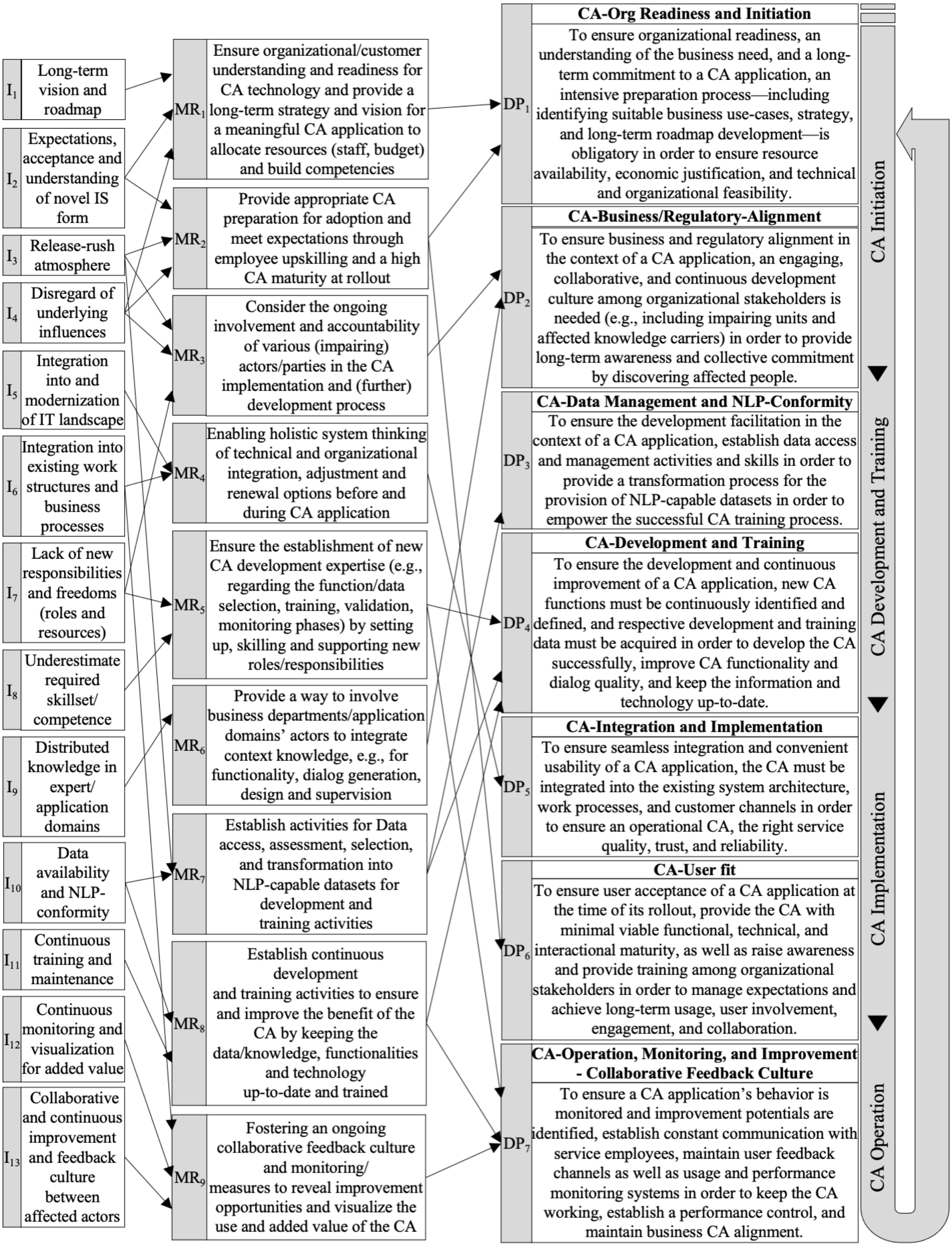

Figure 6: Overview of the Derived DPs from the Publication Lewandowski et al. (2022b) along the Lifecycle Framework of Alter (2013)

5.7 Derivation of a Design Guideline Utilizing Monitoring Data to Improve Conversational Agents in Operation

In the empirical investigation of the lifecycle activities within the INSTANT project, the DSR-oriented studies conducted by Lewandowski et al. (2023b), Lewandowski et al. (2023a), and Heuer et al. (2023) found that a considerable amount of effort is spent after CA deployment, especially in the operation phase. The CA must be continuously evaluated and subsequently trained and improved in a dynamic real-world setting, often characterized by rapid changes and high dynamics, where it is generally impossible to predict how users will interact and what information will be retrieved in the long term (Janssen et al., 2021b; Lewandowski et al., 2023a).

In this context, Heuer et al. (2023) highlight the challenges and limitations of customer service CAs. These stand out during monitoring and analysis in the operation phase to improve user experience and reduce errors. One goal of CA design is to facilitate long-lasting interactions with users and to resolve their requests. Thereby, this work focuses on user dissatisfaction due to communication problems between CAs and users which often lead to unsatisfactory or no solutions. This problem stems partly from the inherent learning curve of CAs, which leads to misunderstandings of user requirements and failure to meet expectations, especially in the early stages of CA projects. CAs face difficulties with requests that fall outside their learned scope, e.g., in the context of intent or entity recognition, compounded by a limited knowledge base that requires constant updating. As a result, CAs may provide incorrect or unsatisfactory answers, leading to user dissatisfaction, especially when rephrasing their queries does not lead to success. Another challenge is their inadequate design, leading to interaction and usage problems.

Related to this, the current CA research landscape lacks an in-depth exploration of practices related to learning from the analysis of monitoring data, such as (failed) dialogs between the user and the CA (Kvale et al., 2019). This understanding could help in a continuous improvement process to shed light on successful and unsuccessful dialogs (Kvale et al., 2019) in order to identify and avoid interaction problems or to incorporate new features/use cases from user requests to improve the quality of CAs and ensure more robust interactions (Heuer et al., 2023).

As a contribution to this field, Heuer et al. (2023) identified **interaction problems** between users and CAs in the operation phase by using a monitoring process, following a developed monitoring guide to analyze and cluster chatlogs of 443 conversations (see **Table 14** and **Section 3.3.2**). The authors then interviewed users about their expectations and requirements and addressed these problems by creating and evaluating mock-up alternatives of possible new solution designs. In an iterative process, the study directly involved users in creating and refining solution design alternatives for the identified interaction problems. This participatory design process resulted in the development of user-centered CA mock-ups and prototypes. In a final selection process, through a quantitative survey with 112 participants, the most popular suggestions (mock-ups) were selected and implemented in a CA prototype. Finally, the resulting CA prototypes were evaluated, which resulted in a significantly improved user experience. On this basis, an abstract **design guideline** was derived and evaluated in focus groups to improve the user experience and interaction quality of CAs based on real-world interaction knowledge and provide descriptive and transferable knowledge for future CA projects.

The final design guideline, informed by empirical findings, monitoring data, survey results, and prototype improvements, addresses critical interaction problems in the live version of the CA, such as findability, welcome messages, dialog control, and fallback management. It is a valuable framework for future development and refinement of CAs in operations (Heuer et al., 2023).

Table 14. Central Interaction Problems and Proposed Design Guideline According to Heuer et al. (2023)

Interaction Problems	Design Guideline (Proposed Dialog Patterns)
Findability	• Fixed and permanently visible CA icon
Welcome Message	• Short • For appropriate target groups, a personalized address, possibly with emojis • Welcome phrase that introduces the CA and offers help • General Data Protection Regulation (GDPR) information as a button/link
Dialog Control	• Suitable mix of free text fields and buttons
Fallback Management	• Make CA limits clear • Build trust in the CA • Do not blame the user *For Fallback 1:* • Clarify misunderstanding • Suggest rewording • Provide the option to forward For Fallback 2: • Clarify the misunderstanding again with different wording and suggest forwarding

Heuer et al. (2023) provide several suggestions for the **findability** of a CA and discuss positioning options *(e.g., "always visible icon on the side of a website"* or *"in a navigation bar";* p. 290), color design, and icon selection. In this context of a website, the icon should always be fixed and permanently visible. Regarding the **welcome message** at the beginning of a conversation between the user and the CA, Heuer et al. (2023) suggest a message form that introduces the CA with a short greeting message, which should, however, disclose that it is a digital assistant and introduce the task areas that the agent covers. In addition, the study participants requested a privacy (e.g., GDPR) notice in the form of a button to keep the message short. Depending on the organizational context, specific linguistic features or humor should be chosen, for which the publication makes extensive suggestions based on the study. In terms of **dialog control,** the paper suggests an appropriate mix of free text fields and buttons. Buttons were particularly suitable for questions with few but clearly defined answer options (e.g., yes/no questions). In the context of **fallback management,** the monitoring analysis in the study reveals that *"of 273 user reactions to a fallback, the conversation was ended in 74 cases, while in 187 cases the query was reformulated. […] The data showed that most users quit after the first or second fallback, and thus the design of a fallback handling strategy would be of paramount importance"* (Heuer et al., 2023, p. 289). First of all, the CA should clarify what skills it possesses and that, initially, the CA is often limited to building up a relationship of trust with the user directly, and making it clear that it is still in a learning state. Regarding the fallbacks, if the CA does not understand a question, the user could be asked to rephrase, which often makes sense in the first fallback, while a redirect is recommended in the second. Otherwise, users will leave completely. The preferred channels were a direct chat transfer or an email contact form; in some cases, a phone call or direct callback was also mentioned (Heuer et al., 2023).

This research contributes by examining the reasons for CA failure from the user perspective and providing detailed design suggestions based on the guideline, supported by the responses from the monitoring data, combined with the statements from the interviews and the trends from the survey of 112 participants. The monitoring data provides deeper insight into the general problems CAs face in live operation, while the survey provides suggestions for implementing each aspect of the design guideline. Finally, the highest-rated suggestions were implemented in a prototype based on the design guideline and evaluated with users, enriching the knowledge base with practical design insights from real-world applications (see **Figure 2**). The central design guideline can be used in CA development for its design, evaluation, and improvement process.

5.8 A Multi-Perspective Evaluation Approach for Continuous Improvement of Conversational Agents in Operation

5.8.1 The Need for Enhanced Conversational Agent Quality Management in Operation

CAs are experiencing a strong hype in research and practice. Many organizations are attracted to their numerous benefits but often fail to realize their full potential in real-world applications because they lack knowledge about evaluating and improving the quality of CAs to sustain them throughout their lifecycle. The research agenda (see **Section 5.5**) reveals that many publications address the design or specific conceptual perspectives of CAs but few address the maturity or quality criteria of CAs and related evaluations. Lewandowski et al. (2022b, p. 12) stated that *"studies need to explore CAs' maturity criteria for measurement to validate the CA in the lifecycle activities beforehand"*, while Lewandowski et al. (2021, p. 10) call for *"[…]guidance on an appropriate chronological order or indicators that define system maturity [and for][…] measurement tools to define system maturity and CA standards for data analysis."* In order to operate a CA successfully, evaluated and in real-world context applied measurements or criteria are needed to guide the alignment of CAs with user needs and help CA teams to sustain them (Følstad et al., 2021; Meyer von Wolff et al., 2022).

In this context, a significant reason limiting the success of CAs is their premature deployment, often driven by high expectations and management pressure, and typically coupled with insufficient knowledge of the CA development process in general and CA quality (measurement) in particular. This practice often leads to non-use, dissent, or complete failure, as highlighted by Janssen et al. (2021b) and Lewandowski et al. (2022b). Further, CAs are only marginally or not continuously evaluated to ensure their successful operation and overall progress in organizations (Janssen et al., 2021b; Meyer von Wolff et al., 2021). Therefore, previous research has proposed continuous evaluation (e.g., via monitoring (Corea et al., 2020) or chatlog data (Kvale et al., 2019)) and improvement processes (Lewandowski et al., 2022b; Meyer von Wolff et al., 2022) to regularly assess their use, quality, and added value (Brandtzaeg & Følstad, 2018; Meyer von Wolff et al., 2022). However, little is known about systematically organizing this operation and improvement process, especially with quality criteria.

To address this gap, Lewandowski et al. (2023b) conducted a DSR project that identified, synthesized, systemized and evaluated findings from the literature and practice to derive a validated and multi-perspective **set of quality criteria for CAs** for applicability and operationality in real-world environments (see **Section 5.8.2**). Building upon these findings, Lewandowski et al. (2023a) extended and re-evaluated both contributed artifacts. Specifically, Lewandowski et al. (2023a) contribute by introducing an extended **procedure model** as an integral part of the instantiation of the quality criteria set within an IT organization, prescribing its application and evaluation activities (see **Section 5.8.3**). This set of criteria and the procedure model form a cyclical, criteria-based evaluation framework that can be initiated by various triggers, providing a methodology for continuous CA improvement. The criteria-based approach can fill the outlined research gap in CA research on lifecycle issues and support the operation phase of CAs (Lewandowski et al., 2022b).

5.8.2 Quality Criteria Set for the Evaluation of Conversational Agents

Lewandowski et al. (2023b) contribute a quality criteria set that allows organizations to continuously evaluate and improve their CAs throughout the lifecycle (see **Table 15**). The set of quality criteria for CAs provides a rigorously elaborated and evaluated prescriptive artifact, conceived through a multi-stage DSR process, that offers applicable knowledge which enriches the academic knowledge base and serves as a design solution for practitioners (e.g., CA development teams) with an adaptable framework for situational instantiations to improve their CAs by applying the derived quality criteria (Drechsler & Hevner, 2018). In addition, this set of criteria provides descriptive knowledge by providing a framework for observation and classification, as well as an overview, new insights, and starting points for further research to evaluate and improve CAs for their long-term success. The quality criteria set is based on a rigorous full-text analysis of 67 relevant publications from an SLR, consisting mainly of journal and conference publications, and the aggregation of several rounds of interviews and subsequent design and evaluation cycles (Lewandowski et al., 2023a).

Given the breadth and complexity of the assembled set of quality criteria, Lewandowski et al. (2023b) developed a multi-level model consisting of three levels (see **Table 15**). This hierarchical structure facilitates holistic or selective application of the set of quality criteria, allowing evaluation of specific (topic-based) areas without having to use the entire set. To enable a systematic and rigorous evaluation process, the set includes a hierarchical structure consisting of 6 meta-criteria, 15 criteria and 33 sub-criteria, including descriptions of their application and cross-references to the original sources. The *meta-criteria* represent the overarching evaluation areas of a CA at the highest level of abstraction. The second-level *criteria* break them down. These can be used, for example, to create responsibilities within a CA team for (meta-)criteria areas, ensuring that accountability is clearly defined and understood. This structure also supports informed decision-making (e.g., prioritizing specific criteria of the CA). Although (meta-)criteria provide logical and structural clarity and classification, they are not sufficiently granular for evaluation purposes. Therefore, *sub-criteria* have been identified at the most granular level, allowing their evaluation by qualitative or quantitative methods.

The set's top-level quality criteria (*meta-criteria*) are divided into **Input**, **Output**, **Anthropomorphism**, **Dialog Control** and **Data Privacy** (see **Table 15**). **Input** criteria for CAs focus on how users create and submit requests, evaluating the various interaction capabilities and the use of familiar communication channels to enhance user comfort (e.g., Feng & Buxmann, 2020; Kowald & Bruns, 2020). It is critical to reflect on and potentially expand communication channels over time, and evaluate and integrate various context-suitable control elements (e.g., text, buttons) to improve dialog flow and user interaction. The context awareness of CAs, including their ability to understand and incorporate previous dialogs to avoid repetitive user input, is another important area of evaluation (e.g., Saenz et al., 2017). Related to this, resumption and return points in the dialog tree are fundamental aspects of evaluation. A well-structured dialog flow helps users provide the correct input, achieve their goals, and avoid deadlocks (e.g., Diederich et al., 2020).

Output refers to criteria related to the CA-generated response provided in return to the user request. In this context, the format of the CA response is essential, including the selection of appropriate visual elements for user- and content-oriented presentation, ensuring high readability, and maintaining consistency in language and terminology to prevent user confusion (e.g., Edirisooriya et al., 2019; Kowald & Bruns, 2020). The output of CAs should transparently disclose their capabilities and limitations to elicit appropriate user expectations that are consistent with the nature of the CA as a learning IS, while ensuring the relevance, meaningfulness, and currency of the information presented to meet users' needs (Diederich et al., 2020). Further, experts suggest, for example, that presenting solutions sorted by relevance and providing justifications for CA responses could increase user trust, and recommend including references to the source of information for transparency (e.g., a clickable link). Additionally, evaluating CA responses for appropriateness, accuracy, and response timing—with an optimal response time identified within two to five seconds—is crucial for user satisfaction, balancing proactivity with potential user interruption (e.g., Edirisooriya et al., 2019; Jiang & Ahuja, 2020).

Anthropomorphism refers to human characteristics, such as emotions, applied to non-human objects (Schuetzler et al., 2021) which can enhance user engagement through the design and evaluation of three aspects: humanlike identity (including profile pictures or avatars, and demographic information such as gender, age, or name), verbal cues (including social dialog skills, emotional expressions, and personalized responses), and non-verbal cues (such as emoticons and typing delay and indicators) (e.g., Schuetzler et al., 2021; Seeger et al., 2021). For CA development teams, it is critical to consider visual presentation and integration elements (position, size, appearance) to ensure that the CA is easily recognizable as the first point of contact on platforms such as websites. However, while these anthropomorphic features can make CAs appear more humanlike and engaging, researchers have also noted that a humanlike CA can be repellent to users (e.g., Grudin & Jacques, 2019). Seeger et al. (2021) pointed out that the different anthropomorphism criteria need to be combined and evaluated practically.

For successful **dialog control**, CAs' understanding of users' requests, along with their intentions and goals, should be evaluated (Clark et al., 2019). However, CAs are learning IS and are, therefore, initially error-prone. In particular, user input in lengthy and complex sentences is a challenge for CAs (Michaud, 2018). Therefore, proactive and reactive dialog management strategies should be employed to avoid, reduce, or recover from failures, including asking users to reformulate requests or providing conversational prompts to guide the interaction and anticipate user needs in regular operation (e.g., Chaves & Gerosa, 2021; Diederich et al., 2020). In scenarios where conversational breakdowns occur (here called: "failure operation"), defining and evaluating repair strategies is crucial for maintaining user trust and ensuring CA success, with criteria including graceful failure acknowledgments, proposing new solutions, or escalating to human service representatives when necessary (e.g., Poser et al., 2021; Poser et al., 2022b).

Further, evaluating the **performance** of CAs represents a strong predictor of CA success (Peras, 2018), intertwining, for example, user satisfaction with effectiveness (e.g., task success rate and the task failure rate) and efficiency criteria (e.g., task completion time, average number of turns, or human handover rate). Also, **data privacy** should be evaluated, including criteria related to the realization and communication of data protection endeavours. One important aspect is to ensure that conversations with the CA are kept as private and anonymous as possible, especially when the CA is dealing with confidential or personal data (e.g., Feng & Buxmann, 2020). In addition, transparency in privacy communications is critical, requiring clear disclosure of the types of user data processed and the provision of privacy policies (e.g., Rajaobelina et al., 2021).

In sum, the quality criteria set contributes to CA research by providing a synthesized and systematized multi-perspective approach to improving the success of contemporary CAs. To achieve this, the publications' authors contributed the quality criteria set derived from strongly dispersed CA research streams (Følstad et al., 2021). A large share of this research focused on specific design and technical issues to elevate the user experience (e.g., Seeger et al., 2021), as these issues were considered the main challenges in the implementation of CAs (Følstad et al., 2018a; Janssen et al., 2021b; van der Goot et al., 2021). However, CAs are inherently complex IS (Maroengsit et al., 2019) with distinctive characteristics that require a comprehensive view and analysis, as failures can arise from multiple (interrelated) factors (Janssen et al., 2021b; Meyer von Wolff et al., 2021). Therefore, the focus were extended from current CA research to a consolidated set of essential quality criteria that should be considered to support the prevention of CA failure. In Section 5.8.3, the procedure model is described and the application of the quality criteria set is explained.

Table 15. CA Quality Criteria Set (Lewandowski et al., 2023a)

Meta-criteria	Criteria	Sub-criteria	Example references
Input	Interaction abilities	Communication channel	(Feng & Buxmann, 2020), Interviews
		Control elements	(Kowald & Bruns, 2020; Li et al., 2020), Interviews
	Context awareness	Dialog-oriented context	(Diederich et al., 2020; Michaud, 2018; Saenz et al., 2017)
		Technical environment	Interviews
Output	Format	Visual elements	(Edirisooriya et al., 2019; Feng & Buxmann, 2020; Kowald & Bruns, 2020), Interviews
		Readability and consistency	
	Content	Transparent capabilities and limitations	(Diederich et al., 2020; Saenz et al., 2017)
		Information retrieval	(Diederich et al., 2020; Edirisooriya et al., 2019), Interviews
		Detail of knowledge	Interviews
		Solution convergence and justification	Interviews
	Calibration	Response appropriateness	(Hu et al., 2018; Jiang & Ahuja, 2020)
		Response accuracy	
	Time	Technical response time	(Edirisooriya et al., 2019; Meyer-Waarden et al., 2020), Interviews
		Balance between proactivity and interruption	(Feng & Buxmann, 2020)
Anthropo-morphism	Humanlike identity	Identity and characteristics	(Schuetzler et al., 2021; Seeger et al., 2021)
		(Humanlike) visual representation	Interviews
	Verbal cues	Emotional expressions	(Saenz et al., 2017; Seeger et al., 2021)
		Chitchat / small talk	(Grudin & Jacques, 2019; Huiyang & Min, 2022; Schuetzler et al., 2021)
		Tailored personality and lexical alignment	
	Non-verbal cues	Emoticons	(Gnewuch et al., 2018; Schuetzler et al., 2021; Seeger et al., 2021), Interviews
		Typing delay and indicator	
Dialog control	Regular operation	Reformulate requests and alternative responses	(Diederich et al., 2020; Saenz et al., 2017), Interviews
		Conversational prompts and suggestions	(Kowald & Bruns, 2020; Li et al., 2020)
	Failure operation	(Proactive & resilient) repair strategies	(Benner et al., 2021; Diederich et al., 2020; Feng & Buxmann, 2020), Interviews
		Fallbacks and handover	(Poser et al., 2021; Poser et al., 2022b; Wintersberger et al., 2020)
Performance	Effectiveness	Task success rate	(Peras, 2018), Interviews
		Task failure rate	
		Retention and feedback rate	Interviews
	Efficiency	Task completion time	(Holmes et al., 2019; Peras, 2018), Interviews
		Number of turns	
		Human handover rate	(Wintersberger et al., 2020), Interviews
Data privacy	Realization and communication	Privacy and anonymity	(Feng & Buxmann, 2020; Janssen et al., 2021b; Lewandowski et al., 2021; Rajaobelina et al., 2021), Interviews
		Transparency	

5.8.3 Procedure Model for Conducting Continuous Evaluations and Improvements of Conversational Agents

In addition to the quality criteria set, Lewandowski et al. (2023b) and Lewandowski et al. (2023a) contribute a **procedure model** that serves as a blueprint for applying these criteria to evaluate CAs, and identify in a systematic process, areas for improvement. This model structures the CA evaluation process, allowing a detailed analysis of the current state of a specific CA, revealing problems and requirements ("potentials for improvement") to identify and address its most relevant improvement aspects (see **Figure 7**). By following the classification of contribution types in DSR of Gregor and Hevner (2013), this research contributes knowledge at different levels (see also **Section 3.1**). The authors contribute to level two by creating an operational artifact in the form of a set of quality criteria (design knowledge). The work also contributes to level one (artifact instantiation) by applying the quality criteria in a real-world context using the procedure model. This model was applied and naturalistically evaluated within an IT organization to improve an existing CA, thereby providing a normative blueprint for practitioners and a starting point for further research. The research integrates design knowledge with practical insights, presenting a structured approach that offers initial insights into the activities, people, and data involved, aiming to improve the **CA operation and improvement phase** (see **Section 5.6**). Further, the quality criteria set and the procedure model can assist in other lifecycle phases of CAs, offering an overview of initial design issues in the **planning and initiation phase** as well as in the **design and development phase**, supporting a comprehensive development process to detect and mitigate potential problems before live integration, thus preventing direct failures.

The procedure model (see **Figure 7**) is based on the supervised instantiation of the quality criteria set using a real-world case in an IT organization to investigate, evaluate, and improve the quality of an existing AI- and text-based CA. The IT organization uses an *ExpertBot* within organizational boundaries to identify, prioritize, and select needed experts. The *ExpertBot* participates in chat conversations and accesses various data sources, such as skill databases, document management systems, and internal chat forums, to provide fitting recommendations for experts and their skills. However, despite several benefits, based on a root cause analysis the authors and the Expert CA team determined that the overall quality and usage rate were insufficient. Therefore, (1) a team of experts with different experience levels and backgrounds regarding CAs and their application area was formed, and (2) a database was created to start the evaluation project (for more details, see Section "Case setting for applying the procedure model"; Lewandowski et al., 2023a). This case setting was the starting point for instantiating the CA quality criteria set through the procedure model.

The procedure model for evaluating and improving CAs is structured into three main phases—general evaluation, in-depth evaluation, and implementation—each with several sub-phases to ensure a detailed and systematic evaluation. The sub-phases allow to (1) provide progressive guidance for each phase of the procedure model, which facilitates the evaluation of CAs; (2) ensure that each aspect of the procedure is thoroughly documented, which is critical to properly evaluating CAs; and (3) provide a more detailed and comprehensive procedure that helps the team ensure a systematic CA evaluation.

In the general evaluation phase, a successive analysis based on quality criteria identifies key problem indicators in a CA and lead to initiating an improvement project. During the in-depth evaluation phase, the quality of a CA was evaluated in collaboration with the IT organization, and the potential for improvement was determined based on the identified problem indicators. In the given case, the authors conducted interviews, along with the corresponding criteria from the set, to assess the problems of CAs and identified six specific improvement potentials. These were then explored through the creation of mockup prototypes which allowed for an informed comparison between the current state version of the CA and the proposed modified CA version(s), with feedback from A/B testing informing which improvements to prioritize. The expert team provided valuable feedback on whether the identified improvements would be beneficial if implemented, or whether they needed to be revised or discarded. Finally, the improvements identified through the mockups and deemed beneficial to the quality of the CA were implemented in a revised live version. These improvements were communicated to the users to ensure their visibility in the organization.

Afterward, the procedure should be repeated to improve the CA long-term, for instance, if problems are identified based on existing data or as part of a general cyclical evaluation to examine the quality of the new CA live version as a whole or in defined segments.

The model promotes documentation of each step, the use of multiple, appropriate evaluation methods, and the involvement of an interdisciplinary team of experts for a multi-perspective evaluation of CA quality. The process addresses more than immediate issues. It also provides a data-driven framework for continuous improvement and adjustment of a CA. By following this structured approach, organizations can systematically evaluate and improve the quality of their CAs, ensuring that they more effectively meet user needs and organizational goals. Refining CA prototypes before final implementation through design tools like Figma and feedback mechanisms such as A/B testing is helpful in this context. Overall, this model provides a replicable and detailed methodology for enhancing CAs through targeted improvements based on quality criteria and user feedback. This work complements other preliminary efforts, such as the evaluation criteria sets of Radziwill and Benton (2017) and Casas et al. (2020), to provide a better understanding of CAs in the improvement process with a system-wide view.

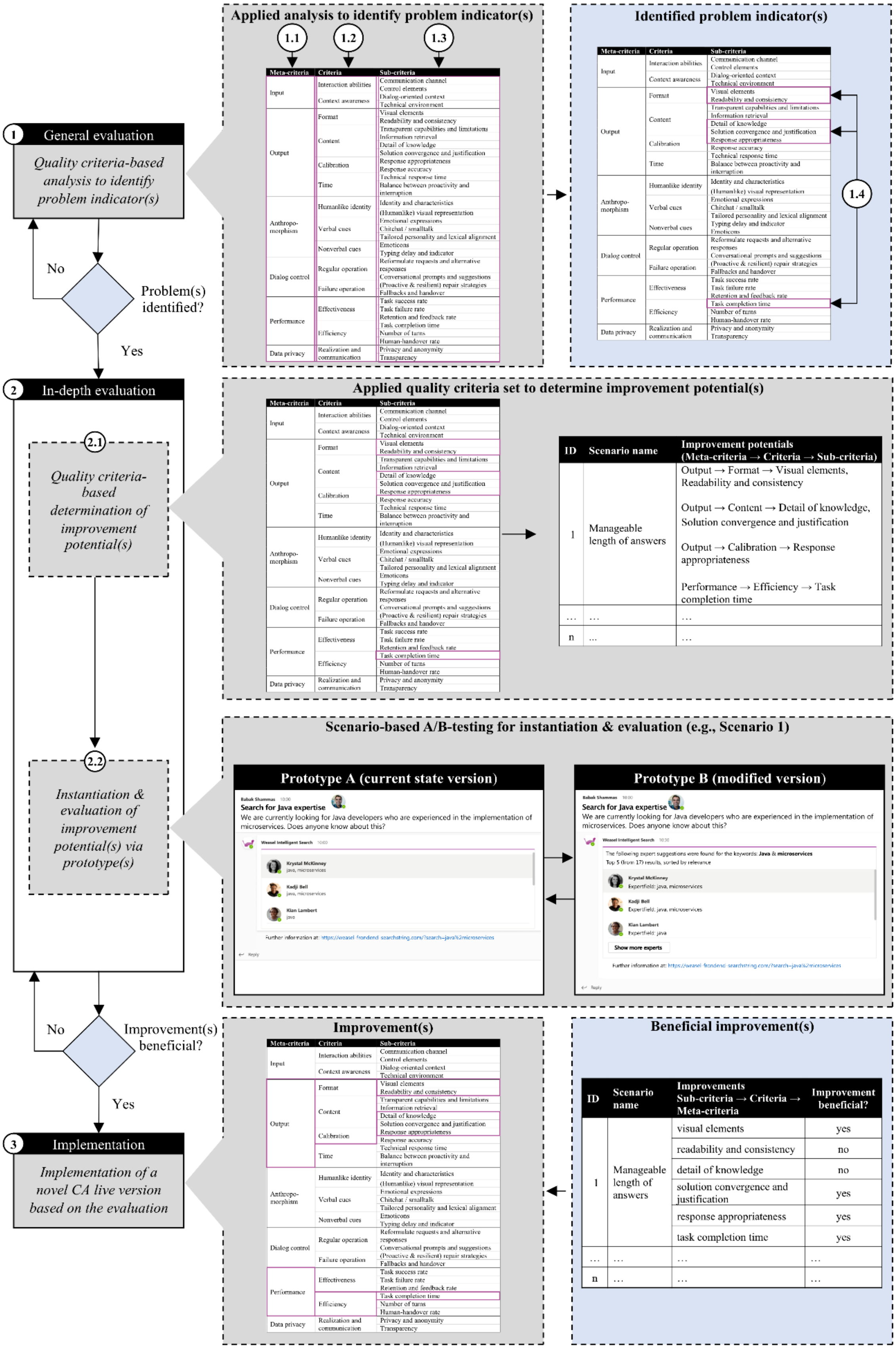

Figure 7: Procedure Model for the Evaluation and Improvement (Lewandowski et al., 2023a)

5.9 Broadening the Perspective to an Organizational View: Proposing A 3D-Model for Guiding Organization-wide AI Transformation

From an organization-wide perspective, AI and related AI-based systems and technologies are perceived as disruptive accelerators (Benbya et al., 2020) and are, therefore, an essential strategic element and a fixed point on the future agenda of many organizations (Sagodi et al., 2022). AI holds various potentials for organizations, such as a new form of value creation and increased business efficiency, leading to competitive advantage and overall growth (Alsheibani et al., 2019b; Alsheibani et al., 2020). However, organizations struggle to unlock these potentials, resulting in many AI initiatives failing early (Uba et al., 2023). The journey to successful AI transformation is challenging, precisely due to a lack of guidance, best practices, and understanding of how to strategically manage, integrate, and (re-)evaluate AI into existing organizational structures and processes. Organizations do not have an overview of where to begin an AI-based transformation (Fukas et al., 2021), combined with exaggerated expectations that lead to highly exploratory and experimental AI transformation efforts.

In this context, Uba et al. (2023) contribute by guiding organizations through the transformative journey of integrating AI, thereby providing a broad perspective by synthesizing the key challenges in establishing AI programs and initiatives to overcome common pitfalls. The publications' authors further highlight the main activities and required capabilities for an organization-wide AI transformation. Thereby, they draw on empirical research, analyzing eleven organizations across industries and sizes, and interviewing more than 30 study participants, including IT executives, senior managers, chief data scientists and other IT professionals, as well as CIOs and CDOs (see **Section 12** for more details). The study presents four transformation types distinguished by different AI transformation stages and journeys (see **Section 5.9.1**). Furthermore, a central contribution of the research is the development and presentation of a 3D-Model to guide organization-wide AI transformation, with a strategic framework to help organizations navigate through the complexities of AI transformation, encompassing three core dimensions, corresponding activities, and concrete recommendations for action for each dimension (see **Section 5.9.2**).

5.9.1 Identifying Transformation Types along the AI Transformation Journey

Based on the interview insights, Uba et al. (2023) introduce four transformation types characterized by different AI transformation stages and journeys: *Explorers*; *Intermediates focusing on process* *optimization*; *Intermediates concentrating on customer value creation*; and *Strategic Visionaries.* The publication illustrates the different types with exemplary cases from the study in order to clarify which attributes characterize each type. *Explorers* are organizations interested in AI but which have little to no experience working with it. *These enterprises* deal with fundamental questions about establishing AI, such as selecting suitable use cases and developing essential expertise and profound data infrastructure. In comparison, *Intermediates* are organizations that have exceeded the point of developing initial use cases and proof-of-concepts. Organizations have at least one established core AI or data science team, a running and often profitable AI solution, and an existing and reusable (infra-)structure to launch additional AI projects. The paper categorizes intermediaries into two groups: Organizations that focus on AI to improve the efficiency of internal processes and organizations that focus on directly impacting the customer experience with AI. In contrast, *Strategic Visionaries* have the highest level of maturity. They explicitly define AI as a part of the company's business strategy, and consider it a key enabler and a competitive advantage for their organization. The case organizations have numerous AI-based systems up and running and are concerned with *"best practices, developing guidelines, setting up a comprehensive governance, and pipelining AI incubation"* (Uba et al., 2023, p. 6134). The types presented are helpful for organizations to assess and classify themselves (to a specific transformation stage) and to plan further steps in their transformation using the 3D-Model.

5.9.2 Introducing the 3-D Model for AI Transformation

Based on the research and the companies' experiences, Uba et al. (2023) further contribute a 3D-Model to guide AI transformation and provide concrete recommendations for each dimension. The findings help navigate, manage, and (re)evaluate AI strategies for enterprise-wide transformation. The 3D-Model comprises three dimensions for strategic action: (1) *Core Capability Building*, (2) *Value Stream Embedding*, and (3) *Organizational Enabling*, spanning the space of possible AI activities and including the recommendations for action (see **Figure 8**).

Core Capability Building (1) emphasizes the importance of foundational AI capabilities, including data infrastructure and management, continuous and collaborative development lifecycle, and tool/platform selection. *Value Stream Embedding* (2) focuses on the seamless integration of AI within existing business processes, knowledge systems, workflows, user interfaces, and customer channels, recommending the facilitation of use case incubation, education of all involved participants on new tasks and responsibilities, and establishing domain responsibility for data.

Organizational Enabling (3) describes the strategic and enterprise-wide integration and establishment of AI, underlining the need for a multifaceted AI governance, centralization of AI expertise, skill-building across all employee levels, and investment in strategic partnerships.

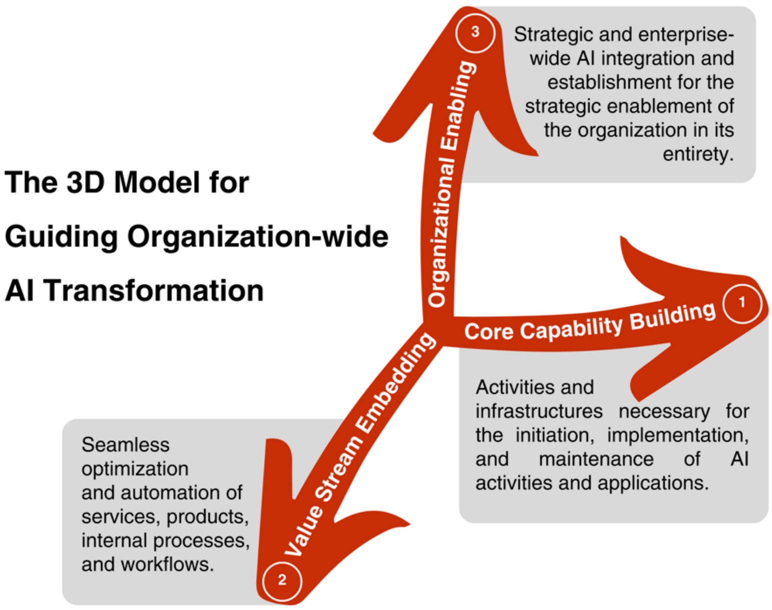

Figure 8: 3D-Model for Guiding AI Transformation (Uba et al., 2023)

The provided findings, including insights on the transformation types and the proposed 3D-Model with the deduced dimensions and aggregated recommendations, serve researchers in finding further starting points for research, and practitioners with valuable guidance and necessary knowledge to assess their current AI practices. By categorizing organizations into transformation types and outlining actionable strategies within each dimension of the 3D-Model, the authors provide the basis to create roadmaps for organizations to assess their current AI capabilities, align AI initiatives with strategic objectives, and overcome common challenges in the AI transformation.

6 Practical Contributions

In addition to its theoretical contributions, this dissertation also has practical contributions for organizations described in the following chapter. DSR projects aim to produce research results that are relevant and applicable to practice in order to contribute directly to solving real-world problems. In this context, DSR inputs information from the contextual environment (e.g., the natural organizational contexts) into the research process (Hevner, 2007). This context information assists in understanding the application domain, including existing structures, people, systems, challenges, and opportunities, to define problems and derive requirements in order to create and evaluate artifacts that ultimately improve the environment (Hevner, 2007).

In this case, the publications included in this dissertation were part of the INSTANT research project (see **Section 3.2.1**). The INSTANT project provided a variety of organizational application contexts, collectively referred to as the "environment" in this dissertation (see **Section 3.2.1**). Thereby, the INSTANT project marked the starting point for the research, as all partners involved lacked an understanding of the CA technology and faced numerous unresolved problems in the management of their current CA projects, for which there were no existing solutions in either academia or practice. Initiated by the identified environmental problems and opportunities, the relevance cycle was iteratively traversed and supplemented with existing knowledge from the rigor cycle to guide design activities in this dissertation. Beyond the extraction of practical knowledge, new artifacts were generated and evaluated with various stakeholders, thus impacting the real-world environment. The research and corporate partners worked closely together in real-world laboratories (labs) to realize an appreciative and effective design for AI-based CAs in customer service. Different CA scenarios were analyzed, tested and evaluated in common field experiments. Concrete examples are prototyping and piloting activities, as well as instantiations and artifact-based evolutions throughout this dissertation (e.g., Heuer et al., 2023; Lewandowski et al., 2023a). In sum, these artifacts were not merely conceptual but concretely instantiated and applied, enhancing the research environment at the corporate partners involved in the INSTANT research project.

In sum, the knowledge generated supports practitioners in organizations and their CA teams in managing CA projects and endeavors more structurally. This dissertation offers a range of orientation knowledge for improving the planning, development, implementation, and long-term sustainability of CAs to prevent their failure. The knowledge contributed is briefly outlined below.

The **systematic conceptualization of CAs and their characteristics** (see **Section 5.2**) provides organizations with a fundamental overview of an emerging type of technology, its advantages, and the design fields that must be addressed. In combination with the overall **application areas and potentials** (see **Section 5.3**) presented in the publications of this dissertation, organizations can consider whether they want to use CAs at all and where they can use them beneficially. The knowledge provided helps organizations identify concrete CA use cases and assess whether expert knowledge is available for the versatile design of CAs and, thus, to plan the first steps for CA ventures. Limited consideration of a single CA perspective, such as a specific technical or conceptual focus, would be counterproductive. Therefore, this knowledge helps create a broad overview right at the outset of a CA project. Further, strategic considerations and basic preparations can be made about organizational requirements such as employee acceptance and adoption, integration with existing governance structures, services and related work routines and company processes, and general readiness of the organization and management (e.g., technical, organizational and environmental factors; Lewandowski et al., 2021).

Building upon this, the **aggregated overview of the challenges** (see **Section 5.4**) provides practitioners with an understanding of the emerging activity fields and the general development hurdles of a CA. As the dissertation outlines, CA development is often done very exploratively and with little process or lifecycle knowledge. However, the application poses entirely new challenges to organizations, which can lead to limited success or the complete failure of CA projects. The overview knowledge helps practitioners to understand early that CAs involve a multi-perspective and highly human-dependent design and development process that requires many interactive, interdisciplinary and participatory activities together, requiring strong cooperation to enable joint continuous development/training and monitoring activities between IT departments and affected business units. In addition, an overview of the challenges helps to take appropriate countermeasures, establish activities, or build knowledge to deal with them.

This dissertation has generally resulted in practice-oriented findings useful for organizations to successfully plan, design/develop, implement, operate, evaluate, and improve CAs. In this context, the **derivation and description of the CA lifecycle** (see **Section 5.6**) and the related design knowledge for dealing with them enabled a procedural, structured and unified guidance for organizations. In sum, this dissertation contributes to the management of CAs in organizations by providing design knowledge for practitioners on how to establish and manage CA lifecycle activities.

Thereby, the provided **design guideline** (see **Section 5.7**) and **multi-perspective evaluation approach** (see **Section 5.8**) present a systematic blueprint to structure the design, evaluation and improvement phases, and the derived artifacts can thereby guide CA teams in various ways to support the successful development, operation, and evolution of CAs. First, the design guideline offers validated recommendations, e.g., for the consideration and arrangement of CA design elements. Second, the combination of the quality criteria set and procedure model allows practitioners to obtain a comprehensive overview of relevant criteria and to narrow down the evaluation of their CA to identify specific problems and improve the overall quality of CAs. Third, the procedure model can serve as a blueprint for CA teams to systematize the evaluation process. The delineation of content and the sequence of relevant steps provides a feasible approach for practitioners to structure their evaluation and improvement activities of existing CAs. In addition, the criteria set can serve as a basis for CA teams to decide whether a CA project should be established and whether requirements are present (e.g., prepared data, an interdisciplinary team) to enable a comprehensive and multi-perspective evaluation of the quality of CAs. Therefore, the execution of evaluation and improvement tasks could be accelerated, consequently counteracting the discontinuation rate of CAs. In addition, the evolution of CAs and related positive effects could not be limited to the CA domain, as their success could foster the overall AI transformation of an organization so that the increased quality and use of CAs can influence other learning and AI-based IS.

However, its application does not necessarily guarantee success in the deployment and continuous improvement of CAs. In pursuit of this goal, the quality criteria and the process model can be viewed as one piece of a greater puzzle (Lewandowski et al., 2023a). To achieve this ambitious goal, on the one hand, organizations must consider, design, and establish numerous activities along the CA lifecycle. On the other hand, by broadening the perspective, general organizational AI prerequisites must be established to enable an organization-wide AI transformation. This dissertation extends the view and introduces a **3D-model to guide organizations through an organization-wide AI transformation** (see **Section 5.9**) to shed light on how AI transformation can be approached and the challenges organizations face. It proposes concrete recommendations for action to navigate, manage, and (re)evaluate their strategy for AI transformation, including insights for practitioners to establish AI capabilities, programs, and initiatives. These insights give practitioners the knowledge they need to evaluate their current practices and develop a roadmap for future AI efforts. As a result, they become AI-savvy organizations that can unlock the potential of AI and sustain an AI-enabled competitive position over the long term (Uba et al., 2023).

7 Limitations

The research conducted within this dissertation has limitations that provide opportunities for further investigation, which are outlined in this chapter.

Limitations of structured literature reviews. Concerning the SLRs, one limitation is that the findings depend on the SLRs' narrowing steps in terms of defining the scope of the review, conceptualizing the topic, including fixed search terms, database and literature selection, aggregation, and authors' judgment. Broadening the scope of the SLRs could lead to more in-depth findings. However, structured methods were used to conduct the SLRs in a valid, transparent, and comprehensible manner (e.g., vom Brocke et al., 2009; vom Brocke et al., 2015). To achieve a more solid academic underpinning, the authors of the publications recommend further research with empirical investigations to gain more in-depth knowledge about CAs in organizations. In particular, the studies by Lewandowski et al. (2021) and Lewandowski et al. (2022a) are based on isolated SLRs that summarize the state-of-the-art in specific research areas in order to derive challenges and activity fields that are addressed in later parts of this dissertation. However, the results were further complemented by qualitative and verified by evaluative research conducted during the dissertation (e.g., Lewandowski et al., 2022b).

Limitations of qualitative data collection and analysis. First, as with any qualitative data collection, the analysis of the collected data may be influenced by the subjective interpretations of the authors, which may distort the meaning of the reported reality (Myers & Newman, 2007). In this dissertation, however, established methods for the structured conduction of interviews and their analysis were used. Furthermore, multiple authors were involved in the coding to allow for multi-perspective discussion and, if possible, intercoder reliability checks (e.g., Lewandowski et al., 2022b) to ensure the validity and higher objectivity of the results (Mayring, 2014). Second, the experts in this study and their domain-specific experiences influence the external validity of the research. The authors have drawn on existing organizations and research project contact networks in this context. The experts were of European origin, and the findings are based on their knowledge. However, many experts work for international companies in various industries, providing a range of experience and sufficient data saturation. In addition, all experts have in-depth CA expertise and/or have worked on CA projects over a long period, so it can be assumed that their experience helped organize the results. Third, it could have been helpful to subdivide the collected data based on expertise or demographic characteristics for deeper insights. All statements were treated and analyzed equally. Fourth, due to the COVID-19 pandemic, all interviews were conducted digitally via video conferencing systems. The digital format may have introduced side effects that would not have occurred had the interviews been conducted on-site and in person. However, the qualitative data analysis did not reveal any noticeable side effects. On the contrary, the digital format even allowed the experts to enrich their statements with digital material.

Limitations of the applied DSR framework. In the context of this dissertation, a rigorous DSR project was conducted by aggregating insights from the literature supplemented by experiences from the practice-based, real-life environment to derive different forms of artifacts (Hevner, 2007). However, this approach is also faced with limitations. First of all, the results were obtained in a natural context, the INSTANT project (see **Section 3.2.1**), and could, therefore, be biased, as they are essentially based and evaluated on the experience of experts from three core organizations. Second, the unstructured nature of the DSR framework is a source of criticism regarding rigor and transferability, as many results are highly dependent on the context considered (Hevner et al., 2004; Kuechler & Vaishnavi, 2008). The transferability of the deduced artifacts to other application environments with CAs of different use cases, other CA teams, and conditions remains to be proven to address further the overarching problem class (Lewandowski et al., 2023a). Third, as Hevner et al. (2004, p. 99) stated, *"in new and emerging applications of technology, the artifact itself represents an experiment."* Not only is the knowledge base limited for design purposes, but there is also little understanding of the nature of the real-world problem, the environment, and possible solutions (Hevner et al., 2004). In this context, the first publications were also very much based on intuition, assumptions and experiments from the authors (Hevner et al., 2004). However, these artifacts were further developed in different iterations (in multiple organizations) and in subsequent studies in this thesis and, therefore, could be further strengthened by the authors. The derived artifacts generally receive more scientific validation, since they were applied in various organizations and CA contexts. Nevertheless, further studies could help to identify further whether the artifacts can be applied to further organizations or if they need to be extended or reorganized based on more perspectives. For example, in the context of Lewandowski et al. (2023a), one objective was to apply a final quality criteria set in a naturalistic evaluation setting to verify whether the set could serve CA teams in evaluating and revealing potential improvements in a procedural and structured way. Although the set was applicable and could meet those objectives, the instantiation referred only to a single CA team in one IT organization from the INSTANT project.

Limitations of naturalistic evaluations. Closely connected, more longitudinal evaluation strategies could substantiate the derived artifacts and support their generalizability to different natural environments (Hevner et al., 2004), which were limited by time constraints (e.g., due to the end of the INSTANT project). The use of even more different methods could lead to more comprehensive results. The publications' evaluation was mainly based on one or two evaluation rounds and methods in naturalistic contexts. More evaluation episodes in different environmental contexts could help determine the artifacts' utility or efficacy in real use. Nevertheless, the developed artifacts in this dissertation were evaluated with three core organizations of the INSTANT project and other supporting organizations in the nearer project context (e.g., Heuer et al., 2023; Lewandowski et al., 2023a). Therefore, these findings can be applied by other CA teams and organizations in their real-world environments and may serve as a starting point for further research.

8 Implications for Further Research

Based on the findings, additional promising and impactful directions for future research have been identified in this cumulative dissertation and are presented in the following chapter. This dissertation has formulated a **research agenda for managing CAs in organizations** (see **Section 5.5**), in which specific topics require more in-depth, complementary or supportive research. While the publications in this dissertation have addressed several research opportunities, there remain unexplored gaps within the research agenda. The agenda generally motivates scholars to conduct studies from organizational, practice-based, or management-oriented perspectives. More research from these perspectives is imperative to improve understanding of the implications of the specific characteristics of CAs, the management challenges they pose, and the activities that can be undertaken to reduce the risk of discontinuation in organizations. In this sense, there is still a need for more DSR-oriented research or entrepreneurial approaches for CAs (Hevner, 2007; Peffers et al., 2007; Ries, 2011) aimed at capturing naturalistic, real-world knowledge and producing evaluated socio-technical artifacts. This knowledge can address the different lifecycle activities of CAs and provide more valid design knowledge for researchers and practitioners.

In general, more research is needed that takes an interdisciplinary approach to CAs to overcome the fragmentation across research streams, disciplines, and application domains that lack a coherent axis of transferability for sustained practical use (Elshan et al., 2022a; Følstad et al., 2021; Li & Suh, 2022). This work thus supports the call by Følstad et al. (2021) that CAs need more overarching research directions to guide researchers, combining knowledge from the progressing interdisciplinary wave of CA research. In this context, interdisciplinary research and the resulting artifacts can help understand the complex nature of CA management. The research agenda of this dissertation serves as a concise research roadmap for the management of CA in organizations and as a compass for researchers interested in deepening their knowledge in these areas.

Concerning the **CA lifecycle**, there are numerous starting points for deepening research. Although fundamental challenges, requirements, and activities have been described in the course of Lewandowski et al. (2022b), Meyer von Wolff et al. (2022) or Winkler et al. (2020), there is often a lack of detailed knowledge about the individual activities, data, roles, skills, diligences, and collaborations that are essential in this context. Further research is needed to underpin this knowledge with concrete recommendations, including frameworks or methods for the planning phase, selection of use cases, platforms during the development phase, or modeling tools for CA-human collaboration and integration into service desk processes. Furthermore, this dissertation highlights the need for detailed strategies for CA go-live preparation and change management in the implementation phase. Lewandowski et al. (2022a) recommend a step-by-step go-live, in which the maturity of the CA is successively increased and the user group is expanded in small steps.

However, further research is needed on how to deal with such dynamic and AI-based systems in organizations. In addition, gaining more detailed knowledge about the monitoring/evaluating of CAs is essential to provide, for example, procedural guidance regarding the continuous improvement of CAs across their lifecycle. Although Lewandowski et al. (2023b) and Lewandowski et al. (2023a) have made significant steps towards a blueprint for evaluation and improvement of CAs in the operation phase, further research is required on how organizations can generate a CA monitoring and evaluation strategy, including aspects such as time points, intervals, databases, criteria prioritization and selection, and appropriate methods. For the purposes of this dissertation and in order to underpin the real-world instantiation, semi-structured interviews and A/B testing were used as qualitative evaluation methods, which are not necessarily suitable for all criteria and situations in CA operations. Overall, a general framework could help organizations and researchers select appropriate evaluation, monitoring, and data analysis methods for relevant areas. As outlined in Lewandowski et al. (2023b) further research is needed to explore alternative or faster ways of performing activities and methods to increase CA quality. There may also be automation potentials with more tool support.

From a broader perspective, further technical progress is expected in the context of CAs, as also described in a study by Schöbel et al. (2023). Recent technical advancements in the field of NLP and ML applications, in particular the **emergence of large language models (LLMs), will lead to a high level of dynamics and several novel research opportunities** in the field of CAs and in the context of text-based HCI. These LLM models have been pre-trained on billions of text samples from specific data sources on the Internet and can generate a wide variety of content types (Brown et al., 2020; Jiang et al., 2022). In particular, with the release of intuitive and conversational interfaces such as OpenAI's ChatGPT or Google's Bard (Jiang et al., 2022; Teubner et al., 2023), these models are becoming widely available to (non-technical) users. These releases imply a remarkable movement in CA research exploring new application scenarios and their potential, also referred to as a new "AI wave" by Schöbel et al. (2023).

Lewandowski et al. (2023a) already raise the question of which knowledge is already affected by this new wave and which requirements result in CAs and their (further) development. A paradigm shift is expected concerning the results, which could lead to a higher dynamic and flexibility in their adaptation, general acceptance, and use. For example, many aspects of the development and design phase could be simplified, as the LLM-based solutions offer a much better understanding of natural language and deliver better results. However, they must still be designed and trained from several perspectives to ensure organizational integration. For example, these solutions still require organizational/business knowledge to function as an integrative platform, thus retaining many activities of the proposed CA lifecycle. They also require process integration in order to work with existing business processes. However, in this dynamic context, a detailed examination of each activity is proposed to determine which challenges and activities still exist, which require a different/revised approach, and which have become obsolete. By exploring these areas, further research could counteract the skepticism of users towards conventional CAs, which are perceived as unnatural, impersonal, or deceptive (Schöbel et al., 2023), and reduce the overall failure of CAs, thereby bringing them further into the mainstream of society, science, and organizations.

9 Publication No. 1: Lewandowski et al. (2021)

Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-Art Analysis of Adopting AI-based Conversational Agents in Organizations: A Systematic Literature Review.

Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

Abstract

AI-based Conversational Agents (CAs), such as chatbots, are becoming increasingly important in organizations and academic research. Beyond their intuitive, (natural) language-based and fast-accessible interface, CAs provide a scalable and cost-effective platform for organizations supporting employees by retrieving, structuring, and analyzing information to assist work processes. CAs represent a novel class of information systems (IS) characterized by increasing levels of intelligence, autonomy, and personality. However, studies taking an organizational perspective on the adoption of CAs remain scarce. We address this research gap by conducting a systematic literature review (SLR) to provide a first and structured overview of managing these systems from a strategic viewpoint, including their impact on work and company processes as well as existing governance structures. We identify organizational, technical and environmental factors and chart an agenda for future research opportunities. Our study contributes to research on CAs and guides practitioners in the adoption of CAs in organizational contexts.

Keywords: AI-based assistants · Conversational agents · Chatbots · Organizational adoption guidelines · Systematic literature review (SLR)

9.1 Introduction

AI-based Conversational Agents, such as chatbots, are becoming increasingly important in organizations. New application areas for CA implementation are emerging due to the increasing digital interconnectedness, growth of information available, and technological advances in ML and computational linguistics (Brandtzaeg & Følstad, 2017; Dale, 2016; Gnewuch et al., 2017). The introduction of enterprise messengers such as Slack or Microsoft Teams enables novel work routines and problem-solving approaches (Stoeckli et al., 2019). In this context, CAs appear as social and AI-based actors transforming employees' interaction with information systems as part of internal corporate communication structures (Maedche et al., 2019; Zierau et al., 2020a). Further, increasing information loads subjects employees to a growing workload and stress (Semmann et al., 2018). CAs bear the potential to assist, solve, or automate tasks in work processes by retrieving and structuring information, and provide employees with cognitive relief by, e.g., identifying solution strategies, providing decision-support, and vocational training through knowledge provision (Meyer von Wolff et al., 2019a; Semmann et al., 2018; Stoeckli et al., 2019).

However, despite the visible potential, a low technical burden for rudimentary setup and increasingly commercial and academic interest (Corea et al., 2020; Dale, 2016; Diederich et al., 2019b; Riikkinen et al., 2018), CAs fall short of expectations and require further research (Corea et al., 2020; Feng & Buxmann, 2020; Meyer von Wolff et al., 2019a). Adopting both speech and text-based CAs within organizations remains in its early stages and lags behind consumer usage (Corea et al., 2020; Feng & Buxmann, 2020; Meyer von Wolff et al., 2019a). While technology companies such as Google possess extensive expertise in the development of AI-based assistance systems (Maedche et al., 2019), "traditional" (and originally non-technology-centric) companies and their employees are only beginning to experience these new forms of intelligent and social information systems.

Although the scientific and practical interest in CAs has grown in recent years in the field of Human-Computer Interaction (HCI), CS and IS (Gnewuch et al., 2017), the research is scattered across various streams, and the findings often remain segregated (Lu et al., 2020; Zierau et al., 2020a). Further, the impact of AI systems in companies compared to conventional IS is insufficiently studied, although numerous companies already implement AI systems in general (L. Wang et al., 2020). Only a few contributions investigate the organizational adoption of CAs (e.g., Corea et al., 2020; Diederich et al., 2019a; Essaied et al., 2020). Existing literature lacks a dedicated organizational-level perspective but instead takes a specific conceptual or technical perspective, often based on laboratory settings (Diederich et al., 2019a; Laumer et al., 2019a; Meyer von Wolff et al., 2020a). These studies investigate possible application areas (e.g., Laumer et al., 2019a; Meyer von Wolff et al., 2020a), current objectives of CA applications (e.g., Brandtzaeg & Følstad, 2017), challenges in using them (e.g., Gnewuch et al., 2017), how to design CAs for higher user-acceptance (e.g., Bittner et al., 2019; Gnewuch et al., 2018), or how to improve the natural language processing capabilities (e.g., Dahl, 2013). Conversely, less is known on adopting this new class of IS in existing organizational processes, governance structures, and work routines and how it differs from other AI-based and traditional IS adoptions. Similarly, no frameworks guide the organizational roll-out and continuous development of novel self-learning and self-communicating systems (Corea et al., 2020; Essaied et al., 2020). First authors call for research regarding the practice-oriented and company-focused adoption of CAs (e.g., Corea et al., 2020). Hence, we formulated the following research question:

RQ: *Which factors need to be taken into account to adopt an AI-based CA in an organization in contrast to other (AI-based) IS?*

To address this research question, we conducted a systematic literature review following the process proposed by vom Brocke et al. (2009). Our research provided a first and aggregative overview of the analyzed literature with a systemized set of factors along the technological-organizational-environmental (TOE) dimensions proposed by Depietro et al. (1990). The results indicate that beyond the aforementioned factors and especially in contrast to the management of other AI-based and traditional IS, a new holistic form of work cooperation and development for CAs' adoption are needed. Due to the nature of CAs as a new class of information system, characterized by increasing levels of intelligence and personality occurring as a novel actor in work routines, services, and processes, these adoption approaches stand out from traditional IS-based adoption approaches. This leads to a research agenda regarding the organizational adoption of CAs.

The paper is structured as follows: First, we introduce related work towards AI-based CAs and technology adoption. Afterward, we present the steps of our SLR. Then, we describe our review process results and discuss the most relevant findings and opportunities for future research. Finally, the last section concludes with a summary of theoretical and practical implications, and our research's limitations and contributions.

9.2 Research Background

9.2.1 AI-based Conversational Agents (CAs)

Over the past years, the field of Artificial Intelligence (AI) has evolved from a technical trend to a ubiquitous phenomenon in our everyday life (Maedche et al., 2019). Nowadays, AI-based systems diffuse various application domains and contribute to multiple innovations (L. Wang et al., 2020). AI capabilities cover perception, learning and acting, leading to widespread applications of AI technologies (Bawack et al., 2019). One of today's application domains is AI-based CAs. Although AI is experiencing unprecedented growth, the idea of communicating with computers via natural language has already existed for several decades. In 1966, Weizenbaum took initial steps towards a natural language text-interface between humans and computers with *ELIZA* (Weizenbaum, 1966). Since then, numerous CAs have been developed (Brandtzaeg & Følstad, 2017; Gnewuch et al., 2017). Initially, these systems' speech comprehension remained rather rudimentary and lacking intelligence (Diederich et al., 2019a). However, due to growing digital interconnectedness and technical advances, CAs have diffused to private households and companies (Berg, 2015; Dale, 2016; Feng & Buxmann, 2020). Reasons for the diffusion of CAs include the intuitive communication channel for employees and a low technical burden to set up a rudimentary CA implementation for companies (Diederich et al., 2019b; Riikkinen et al., 2018; Xu et al., 2017). Consequently, CAs have appeared as entirely new social actors in recent years, e.g., in enterprise messengers or voice-based assistants in a broader wave of companies' ongoing digitization and agile transformation, leading to new forms of collaboration (Maedche et al., 2019). CAs promote a new form of flexibility, quality, speed, and personalization. They are a scalable (24/7) and cost-effective solution, reducing the number of tasks manually executed by employees (Gnewuch et al., 2017; Stoeckli et al., 2019).

Due to their increasing popularity and the growing academic interest, researchers and practitioners have proposed various definitions and taxonomies of CAs (Feng & Buxmann, 2020; Meyer von Wolff et al., 2019a). CAs can be divided, for example, into text-based CAs—usually referred to as chatbots—or speech-based CAs such as smart assistants (Gnewuch et al., 2017; Winkler & Söllner, 2018). From a technical perspective, the distinction is marginal as speech-based input can easily be transferred to text-based input and vice versa (Diederich et al., 2019b). However, CAs can be defined as *"an application system that provides a natural language user interface for the human-computer-integration. It usually uses artificial intelligence and integrates multiple (enterprise) data sources (like databases or applications) to automate tasks or assist users [employees] in their (work)* *activities"* (Meyer von Wolff et al., 2019a, p. 96). In this study, we utilize the term CA to describe all AI-based software systems that communicate with users, both employees and customers, via a natural language interface provided by NLP/NLU technologies, such as via CA frameworks like RASA.ai including an intelligent communication and built-in self-learning component.

The main difference and novelty of AI-based CAs compared to conventional IS consist of two aspects: First, employees interact with the system via a natural language as a new socio-technical application class (Maedche et al., 2019). Second, *"the assistant's knowledgeability and human-like behavior, often summarized as artificial intelligence"* (Knote et al., 2019, p. 2025), bear a great potential to assist, solve, or automate employees' tasks intuitively. Contradictorily to the classification of AI-based CAs as IS from a technological perspective, the literature reveals that CAs' organizational adoption must be regarded in fundamentally different ways (e.g., Corea et al., 2020; Essaied et al., 2020): First, CAs are a new kind of IS augmented by a social user-centric and intelligent component extending companies' IT landscape (Stoeckli et al., 2019). Opposed to other IS and AI-based Systems, they occur as *social actors* with increasing intelligence, autonomy, and personality. They interact with employees in various new use cases and ways, learning from the collaboration and increasingly make their own decisions (Seeber et al., 2020). Users can establish relationships with CAs as "Teammates" (Bittner et al., 2019) and benefit from their ability to integrate information from different data sources to highly structured knowledge (Maedche et al., 2019; Meyer von Wolff et al., 2019a). However, to the best of our knowledge, research that examines the integration of such new *social actors* into existing business processes, actors and stakeholder structures, and enterprise workflows are scarce. Research possesses potentials for the design of company- and user-based processes and establishes a way to empower employees to act as "CA experts" (Zierau et al., 2020a).

Second, CAs can be classified as a new variety of *unfinished self-learning IS* depending on ongoing (software) development. Brendel et al. (2020a, p. 2) describing that adopting *"a CA in a service system still represents a major challenge and becomes increasingly complex as high user expectations, such as adaptive interaction behavior, can often not be fulfilled."* CAs often have limited skills initially, and learning progress depends on the application area and actors' engagement to train these systems. The literature already suggests that CAs, e.g., in the context of enterprise messengers (Stoeckli et al., 2019), take in a bottom-up and group-oriented perspective of voluntary participation instead of top-down enforcement. Consequently, besides a positive attitude toward CAs, including enjoyment, usefulness, trust and perceived intelligence (Pillai & Sivathanu, 2020; Zierau et al., 2020b), ambition is needed to take part in a continuous improvement process

(Stieglitz et al., 2018) instead of adverse psychological outcomes, such as a felt loss of autonomy or job insecurity in the context of AI.

Finally, CAs are AI-based IS. Although AI implementation in companies is omnipresent in research and practice, no systematic understanding of AI technologies exists concerning their definitions, application potentialities, and limitations for companies (Bawack et al., 2019). In this regard, there is a lack of research providing a holistic overview or an overarching framework of success factors in the context of long-term adoption of AI systems (e.g., regarding "the adoption like the implementation of the technology into organizational processes and governance structures" (Pumplun et al., 2019). Essaied et al. (2020) describe that studies made first investigations in the organizational adoption of AI without focusing on any particular type of AI solution. However, exposed aspects of the overarching AI research can be transferred to CAs for organizational adoption (e.g., learnings from privacy and ethics), but others cannot, as CAs are a separate application class with their own characteristics.

9.2.2 Technology Adoption Research and Framework

Despite CAs' increasing diffusion in companies, less is known about factors that need to be considered for the organizational implementation and adoption of this novel class of AI-based information systems. Similarly, no approaches exist guiding practitioners on how to manage this class of IS. Concerning the adoption, CA research mainly focuses on an individual user level, either on perceived trust and affordance theory (e.g., Zierau et al. (2020b), Stoeckli et al. (2019)) or in the wider context of IS acceptance theories, for instance, the "Technology Adoption Model" (TAM, e.g., Pillai and Sivathanu (2020)), "Adoption Use and Impact Framework" (AUI, e.g., Bawack et al. (2019)), "Unified Theory of Acceptance and Use of Technology" (UTAUT, e.g., Laumer et al. (2019b)).

However, the studies above focus on technology adoption and usage theories frameworks and present no feasible and structured overview of organizational-level impacts. A structured overview including a discussion ranging from the adoption into existing organizational service systems, company-based and user-based processes, actor and stakeholder structures, or overarching enterprise workflows and governance structures could not be identified. In this concern, only two research contributions could be identified dealing with organizational CA adoption from a more superficial and aggregative perspective: The authors Essaied et al. (2020) made first steps, explaining the organizational adoption of AI-based CAs by insurance companies regarding the role of "organizing vision" and "technological frame". Additionally, Corea et al. (2020) deriving guidelines for organizations and state factors that should be considered in adopting chatbots from a practice-based and managerial viewpoint. Nevertheless, they do not integrate their findings into a theoretical adoption framework. To the best of our knowledge, no literature review exists that aggregates the scattered literature landscape. We close this research gap and provide overarching factors that include our findings into a theoretical framework, presenting a starting point for giving an overview and guiding future organizational-focused studies.

This study builds upon the Technology-Organization-Environment (TOE) framework by Depietro et al. (1990). The TOE framework is an organization-level theory for structuring constraints and opportunities that influence the adoption of an IT innovation within a company. TOE has been widely applied in other technological domains, such as in "cloud computing" as well as "knowledge management systems" (Pumplun et al., 2019). In this study, the framework is applied for generating insights for technical and organizational factors that influence the adoption of CAs—for example, regarding changes for employee's communication structure or their skills. The framework additionally considering environmental factors such as regulation that constrain technology adoption, such as ethical and privacy-related factors in the context of AI-based systems. We selected TOE as a conceptual starting point to code and structure the identified contributions along the TOE dimensions. In addition, by identifying and aggregating factors, we would like to emphasize, in a contrasting line of argument, in which contexts contemporary knowledge about IS adoption may not apply and enhance the TOE Framework by new components.

9.3 Literature Review Process

To identify factors for the organizational adoption of CAs *(RQ)*, we conduct a systematic literature review (SLR) in five steps following vom Brocke et al. (2009).

9.3.1 Definition of Review Scope

First, we define our SLR scope based on the taxonomy proposed by Cooper (1988). We focus on research outcomes, practices and applications of CAs *(Focus)*. Our goal is to identify, integrate and aggregate central issues in how CAs can be implemented and adopted in an organization in contrast to other IS *(Goal)*. We take up a neutral position to delineate existing scientific contributions and synthesize selective, representative research outcomes *(Perspective and Coverage)*. A conceptual organization is chosen to cluster the existing research contributions *(Organization)*. Furthermore, the literature review is targeted at an audience holding a specialized background knowledge in IS research and practitioners dealing with the introduction of CAs in an organizational context *(Audience)*.

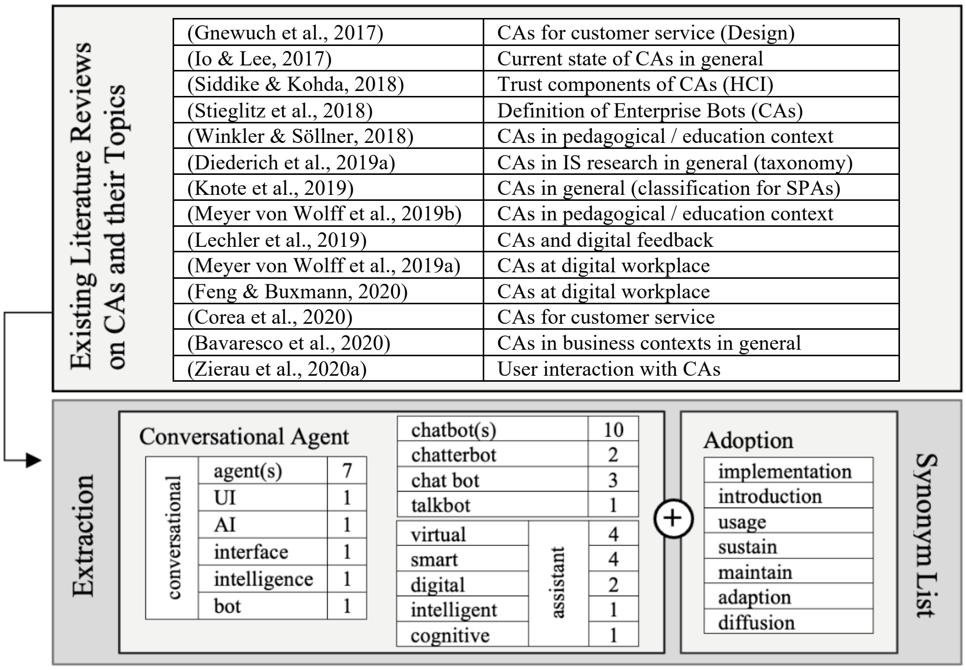

Figure 1. Keyword extraction process

9.3.2 Conceptualization of the Topic

In the second phase, the conceptualization of CAs and their organizational adoption, we rely on the initial definitions and terms introduced in the previous Section **Research Background**. Building upon this foundation, we conduct a first unstructured search considering literature reviews covering CAs from diverse perspectives (as outlined in **Figure 1**) to assemble frequently used synonyms for the term *conversational agent*.

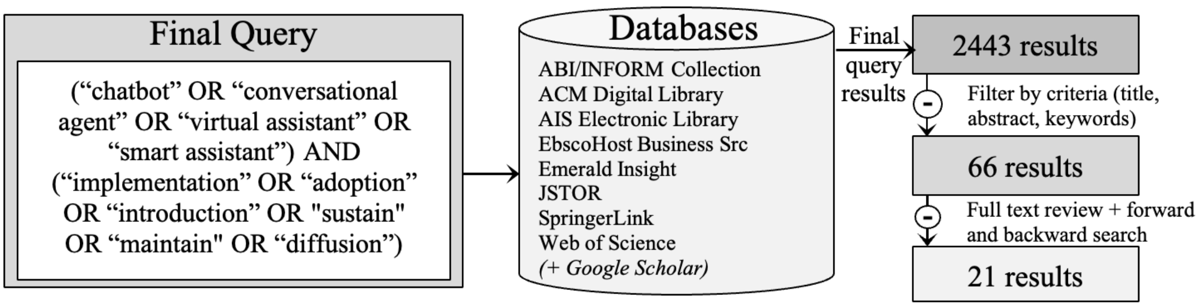

Figure 2. Research process

In addition, in order to deduce possible keywords for the topic of the implementation and adoption of CAs, we construct a search term list according Brink (2013) consisting of synonyms. In the next phase, "literature search", both keyword collections were utilized for creating and refining our search queries. Afterward, the individual search terms were iteratively improved and supplemented. The final synonym list is shown in **Figure 1**.

The first literature search in databases (listed in **Figure 2**) revealed that existing literature reviews investigating the topic are not focused on adopting CAs in an organizational context compared to other IS. The existing reviews deal predominantly with definitions, (historical) overviews, taxonomies and DPs for CAs or their application in specific domains, including advantages and challenges. None of the identified studies provides insights into the state of research of an organizational adoption.

9.3.3 Literature Search, Analysis and Research Agenda

To determine the current research state, we conducted a keyword search according to the review scope in December 2020. Ensuring a proper level of quality towards the selected relevant publications in advance, we include only peer-reviewed articles in selected databases (DB). **Figure 2** displays the search query that was iteratively derived from the conceptualization and the analyzed DBs.

We limit our search to the period between 2015 and 2020, due to the increasing interest in academics and the advances on CAs in recent years (Feng & Buxmann, 2020). After a full-text search, we obtain 2443 contributions. We filter articles based on their title, keywords and abstracts and include or exclude articles which (a) concern the adoption of CAs in organizations as opposed to the pure consideration of the interaction design, technology acceptance, usage theories or technical implementation and (b) concern an organizational setting as opposed to other settings, e.g., a personal setting. The resulting corpus embraces 66 contributions. Lastly, a full-text review, the exclusion of duplicates and a forward and backward search including using Google Scholar were conducted, adding further contributions to the set of relevant articles. The final corpus of relevant articles contains 21 articles, which we analyze and synthesize. To do so, we code the identified literature along the TOE dimensions, following an inductive approach. TOE sets the starting point for determining an initial list of master codes and descriptions. The initial set of codes was continuously adapted independently by two researchers in *MAXQDA software* on the articles. Afterward, both researchers create a concept matrix according to Webster and Watson (2002) to summarize their insights, which were discussed and aggregated in a next joint step, to deduce the model's components presented in the following Section. Finally, we identify emergent trends and research gaps that will be discussed in the Section **Discussion.**

9.4 Implementation and Adoption Factors for CAs

In summary, the analyzed articles assert that many studies on CAs focus on one specific aspect, such as the design or the technical implementation. Extant literature possesses a lack of contributions regarding organizational-level approaches for adopting CAs, for example, a framework that guides them into organizational processes and governance structures. Some articles deal with a general AI adoption and extract first conclusions (e.g., Pumplun et al., 2019), however, without focusing on any particular type of AI solution (Essaied et al., 2020). Although initial AI findings can be transferred to CA adoption, e.g., including "Organizational Vision" and "Technological Frame" (Essaied et al., 2020) or learnings from privacy and ethics (e.g., Kruse et al., 2019), to the best of our knowledge, no framework has been established that provide and discuss an aggregative overview of learnings regarding the adoption of CAs in contrast to other (AI-based) IS systems. To answer the RQ with the literature corpus, we structure the identified factors for a CA adoption following the TOE framework (see **Figure 3**). Subsequently, we discuss the most relevant findings for successfully adopting CAs in the section below.

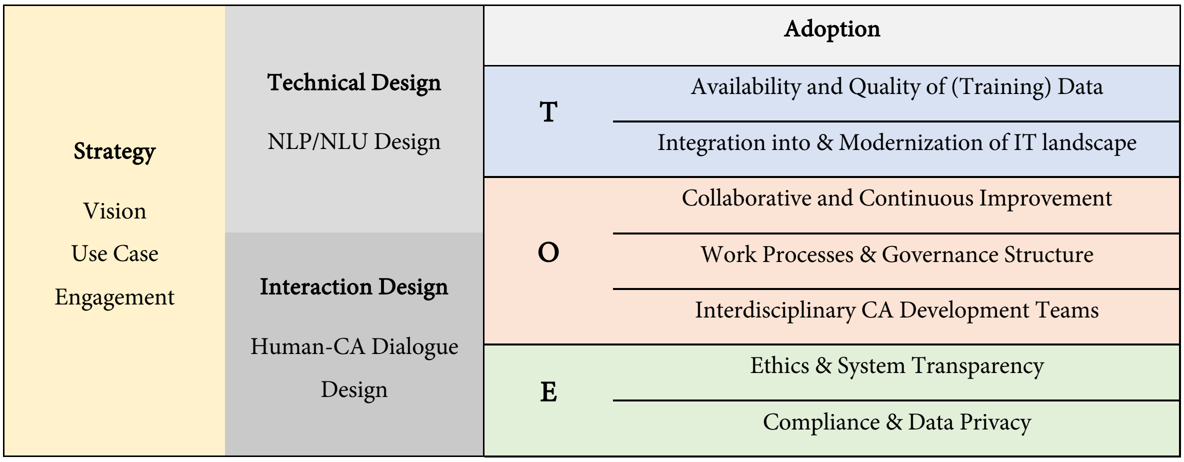

Figure 3. Adoption Factors for Conversational Agents

9.4.1 Foundation of CA Adoption

Strategy: The adoption of CAs needs, besides top management support to provide the required resources and to foster progress for adoption, a clear strategy and a new form of "organizational vision", including suitable use cases and understanding of employee readiness for CA usage and interaction (Essaied et al., 2020; Pillai & Sivathanu, 2020). On the one hand, the CA literature describes the need for a company-wide understanding of meaningful use cases and the CAs' technology capabilities ("clear problem understanding") to build trustful expectations and provide resources in the long term (Corea et al., 2020; Meyer von Wolff et al., 2020b). On the other hand, it is essential to engage employees to use this novel *social actor* as they are familiar or skeptical to interact via natural language with an *unfinished self-learning IS* (Stieglitz et al., 2018). The CA adoption relates to the employees' attitude towards AI-based CAs (Meyer von Wolff et al. 2020b; Stieglitz et al. 2018). Many employees are not familiar with CAs and often have a skeptical or adverse attitude to use them for various reasons, such as loss of autonomy, negative psychological outcomes as job insecurity or privacy and security concerns (Laumer et al., 2019a; Lu et al., 2020). To enable employee's willingness to adopt, includes a CA-related up-skilling and the explanation of employee benefits such as reduced routine work, enhanced productivity and job satisfaction through augmentation and cognitive relief (Lu et al., 2020).

Technical Design: Contrary to all expectations, many CAs cause frustration and have already disappeared from companies (Gnewuch et al., 2017). CAs often did not understand text inputs because of inadequate NLP capabilities (Brendel et al., 2020a). In addition to a clear strategy, the design of the NLU component and continuous training sets the foundation for adoption that expands CA development from traditional IS development as a project setting to a lifecycle environment (Abdellatif et al., 2020).

Interaction Design: CAs support employees in the process and task execution, providing a novel form of qualitative assistance towards daily work tasks (Brendel et al., 2020a; Elshan & Ebel, 2020). Consequently, they occur as new *social actors*, exhibit human-like characteristics, interact with employees, learn from the collaboration, and increasingly make their own decisions (Seeber et al., 2020). Besides the technical foundation, our SLR revealed that a human-like interaction design is required to foster the adoption of CAs, leading to better task performance and effective and trustful user-centric interaction (e.g., Gnewuch et al., 2017; Seeber et al., 2020).

In summary, the foundation comprises a clear strategy and vision, and CA use cases, which consider a distinct problem domain, technological possibilities and organizational readiness to adopt this new type of IS. Consequently, the technical and interaction design represent essential prerequisites for successful adoption. However, the SLR revealed that extensive research exists regarding the improvement of technical aspects (e.g., "NLP algorithms" and "architectures" (Gnewuch et al., 2017; Hu et al., 2018) or towards the design of CAs, such as the user-dialogue or design elements including human-like characteristics (e.g., avatar design, small talk, typing dots) and their effect on selected user perceptions and adoption (Jain et al., 2018; Janssen et al., 2020; Zierau et al., 2020a). Nonetheless, the long-term adoption of CA technology innovation involves more factors, which are presented below.

9.4.2 Technical Adoption Factors for CAs

Compared to traditional IS adoption, data availability and quality are of importance for adopting almost all AI applications (Pumplun et al., 2019). For CA development, a wealth of **(training) data availability and quality** is required to improve the NLP component representing the core of a good CA implementation (Sousa et al., 2019). For this purpose, the *development of NLP-ready data sets* can be started in a pre-step. Although companies often possess large amounts of available data regarding the selected use case, it must be clarified where sufficient data with an appropriate quality originates and must be transformed into dialogue-capable data (Meyer von Wolff et al., 2020b). The additional effort to train NLU components distinguishes CAs from other AI systems (e.g., analytics). Since CAs are *unfinished self-learning IS*, their success depends on ongoing data retrieval in the context of model and dialogue training and knowledge expansion (Meyer von Wolff et al., 2020b), describing a *shift from traditional software programming to more software training* (Zierau et al., 2020a).

To ensure that a CA presents a convenient platform that allows employees to access knowledge and perform tasks, it must be **integrated into the back-end and front-end systems** (Meyer von Wolff et al., 2020a; Pfeiffer, 2020). This integration ranges from enterprise messengers or self-service platforms (Meyer von Wolff et al., 2020b) to corporate landscape by connecting the CAs with other IS via internal and external APIs (Gnewuch et al., 2017; Sousa et al., 2019) to account for seamless user experience (Pereira & Díaz, 2018). In a practical realization, the CA represents a unified overarching platform that integrates different information bases seamlessly into new knowledge (Corea et al., 2020). To ensure up-to-date performance, a **transformation towards a modern IT architecture** is required. Instead of enriching traditional enterprise IS with CAs, a more radical step can rebuild the IT architecture to avoid bimodal structures (Stoeckli et al., 2019). Some authors suggest that AI-based systems, such as CAs require, e.g., a *microservice architecture*, to avoid complex and slow data interfaces (Kruse et al., 2019). The technical modernization can be realized in the context of an overarching AI strategy to achieve even higher task automation, such as Robotic Process Automation (Burns & Igou, 2019; Pfeiffer, 2020).

9.4.3 Organizational Adoption Factors for CAs

In contrast to traditional IS, which are mainly top-down driven (Stoeckli et al., 2019), the interaction via natural language with an *AI-based socio-technical actor* is an entirely new situation for the employees (Stieglitz et al., 2018). As CAs are *unfinished self-learning IS*, they have limited skills initially, often leading to a situation where high user expectations cannot be fulfilled (Gnewuch et al., 2017). To overcome this problem, CA adoption requires **new collaborative development and improvement approaches** where employees, the CA and developers continuously interact with each other. In general, AI adoption literature reveals that AI projects cannot be outsourced or exclusively operated in a separate lab isolated from the rest of the organization (e.g., Pumplun et al., 2019). Regarding CA adoption, most of the analyzed articles recommend a bottom-up CA adoption. CAs are tailored stepwise to work tasks, enhancing employees' skills and helping them in their daily routines (e.g., Stoeckli et al., 2019). Besides building a positive attitude toward CAs such as enjoyment, trust, and perceived intelligence (Pillai & Sivathanu, 2020; Zierau et al., 2020b), ambition is a prerequisite to a *continuous improvement process* (Stieglitz et al., 2018). The long-term adoption of these self-learning agents requires proactive feedback and error correction, and an agile mindset with a high degree of personal responsibility (de Lacerda & Aguiar, 2019; Pumplun et al., 2019; Stoeckli et al., 2019). In this context, the *training of employees* regarding CAs is one of the main factors for the successful development of CAs (Burns & Igou, 2019; Meyer von Wolff et al., 2020b). Building on the fact that employees possess the most knowledge of a CA's specific application domain (Meyer von Wolff et al., 2020b), integrating them in the development and continuous improvement process determines CA quality and success. In contrast to conventional IS, employees occur not only as software users. They are ongoing *"knowledge integrators"* determining the success through participation and *"caretaking" instead of top-down enforcement* (e.g., Lu et al., 2020; Stoeckli et al., 2019).

Further, successful CA adoption requires **integration into existing governance structures and business work processes (Corea et al., 2020).** To explore technological capabilities, conversation scenarios, and user acceptance with CAs, lab settings can be used as a starting point (Meyer von Wolff et al., 2020b). However, to prevent the development of "black boxes" isolated from the rest of the company, the adoption requires examining affected processes, including context awareness and compatibility with existing work structures to introduce *process-aware CAs instead of simple dialogue systems* (Corea et al., 2020). One mechanism is the *intertwined design of technology, interaction and business work processes*. In fact, that CAs appear as new *social actors*, initially performing simple and repetitive tasks such as pre-assessing requests, however, there will still be unpredictable requests that CAs cannot answer, which requires human intervention and, therefore, a “handoff event” process design (Corea et al., 2020; Poser et al., 2021). Through the application domain worker's takeover, the CA can be gradually improved through continuous training together with the developers.

Building on this integrated perspective, the literature suggests establishing **interdisciplinary CA development and improvement teams** (Abdellatif et al., 2020; de Lacerda & Aguiar, 2019). CA development teams need various specialists, such as conversation and UX designers, data scientists, DevOps engineers, data protection, ML, and security experts, which exceed traditional software development teams (Abdellatif et al., 2020; de Lacerda & Aguiar, 2019). In addition to technical expertise, the CA development team must possess soft and collaborative skills. As the success of a CA adoption depends on *its near-field training capabilities*, a deep understanding of the problem domain and the availability of specific training data are required (Følstad & Brandtzæg, 2017).

9.4.4 Environmental Adoption Factors for CAs

For many employees, both on the user and developer side, CAs are described as “black boxes,” particularly if proprietary frameworks are used (Abdellatif et al., 2020; Maedche et al., 2019). Especially in the context of AI-based CAs, examples such as Microsoft Tay (Lee, 2016) or Alexa are often highlighted as well-known examples for violating a code of conduct or intelligent virtual interfaces that are “always listening” and recording conversations (Burns & Igou, 2019; Manseau, 2020). This leads to a diminished reputation and trust at the onset of adoption if users “perceive a lack of system interpretability” (Maedche et al., 2019). To increase employees' trust in CAs, it is also essential to create **system transparency and explain how self-learning CAs work** (e.g., education and enlightenment of the employees) (Laumer et al., 2019a; Maedche et al., 2019). In this context, some articles also suggest instituting **an *“ethical code of conduct”* to build trust in their usage (e.g., Seeber et al., 2020).** Additional mechanisms to increase transparency and trust are legal regulations, such as **data protection guidelines** (e.g., Rodríguez Cardona et al. (2019); Burns and Igou (2019)). AI adoption literature states that AI operation requires massive data analytics, often processed in decentralized data centers. Thus, protection issues have been identified, such as unauthorized access to business data (e.g., Kruse et al., 2019). However, CAs have unique characteristics: Due to their dialogue-based nature, they are equipped with an intelligent conversational memory such as conversation user chats to improve the CAs’ capabilities (Sousa et al., 2019). Because CAs can map whole companies' communication processes (Janssen et al., 2020), they are often under special surveillance through data-protection. The GDPR and the employees' council must be engaged in the adoption process. For example, employees' chats are monitored by metrics to improve CAs' interactions (e.g., Pereira and Díaz (2018)). This leads to employees' skeptical attitude, often accompanied by a heightened perception of privacy risks (Laumer et al., 2019a; Manseau, 2020). Extant studies recommend *clear data protection and usage policies* for a CA introduction, indicating how and where personal data are processed (Burns & Igou, 2019; Corea et al., 2020; Manseau, 2020). A balanced IT architecture, consisting of external providers and in-house operation for personal data, helps keep sensitive data within the organization's responsibility (Meyer von Wolff et al., 2020b).

9.5 Discussion

Compared to conventional IS, the impact of AI systems from an organizational adoption perspective is insufficiently studied (L. Wang et al., 2020). Only a few articles deal with AI adoption in general on an abstract level (e.g., Kruse et al., 2019; Nascimento et al., 2018; Pumplun et al., 2019). Moreover, although some of these findings are valid for CA adoption, our SLR emphasizes that the adoption of AI-based CAs in organizations needs a dedicated organizational perspective due to specific characteristics. CAs represent a novel form of IS occurring as *intelligent, autonomous social actors* and *unfinished self-learning IS,* which depends on *ongoing collaborative development and improvement approaches*. However, no studies have focused on organizational adoption of specific AI solutions, such as CAs (Essaied et al., 2020). Our research addresses this gap, outlines the current state-of-the-art within the segregated academic literature for CA adoption. This provides an aggregated perspective of learnings in contrast to general AI-based- and traditional IS adoptions.

We were able to deduce coherent factors that contribute to the overall utilization of CAs in an organization and guide practitioners in CA adoption. In addition to challenges that need to be considered in adopting traditional IS (Meyer von Wolff et al., 2020b), CAs bring supplementary hurdles that cannot be solved in a purely top-down, large-scale and project-orientated manner (Stoeckli et al., 2019). CAs gained adverse attitudes in recent years among employees due to a limited language understanding and skill level. Furthermore, employees have ethical and data protection concerns, including the heightened perception of privacy risks, leading to CAs non-use. At the beginning of the introduction, concerns often overwhelm CAs benefits, leading to an obstacle since CAs can only be improved with extensive and continuous training effort. While traditional IS adoption is influenced, for example, by top management support, resource allocation, and the integration into the existing processes and IT service landscape (Depietro et al., 1990), the focus for

CAs shifts towards a more *employee-depended and data-centered adoption*. CA success requires a lot of persistence and long-term arrangements the management should actively promote, including collaborative and continuous development approaches between IT departments and affected business units. Consequently, the adoption of a CA constitutes a disruptive change for employees (Manseau, 2020). As employees occur not only as software users, rather as ongoing *"knowledge integrators"* determining the success through participation and *"caretaking"* (e.g., Lu et al., 2020; Stoeckli et al., 2019), new approaches are needed to engage employees in the ongoing development process. One mechanism is demonstrating the enhanced productivity and cognitive relief of CAs by providing and demonstrating progressively new functionalities towards employees. To establish a CA strategy, we recommend that "*knowledge integrators"* be determined to guide the CA adoption. In this context, more *design-oriented knowledge such as patterns is needed* that lead the collaborative adoption cycle. The identified factors give no guidance on an appropriate chronological order or indicators that define system maturity. Further studies must elaborate processes and models involving developers, "*knowledge integrators"*, and other employees to enable a step-by-step introduction of CAs. This includes a chronological order of the identified factors, measurement tools to define system maturity and CA standards for data analysis.

Besides, as CAs are self-learning IS and map whole companies' communication processes (Janssen et al., 2020), precise usage, ethics and privacy standards must be established as a baseline to generate responsibility, transparency and clear commitments towards employees to prevent adverse feelings. Although research has already been conducted in the broader field of AI and ethics and discussions at an abstract level are taking place in the public (Jobin et al., 2019), our SLR reveals a lack of research on CA-specific guidelines. As NLP training required as much wealth training data as possible (Sousa et al., 2019), it is of interest how to deal with the data acquisition with concern to data minimalism (e.g., which amount of data is needed to provide a good NLP training without violating the privacy of the employees).

Finally, most literature discussed CAs regarding an isolated conceptual or technical perspective, often in laboratory settings and within a specific research domain, such as HCI, CS, or IS (Diederich et al., 2019a; Janssen et al., 2020). Since many CA projects fail because CAs emerge from lab-settings, and the problem to be solved is imprecise or isolated from real processes (Corea et al., 2020), the need for a collective awareness is reinforced. For CA adoption, an integrated perspective of technological, organizational, and environmental factors within a naturalistic context is required. More design science-oriented research or entrepreneurial approaches are needed (Peffers et al., 2007; Ries, 2011) to pilot AI-based CAs in socio-technical environments (Briggs et al., 2019). For instance, a handoff design (e.g., Poser et al., 2021) between a CA and employee will require expertise to develop the technical solution and novel business processes. Therefore, we recommend interdisciplinary studies, considering a comprehensive design perspective (e.g., business process and CAs' dialogue design).

9.6 Conclusion

AI-based CAs are becoming increasingly important in organizations and research, resulting in new application areas and research studies. Nonetheless, the current body of research knowledge is highly fragmented and scattered across various research streams such as IS, HCI, or CS domains with often specific CA applications (Zierau et al., 2020a). These applications often disregard CAs long-term success of adoption (Brandtzaeg & Følstad, 2017; Dale, 2016; Gnewuch et al., 2018). To address this research gap, we conduct a SLR to examine the state-of-the-art toward adopting CAs in organizations. Our work contributes to the study of CAs by providing researchers and practitioners with an aggregative perspective of technological, organizational and environmental factors that need to be considered. In addition to this analysis of the literature corpus, we discuss observations regarding differentiation to traditional IS and general AI-based systems.

In sum, the introduction of **CAs is much more than a "classic software introduction."** First, compliance and system transparency are of importance, owing to this new type of IS. Second, employees have to be trained and convinced to overcome the barrier of system usage. Third, employees must understand that CAs will remain an *unfinished self-learning IS* for a long time. The adoption success depends on their cooperation and acceptance of their role as continuous "*knowledge integrators*". Both the management and employees must understand CAs as an entirely new type of IS. Training is a crucial factor for the successful adoption of CAs in organizations. Therefore, it is essential to motivate employees to engage and educate new improvement approaches. Consequently, a holistic approach and form of work cooperation are required. An isolated development or outsourcing could lead to failure. This study is faced with some limitations. First, we identify few contributions concerning factors organizations need to consider in case of a CA adoption. Second, some of the identified literature deals with AI technology adoption in general and requires further research inquiries to validate its transferability to self-learning systems like CAs. Finally, we recognize that this study's results are dependent on the literature selection, integration and judgment of the authors. To achieve a more solid academic underpinning, we recommend further research with empirical investigations to gain more in-depth knowledge on factors for successful adoption and long-term usage scenarios of CAs in companies.

9.7 Acknowledgements

The research was financed with funding provided by the German Federal Ministry of Education and Research and the European Social Fund under the "Future of Work" program (INSTANT-BOEHMANN, funding reference 02L18A111).

9.8 References

Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020). Challenges in chatbot development: A study of stack overflow posts. International Conference on Mining Software Repositories (MSR).

Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., & Vanzin, M. (2020). Conversational agents in business: A systematic literature review and future research directions. Computer Science Review, 36.

Bawack, R. E., Fosso Wamba, S., & Carillo, K. (2019). Artificial Intelligence in Practice: Implications for IS Research. Americas Conference on Information Systems (AMCIS).

Berg, M. M. (2015). NADIA: A simplified approach towards the development of natural dialogue systems. International Conference on Applications of Natural Language to Information Systems.

Bittner, E. A. C., Oeste-Reiß, S., & Leimeister, J. M. (2019). Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. International Conference on Internet Science (INSCI), Cham.

Brendel, A. B., Greve, M., Diederich, S., Bührke, J., & Kolbe, L. M. (2020). You are an Idiot!–How Conversational Agent Communication Patterns Influence Frustration and Harassment. Americas Conference on Information Systems (AMCIS).

Briggs, R. O. B., Tilo, Schwabe, G., & Tuunanen, T. (2019). Advancing Design Science Research with Solution-based Probing. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Brink, A. (2013). Anfertigung wissenschaftlicher Arbeiten: Ein prozessorientierter Leitfaden zur Erstellung von Bachelor-, Master-und Diplomarbeiten (5th ed.). Springer Gabler, Wiesbaden.

Burns, M. B., & Igou, A. (2019). "Alexa, write an audit opinion": Adopting intelligent virtual assistants in accounting workplaces. Journal of Emerging Technologies in Accounting, 16(1), 81-92.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104–126.

Corea, C., Delfmann, P., & Nagel, S. (2020). Towards intelligent chatbots for customer care-practice-based requirements for a research agenda. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Dahl, D. A. (2013). Natural language processing: past, present and future. In Mobile speech and advanced natural language solutions (pp. 49-73). Springer.

Dale, R. (2016). The return of the chatbots. Natural Language Engineering, 22(5), 811-817. https://www.doi.org/10.1017/S1351324916000243

de Lacerda, A. R. T., & Aguiar, C. S. R. (2019). FLOSS FAQ chatbot project reuse - how to allow nonexperts to develop a chatbot. International Symposium on Open Collaboration, Skövde, Sweden.

Depietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. The processes of technological innovation, 199(0), 151-175.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019a). On conversational agents in information systems research: Analyzing the past to guide future work. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019b). Towards a taxonomy of platforms for conversational agent design. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Elshan, E., & Ebel, P. (2020). Let's Team Up: Designing Conversational Agents as Teammates. International Conference on Information Systems (ICIS).

Essaied, K., Marsan, J., Bazomanza, W. N., Tewamba, H. J. N., & Yombia, S. M. K. (2020). Organizational Decision to Adopt Chatbot Technology: The Role of Organizing Vision and Technological Frame. Americas Conference on Information Systems (AMCIS).

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42. https://dl.acm.org/doi/fullHtml/10.1145/3085558

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. European Conference on Information Systems (ECIS), Portsmouth, United Kingdom.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. International Conference on Information Systems (ICIS), Seoul, Korea.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. Conference on Human Factors in Computing Systems (CHI), Montréal, Canada. http://dx.doi.org/10.1145/3173574.3173989

Io, H. N., & Lee, C. B. (2017). Chatbots and Conversational Agents: A Bibliometric Analysis. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. Designing Interactive Systems Conference (DIS), New York, NY, USA. https://doi.org/10.1145/3196709.3196735

Janssen, A., Passlick, J., Cardona, D. R., & Breitner, M. H. (2020). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. Business & Information Systems Engineering, 62(3), 211–225. https://doi.org/10.1007/s12599-020-00644-1

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying smart personal assistants: an empirical cluster analysis. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Laumer, S., Gubler, F. T., Racheva, A., & Maier, C. (2019a). Use Cases for Conversational Agents: An Interview-based Study. Americas Conference on Information Systems (AMCIS), Cancun, Mexico.

Laumer, S., Maier, C., & Gubler, F. T. (2019b). Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis. European Conference on Information Systems (ECIS).

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. European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden.

Lee, P. (2016). Learning from Tay's introduction. Lee, Peter. Retrieved 20.08.2020 from https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? Journal of Service Theory and Practice, 30(3).

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. Business & Information Systems Engineering, 61(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Manseau, J. (2020). AI in the Workplace: A Qualitative Analysis of Intelligent Employee Assistants. Americas Conference on Information Systems (AMCIS).

Meyer von Wolff, R., Hobert, S., Masuch, K., & Schumann, M. (2020a). 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), 3. https://doi.org/10.17705/1pais.12203

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2019a). How May I Help You?–State of the Art and Open Research Questions for Chatbots at the Digital Workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2020b). Einsatz von Chatbots am digitalen Büroarbeitsplatz–Eine praxisorientierte Betrachtung von Einsatzbereichen, Wirkungen und Handlungsempfehlungen. HMD Praxis der Wirtschaftsinformatik(57), 413-431.

Meyer von Wolff, R., Masuch, K., Hobert, S., & Schumann, M. (2019b). What Do You Need Today?- An Empirical Systematization of Application Areas for Chatbots at Digital Workplaces. Americas Conference on Information Systems (AMCIS), Cancun, Mexico.

Nascimento, A. M., da Cunha, M. A. V. C., de Souza Meirelles, F., Scornavacca Jr, E., & de Melo, V. V. (2018). A Literature Analysis of Research on Artificial Intelligence in Management Information System (MIS). Americas Conference on Information Systems (AMCIS), New Orleans, USA.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of management information systems, 24(3), 45-77. http://www.tandfonline.com/doi/abs/10.2753/MIS0742-1222240302

Pereira, J., & Díaz, Ó. (2018). Chatbot dimensions that matter: Lessons from the trenches. International Conference on Web Engineering.

Pfeiffer, J. (2020). Interview with Omer Biran and Michael Halfmann on “Conversational Agents at SAP”. Business & Information Systems Engineering, 62(3), 249-252.

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. International Journal of Contemporary Hospitality Management, 32(10). https://doi.org/10.1108/IJCHM-04-2020-0259

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A new organizational chassis for artificial intelligence - Exploring organizational readiness factors. European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden.

Ries, E. (2011). The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. Crown Publishing Group.

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. International Journal of Bank Marketing, 36(6).

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. Americas Conference on Information Systems (AMCIS).

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. Internet Research, 30(1), 1-18.

Semmann, M., Grotherr, C., Vogel, P., Bittner, E., Biemann, C., & Böhmann, T. (2018). Intelligent Collaboration of Humans and Language-Based Assistants (INSTANT). International Conference on Information Systems (ICIS), San Francisco, USA.

Siddike, M. A. K., & Kohda, Y. (2018). Towards a Framework of Trust Determinants in People and Cognitive Assistants Interactions. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Sousa, D. N., Brito, M. A., & Argainha, C. (2019). Virtual Customer Service: Building your Chatbot. International Conference on Business and Information Management (ICBIM), Paris, France.

Stieglitz, S., Brachten, F., & Kissmer, T. (2018). Defining Bots in an Enterprise Context. International Conference on Information Systems (ICIS), San Francisco, USA.

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2019). How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electronic Markets, 30, 369–403. https://doi.org/10.1007/s12525-019-00359-6 vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. European Conference on Information Systems (ECIS), Verona, Italy.

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2020). Effects of voice-based AI in customer service: Evidence from a natural experiment. International Conference on Information Systems (ICIS), India.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, 26(2), 8-23. <Go to ISI>://WOS:000176079000002

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

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. Academy of Management Annual Meeting (AOM), Chicago, USA.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. Conference on Human Factors in Computing Systems (CHI), New York, NY, United States. https://doi.org/10.1145/3025453.3025496

Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020a). A review of the empirical literature on conversational agents and future research directions. International Conference on Information Systems (ICIS), India.

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020b). Towards developing trust-supporting design features for AI-based chatbots in customer service. International Conference on Information Systems (ICIS), India.

10 Publication No. 2: Lewandowski et al. (2022a)

Lewandowski, T., Grotherr, C., & Böhmann, T. (2022). Managing Artificial Intelligence Systems for Value Co-creation: The Case of Conversational Agents and Natural Language Assistants

In: Edvardsson, B., and Tronvoll B. (Eds.), The Palgrave Handbook of Service Management (pp. 945-966). Springer International Publishing, Cham.

Abstract

Conversational Agents (CA) are a form of Artificial Intelligence that is increasingly used to support and automate service encounters. CAs are cost-effective service actors which enable new forms of service provisioning and value co-creation scenarios. Despite their potential, organizations struggle to leverage the potential of CAs in real-life settings. We analyze the nascent literature and give insights from a design science research project on the implementation of CAs in a service setting to identify challenges in the design, implementation, and operation of CAs in service systems. Using the lens of a multilevel framework for service systems, we present insights on how CAs can be designed and managed for value co-creation.

Keywords: Service systems engineering · Multilevel framework · Artificial intelligence systems · Conversational agents · Chatbots · Customer service

10.1 Introduction: AI in Service

In recent years, digitalization has spurred service innovation in manifold ways (Barrett et al., 2015; Chandler et al., 2019). Key drivers of this innovation are technological advances such as augmented reality (AR), machine-2-machine interactions and artificial intelligence (Lusch & Nambisan, 2015). Particularly, AI has had a lasting impact in many domains and organizations (L. Wang et al., 2020), as, for example, enhancing and automating technology in service encounters (Ostrom et al., 2019). The use of AI provides a new perspective to service contexts, essentially to *"provide value in service environments through flexible adaptation enabled by sensing, reasoning, conceptual learning, decision-making and actions"* (Bock et al., 2020, p. 317). The ongoing advancements in AI in the next few years will virtually transform all service sectors. They could even lead to AI-based systems becoming the most prevalent actors in service interaction (Xiao & Kumar, 2019). As a result, entirely new scenarios for value co-creation are conceivable (Bock et al., 2020).

Conversational Agents are one specific and prominent case of AI in service. CAs are automated, scalable, cost-effective service systems delivering services to actors through textual or auditory means (Bock et al., 2020), which enable new forms of service interaction and value co-creation scenarios. AI-based CAs can transform service encounters from human-centric to technology-dominant (Castillo et al., 2020). CAs influence conventional service offerings and enable individual and convenient interaction forms (Klaus & Zaichkowsky, 2020). From a service employees' perspective, CAs bear the potential to automate, augment and assist service interactions in, e.g., human-centered tasks by identifying new solution strategies, providing decision-making support, or problem-solving. In this regard, CAs possess significant potentials since service and interaction workers are exposed to occupational stress due to constantly increasing requests combined with the massive rise of information load (Semmann et al., 2018). From a customers' perspective, CAs also occur as a novel actor in the foreground in various customer support settings, which promote a new form of speed and personalize customer relationships. Consequently, CAs can appear as service actors between the provider and the customer, allowing novel value co-creation scenarios.

Despite the aforementioned potentials, increasing practical interest and high popularity of CAs in various domains of research in recent years, many contributions investigate CAs solely from a technical perspective, e.g., how to improve the natural language processing (NLP) component, or from a specific conceptual point of view, e.g., how to design a concrete dialogue flow often prototyped in a pre-service encounter stage, such as in lab and greenfield environments (Lu et al., 2020). However, as soon as they are instantiated in concrete service settings (e.g., as intelligent technology in the background or as a self-service platform in the frontline), they could often not meet expectations and already disappeared (Gnewuch et al., 2017). Thus, although a growing number of companies are adopting AI-based CAs in service settings, academics lag in studying its implications for service science (Bock et al., 2020; Lu et al., 2020).

Consequently, AI-based systems play a crucial role in digital-enabled service innovations; however, they cannot be discussed separately from the context such as the organization and environment. The service-dominant (SD)-logic perspective emphasizes the importance of analyzing actors such as customers, employees, and CAs within value co-creation processes. Following this view, CAs can be understood as socio-technical actors and active co-creators which implies a shift in designing such solutions beyond technological requirements.

The paper aims to broaden CA design's perspective beyond its currently technological dominant perspective by applying a service systems perspective. Our insights are based on the nascent literature on CAs and on insights gained through a design science research project on the implementation of CAs in real-life service settings. The research project seeks to develop and pilot novel interaction processes between customers, employees, and CAs (Semmann et al., 2018).

We structure our insights using Grotherr et al. (2018) multi-level framework for service system design. This framework builds on the tiered understanding of value co-creation and actor engagement posited by Storbacka et al. (2016) that links micro-level engagement activities to macro-level phenomena such as value co-creation and the associated institutional arrangements. We apply the Grotherr et al. (2018) framework as a comprehensive perspective to address technological design, work-/service design and institutional design in the context of CAs. This multilevel lens on service system design integrates DSR on CAs with service research.

10.2 Research Background

10.2.1 AI-based Systems

The advances in information technologies (IT) as an enabler and contributor within a highly dynamic environment are key characteristics in service innovation (Barrett et al., 2015). One innovative IT field that has evolved from a technological trend to a ubiquitous phenomenon in the service landscape is AI-based systems. Although AI has started to dominate our daily lives and academic interest has grown considerably, there is neither in science nor in practice a consistent definition of the term. Further, compared to traditional information systems (IS), the impact of AI systems in enterprise settings and for service systems is insufficiently studied (Bock et al., 2020; L. Wang et al., 2020). Typically, service researchers understand AI as a generic concept for a set of technologies capable of mimicking human behavior and learning how to solve tasks usually performed by human intelligence (Castillo et al., 2020). AI systems are often described as algorithms that operate not rule-based but, similar to the human brain, use cognitive or conversational functions and interact with a large amount of data. Nowadays, AI-based systems diffuse various application domains and contribute to multiple innovations (L. Wang et al., 2020). Service AI's field comprises configurations of technologies to provide value in internal and external service environments through flexible capabilities that cover perception and sensing, learning and acting (Bock et al., 2020). AI technologies include biometrics (e.g., computer vision), robotics, machine, and deep learning, as well as natural language processing (NLP).

10.2.2 AI-Based Conversational Agents

Conversational Agents are one specific and prominent case of AI in service, describing intelligent, automated and intangible systems delivering services to actors through natural language (Bock et al., 2020). Although the technical possibilities are not comparable with the current potentials in AI and data processing, the idea of communicating with computers has already existed for several years. Weizenbaum (1966) has previously taken initial steps towards an NLP component or, respectively, text-interface between humans and computers with *ELIZA*, a system generating responses to text inputs simulating a psychotherapist in a therapy session. Since then, various CAs with increasing capabilities and intelligence were developed (Brandtzaeg & Følstad, 2017), which will continue to rise in the upcoming years in service settings. At present, CAs exist in domains like customer service, marketing, entertainment and education and become significant in private households as well as in professional workplace contexts (Feng & Buxmann, 2020; Gnewuch et al., 2018). CAs promote a new form of flexibility, quality, speed, and personal ways to accomplish work tasks, access content and services (Wilson, 2018). A transformation towards convenient, automated, multi-lingual, globally available, 24/7 support channels are already conceivable today (Følstad et al., 2018a; Gnewuch et al., 2017).

As a result of their emerging interest in IS and service research, numerous designations, taxonomies and concepts have been formulated over the years. In service literature, CAs are also known under synonyms such as service robots (e.g., Lu et al., 2020; Wirtz et al., 2021), chatbots (e.g., Castillo et al., 2020), virtual assistants (e.g., Bock et al., 2020), or voice bots (e.g., Klaus & Zaichkowsky, 2020).

The subdivision is often made based on two dimensions ("Primary Mode of Communication", Gnewuch et al., 2017): The first class comprises *text-based CAs*, commonly known under synonyms such as chatbots or chatterbots (e.g., *ELIZA* or *Cleverbot*), while the second class embraces *speech-based CAs* as virtual or smart assistants (e.g., *Amazon Alexa* or *Apple's Siri*). However, the main concepts have remained principally the same due to a similar underlying architecture, and in many academic publications, no distinction is made (Meyer von Wolff et al., 2019a). Conversational agents can be defined as *"system-based autonomous and adaptable interfaces that interact, communicate, and deliver service to an organization's customers"* (Wirtz et al., 2018, p. 909). The service interface is built on NLP technologies, including intelligent communication and a built-in ML component, allowing the user to communicate via human languages. In this context, the CA become the dominant interaction partner and a new actor in the co-creation of value by representing the visible and customer-facing interface of large and integrated service systems (Wirtz et al., 2018). Thereby, they integrate "*multiple data sources (like databases or applications) to automate tasks or assist users [e.g., internal employees or external customers] in their (work) activities"* (Meyer von Wolff et al., 2019a, p. 96). For example, instead of consulting a support hotline, an employee can submit the support request via natural language to a CA directly, which serves as an instantaneous assistant actor by scanning diverse knowledge and data sources in the background and providing answers to requests.

Managing these systems comes with novel challenges that are different to traditional IT systems used in service organizations. CAs have distinct characteristics that differentiate them from traditional IS as well as from other AI-based technologies for service provisioning:

First, CAs are *social actors*. These AI-based systems influence conventional service offerings and enable new individual and convenient socio-technical interactions (Klaus & Zaichkowsky, 2020). Often, they undertake a social position in the service delivery process in terms of being consulted as "*user's friend and helper,*" providing quick and accurate solutions to customer requests via natural language (Bock et al., 2020). Developers often design chatbots very human-like. They were endowed with social features, provided with names, avatars, and communicative behaviors to attract users' attention and simulate a natural conversation (McTear et al., 2016). CAs learn from previous collaborations and increasingly make their own decisions augmented by a user-centric and intelligent component, extending service landscapes (Seeber et al., 2020; Stoeckli et al., 2019). Compared to conventional service provision in customer service settings, which consisted of a dyadic interaction between a customer and a service provider (representing the "face" of the organization), CAs will progressively represent the prevailing *customer-facing part of an extensive* *and integrated service system* (Ostrom et al., 2019; Wirtz et al., 2018). Therefore, CAs will transform all service sectors and need a new viewpoint on service management practices (Ostrom et al., 2019).

Second, CAs can be classified as *unfinished and learning IS*. CAs have few skills at the outset and can only engage in light-weight and simple initial tasks tended to be low in their cognitive and emotional complexity (Wirtz et al., 2021), while expectations among managers, employees, and customers are extremely high. However, AI-based CAs can be continuously trained and enhanced whereby they obtain access to increased amounts of data and are connected to diverse sources and systems in the IT and service landscape (Castillo et al., 2020; Xiao & Kumar, 2019). CAs benefit from a scaling effect. Progressively, allowing them to make more recommendations, decisions, and actions with little or no human intervention (Xiao & Kumar, 2019). However, until this state can be achieved, a new understanding and engagement of all service-involved actors are needed. CAs learning process depends on the commitment of the individual service actors (customer and client-side) since the CA will only improve when used. Compared to conventional IS, which are mainly instituted to support service delivery, CAs come to the forefront as an actor in a field of tension: On the one hand, ambition is needed to take part in a continuous improvement process (Stieglitz et al., 2018), and on the other hand, customers are skeptical about CAs use (due to, e.g., initially limited capabilities) and service employees can develop negative attitudes towards CAs (e.g., due to loss of autonomy or job insecurity). Since customers hold nearly similar expectations considering the service provision, e.g., regarding the service levels (Castillo et al., 2020), one question is how to manage CAs limitations directly from the beginning. CA-caused service failures could decrease service quality, resulting in customer resource loss ("value co-destruction" instead of collaboratively and interactively value co-creation). Literature calls for research on how to manage these *new social and unfinished form of IS* to hinder *"failed chatbots"* and, in this context, how to engage and develop employees and customers *"to play their enabler, innovator, coordinator, and facilitator roles in modern service encounters"* (Lu et al., 2020, p. 380).

Particularly these findings on the challenges and risks of implementing CAs in service settings highlight the need to approach the design of CAs from a broader socio-technical perspective. Therefore, we will present a framework of these management challenges and derive research implications on managing AI systems in service organizations for value co-creation in the next section.

10.3 Designing and Managing Conversational Agents in Service Systems: A Multilevel Framework

10.3.1 Multilevel Design Framework for Service Systems

To achieve a more comprehensive understanding of the design challenges for CAs in service settings, we make use of Grotherr et al.'s multi-level framework for service systems design (Grotherr et al., 2018). This framework bridges the gap between abstract value co-creation and observable actor engagement with design elements on multiple levels. With this framework, we seek to contribute to transdisciplinary research discourse for the design of digital service systems by providing a foundational work for the emergence of next-level design theory for CAs.

Over the last few years, there have been significant shifts in digital enabled business models and service systems. First, there is a movement away from a traditional perspective on services as single entities toward seeing value creation as a co-creative endeavor of multiple actors, resources, and systems of services. Second, opportunities can be found in the study of service systems which are (1) technology-enabled, (2) actor-centered, and (3) shaped by *institutions*. Nonetheless, organizations are challenged to design competitive service systems in a dynamic market. On the one hand, customer demands are dynamic, and due to the rapid growth of technological advancements, new digital innovations emerge. On the other hand, there is also a need to take *institutions* and their shaping and transforming role into account (Vargo et al., 2015). In other words, developing service systems implies two central aspects which must be reflected by elaborating existing design approaches: (1) to cope with volatile environments and (2) to take a perspective on the design of service systems reflecting socio-technical artifacts as well as *institutions*.

In this regard, service systems become meaningful when actors engage, mobilize their resources, and integrate them for value co-creation. To understand unexpected resource constraints or lack of cooperation, a value-in-context mindset is essential (Chandler & Vargo, 2011). However, the observability and measurability of the value co-creation process in service system design appear highly challenging (Storbacka et al., 2016). One solution approach that has gained acceptance in service research in recent years is the focus on actor engagement as a micro-foundation that is observable and measurable, and thus manageable. Actor engagement takes place on the *engagement platform* on the micro-level. These engagement practices represent actors' disposition to engage, leads to engagement activities, and are characterized by observable engagement properties (temporal, relational, informational) (Storbacka et al., 2016).

In this context, Grotherr et al. (2018) proposed a *multilevel design framework for service systems* that can facilitate a service system's analysis and design of components for actor engagement on the macro-meso-micro-level. Contemporary service design theories are built on stability assumptions, such as defining a priori problem. These traditional approaches lack consideration of dynamics and do not provide realistic means for understanding human actions in their environment. The shift from *plan-oriented process models* towards *path-dependent design systems* builds a substantial basis for exploring and exploiting digital, actor-centered service systems. Moreover, by applying a *multilevel perspective*, *institutions*, *technology*, *actors* and *resources* are captured in the design process. This approach emphasizes the design of individuals interaction facilitated with technological advancements on micro-level and broad adjustment of prevalent *institutions* on macro-level.

On the one hand, the framework helps drill down from an abstract perspective of value co-creation and guides value propositions to observable actor engagement. On the other hand, to aggregate the observational results and drill up to implications for the service systems design (Grotherr et al., 2018). On the other hand, to cope with various design interventions' complexity, the *multilevel design framework for service systems* consists of two intertwined design cycles: (1) *institutional design* and (2) *engagement design*. The *engagement design* comprises interactional and socio-technical components, which facilitate actor engagement on the micro-and meso-level. The *institutional design* refers to reflections made on the meso-level, which have implications to value propositions and the service systems' institutional environment (see **Figure 1**).

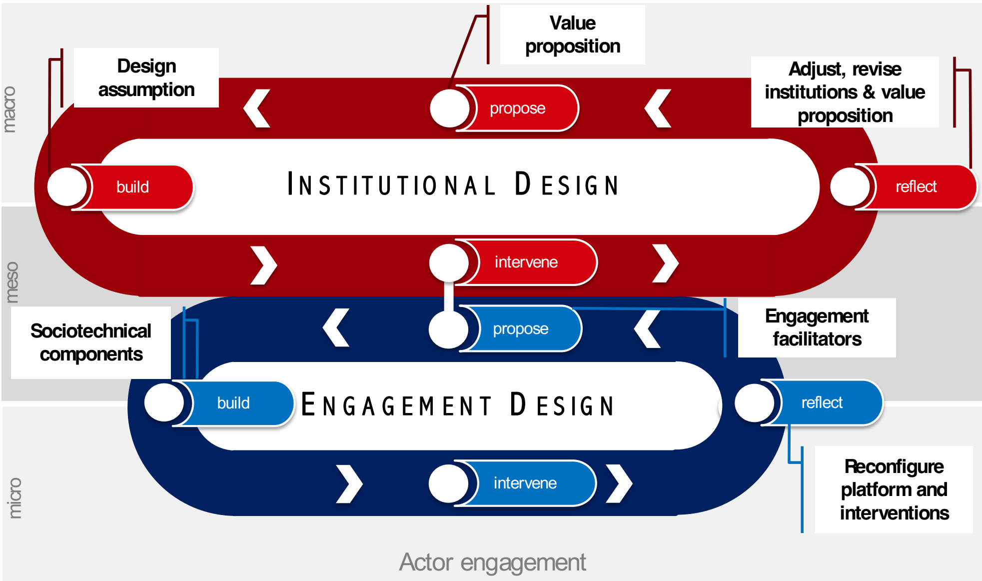

Figure 1. Multilevel Design Framework for Service Systems (based on Grotherr et al., 2018)

Within a service (eco)system, an *engagement platform* facilitates the interaction between actors on the meso-level (Breidbach et al., 2014). They are intermediaries that enable organizations to co-create value with the customer by bringing actors and their resources together (e.g., Storbacka et al., 2016). Consequently, *engagement platforms* enable the process of resource mobilization and integration. Within our research project, the *engagement platform* exemplifies the AI-based CA and required design activities will be applied following the *engagement and institutional design* of the *multilevel design framework* (see **Figure 1**).

10.3.2 Engagement Design with Conversational Agents

Conversational Agents as a New Engagement Platform that Differentiates from Former Customer (Self) Services

CAs represent a novel type of platform for customer service. Compared to traditional customer service systems, where service exchange occurs between the customer and actors on the provider side via telephone or e-mail, CAs appear as a new automated and convenient customer channel and central platform (Espig et al., 2019; Gnewuch et al., 2017). Customers benefit from integrating information from different data sources to highly structured knowledge to deliver answers to service requests (Meyer von Wolff et al., 2019a). However, CAs do not merely appear as another passive and intermediary (engagement) platform that intelligently integrates data and thus represents another information channel of an extensive and integrated service system. Instead, a CA outlines an active social actor that solves customer problems naturally, and is dialog-based and intuitive (Gnewuch et al., 2017). CAs can offer customers a stable, homogeneous service at a low cost, which in the best case has no biases but can also map emotions and offer individual problem solutions (Wirtz et al., 2021).

Thus, they engage in resource integration by establishing relationships in different interaction scenarios. In the near future, CAs will *"increasingly fulfill the role of service employees and substitute tasks historically performed by human service personnel" (Gnewuch et al., 2017, p. 4)*, leading to new co-creation scenarios on the engagement level. Conversational agents address and solve one of the core challenges of former customer service settings, in which it was virtually impossible to offer a more efficient and cost-effective service without at the same time compromising on the quality (e.g., personalization, individualization, capabilities) (Gnewuch et al., 2017). In addition to existing self-service technologies (SSTs) for customer service (e.g., websites, portals (FAQs) or apps), CAs will be available to customers as an even more natural and constantly accessible technology-based channel and actor to perform various internal and external customer service forms. Compared to conventional self-service technologies (SSTs), they allow customers individual and flexible interactions and customer journeys (Wirtz et al., 2021).

Within our research project, different types of CAs are used as a central entry point and platform for IT, business, and product support, available for internal and external customers via various channels (e.g., enterprise messengers, internal self-service platforms and request trackers). In addition to offering external product support (e.g., regarding questions about ordering and deliveries as well as complaint management), CAs are used for internal customers to answer different service requests, such as requests for information (e.g., regarding the configuration and use of software), incidents (e.g., password resets), change requests (e.g., request for a new e-mail address) and service catalog items (e.g., setting up new accounts for employees).

CAs as Novel Social Actors Require New Design Approaches and Knowledge

Since CAs represent a new type of *learning and social actor*, new design knowledge and approaches are needed to instantiate them in service settings. CAs affect customers' benevolence toward the provider by providing engaging, positively valanced advisory experiences for customers and leveraging the relationship-building potential with simulated one-to-one advisory interaction leading to a positive perception of service (Hildebrand & Bergner, 2020). However, although an initial version of a CA can be quickly developed and provisioned, and therefore numerous agents have been instantiated in service settings in recent years, most of them could not meet expectations (Gnewuch et al., 2017). For both customers and providers, initial expectations are too high and CAs are insufficiently designed and aligned to existing service settings. This can lead to service failures and subsequently to customers' confusion and dissatisfaction or, in the worst case, even to value co-destruction, thus negatively impacting the customer relationship and the company's image (Castillo et al., 2020). Consequently, CAs require *engagement design* and software development knowledge, which goes far beyond traditional IS and service design approaches, as highlighted in the following.

Challenges Regarding the Engagement Design for CAs

The most crucial aspect of creating a good customer acceptance is to enable actors to deal with CAs seamlessly and efficiently for resource integration (Wirtz et al., 2018). The aspects described are reflected in the *engagement design*, which captures the ongoing design, development, and enhancement of socio-technical components such as engagement platforms. This encompasses the activities necessary to enable a CA's evolution from its initial value proposition into *Technical Design*, *Interaction Design* and *Service Design* activities in the CA context.

Technical Design

First, Technical Design is essential as many CAs fail to understand customer inputs due to inadequate NLP capabilities. As CAs are *unfinished learning IS,* new approaches considering the dynamics are obligatory, and need to be tested, evaluated, and reflected on the micro level. The technical design sets the basis for adequate service, including the NLP component's design, which anchors the foundation for understanding customers and enabling resource integration and value co-creation. Once an organization has decided to embrace a CA as service channel and actor, technology selection becomes a key consideration (Schuetzler et al., 2021). In this context, machine learning CAs have become prevalent in recent years, possessing a growing number of capabilities, and handling increasingly complex dialogs. According to Schuetzler et al. (2021) the vast majority of CAs in use today are supervised machines learning to accomplish their task. They use various sample messages to train the machine learning algorithm to distinguish different intents and allow it to properly understand the intent of a new message (Schuetzler et al., 2021). However, there exist numerous CA frameworks, technologies, bot builders, and configuration possibilities (e.g., Google DialogFlow, RASA.ai or the Microsoft Bot framework, Abdellatif et al., 2020; Sousa et al., 2019) that need to be thoroughly examined for their capabilities, customizability, and integration before being instantiated.

Building upon this, training is vital since (1) CAs possess limited abilities initially and (2) service environments constantly change. In comparison to conventional IS, AI-based systems need ongoing development and improvement of technical aspects (e.g., "NLP algorithms"), including, for example, supervised coaching and testing, to train their capabilities and stay current with changing tastes and technology in service systems (Xiao & Kumar, 2019). Actors' ambition is needed to understand that CAs possess limited skills initially and take part in a continuous learning and improvement process, expanding CA development *from traditional development as a project setting to a lifecycle environment, in which new functions (intents and entities) are continuously proposed, build (trained), reflected and improved.*

Interaction Design

Second, Interaction Design is an essential prerequisite for socio-technical artifacts. CAs occur as new *social actors* in the front line, exhibit human-like characteristics, interact with customers, and solve problems. Because CAs will increasingly accomplish more tasks as their capabilities evolve, many customers will build relationships with them. A customer conversation (e.g., complaint conversation) is characterized by, e.g., small talk between the customer and the service employee, where employees can react and respond to the customer's emotions (e.g., frustration). As the customer conversation is progressively performed by the CA, which learn from the collaboration and increasingly make their own decisions, the IS and service literature recommends a cooperative and anthropomorphic interaction design to foster actor engagement (e.g., Gnewuch et al., 2017; Hildebrand & Bergner, 2020; Van Pinxteren et al., 2020). In this context, organizations need to decide how they want customers to perceive CAs, which can be controlled by selecting various forms of social and anthropomorphic cues (Schuetzler et al., 2021).

For example, Schuetzler et al. (2021) distinguish between three types of implicit cue and signal types that can facilitate a humanlike interaction of a conversational agent: First, *identity cues* can be integrated into CAs' design to provide a pleasing humanlike visual appearance, e.g., through selecting a human name, avatars, and self-references (such as "I like") in the dialog with the customer (Schuetzler et al., 2021). Second, *non-verbal cues* represent another opportunity to make the conversation design more natural and realistic, using typing delays, typing dots, or emoticons (Schuetzler et al., 2021). Third, *verbal cues* can be adopted to provide a natural dialog leading to satisfaction, trust, and emotional closeness (Schuetzler et al., 2021). *Verbal cues* include simulations of characteristics of natural dialogs, such as vast vocabulary and variability in the language.

In sum, a good speech comprehension and dialog design, as well as a comfortable and natural customer experience, can lead to a competitive advantage against competing providers. However, in the context of creating humanlike CAs, it is also necessary to evaluate which range and types of social cues and empathic design options are adopted, and it depends not only on, for example, the service type, the company goals, and external image, but also on the capabilities of the CA and the customers. In this regard, different socio-technical design options for the engagement platform must be proposed, implemented, and continuously reflected, evaluated, and improved.

Service Design

Third, Service Design is appropriate to enable actor engagement. Customer service is a highly standardized practice in traditional customer support service systems, with predefined processes, roles, and task responsibilities, including resource acquisition and handoffs. Customers are often guided clearly, step-by-step through a predetermined process. As the CA now performs simple tasks, the CA must be integrated smoothly into the simultaneous service process operation. The most challenging aspect denotes that the systems are in learning progress. Although the picture is often drawn that CAs will completely replace service employees and change entire service landscapes to better meet a firm's strategic goals, e.g., profit through automation, different transition stages and handover routines will exist due to limited skills (e.g., Poser et al., 2021; Wintersberger et al., 2020). The transition stages include diverse augmentation and relief scenarios, where the CA take over "lower" (easier for AI) tasks initially, starting with more knowledge retrieval and analytical tasks, before moving up to higher intelligence tasks (e.g., needing intuition and empathy) (Huang & Rust, 2018).

Furthermore, in this scenario, customers lose their opportunity to receive "human service", of that not all types of customers are ready. As a consequence, value is not only created in dyadic relations of a customer and a CA but amongst different configurations and combinations of actors such as employees, customers and CAs integrating different resources knowledge and solving problems. Therefore, besides the classic human-CA interaction, in which the CA occur as a central interface in a service context, hybrid intermediate designs are emerging (e.g., "humans-to-(human & machine) actor combinations", Storbacka et al., 2016).

One specific constellation examined in the research project involves the *"Hybrid Service Recovery Strategy"* (Poser et al., 2021). Since CAs have limited abilities at the outset and perform only simple and repetitive tasks (e.g., pre-assessing or easy manageable requests), dealing with the situation when the CA's abilities are exceeded is crucial for the live-support process to prevent service failures and dissatisfied or even loss of customers (Wintersberger et al., 2020). For example, at the beginning of the CA introduction, problems may arise due to the NLU (NLU) or dialog management component's limited capabilities, leading to a situation where input is misinterpreted, the dialog process is hindered (intent or entity detection), or information retrieval or task execution is prevented (Poser et al., 2021). Further, few functions/tasks (intents and entities) are implemented and trained, causing users to reach the limits of the technology after only a few interactions. Therefore, some authors propose to indicate to the user the range of functions a CA possesses (e.g., Schuetzler et al., 2021), which is, however, difficult to display due to the compact language-based interface.

Instead of value co-creation, this results in value co-destruction, as customer queries can be misinterpreted, resulting in incorrect responses or no responses at all (Poser et al., 2021). To remedy this deficiency, procedures and processes should be determined when a "handoff event" (human intervention) occurs (Wintersberger et al., 2020). Our research project found the introduction and design of seamless handovers to be a crucial and challenging design aspect for *user engagement*. A CA that fails to respond to even first user queries and further fails to uphold the service leads to user frustration, rapidly becomes a bad image, and is no longer used. Users then revert quickly to conventional service channels, and their engagement is complex to re-achieve.

In order to enable efficient handovers of inappropriately (false positive) answered or unanswered requests, the CA must be designed to gather information around the request beforehand and be able to identify and transfer the information (relevant set of information extracted from the conversation in a workable format) to the right entity in order to ensure further processing (Poser et al., 2021). In the context of our research project, we investigated different types of handover implementations in CAs in real-life service settings. Two fallback strategies to ensure service continuity and recovery embrace asynchronous and synchronous handovers. An example of an asynchronous handover is creating a ticket by the CA when its capabilities are exceeded. In this case, the CA identifies, collects, and analyzes all necessary information from the chat interactions and generates and routes a ticket to a human assistant (e.g., Poser et al., 2021). An even more fluid handover scenario depicts synchronous handovers, where the CA is linked to a live chat. In this case, the service employee takes over the interaction and can access the chat history to solve the request as fast as possible (e.g., Schuetzler et al., 2021; Wintersberger et al., 2020).

10.3.3 Institutional Design

Intervention into Actors' Environment for Capturing Prevailing Institutions

It is relevant to consider that *CA management goes far beyond the engagement design*. From a service systems perspective, CAs are service systems embedded and connected with other service systems. These systems define the boundaries and context, such as institutional arrangements, organizational structures and principles that facilitate the exchange and integration of resources (Lusch & Nambisan, 2015). *Institutions* are defined as *"humanly devised rules, norms, and beliefs that enable and constrain action and make social life at least somewhat predictable and meaningful"* (Vargo & Lusch, 2016, p. 11). Institutional arrangements are *"interrelated sets of institutions that together constitute a relatively coherent assemblage that facilitates [the] coordination of activity in value-co creating service ecosystems"* (Vargo & Lusch, 2016, p. 18). Thus, actors' disposition to engage (positive, negative, ambivalent) is therefore determined by social norms and shared beliefs (Li et al., 2018).

Consequently, CAs necessitate being *improved outside of labs within real-world environments*. Much research investigates CAs solely from a purely technical or interactional perspective in lab and greenfield environments. However, as soon as they are instantiated in concrete service settings, they often could not deliver their value proposition to customers (Castillo et al., 2020). Moreover, CAs certainly possess the potential *"to replace human workers in many service functions, but when it comes to customer service that involves intensive interactions with customers, it's never a purely technical issue"* (Xiao & Kumar, 2019, p. 22). In this regard, service systems design becomes meaningful to observe value co-creation in brownfield environments; however, it arises with several design challenges.

Although several service innovation approaches exist, service systems' redesign is usually more complicated than starting from scratch (Helkkula et al., 2018). First, as resources are scarce (Murphy, 2007), service systems designers have to start with what actors and resources are available (resource mobilization). Second, actors relate to single objects such as the CA platform and the overall context, such as social context and institutional logic. This may lead to several implicit (norms, values, roles) and explicit (laws, compliance) dualisms for actors and impacts resource integration and mobilization. For instance, actors engage simultaneously in various service systems with multiple institutional arrangements, leading to role conflicts, as it is *"often not possible, feasible, or necessary for an actor to accept all value propositions"* (Chandler & Lusch, 2014, p. 6). This can lead to actor disengagement and, in the worst case, to negative engagement properties, which intend to affect negatively other actors or resources, leading to value co-destruction (Echeverri & Skålén, 2011). Therefore, service systems designers have to engage in institutional context to understand the values and norms of multiple, overlapping service systems which affect actors needs and motivations and subsequent organizational mechanisms that enable a state of institutional arrangement and logic, which are considered in the *institutional design* (see **Figure 1**).

Challenges Regarding the Institutional Design for CAs

In the CA context, *institutional design* encompasses the activities necessary to address two challenges that have to be captured in the service systems design process: (1) resource mobilization and integration remain challenging as resources in existing environments are challenging to control, and (2) actor engagement can vary regarding time and contribution. To enable CAs' *institutional design* diverse activities and areas need to be designed and managed to facilitate *engagement design* from macro-level and thus value co-creation. These elements include in forms of (1) data governance, (2) privacy and security, as well as (3) ethics and monitoring.

Data Governance

Data availability and quality play an essential role and require organizational attention and consent. The CA represents a novel kind of interface towards the customer, representing the "face" of the company and the company's knowledge in contexts of, for example, product counseling or problem-solving management. However, although companies possess vast amounts of data from the existing service environment, e.g., in terms of internal databases and (legacy) systems (e.g., ticket systems), it is often not to be deprecated to acquire data. Further, the data often correspond to poor quality and necessitate to be adjusted in order to transform them into dialogue-capable datasets to train the NLP component for the specific use cases. The additional effort to train and maintain NLP components distinguishes CA management from traditional IT systems used in service organizations. The engagement on the micro level depends on the ongoing retrieval of *NLP-ready data sets* in the context of model and dialogue training and knowledge expansion of the CA. The design needs preparation as part of the *institutional design* in terms of data acquisition, constant intervention, and reflection of the use cases to acquire new data sets, transpose them "dialogue-ready," and transfer them to "real knowledge," presented in the conversation with the customer. Adjustment due to constantly changing service environments and technology in service systems is needed (Xiao & Kumar, 2019).

Privacy and Security

Institutions on the macro-level comprise regulatory requirements, such as laws and rules, and other requirements imposed on the company by external influences. In the case of CAs, the main concerns that can arise are (1) (data) privacy concerns, along with (2) security concerns that may harm their use on both the customer and the employee side. According to the service literature, CAs need a new form of governance to deal with potential regulations early and accommodate customers with clear data protection policies (Bock et al., 2020). First, since CAs often elicit a negative attitude, through examples from private context, such as Amazon Alexa as "always listening and recording data leeches," it is also essential to create *system transparency and explain how learning CAs work* in combination with *clear data protection guidelines* toward employees and customers. The research contributions show that the lack of transparency regarding data protection negatively influences service customers' acceptance and willingness to use personalized AI services (Ostrom et al., 2019).

Ethics and Monitoring

"Most organizations are governed by ethical values, codes and compliance. The same should be required by AI" (Bock et al., 2020). Beyond the benefits of implementing CAs for customers and employees, CAs also entail risk areas for a company's reputation. Hence, governance is needed to ensure a significant consideration of AI-based service systems' *ethical design* (Bock et al., 2020). Since CAs can sense, process, and record the world around them, learn and thus can misbehave (e.g., like Microsoft Tay) or biases conversations, it requires ethics standards and monitoring that uncover the risks associated with AI (Wirtz et al., 2018). Besides, as CAs are in a permanently changing and training process, there is a need to handle changes and errors in collaboration and accountability structures (e.g., mainly when they affect the direct customer conversation).

Employee Readiness and Engagement as Knowledge Integrators

As part of the *institutional design*, general CA acceptance must be created to build trustful expectations toward the service employees. In the beginning, service employees often possess adverse or skeptical attitudes regarding the cooperation with the CA due to different reasons (e.g., loss of autonomy, job insecurity, or privacy and security concerns), which could lead to non-endorsement and therefore to service failures (Lu et al., 2020). For this reason, employees need to be picked up and trained early. Instead of letting negative attitudes take hold, the CA should be motivated as a new social actor, relieving overworked employees of tasks and leads to enhanced productivity and job satisfaction (Lu et al., 2020). Furthermore, the CA introduction requires new forms of collaboration.

In addition, management needs to foster new collaborative development and improvement approaches where employees, the CA, and developers continuously interact. Compared to a traditional IS, CAs as *unfinished and learning IS* exhibit the distinction that they inherit limited capabilities at the beginning of the roll-out. This leads to a new situation for the employees, as they can not only be specified as software users but also as *knowledge integrators*. Relying on the circumstance that service employees possess most of the knowledge related to service operations, e.g., solving customer requests or giving product advice, they should be intensively involved in the development and design process. The case of CAs demonstrates that service systems cannot generate value by themselves. They need to engage others to offer value propositions.

Shaping Value Propositions and Business Models Through AI-Based Systems

Finally, CAs' introduction in the service frontline transforms interaction touchpoints with the customer and, thus, entire customer journeys and the value-proposition leading towards *AI-shaped business models,* requiring new management approaches and *market competitive service designs*. Business models bridge technological and market innovations, emphasize a service systems-centric approach to *"how firms do business"* (Peters et al., 2016, p. 140). In this context, CAs change the value-co creation process and value proposition by (1) representing a new customer channel and (2) establishing new forms of customer relationships (e.g., changing the way how the customer perceives the provider/company/brand positioning) leading to new forms of revenue streams and reduced costs in the long-term. In addition to the service system's design at the micro-meso-macro level, it is essential to consider, analyze, and classify the business model to adapt it to current conditions continually.

10.4 Conclusion

Numerous companies already implement AI-based systems (L. Wang et al., 2020) as enhancing, augmenting, or automating technology in service encounters (Ostrom et al., 2019). AI-based systems can be both an outcome and a facilitator for value co-creation and service innovation. The multilevel perspective on designing such AI systems as exemplified by CAs provides a more comprehensive view of the design challenges for such technologies in service contexts. The case of CAs shows that service systems design needs to facilitate learning cycles on the individual micro-level and on the institutional macro-level, to succeed in increasingly dynamic environments. To realize value, changes in actors' practices and *institutions* have to be integrated with each other. Moreover, changes in one service system's *institutions* must be integrated and aligned with other *institutions* into a broader service ecosystem context (Vargo & Lusch, 2016). This perspective is particularly valuable for the transformation of extant service systems with AI. The multilevel framework highlights the interdependencies of (re)designing technologies, work processes, and service interactions, as well as institutional arrangements framing for achieving a beneficial design and use of AI. We contribute to research on AI in service science and guide practitioners in designing service innovations in the context of CAs.

10.5 References

Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020). Challenges in chatbot development: A study of stack overflow posts International Conference on Mining Software Repositories (MSR).

Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. (2015). Service innovation in the digital age: key contributions and future directions. MIS quarterly, 39(1), 135-154.

Bock, D. E., Wolter, J. S., & Ferrell, O. C. (2020). Artificial intelligence: Disrupting what we know about services. Journal of Services Marketing, 34(3).

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots International Conference on Internet Science (INSCI), Cham.

Breidbach, C., Brodie, R., & Hollebeek, L. (2014). Beyond virtuality: from engagement platforms to engagement ecosystems. Managing Service Quality, 24(6), 592-611.

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. The Service Industries Journal, 41(13-14).

Chandler, J. D., Danatzis, I., Wernicke, C., Akaka, M. A., & Reynolds, D. (2019). How Does Innovation Emerge in a Service Ecosystem? Journal of Service Research, 22(1), 75-89. https://journals.sagepub.com/doi/abs/10.1177/1094670518797479

Chandler, J. D., & Lusch, R. F. (2014). Service Systems. Journal of Service Research, 18(1), 6-22. https://doi.org/10.1177/1094670514537709

Chandler, J. D., & Vargo, S. L. (2011). Contextualization and value-in-context: How context frames exchange. Marketing Theory, 11(1), 35-49. https://doi.org/10.1177/1470593110393713

Echeverri, P., & Skålén, P. (2011). Co-creation and co-destruction: A practice-theory based study of interactive value formation. Marketing Theory, 11(3), 351-373.

Espig, A., Klimpel, N., Rödenbeck, F., & Auth, G. (2019). Bewertung des Kundennutzens von Chatbots für den Einsatz im Servicedesk International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study International Conference on Internet Science (INSCI).

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction European Conference on Information Systems (ECIS), Portsmouth, United Kingdom.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service International Conference on Information Systems (ICIS), Seoul, Korea.

Grotherr, C., Semmann, M., & Böhmann, T. (2018). Using Microfoundations of Value Co-Creation to Guide Service Systems Design – A Multilevel Design Framework. International Conference on Information Systems (ICIS), San Francisco, California, USA.

Helkkula, A., Kowalkowski, C., & Tronvoll, B. (2018). Archetypes of Service Innovation: Implications for Value Cocreation. J Serv Res, 21(3), 284-301. https://doi.org/10.1177/1094670517746776

Hildebrand, C., & Bergner, A. (2020). Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. Journal of the Academy of Marketing Science, 1-18.

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. Journal of Service Research, 21(2), 155-172.

Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: a services marketing research agenda. Journal of Services Marketing, 34(3), 389-398. https://doi.org/10.1108/jsm-01-2019-0043

Li, L. P., Juric, B., & Brodie, R. J. (2018). Actor engagement valence: conceptual foundations, propositions and research directions. Journal of Service Management, 29(3), 491-516.

Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? Journal of Service Theory and Practice, 30(3).

Lusch, R. F., & Nambisan, S. (2015). Service Innovation: A Service-Dominant Logic Perspective. MIS quarterly, 39(1), 155-175.

McTear, M. F., Callejas, Z., & Griol, D. (2016). The conversational interface (Vol. 6). Springer.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2019). How May I Help You?–State of the Art and Open Research Questions for Chatbots at the Digital Workplace Hawaii International Conference on System Sciences (HICSS), Hawaii.

Murphy, P. (2007). 'You Are Wasting My Time': Why Limits on Connectivity Are Essential for Economies of Creativity. University of Auckland Business Review, 9(2), 16.

Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer acceptance of AI in service encounters: understanding antecedents and consequences. In Handbook of Service Science, Volume II (pp. 77-103). Springer.

Peters, C., Maglio, P., Badinelli, R., Harmon, R. R., Maull, R., Spohrer, J. C., Tuunanen, T., Vargo, S. L., Welser, J. J., & Demirkan, H. (2016). Emerging digital frontiers for service innovation. Communications of the Association for Information Systems: CAIS, 1(39).

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding whether and how to deploy chatbots. MIS Quarterly Executive, 20(1), 4.

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. Internet Research, 30(1), 1-18.

Semmann, M., Grotherr, C., Vogel, P., Bittner, E., Biemann, C., & Böhmann, T. (2018). Intelligent Collaboration of Humans and Language-Based Assistants (INSTANT) International Conference on Information Systems (ICIS), San Francisco, USA.

Sousa, D. N., Brito, M. A., & Argainha, C. (2019). Virtual Customer Service: Building your Chatbot International Conference on Business and Information Management (ICBIM), Paris, France.

Stieglitz, S., Brachten, F., & Kissmer, T. (2018). Defining Bots in an Enterprise Context International Conference on Information Systems (ICIS), San Francisco, USA.

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2019). How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electronic Markets, 30, 1–35.

Storbacka, K., Brodie, R. J., Böhmann, T., Maglio, P. P., & Nenonen, S. (2016). Actor engagement as a microfoundation for value co-creation. Journal of Business Research, 69(8), 3008-3017.

Van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. A. M. (2020). Human-like communication in conversational agents: a literature review and research agenda. Journal of Service Management, 31(2), 203-225. https://doi.org/10.1108/josm-06-2019-0175

Vargo, & Lusch. (2015). Institutions and axioms: an extension and update of service-dominant logic. Journal of the Academy of Marketing Science, 44(1), 5-23.

Vargo, S. L., & Lusch, R. F. (2016). Institutions and axioms: an extension and update of service-dominant logic [journal article]. Journal of the Academy of Marketing Science, 44(1), 5-23. https://doi.org/10.1007/s11747-015-0456-3

Vargo, S. L., Wieland, H., & Akaka, M. A. (2015). Innovation through institutionalization: A service ecosystems perspective. Industrial Marketing Management, 44, 63-72. https://doi.org/10.1016/j.indmarman.2014.10.008

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2020). Effects of Voice-Based AI in Customer Service: Evidence from a Natural Experiment International Conference on Information Systems (ICIS).

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

Wilson, J. H. D., Paul R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. Havard Business Review. Retrieved 18.07.2020 from https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

Wintersberger, P., Klotz, T., & Riener, A. (2020). Tell me more: Transparency and time-fillers to optimize chatbots' waiting time experience Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, Tallinn, Estonia.

Wirtz, J., Kunz, W., & Paluch, S. (2021). The Service Revolution, Intelligent Automation and Service Robots. European Business Review.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. Journal of Service Management, 29(5), 907-931. https://doi.org/10.1108/josm-04-2018-0119

Xiao, L., & Kumar, V. (2019). Robotics for customer service: A useful complement or an ultimate substitute? Journal of Service Research, 24(1), 9–29.

11 Publication No. 3: Lewandowski et al. (2022b)

Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022). Design Knowledge for the Lifecycle Management of Conversational Agents

Proceedings of the International Conference on Wirtschaftsinformatik (WI), Nürnberg (Germany), A Virtual Conference.

Abstract

Organizations spend extensive resources on artificial intelligence (AI) solutions in customer service in order to remain customer-focused and competitive. A rising language-based application of AI emerges in the context of conversational agents (CAs), such as chatbots, which represent increasingly intelligent, autonomous, scalable, and cost-effective service platforms. However, AI-based CAs bring new organizational challenges. They are underrepresented in current research, leading to many unanswered questions and research potential regarding the management of their introduction, operation, and improvement. To address this issue, we provide design knowledge that considers the organizational perspective of CAs. Therefore, we conducted a systematic literature review (SLR) and qualitative interview study to reveal and analyze individual issues and challenges, develop meta-requirements, and finally, use them to create design principles. We contribute to the emerging field of CAs that has previously focused mainly on the individual, behavioral, interactional, or technical design.

Keywords: AI-based assistants · Conversational agents · Chatbots · Design principles · Interview study

11.1 Introduction

Organizations invest extensive resources in customer service in order to remain customer-focused and competitive (Corea et al., 2020). Customer service is important in determining critical service outcomes such as satisfaction and loyalty (Bitner & Wang, 2014; Ostrom et al., 2019). However, technological advancements and the growth of information are reshaping the work of service employees (Semmann et al., 2018). Prevailing challenges include a high volume and complexity of inquiries and rising customer expectations regarding service quality (Corea et al., 2020; Hu et al., 2018). Consequently, service employees face high-stress situations, ultimately inferior service quality (Semmann et al., 2018).

Advances in natural language processing (NLP), machine learning (ML), and general AI have spurred service innovations and promoted possibilities for designing intelligent, human-machine user interfaces (UI) (Diederich et al., 2019b; Gnewuch et al., 2017). CAs represent one specific application of AI: communicating with customers via natural language commands (Feng & Buxmann, 2020). Typical examples include chatbots in messaging applications, such as in MS Teams (Stoeckli et al., 2019). CAs are scalable and cost-effective, bearing the potential to automate, augment, and assist service interactions by identifying solution strategies and providing decision-making and problem-solving support (Diederich et al., 2019b; Gnewuch et al., 2017; Semmann et al., 2018). They can assist employees in service encounters with cognitive relief by facilitating the performance of specific tasks (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). Further, CAs are convenient channels for customers (Klaus & Zaichkowsky, 2020; Zierau et al., 2020b). Customers are expected to resolve issues themselves via this novel UI before reaching out to customer service employees (Castillo et al., 2020). However, despite an increasing interest from researchers and practitioners regarding the potential of CAs in service encounters and workplaces—evident by new research studies (Janssen et al., 2020)—many CAs fall short of expectations (Gnewuch et al., 2017). Furthermore, organizational adoption of CAs lags behind consumer usage (Corea et al., 2020; Feng & Buxmann, 2020; Meyer von Wolff et al., 2020a). CAs represent a novel subtype of AI-based information systems (IS) with distinct characteristics (Zierau et al., 2020c) such as being autonomous social actors (Maedche et al., 2019) while learning and being intelligent (Zierau et al., 2020c). Their successful adoption depends on organizational arrangements, including collaborative and continuous training, and development approaches involving efforts by IT, business, and service professionals (Lewandowski et al., 2021). Further, CAs demand novel employee and data-focused management approaches (Lewandowski et al., 2021).

In this context, extant research into CAs is primarily focused on individual (e.g., trust issues), conceptual (e.g., interaction design), or technical design aspects (e.g., NLP algorithms) (Diederich et al., 2019a; Janssen et al., 2020; Lu et al., 2020; Premathilake et al., 2021; Zierau et al., 2020a). Conversely, less is known regarding the management of CA applications in organizational contexts (Corea et al., 2020; Meyer von Wolff et al., 2021) and studies investigating CA applications often ignore their long-term success (Corea et al., 2020; Rodríguez Cardona et al., 2019). Closely related to this, research regarding the strategic management of CAs' introduction, operation, and improvement is scarce (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). However, the successful introduction and management of CAs depends on clear operation and maintenance processes, and diligences (Kvale et al., 2019). Guidance in integrating CAs in existing organizational processes, governance structures, and work routines, as well as how their adoption differs from other AI-based and conventional IS is limited (Lewandowski et al., 2021). First authors call for research on how organizations can most effectively implement/deploy (Janssen et al., 2020; Schuetzler et al., 2021), adopt (Essaied et al., 2020), manage (Corea et al., 2020; Meyer von Wolff et al., 2021) and maintain CAs (Kvale et al., 2019). While existing studies reveal initial issues and factors that influence the successful adoption of AI-based systems (e.g., Kruse et al., 2019; Pumplun et al., 2019) and CAs (e.g., Corea et al., 2020; Meyer von Wolff et al., 2021; Schuetzler et al., 2021)), research does not yet provide procedural guidance regarding the organizational rollout and continuous improvement of CAs across their lifecycle. Thereby, an understanding of CAs' lifecycle management (LCM) can provide a structured, unified view of this dynamic and novel IS, and link resources in order to ensure a reliable, consistent, and cost-effective handling of planned and unplanned changes based on previous issues (Alter, 2013). Based on this research gap, we formulated the following guiding research question (RQ):

RQ: *How to manage the lifecycle of conversational agents?*

We addressed this RQ by first developing prescriptive and supportive design knowledge following the process of Gregor et al. (2020) and Möller et al. (2020) to manage CAs' lifecycle. Drawing upon the results of the SLR, we conducted an empirical interview study to identify issues regarding the implementation, adoption, and LCM of CAs. Based on these issues, we defined meta-requirements and derived design principles (DPs) under consideration of the work system life cycle model (WSLC) of Alter (2013) as a supportive design frame. This article is structured as follows: Section 2 outlines the research background on LCM and customer service CAs. In Section 3, we present our research methodology. Subsequently, in Section 4, we present the findings of our study, including an overview of issues, meta-requirements, and the DPs. Finally, we discuss our findings in Section 5, and conclude with a summary of our limitations and contributions in Section 6.

11.2 Research Background

11.2.1 Lifecycle Management

In scholarship, several models exist for LCM, such as the work system LCM, IS LCM, or software/product LCM (Alter, 2001, 2013). Thereby, it is often unclear which models pertain to which topic and how the proposed phases vary (Alter, 2001; Niemann et al., 2009). Nonetheless, LCM models elicit a shared consensus, and usually includes a phase-based/iterative view of systems to understand issues that occurred in the past to guide a more successful course for the future (Alter, 2001). LCM models often rely on a broad view that integrates organizational (e.g., the change process), management-driven (e.g., view on the process, participants, and information), innovation-driven, and technical views, and thus provides a holistic view of socio-technical systems (Alter, 2001, 2013) and promoting, e.g., system thinking (Alter, 2004; Checkland, 1999).

LCM models originate from the field of software engineering (e.g., system development lifecycle; Alter, 2001) and usually comprise a process from requirements analysis to the maintenance of IS (Niemann et al., 2009). In this context, Niemann et al. (2009) have compared software and service LCM approaches from practice and academics. They found that software LCM models predominantly have parts of the "Plan/Analysis," "Requirements definition," "Design," "Development," "Test/Deployment," "Run/Operation," and "Improvement" phases.

However, software LCM approaches are strongly technology- and development process-focused and often de-emphasize management-oriented viewpoints as the initiation, preparation, implementation, and change in an organization (Alter, 2001). In this context, one specific LCM framework—*"encouraging a balanced view that includes the organizational and technological viewpoints"* (Alter, 2001, p. 3)—is the WSLC of Alter (2013). The WSLC is based on the work system framework and is comprised of the phases of initiation, development, implementation, and operation/maintenance (Alter, 2001). We build our study upon this model as it encompasses most existing LCM models for IS, processes and projects (Alter, 2001), and provides within its iterative and adaptive frame a more holistic view on an IS lifecycle in organizations, with consideration for several influences on IS. In this context, the WSLC provides a good analysis and design frame (Bock et al., 2014) for the step-by-step management of CAs as novel form of AI-based IS in organizations, since their management raises many issues and no approaches exist guiding practitioners on how to manage this class of IS in their lifecycle (Lewandowski et al., 2021). Further, CAs need an integrated, collaborative, socio-technical, and interdisciplinary view (Lewandowski et al., 2021) instead of a *"system-as-technical artifact perspective"* (Alter, 2013, p. 74), as the WSLC model also embraced (Alter, 2013).

11.2.2 CAs in Customer Service

Customer service encounters represent the prevalent channel used in service-oriented business models (Gnewuch et al., 2017; Ostrom et al., 2019) to supply information, and provide advice and support between providers and customers (Janssen et al., 2021a). For measuring the performance of the customer service provider, service quality is an important concept (Gronroos, 1988; Johnston, 1995), defined as the outcome of a comparison between expectations of service and what is perceived to be received (Parasuraman et al., 1985). A significant challenge for conventional customer service is improving efficiency and reducing resources without compromising the quality of service (Frei, 2006; Gnewuch et al., 2017). Thereby, customer service is often the most resource-intensive department within an organization (Cui et al., 2017). Many service requests are currently handled manually, which is time-consuming and leads to a high error rate, whereby user expectations can often not be fulfilled (Xu et al., 2017).

In this context, CAs are evolving to become the dominant customer service channel (Zierau et al., 2020b), representing a class of IS that is capable of *"interpret[ing] and respond[ing] to statements made by users in ordinary natural language"* (Lester, 2004, p. 1). As CAs possess the potential to relieve service encounters by automating, augmenting, and assisting service interactions (Diederich et al., 2019b; Semmann et al., 2018), by, e.g., a 24/7 available CA instead of waiting for an email response (Zierau et al., 2020b), they generate widespread attention (Gnewuch et al., 2017). CAs are increasingly popular in research and practice (Feine et al., 2019a; Zierau et al., 2020a) for their ability to improve service efficiency, experience, and quality (Zierau et al., 2020b), and are being labeled as, e.g., chatbots or conversational intelligence in publications (McTear, 2018; Shah et al., 2016; Shawar & Atwell, 2007). While early CAs were limited to defined sets of conversations (McTear, 2018; Shah et al., 2016; Shawar & Atwell, 2007), present-day CAs are sufficiently intelligent for application in organizations (Io & Lee, 2017), due to improvements in NLP and ML (McTear, 2018; Shah et al., 2016; Shawar & Atwell, 2007). In current service encounters, CAs are playing an active role, service employees have conventionally performed (Gnewuch et al., 2017; Herrera et al., 2019).

Our research focuses on text- and AI-based CAs, often referred to as AI-based chatbots in customer service (e.g., Zierau et al., 2020b), due to the opportunities to reach many customers via text-based CA. Moreover, we selected the customer service context as it allows us to study the management of CAs in a context in which they currently attract much attention, even though they have been applied for this purpose without scientific guidance in the past (Mimoun et al., 2012). In this context, research on how CAs can be introduced in customer service and its organizations is still scarce (Janssen et al., 2021a; Lewandowski et al., 2021). However, CA applications pose various new challenges for organizations (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). AI-based CAs represent a novel type of IS (Zierau et al., 2020a) by, e.g., being social, unfinished, and learning (Lewandowski et al., 2021), and therefore, they demand new approaches and research regarding their implementation and LCM (Corea et al., 2020; Lewandowski et al., 2021; Meyer von Wolff et al., 2021). While current technical limitations could be resolved thanks to ongoing technological advances, the lack of knowledge related to organizational design aspects represents an issue needing investigation (Corea et al., 2020; Meyer von Wolff et al., 2021).

11.3 Research Methodology

11.3.1 Goal and Study Design to Derive Design Knowledge

This article aims to provide design knowledge that helps organizations manage CAs' lifecycles, presented in the form of issues, requirements, and DPs. The DPs originate from (1) an SLR, and (2) a qualitative interview study with 17 experts on CAs in customer service. (1) The SLR followed the five-step process by vom Brocke et al. (2009) which we conducted in the preliminary of this study (Lewandowski et al., 2021). It revealed several issues from the nascent CA literature that impacts the adoption and management of CAs as opposed to general AI-based and traditional IS applications (Lewandowski et al., 2021). (2) Based on these findings, we conducted semi-structured expert interviews according to (Gläser & Laudel, 2009; Meuser & Nagel, 2009a; Myers, 2019), which allowed a more detailed investigation. The SLR and interview study provided the basis for developing consolidated meta-requirements used to derive DPs. In the following, we present the phases of the empirical research procedure and the steps to derive the DPs in detail.

11.3.2 Data Collection and Analysis

To gather qualitative data about issues and meta-requirements regarding the CA lifecycle, we started with a preparation consisting of two steps. First, we developed a semi-structured interview guide to ensure a systematic procedure and comparably gathered data (Meuser & Nagel, 2009a). The interview questions were formulated based on a preliminary *theoretical reasoning stage* according to the process of Gläser and Laudel (2009), embracing the consideration of the nascent state of the literature identified with the SLR (Lewandowski et al., 2021), the research gap, and the goals of our study (e.g., expansion of the current body of knowledge on CAs). The participants were asked about the following topics: (1) general experience with CAs and current CA projects (roadmap), reasons/use cases to adopt CAs for customer service (initial situation); (2) general prerequisites for an organization to introduce CAs; (3) challenges in their application (e.g., development and training), use, and acceptance; (4) requirements for a successful application and management; and (5) challenges, requirements, and steps for a continuous improvement process (e.g., activities, tools, and stakeholders/actors that need to be involved).

Second, we determine potential interview partners for the study, intending to understand the application, and management of CAs in customer service. Therefore, we consider several practitioners from diverse areas as *experts* (according to Meuser and Nagel (2009a)), such as executives, product owners, AI/ML/CA experts, and consultants with professional experience and different *contextual knowledge* (Meuser & Nagel, 2009a) in the course of CA projects. We acquired the experts across a dual-stage process. First, we selected experts from our CA research project in customer service. Second, we have access to a broad corporate network of practitioners covering many industries (e.g., banking, consumer goods, e-commerce, transport, manufacturing) from which we have requested and selected CA experts according to the criteria mentioned above. We conducted 20 interviews with 17 experts (see **Table 1**) that lasted between 24 and 67 minutes (mean = 49.95 minutes).

Table 1. Overview of interview study participants

No.	Role	Duration (h)	No.	Role	Duration (h)
01	Project Manager AI/ML	0:56	05(+04)	CA Trainer & Consultant	1:02
02	Manager - AI Innovation	0:52	09(+04)	Consultant & AI Software Developer	1:07
03-06	IT Service Delivery Team	1:01	10	Chief Marketing Officer CA Supplier	0:38
07	Software Project Manager	0:57	11	Team Member in a CA project	0:24
08	Technical CA Consultant	0:53	12	AI Supervisor (CA implementation)	1:03
04	Consultant	1:05	13	Application Integration Professional	0:38
01	Project Manager AI/ML	1:01	14	Product Owner in a CA project	0:50
04	Consultant	0:52	15	Product Owner in a CA project	0:27
04	Consultant	0:55	16	Technical Consultant	0:35
05(+04)	CA Trainer & Consultant	0:55	17	Customer Success Manager	0:28

The interviews were conducted via conference systems, and recorded and transcribed for data analysis until we could not generate any further insights, according to the theoretical saturation by Glaser and Strauss (2006). For the data extraction and analysis, we followed the instructions of Mayring (2014) and Rädiker and Kuckartz (2019). We conducted a qualitative content analysis using *MAXQDA software*. According to the *intercoder reliability check,* two independent researchers continuously compared and adapted an initial set of codes (issues) to ensure the validity of the results (Mayring, 2014). Afterward, based on the coded material, we identified 57 initial mutual issues, which were discussed and clustered with three researchers into 13 issues to help derive meta-requirements and, subsequently, DPs.

11.3.3 Design Principle Generation

A DP can be described as a *"fundamental rule […] [derived from] extensive experience and/or empirical evidence, which provides design process guidance to increase the chance of reaching a successful solution"* (Fu et al., 2015, p. 2). Our study adopts guidance of Gregor et al. (2020) and Möller et al. (2020), describing the formulation of DPs as an essential pre-step and description of abstract propositions for complex artifacts to allow their validated design. Thereby, rigorously formulated DPs can organize the designing of IS artifacts from a higher "meta-level" and, thus, help and improve, e.g., IS development, application, and management processes (Cronholm & Göbel, 2018; Gregor, 2002; Gregor et al., 2020; Möller et al., 2020). The DPs are often derived based on prior knowledge from literature and statements from experts or observations (Gregor et al., 2020). The term follows a dual nature, since DPs can, e.g., guide a process of designing an artifact or describe software functionalities (Möller et al., 2020). Our study derives DPs to generate prescriptive design knowledge that is *"intended to be manifested or encapsulated in an artifact, method, [or] process"* (Gregor, 2002, p. 17) (here: denoted as a first approach) to manage CAs' lifecycle. Following the development taxonomy of Möller et al. (2020), we developed (1) supportive DPs from (2) a qualitative study (3) to identify issues from the current literature, and then coded and analyzed the interview study (4) in order to derive meta-requirements (Section **4.1)** (5) to formulate DPs in the next step (Section **4.2)** (6) based on the formulation template of Gregor et al. (2020). In this regard, a DP serves a precise goal, context, and mechanism and is grounded in its derivation by the relationships among DP elements (Gregor et al., 2020). Thereby, we followed the first six process steps of Möller et al. (2020) for DP Development.

11.4 Results

11.4.1 Issues and Meta-Requirements

We identified 13 issues (**I**) and formulated 9 meta-requirements (**MR**) (see **Table 2).**

Issue $\mathbf{I_1}$ refers to a missing committed long-term vision and roadmap and, thus, a lack of addressing a clear-cut, valuable, and scalable business problem, resources, and (management) support. Experts stated that CA development often runs *"parallel to day-to-day business and the biggest challenges are more organizational than technical" (E_2).* From the literature, (Lewandowski et al., 2021) describes the need for a long-term vision and commitment. (Meyer von Wolff et al., 2020b, 2021) addresses the missing agenda and underestimated effort.

$\mathbf{I_2}$ deals with insufficient knowledge, wrong expectations, and missing acceptance of the CA as novel IS, e.g., due to the new UI. The experts stated: *"we did just go live to test how [the CA] resonates, but people just used it as a search engine" (E_3)* or *"the introduction is critical, you have one shot with the CA, or everything is lost" (E_7).* E_{17} supports this issue: *"Some [...] overestimate CAs - Once it's set up, the bot works perfectly […]. That's how they imagine it" (E_{17}).* Similarly, López et al. (2018) and Schuetzler et al. (2021) identified these issues *("If a chatbot does not live up to expectations, users get frustrated"* (Schuetzler et al., 2021, p. 5)) as well as (Corea et al., 2020; Feng & Buxmann, 2020; Meyer von Wolff et al., 2019a). Based on $\mathbf{I_{1,2}}$, $\mathbf{MR_1}$ emphasizes the provision of a roadmap for org-readiness and vision, including allocating resources (budget, staff), and enabling the organization and customers to understand the capabilities of the CA and minimize adverse effects due to limited understanding, skill level and wrong expectations.

Table 2. Overview of the aggregated issues

ID	Title	Description	Source
I_1	**Long-term vision and roadmap**	The CA deployment does not have a long-term committed vision and roadmap, due to, e.g., a lack of addressing a valuable and scalable business problem, resulting in a lack of resources and support at all levels.	E_{1-5}, $E_{10,13,}$ $E_{15,}$ (Lewandowski et al., 2021; Meyer von Wolff et al., 2020b, 2021; Schuetzler et al., 2021)
I_2	**Expectations of novel IS**	The organization has insufficient knowledge, wrong expectations, or lack of acceptance, (employee/user) readiness, and skills when using CAs.	E_{1-5}, E_7, $E_{13,15}$, (Corea et al., 2020; Lu et al., 2020; Meyer von Wolff et al., 2019a; Schuetzler et al., 2021; Xiao & Kumar, 2019)
I_3	**Release-rush atmosphere**	The preparation effort is underestimated in terms of maturity (quality of data, technology selection, NLP, dialog design, functionality), and CA may thus go live too early, leading to long-term non-use.	E_2, E_{4-8}, (Brendel et al., 2020a; Schuetzler et al., 2021; Sousa et al., 2019)
I_4	**Disregard of underlying influences**	When using CAs, legal (incl. IT security, compliance, data protection and data analysis (in the cloud)), ethical issues (e.g., system transparency) and organizational issues (lack of trust and aversion) are underestimated.	$E_{1,2}$, $E_{4,5}$, E_8, (Corea et al., 2020; Lewandowski et al., 2021; Maedche et al., 2019; Meyer von Wolff et al., 2020b, 2021; Rodríguez Cardona et al., 2019)
I_5	**Integration and modernization of IT landscape**	On the technical side, CAs are developed detached from real structures (e.g., from existing architectures, and (frontend/backend) systems, data sources) and/or a modernization of the IT architecture is not considered (e.g., provision of APIs).	$E_{1,2}$, $E_{4,5}$, E_8, $E_{17,}$ (Burns & Igou, 2019; Gnewuch et al., 2017; Kruse et al., 2019; Meyer von Wolff et al., 2021; Sousa et al., 2019)
I_6	**Integration into work structures and processes**	On the business side, the integration of CAs into already existing workflows and business processes is overlooked and CAs are developed detached from existing processes (e.g., feedback cycles and handovers).	E_2, E_4, $E_{8,}$ E_{10}, E_{17}, (Corea et al., 2020; Poser et al., 2021; Zierau et al., 2020c)
I_7	**Lack of new responsibilities, freedoms**	Further development of a CA requires the continuous involvement of company stakeholders from diverse areas (e.g., works council) as well as creating new roles/freedoms to ensure development efforts (e.g., data, sampling, analysis, training, managing intents, monitoring).	$E_{2,}$ $E_{11,12}$, $E_{16,}$ (Kvale et al., 2019; Lewandowski et al., 2021; Meyer von Wolff et al., 2021; Zierau et al., 2020a)
I_8	**Underestimation of required competences**	Companies underestimate the required developer expertise, the development of new competence fields (trainers, modelers), e.g., resulting in possible lock-in effects to CA (platform) providers and their frameworks.	E_2, E_{12}, E_{14}, , (Abdellatif et al., 2020; de Lacerda & Aguiar, 2019; Kvale et al., 2019)
I_9	**Distributed knowledge in expert domains**	The CA deployment lacks the knowledge of the expert domains in the support for the use case for successful operation; experts do not have the capacity to provide training data in addition to the daily business.	E_{3-5}, E_{7-11}, $E_{13-17,}$ (Brandtzaeg & Følstad, 2017; Kvale et al., 2019; Lewandowski et al., 2021;

			Meyer von Wolff et al., 2021; Stoeckli et al., 2019)
I_{10}	**Data availability and NLP-conformity**	CA deployment relates to data management, which is underestimated in terms of accessing and integrating heterogeneous data sources and processes these into high-quality NLP-data sets that can be used for training.	$E_{4\text{-}9}$, E_{16}, (Corea et al., 2020; Lewandowski et al., 2021; Meyer von Wolff et al., 2020b, 2021)
I_{11}	**Continuous training and maintenance**	The CA does not receive continuous further development and training, although the knowledge, technology, and data would have to be constantly kept up to date, analyzed, trained, and feedback collected to ensure utility.	$E_{1\text{-}2}$, E_4, E_8, $E_{16,17}$, (Kvale et al., 2019)
I_{12}	**Continuous monitoring and visualization**	The CA deployment does not have a continuous monitoring process to demonstrate the behavior and benefits of the deployment to the organization, resulting in missing acceptance and little participation.	$E_{2\text{-}5}$, $E_{7\text{-}8}$, $E_{10\text{-}12}$, $E_{14\text{-}17}$, (Corea et al., 2020)
I_{13}	**Continuous improvement culture**	The organization has poor feedback and lacks a communication culture, which is much needed for the continued development of a CA, as diverse knowledge is needed at different stages of development.	$E_{1\text{-}2}$, E_4, E_5, $E_{10\text{-}13}$, $E_{15\text{-}17}$

Further, we discovered that the preparation effort is underestimated concerning the maturity of the CA. This includes technology selection, data preparation, interaction design, and functionality building. Therefore, the CA may go live too early (e.g., driven by management pressure), leading to the CA's non-usage, and sometimes, a permanent dissent, summarized with $\mathbf{I_3}$. Also, several authors underline this issue (e.g., Schuetzler et al., 2021) for the right technology selection, interaction design, and social cues (Meyer von Wolff et al., 2020b; Sousa et al., 2019), and the NLP data preparation (Brendel et al., 2020a) for functional maturity. The interview respondents explained, *"[we] have to design [CAs] from diverse perspectives, […] otherwise you can lose the user completely" (E_2)* or *"We went early go-live. But the people only thought the CA could not do anything. This led to a lasting low acceptance of the bot" (E_4).*

In the context of $\mathbf{I_4}$, environmental issues were identified. For CA application, several legal, security, ethical, and organizational issues needed to be considered, especially data protection efforts (e.g., Lewandowski et al., 2021; Rodríguez Cardona et al., 2019) and system transparency (e.g., how the CA works). Experts stated several challenges: *"If someone uses the CA, the chat gets logged, and possibly every conversation could be recorded and analyzed including sensitive information" (E_2)* or *"[The CA is] only allowed to communicate about personal data if the user has been authenticated" (E_8).* The issues $\mathbf{I_{2,3,4}}$ contributed to $\mathbf{MR_2}$, which targets an appropriate CA preparation and ensures that the expectations are met, e.g., by employee training and an appealing, committed CA maturity at rollout.

Per **MR_3**, the CA implementation and (further) development needs to consider the involvement of various (perhaps impairing) actors. This requirement results from issues **I_3** and **I_4**. The missing involvement and underestimation of underpinning parties (e.g., the data protection department or the worker's council) can lead to a non-usage or closure of the CA project before it has fully arrived within the organization.

For MR_4, **I_5** and **I_6** were the input. **I_5** refers to the fact that CA development happens detached from actual IT structures (e.g., from existing architectures, systems (frontend/backend) and modernization of the IT is not regarded. Experts describe that CA development requires *"complex things outside the core technology, technical integrations with backend systems" (E_{15})*. An integration of the CA into relevant systems (Gnewuch et al., 2017; Meyer von Wolff et al., 2020a) and the handling of data from various systems to create a seamless orchestration point for customer service should be considered (Corea et al., 2020).

Further, **I_6** addresses the lack of integration into governance, work structures, and business processes. The literature emphasizes that for a successful CA application, an integration into current processes is obligatory *("process-aware CAs,"* (Corea et al., 2020, p. 5823), including handovers to the service desk or human-in-the-loop concepts (Corea et al., 2020; Poser et al., 2021; Zierau et al., 2020c). Experts raised several problems: *"We need to get the user to look first at the CA and afterwards at the usual service desk […], therefore CAs must be integrated in existing processes." (E_3)* Furthermore *"a direct human handover would be nice, if the CA is unable to handle the request." (E_4)*. **I_5** and **I_6** led to **MR_4**, defined as holistic system thinking of technical and organizational integration and renewal options. However, there is a lack of responsibility, roles, and freedoms for ensuring underestimated development efforts get underway (**I_7**). The interviewees argued for new roles like a "CA trainer." E2: *"We need one full-time person for only training and implementing use-cases."*

I_8 addresses the undervalue of the required expertise for CA development, including a lack of time to develop CAs further. CAs' development often disregards novel competencies and responsibilities (e.g., for data preparation, training, monitoring), often leading to "lock-in" effects on CA (platform) providers. In general, CAs tend to work like black-boxes and require new developer expertise (Abdellatif et al., 2020; de Lacerda & Aguiar, 2019). These two issues led to **MR_5**. Further, **I_9** comprises that the CA deployment disregards the knowledge of the expert domains (e.g., concrete knowledge of use cases, conversations, and processes). In this context, experts cannot provide training data in addition to their daily tasks without relief. Thereby, CAs need training as unfinished IS and depend on knowledge provision (Lewandowski et al., 2021). It is crucial to integrate the domain experts into the development process (Lu et al., 2020; Meyer von Wolff et al., 2020b; Stoeckli et al., 2019). Experts stated: *"Not every developer has know-how about the processes. The business units need to get continuously involved." (E_4)* or *"we analyze the chats with the customer service […]. For example, this wording doesn't fit [...] the conversation flow, [this] needs to be redesigned because it's too complicated" (E_{17}).* This issue leads to **MR_6**, which states that it is necessary to involve domain experts to integrate "real" knowledge, e.g., for functionality/dialog generation and design.

I_{10} illustrates that a CA application does not concern data management activities. CAs' training depends on the access and preparation of many (often heterogeneous, unstructured) data sources that are difficult to integrate and process into high-quality data sets for training activities. Several authors emphasize data availability, preparation, actuality, and NLP-conformity ("creation of a knowledge base") (Meyer von Wolff et al., 2020b; Pumplun et al., 2019; Sousa et al., 2019; Zierau et al., 2020a). Similarly, *E_4* describe: *"One challenge is the homogenization of the data" (E4)* and *"several of knowledge data in the different business units, […] difficult to integrate them for the data processing and keep it up-to-date" (E4).* The additional effort to train NLP components distinguishes CAs from other AI systems. Consequently, **MR_7** requires to establish activities for data access, assessment, selection, and preparation.

I_{11} addresses a CA does not receive continuous training. However, the data and technology, need to be constantly analyzed (e.g., with chat logs analysis), updated and trained, or otherwise, *"acceptance problems or legal effects could be the consequence"* (Meyer von Wolff et al., 2021, p. 7, p. 7). In addition, feedback needs to be collected to ensure utility and relevance. *E_{17}* describes, *"CA will quickly get outdated, […] user questions and the content are changing [...]. Emphasize the topic of continuous improvement, training [...] that's […] our biggest problem."* (E_{17}). The described issues **$I_{10\text{-}11}$** contributed to **MR_8**.

For **MR_9** the issues **$I_{12,13}$** influenced. **I_{12}** describes the CA application not to have a continuous monitoring for demonstrating behavior (e.g., chatlog analysis) of the CA to the supported domains (e.g., metrics/dashboards). Expert states *"It's important that there is monitoring to decide which [..] functions run well." (E_8)* or *"the business units [need] to see which knowledge articles are good and which need improvement" (E_4).* Interviewees identified that organizations often have poor feedback and communication culture in CAs' development, lead to **I_{13}**. There are diverse knowledge and feedback needs: *E_{16}* describes, *"we accompany the whole thing with training, feedback […]. This* *includes […] continuous improvement. [It's] not a one-time thing […], it is permanent. Continuous tasks, […], training of the bot, quality assurance, monitoring." (E_{16}).*

11.4.2 Design Principles

Based on the coded text passages, we have identified 13 issues and formulated 9 **MRs,** which were used to derive 7 prescriptive **DPs** to guide and manage CAs' initiation and further development lifecycle. The DP development is outlined in Section 3.3. The **DPs** are depicted in **Figure 1**, including the mapping from issues to **MRs** to **DPs**.

CA Initiation: $\mathbf{DP_1}$ aims to guide the initiation and strategic preparation of the introduction to CAs to ensure organizational, and customer readiness, engagement, and long-term commitment regarding this novel IS form ($\mathbf{MR_1}$ and $\mathbf{MR_2}$). E_1 states, for example, that not every form of company is suited for a CA application. With readiness ensured, the CA application comprises an extensive and often undervalued initiation process. The CA must address an apparent, scalable business problem and vision, ensuring that the CA is *"more than another proof-of-concept" (E_4).* Formulating a roadmap supports establishing a CA team ($\mathbf{MR_5}$), and expectations regarding development time to ensure that the CA application gets enough effort. Further, a CA application needs right from the initiation the establishment of a collaborative, and continuous development culture ($\mathbf{DP_2}$). The consideration of regulatory and ethical issues (Seeber et al., 2020; Zierau et al., 2020b), and expert knowledge need to be modeled in the CA is highly relevant. For example, the team around $E_{4\text{-}6}$ offer specialists (e.g., support employees) a middleware on which they can create knowledge articles and dialog data sets to later train the CA. For later development activity, the involvement of (impairing) stakeholders is indispensable for establishing long-term commitments ($\mathbf{MR_{3,6,9}}$).

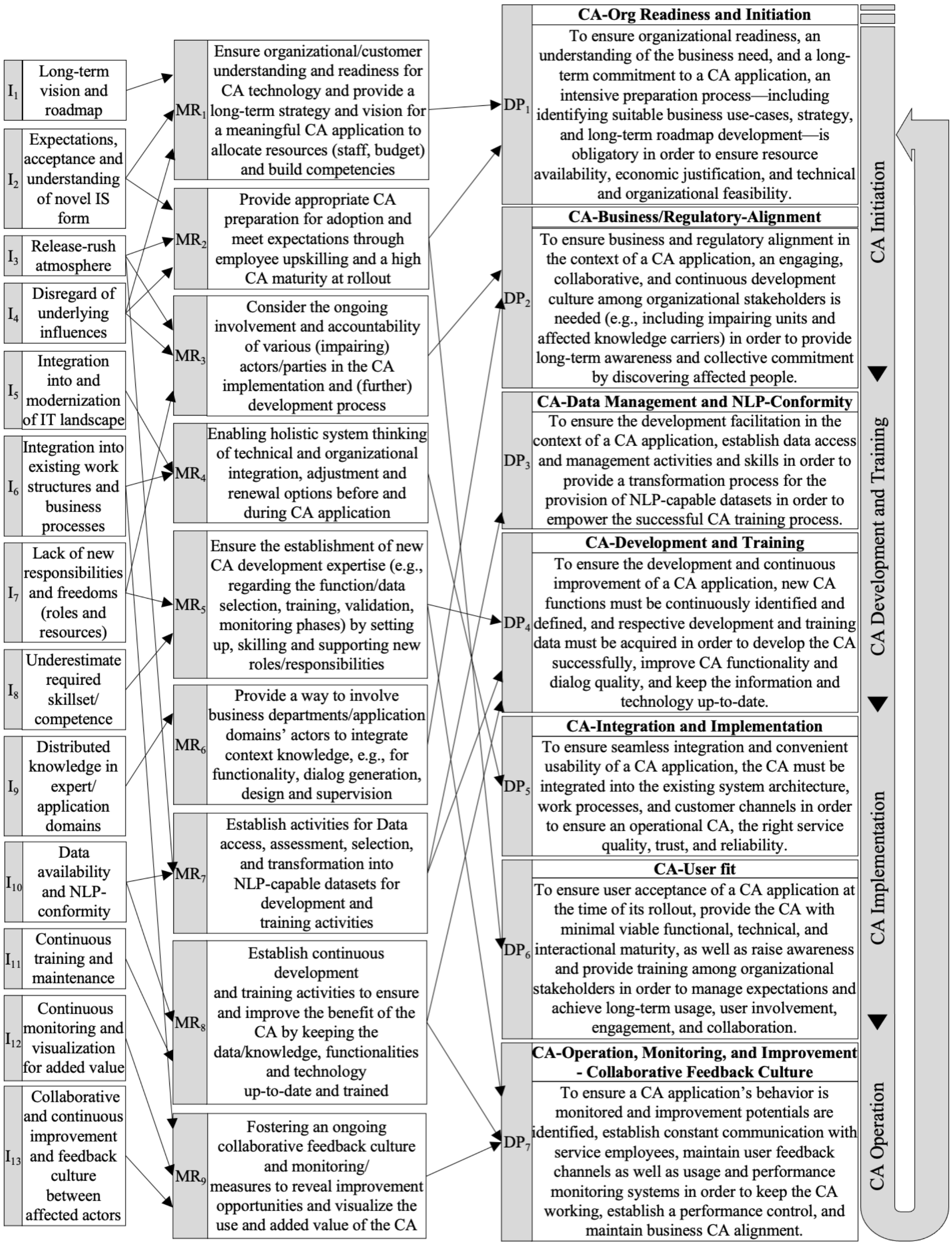

Figure 3. Overview of the derived DPs according to Gregor et al. (2020) and the design frame of Alter (2013)

CA Development and Training: To empower CA development and training activities, a CA application requires practicing preparatory data management activities to provide/formulate, e.g., NLP-capable datasets, as depicted in **DP_3**. Strongly related is **DP_4**: In addition to **DP_2** (e.g., knowledge carriers), a continuous interplay between CA development, data access, selection, and preparation activities (**DP_3**) is needed to identify CA functions and keep the dialog and technology up to date. Companies applying a CA must be aware that it is *"a continuous software development process in which numerous hurdles can arise" (E_{17})*, (e.g., during extending functions, with poorly documented, not NLP-ready knowledge, calling for AI trainers) *($E_{4,14}$)*.

CA Implementation: **DP_5** prerequisites the integration of CA in technical and organizational structures to ensure usability and cognitive relief for service employees (**MR_4**). Work integration is necessary to guide the user effectively and efficiently through the process and, if required, to get in touch with a service representative (Poser et al., 2021; Zierau et al., 2020b). **DP_6** strives to target CA and organizational preparation to ensure seamless integration. CA-related education, and user preparation should be managed pre-rollout to fulfill expectations (Lewandowski et al., 2021; Meyer von Wolff et al., 2021). Moreover, a high level of maturity (functional, technical, and interactional) should ensure long-term involvement. E_2 and E_3 recommend a successive CA launch with gradual approval of small user groups in which functions are improved (e.g., dialog design and NLP behavior) to avoid limited maturity (**MR_2**).

CA Operation and Control: Finally, per **DP_7**, a CA application demands the establishment of ongoing monitoring activities, including novel skills and roles (**MR_5**) to uncover the actual CA behavior toward end-users and thus the potential for improvements (**MR_8**). **DP_7** may be instantiated by providing the user with diverse feedback options in the interaction (free text, star rating/button, questionnaire, forwarding), collaboration with service employees and frequent monitoring activities (usage indicators, chatlog/request analysis), as recommended by $E_{8,16}$.

11.5 Discussion

Although CAs are an emerging AI-based IS for customer service, resulting in various use cases and research studies (Janssen et al., 2020), CA applications often neglect long-term success (Corea et al., 2020; Rodríguez Cardona et al., 2019) and inhibiting influences in companies (Meyer von Wolff et al., 2021). Current knowledge on CAs focuses on individual, conceptual, or technical design perspectives (Diederich et al., 2019a; Janssen et al., 2020; Lu et al., 2020; Premathilake et al., 2021; Zierau et al., 2020a). However, our research revealed that CAs fail due to organizational and employee-dependent issues in CAs' lifecycle. First authors already call for a *"switch from CA design research to [...][a] management view [. . .][, since] organizational and individual issues have the highest influence"* (Meyer von Wolff et al., 2021, p. 12f., p. 12f.) and for *"practice-based requirements[, which] can provide insights that may not have been captured in scientific literature"* (Corea et al., 2020, p. 5827). We address this gap contributing to CAs' management in organizations by providing design knowledge for practitioners on how to establish and manage first CA lifecycle activities. Our research supports previous CA contributions (Lewandowski et al., 2021; Meyer von Wolff et al., 2021) that emphasize that although some core issues in conventional IS management are similarly present in the CA lifecycle, CAs need a dedicated perspective due to more specific characteristics: First, the impact of AI from an organizational perspective has been insufficiently studied (Y. Wang et al., 2020), although various AI applications require dedicated in-depth research for leveraging AI's business value (Jöhnk et al., 2021). Few articles explore AI adoption factors (Jöhnk et al., 2021; Kruse et al., 2019; Pumplun et al., 2019). Related, research does not address issues for managing the CA lifecycle and how the CA LCM activities differ from previous LCM frames, such as Alter (2013). CAs' management has numerous novel activities that other AI applications do not possess (e.g., image recognition), and usually tend to be more data-model and IT-department-centric. Some of the issues in AI literature (e.g., long-term management support or data quality; Jöhnk et al., 2021; Pumplun et al., 2019) align with CA management issues. However, CAs as learning, dialog-based, and social IS (Maedche et al., 2019) possess a strongly human-dependent lifecycle and depend on new collaborations, and common continuous development/training and monitoring activities between IT departments and affected business units **(**$\mathbf{DP_7}$) (Lewandowski et al., 2021). CA training requires new roles and interdisciplinary team structures to perform tasks such as preparation of NLP-ready data sets, managing intents, and writing compelling conversations, while also being aware of organizational influences and enduring communication with domain experts **(**$\mathbf{DP_2}$**)** (Kvale et al., 2019), who also need freedom ($\mathbf{DP_4}$). Yet, no research describes the individual activities, diligence, skills, means of communication, or relations with domain experts in a CA lifecycle, which is a follow-up topic needing more in-depth investigation.

Second, a CA application must consider an integration into existing company and IT structures for a seamless user experience ($\mathbf{DP_5}$). Contributions (Poser et al., 2021) present the first approaches to integrating CAs in service desk processes. However, our results show that integration with actual company tasks is a scarcely considered aspect in research.

Finally, CA applications need an initiation and integration process besides the pure development (Alter, 2013) to ensure org-readiness for facilitating the business problem–CA fit (**DP$_1$**) and ensuring user adoption at the CA rollout (e.g., with sufficient CA maturity and not alienating users) (**DP$_6$**). However, attitudes toward CAs may be negative due to limited skills and poor initial integration. Many articles address CAs' design but few deal with an overarching maturity. Further studies need to explore CAs' maturity criteria for measurement to validate the CA in the lifecycle activities beforehand.

11.6 Conclusion and Limitations

AI-based CAs accelerate customer-focused and competitive customer service, leading to new applications and research studies. However, current research disregards CAs' lifecycle management, although the application poses entirely new challenges for companies. We contribute by conducting an SLR and an empirical interview study with CA experts to reveal issues and provide design knowledge to manage the CA lifecycle.

This study is faced with some limitations. First, the European experts in this study and their domain-specific experiences influence the study's external validity. In this context, we have drawn on existing company and research project contact networks. However, many experts work at international companies from diverse industries and offer various experiences and sufficient data saturation (Corea et al., 2020; Guest et al., 2006). Particularly, our derived design knowledge is dependent on a concrete instantiation. By suggesting the DPs, we contribute to managing the CA lifecycle, but the DPs require contextualization for the individual use-case. In this context, the next step would be to first evaluate, and then improve and instantiate the DPs in a concrete research project with corporate partners.

11.7 Acknowledgements

The research was financed by the German Federal Ministry of Education and Research and the European Social Fund - "Future of Work" program (INSTANT - 02L18A111).

11.8 References

Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020). Challenges in chatbot development: A study of stack overflow posts. International Conference on Mining Software Repositories (MSR).

Alter, S. (2001). Which Life Cycle—Work System, Information System, or Software? Communications of the association for Information Systems, 7(1), 17.

Alter, S. (2004). Desperately seeking systems thinking in the information systems discipline. International Conference on Information Systems (ICIS), Washington, DC.

Alter, S. (2013). Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. Journal of the Association for Information systems, 14(2), 72-121.

Bitner, M. J., & Wang, H. S. (2014). Service encounters in service marketing research. In Handbook of service marketing research. Edward Elgar Publishing.

Bock, A., Kaczmarek, M., Overbeek, S., & Heß, M. (2014). A comparative analysis of selected enterprise modeling approaches. Working Conference on The Practice of Enterprise Modeling - IFIP International Federation for Information Processing.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. International Conference on Internet Science (INSCI), Cham.

Brendel, A. B., Greve, M., Diederich, S., Bührke, J., & Kolbe, L. M. (2020). You are an Idiot!–How Conversational Agent Communication Patterns Influence Frustration and Harassment. Americas Conference on Information Systems (AMCIS).

Burns, M. B., & Igou, A. (2019). "Alexa, write an audit opinion": Adopting intelligent virtual assistants in accounting workplaces. Journal of Emerging Technologies in Accounting, 16(1), 81-92.

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. The Service Industries Journal, 41(13-14).

Checkland, P. (1999). Systems thinking. Rethinking management information systems, 45-56.

Corea, C., Delfmann, P., & Nagel, S. (2020). Towards intelligent chatbots for customer care-practice-based requirements for a research agenda. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Cronholm, S., & Göbel, H. (2018). Guidelines supporting the formulation of design principles. Australasian Conference on Information Systems (ACIS), Sydney.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A customer service chatbot for e-commerce websites. Meeting of the Association for Computational Linguistics-System Demonstrations, Vancouver, Canada. https://doi.org/10.18653/v1/P17-4017

de Lacerda, A. R. T., & Aguiar, C. S. R. (2019). FLOSS FAQ chatbot project reuse - how to allow nonexperts to develop a chatbot. International Symposium on Open Collaboration, Skövde, Sweden.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019a). On conversational agents in information systems research: Analyzing the past to guide future work. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019b). Towards a taxonomy of platforms for conversational agent design. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Essaied, K., Marsan, J., Bazomanza, W. N., Tewamba, H. J. N., & Yombia, S. M. K. (2020). Organizational Decision to Adopt Chatbot Technology: The Role of Organizing Vision and Technological Frame. Americas Conference on Information Systems (AMCIS).

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. https://doi.org/10.1016/j.ijhcs.2019.07.009

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Frei, F. X. (2006). Breaking the trade-off between efficiency and service. Harvard business review, 84(11), 92.

Fu, K. K., Yang, M. C., & Wood, K. L. (2015). Design principles: The foundation of design. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

Glaser, B. G., & Strauss, A. L. (2006). The discovery of grounded theory: Strategies for qualitative research (Reprinted ed.). New Brunswick London: Aldine Transaction.

Gläser, J., & Laudel, G. (2009). Experteninterviews und qualitative Inhaltsanalyse: Als Instrumente rekonstruierender Untersuchungen (3rd ed.). VS Verlag für Sozialwissenschaften, Wiesbaden.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. International Conference on Information Systems (ICIS), Seoul, Korea.

Gregor, S. (2002). Design theory in information systems. Australasian Journal of Information Systems, 10(1), 1.

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. Journal of the Association for Information systems, 21(6), 2.

Gronroos, C. (1988). Service quality: The six criteria of good perceived service. Review of business, 9(3), 10.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field methods, 18(1), 59-82.

Herrera, A., Yaguachi, L., & Piedra, N. (2019). Building conversational interface for customer support applied to open campus an open online course provider. International Conference on Advanced Learning Technologies (ICALT), Maceio, Brazil. https://doi.org/10.1109/ICALT.2019.00011

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. Conference on Human Factors in Computing Systems (CHI), Montréal, Canada. http://dx.doi.org/10.1145/3173574.3173989

Io, H. N., & Lee, C. B. (2017). Chatbots and Conversational Agents: A Bibliometric Analysis. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).

Janssen, A., Cardona, D. R., & Breitner, M. H. (2020a). More than FAQ! Chatbot Taxonomy for Business-to-Business Customer Services. International Workshop on Chatbot Research and Design.

Janssen, A., Passlick, J., Cardona, D. R., & Breitner, M. H. (2020b). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. Business & Information Systems Engineering, 62(3), 211–225. https://doi.org/10.1007/s12599-020-00644-1

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. Business & Information Systems Engineering, 63(1), 5–20.

Johnston, R. (1995). The determinants of service quality: satisfiers and dissatisfiers. International journal of service industry management.

Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: a services marketing research agenda. Journal of Services Marketing, 34(3), 389-398. https://doi.org/10.1108/jsm-01-2019-0043

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2019). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. International Workshop on Chatbot Research and Design, Springer, Cham. https://doi.org/10.1007/978-3-030-39540-7_13

Lester, J., Branting, K., Mott, B. (2004). Conversational Agents. The Practical Handbook of Internet Computing.

Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-art analysis of adopting AI-based conversational agents in organizations: A systematic literature review. Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

López, G., Quesada, L., Guerrero, L. (2017). Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. International Conference on Applied Human Factors and Ergonomics, Los Angelese, CA, USA.

Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? Journal of Service Theory and Practice, 30(3).

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. Business & Information Systems Engineering, 61(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Mayring, P. (2014). Qualitative content analysis. Theoretical foundation, basic procedures and software solution.

McTear, M. (2018). Conversation modelling for chatbots: current approaches and future directions. Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung, 175-185.

Meuser, M., & Nagel, U. (2009). Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage. In S. Pickel, G. Pickel, H.-J. Lauth, & D. Jahn (Eds.), Methoden der vergleichenden Politik- und Sozialwissenschaft (pp. 465-479). VS Verlag für Sozialwissenschaften.

Meyer von Wolff, R., Hobert, S., Masuch, K., & Schumann, M. (2020a). 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), 3. https://doi.org/10.17705/1pais.12203

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2019). How May I Help You?–State of the Art and Open Research Questions for Chatbots at the Digital Workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2020b). Einsatz von Chatbots am digitalen Büroarbeitsplatz–Eine praxisorientierte Betrachtung von Einsatzbereichen, Wirkungen und Handlungsempfehlungen. HMD Praxis der Wirtschaftsinformatik(57), 413-431.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2021). Sorry, I can't understand you! – Influencing factors and challenges of chatbots at digital workplaces. International Conference on Wirtschaftsinformatik (WI), Essen, Germany.

Mimoun, M. S. B., Poncin, I., & Garnier, M. (2012). Case study—Embodied virtual agents: An analysis on reasons for failure. Journal of Retailing and Consumer services, 19(6), 605-612.

Möller, F., Guggenberger, T. M., & Otto, B. (2020). Towards a Method for Design Principle Development in Information Systems. International Conference on Design Science Research in Information Systems and Technology.

Myers, M. D. (2019). Qualitative research in business and management. Sage.

Niemann, M., Appel, M., Repp, N., & Steinmetz, R. (2009). Towards a Consistent Service Lifecycle Model in Service Governance. Americas Conference on Information Systems (AMCIS).

Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer acceptance of AI in service encounters: understanding antecedents and consequences. In Handbook of Service Science, Volume II (pp. 77-103). Springer.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. Journal of Marketing, 49(4), 41-50.

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Premathilake, G. W., Li, H., Liu, Y., & Helander, N. (2021). A review of the empirical literature on service robots in information systems literature. Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A new organizational chassis for artificial intelligence - Exploring organizational readiness factors. European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden.

Rädiker, S., & Kuckartz, U. (2019). Analyse qualitativer Daten mit MAXQDA. Springer.

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. Americas Conference on Information Systems (AMCIS).

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding whether and how to deploy chatbots. MIS Quarterly Executive, 20(1), 4. https://doi.org/10.17705/2msqe.00039

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. Internet Research, 30(1), 1-18.

Semmann, M., Grotherr, C., Vogel, P., Bittner, E., Biemann, C., & Böhmann, T. (2018). Intelligent Collaboration of Humans and Language-Based Assistants (INSTANT). International Conference on Information Systems (ICIS), San Francisco, USA.

Shah, H., Warwick, K., Vallverdú, J., & Wu, D. (2016). Can machines talk? Comparison of Eliza with modern dialogue systems. Computers in human behavior, 58, 278-295.

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they Really Useful? LDV-Forum.

Sousa, D. N., Brito, M. A., & Argainha, C. (2019). Virtual Customer Service: Building your Chatbot. International Conference on Business and Information Management (ICBIM), Paris, France.

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2019). How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electronic Markets, 30, 369–403. https://doi.org/10.1007/s12525-019-00359-6

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. European Conference on Information Systems (ECIS), Verona, Italy.

Wang, Y., Xiong, M., & Olya, H. (2020). Toward an understanding of responsible artificial intelligence practices. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Xiao, L., & Kumar, V. (2019). Robotics for customer service: A useful complement or an ultimate substitute? Journal of Service Research, 24(1), 9–29.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. Conference on Human Factors in Computing Systems (CHI), New York, NY, United States. https://doi.org/10.1145/3025453.3025496

Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020a). A review of the empirical literature on conversational agents and future research directions. International Conference on Information Systems (ICIS), India.

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020b). Towards developing trust-supporting design features for AI-based chatbots in customer service. International Conference on Information Systems (ICIS), India.

Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., & Leimeister, J. M. (2020c). The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. International Conference on Information Systems (ICIS), India.

12 Publication No. 4: Uba et al. (2023)

Uba, C., Lewandowski, T. & Böhmann, T. (2023). The AI-based Transformation of Organizations: The 3D-Model for Guiding Enterprise-wide AI Change

Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS), Hawaii (USA).

Abstract

Artificial Intelligence (AI) is increasingly gaining importance for organizations due to its immense potential for value creation and growth. However, companies struggle to tap this potential, as many AI projects fail in the early stages because of lacking guidance and best practices. To shed light on how AI adoption and transformation can be approached and what challenges organizations face, we analyzed eleven organizations of varying sizes and industries. Drawn on these insights, we identify four transformation types distinguished by different AI transformation stages and journeys. Furthermore, we develop a 3D-Model to guide enterprise-wide AI change and propose concrete recommendations for action on each dimension. Our findings help practitioners navigate, manage, and (re)evaluate their AI strategy for an enterprise-wide transformation.

Keywords: Artificial intelligence · AI transformation · AI adoption, Multi-case study · Practice-based IS research

12.1 Introduction

AI is a significantly disruptive technology for organizations (Benbya et al., 2020). At this point, 37% of global companies have incorporated AI into their businesses and products (Jovanovic, 2023), marking AI as a fast-growing technology and a fixed point on many more organizations' future agendas (Sagodi et al., 2022).

AI heralds various potentials for organizations, including increased revenues, improved customer interactions, and boosted business efficiencies (Alsheibani et al., 2020). Due to the varied application possibilities, AI is increasingly incorporated as a crucial strategic, innovative, and therefore, IT transformational element in organizations to achieve a competitive advantage (Alsheibani et al., 2019b).

Despite its potential, AI's management and strategic involvement are seen as a challenge in the recent academic discourse for practitioners (Fukas et al., 2021). Organizations have no strategic overview of where to start an AI-based transformation (Fukas et al., 2021). Consequently, AI adoption for many is still in its infancy, and organizations struggle to incorporate AI into their product and IT (service) landscape (Laut et al., 2021; Pandl et al., 2021). Currently, only 5% have comprehensively integrated AI (Pumplun et al., 2019), while a recent survey outlines that 65% of executives perceived no immediate improved value relating to their AI endeavors (Pandl et al., 2021). In this vein, AI in organizations is frequently closely connected with disappointed and exaggerated expectations. AI projects presently are highly explorative and remain experimental, often even already failing as pre-production proof-of-concepts (Benbya et al., 2020).

Organizations are increasingly aware that AI management is different from traditional enterprise IT endeavors and novel approaches are needed to sustain AI-based technologies (Berente et al., 2021). This is because AI comprises a complex bundle of technologies and applications, necessitating a new holistic understanding by managers of how to communicate, lead, coordinate, and control them (Berente et al., 2021). Additionally, it is also due to the different technological properties AI possesses in comparison to conventional information systems (IS), such as, for example, being learning systems with black box characteristics and context-sensitivity (Sagodi et al., 2022). As a consequence of these challenges, organizations need to build capabilities for mastering new AI management activities, such as establishing data security and management, AI governance, AI strategic alignment, regulatory approvals for AI-based decisions, and ethical scrutiny of learning systems (Jöhnk et al., 2021; Kruse et al., 2019; Pumplun et al., 2019).

Generally, current research has predominantly focused on AI systems' general design and applications or underlying technological advancements (Nguyen et al., 2022; Pumplun et al., 2019). Research from the AI management perspective has been focused on initial AI maturity models as well as structural and psychological prerequisites (Eitle et al., 2022). Further, on the AI management side, organizational readiness and adoption factors have also been studied (e.g., Pumplun et al., 2019).

However, helping organizations systematically develop AI capabilities is still a scarce field of knowledge in research and practice. This is unsatisfactory, given that various organizations face the challenge of establishing enterprise-wide AI programs and initiatives (Eitle et al., 2022). First authors already highlight the importance of thinking broadly when laying the foundation for AI transformation (e.g., Fridgen et al., 2022). Hence, we answer the following research question (RQ):

RQ: *What are the key activities for driving enterprise-wide AI change and capabilities?*

To answer this RQ, we conducted a multi-case in-depth study. We collected data from eleven organizations implementing AI, ranging from manufacturers to service providers. We draw on our insights to illustrate organizations' different levels and approaches regarding enterprise-wide AI adoption and transformation. Our article showcases three AI transformation dimensions organizations can pursue, containing a broad overview of possible strategic directions and corresponding recommendations to guide AI transformation effectively.

12.2 Conceptual Background

AI has recently gained much attention in organizations by comprising a set of technologies able to sense, reason, and facilitate conceptual learning and decision-making (Bock et al., 2020). Due to AI's variety of technologies and skills—resulting in several application cases that have changed over time—there exists neither in practice nor science a consensus on the exact meaning of the umbrella term "AI" (Alsheibani et al., 2019b; Nguyen et al., 2022).

Generally, researchers define AI as a generic concept for technologies capable of mimicking human behavior and learning how to solve tasks usually performed by human intelligence (Castillo et al., 2020). In this sense, AI differs from conventional IS by being able to learn and make decisions generally based on input data rather than predefined or deterministic rules (Crowston & Bolici, 2019). While early AI attempts were restricted by limited computing power and data, contemporary AI exemplifies greater autonomy and more profound learning capacity, as it can use cognitive or conversational functions and interact with an immense amount of data (Baird & Maruping, 2021;

Berente et al., 2021). As a result, AI technologies possess tremendous potential for organizations and offer a transformative role in various sectors and industries, for instance, by reinventing business models, augmenting or automating work, and providing performance improvements for organizations in general (Collins et al., 2021). AI has applications in manifold application domains, such as chatbots utilizing natural language processing, facial recognition employing image processing, and recommender systems fueled by machine learning (ML) algorithms.

However, despite the general potential of AI and the steep increase of AI applications in organizations, it becomes clear that managing AI *"is unlike information technology (IT) management in the past"* (Berente et al., 2021, p. 2). AI will not simply fit into previous concepts of managing traditional IT technologies. This leads to a situation in which organizations or respectively organizational decision-makers need to adapt their behavior, reinterpret their approach, and understand relevant nuances of AI capabilities and their continuous strategic management (Fridgen et al., 2022). In addition to already investigated fundamental readiness and adoption factors to ensure a secure foundation of AI technologies (cf., Jöhnk et al., 2021; Pumplun et al., 2019), practitioners need a holistic view of AI application as an organizational transformation involving multiple new activities and engagements that need to be controlled and directed. In this context, research on AI management and transformational change is scarce. The business and strategy-oriented understanding of the management and long-term value-adding implementation of AI for enterprise-wide change is still new to researchers and organizations even though it is a vital capability in the future (Fukas et al., 2021; Sagodi et al., 2022). Many organizations appear to be at the stage where they are attempting to create a business case for AI. It is stated that many present-day AI initiatives and strategies fail, leading to a more pessimistic outlook (Alsheibani et al., 2019a; Sagodi et al., 2022). To combat this sentiment and help organizations further develop their initiatives, our research gives guidance on how to implement AI as an organization-wide change to help generate its proposed value and sustain AI efforts.

12.3 Research Design

Our research goal is to understand how different companies with diverse transformation levels approach AI implementation and to examine current best practices and challenges. To obtain a broad picture, we conducted a multi-case study including eleven cases ranging from e-commerce and manufacturing organizations to insurance providers and media companies (Yin, 2003). We purposefully investigated organizations of varying sizes to encompass various AI transformation stages and approaches. Moreover, we selected over 30 study participants, including IT executives, senior managers, chief data scientists, and other IT experts, as well as Chief Information Officers (CIO) and Chief Digital Officers (CDO) (see Appendix for more details).

Our study data was collected in a two-step procedure: First, we conducted six focus group sessions, which ranged from 1h to 1h 15min, with IT executives and experts of four different organizations using video conferencing tools. During these sessions that took place between January and August 2021, organizations in turn presented their AI strategy and their challenges in adopting AI. Afterward, the sessions concluded with open discussions among all company representatives and researchers. Based on these discussions and the material presented, we derived common fields of AI activity, as depicted in **Table 1**. These fields were iteratively validated in the upcoming focus group sessions to make additions and discuss critique. Second, we conducted seven semi-structured interviews, each with organizations not represented in the focus groups. The semi-structured interviews allowed for adaptability while enabling us to structurally incorporate insights from the focus groups (Myers & Newman, 2007). The interviews took place in August 2021 and lasted a little over 50 minutes on average. Drawing on the observations from the focus groups, we made certain to generally cover the derived fields of activity, enabling us to better compare and categorize the studied cases afterward.

Table 1. Focus group result – Fields of AI activity

1	Strategy & Governance
2	Development Lifecycle
3	Data Management
4	Tools & Platforms
5	Process & Work Design
6	Service Design
7	Capability Building
8	Ecosystem Integration

Two researchers coded the recorded and transcribed interviews along with protocols of the focus group sessions and additional company materials (e.g., slides, internal documents) separately in

MAXQDA. We used open, axial, and selective coding to examine interesting aspects, find relationships between these aspects and finally identify aspects explicitly relating to AI strategy and journey (Corbin & Strauss, 1990). During the entire coding process, we repeatedly discussed our codes and interpretations of the material to ensure our results' consistency and validity.

12.4 Types & Selected Example Cases

In our study, we identified four organization types distinguished by different AI transformation stages and journeys. Namely, *Explorers*, *Intermediates with a focus on process optimization, Intermediates with a focus on customer value creation,* and *Strategic Visionaries*. In the following, we describe what criteria characterize each type and then for each, present an exemplary case highlighting the type's AI approach as well as their findings, realizations, and learnings.

12.4.1 Explorers

Explorers are companies that are interested in AI but possess little to no experience in dealing with it. They are curious to discover how AI can be employed as a beneficial technology in their organization to create value for internal or external applications. We characterize them as *Explorers* as they are still in the beginning stages, figuring out precisely what AI entails and exploring which use cases might be suitable to gain first practical insights and experiences. Our study shows that *Explorers* score relatively low in terms of overall digital maturity and that they are usually active in traditional industries that are not well known for their digital affinity. Besides finding new use cases, *Explorers'* main challenges are building up the fundamental expertise to get started and sustain their AI efforts, as well as establishing a profound data infrastructure fueling these efforts.

Case Example Explorer

An *Explorer* case example anonymously referred to as BROKER (Case ID 11), is a business insurance broker and employer to over 1,000 people. It provides services on businesses' insurance needs and risk management.

In the insurance industry, there is a plethora of documents like policies, contracts, and reports that need constant analysis and evaluation. For instance, benchmarking insurance offers is a manual and document-intensive task which needs a lot of time and expert knowledge. To improve and automate this process, BROKER's pilot AI project set the goal to automatically turn document content into structured data to subsequently automatically benchmark different insurance offers.

BROKER teamed up with an external technology partner who contributed AI skills which BROKER was lacking at the time.

Due to the novel nature of the project as well as the non-deterministic nature of AI, BROKER quickly realized that the prospect of success was not entirely clear. Time, people, and resources would need to be invested even if the company did not know "*if it is generally even possible to solve the assignment with the available technology,"* as BROKER's Digital Transformation Manager put it. As a result, the company consciously set its project objective beyond solely implementing an AI-based system. BROKER's CDO described their approach as follows: *"A goal is to especially stake out the technology's general performance and test the collaboration with such a partner. Which in a classic project you would not like to see as an objective."*

By outsourcing AI development, BROKER was able to quickly get started on the AI project. This is in line with BROKER's general bottom-up AI approach, where AI skills are not built in-house but outsourced to external IT providers or, if necessary, incorporated by hiring employees when capability gaps appear.

An early realization BROKER had, is the changed role the domain experts of the functional teams play in the ideation, development, and operation of AI. Specifically, their indispensable part in validating the system's accuracy which necessitates a deeper understanding of the technology used in the project. As the CDO explained: *„In the course of the project, we demand the department in a different way. We concern them a lot with what the technology is doing just at [that moment]."*

Questioned on data management regarding data responsibility, infrastructure, and strategy BROKER's CDO stated: *"We are now at the point where we're asking [ourselves]: What do we have to establish? What do we actually need?"* Facing this challenge by fully assessing all requirements and freeing up resources is an ongoing field of activity for BROKER.

12.4.2 Intermediates

Intermediates are companies that are beyond the stage of developing proof-of-concepts. They have successfully implemented at least one complete AI system that is up and running. Additionally, they have at least one core AI or data science team, where the company's current collective AI expertise is concentrated. The core AI team is commonly responsible for selecting use cases and managing the project's ideation and incubation phase. Further, the development of AI systems or the management of third-party solutions usually also falls within the responsibility of the organization's core AI team. Generally, *Intermediates* have a data infrastructure that facilitates the development of AI projects, although for some setting up a fully satisfactory data infrastructure is an ongoing process.

On their way to further build expertise and develop new solutions, we found two types of *Intermediates:* (1) Those who focus on using AI to improve the efficiency of internal processes, and (2) those with focus on directly impacting the customer's experience. Although for some *Intermediates* there might be some overlap, we found that most chose either one or the other approach in their AI journey.

Intermediates with a focus on internal process optimization

Many reasons exist for *Intermediates* to focus on internal process optimization. Firms with internal processes characterized by being resource and time-intensive or companies with minimal customer interaction (e.g., manufacturing industry) are more likely to fall in this category. Moreover, we found that organizations, being heavily regulated regarding data use or those working with sensitive data, usually also focus on internal process optimization. This might be due to legal or ethical difficulties connected to using AI for customer-facing applications.

Case Example Intermediate with a focus on internal process optimization

One *Intermediate* case *focused on internal processes* is a statuary health insurance provider in Germany, anonymously referred to as INSURER (Case ID 05). The organization insures several million people and is an employer to over 10,000 employees.

INSURER has successfully developed multiple AI applications, which tremendously improved work and business processes. Notable cases are an inbox classification application, which assigns mails to the responsible employees, a hospital invoice auditing application, and an image processing application for recognizing different stamps on documents.

All of INSURER's AI projects are built and developed in-house. This conscious decision was motivated by the company's wish to not have its AI ambitions be directed by and dependent on external actors. To realize this, significant effort was put into building a core AI team, even before the first use cases were developed. The core AI team was set up to function as an incubator and a pipeline for AI ideas and AI development. INSURER's head of IT innovations described the function of this team as follows: *"[it's] a group of people that always appears when an idea or the request for some brainstorming comes up in a business division […] This squad then starts to think, to analyze […] and to make a model, but most importantly also starts to implement."*

To cover this wide array of tasks, INSURER turned away from establishing a strict role differentiation often found in AI teams (e.g., data scientists, analysts, or ML engineers). On the contrary, INSURER decided to encourage and even expect AI team members to broaden their skill set while keeping their specialization, allowing for a more flexible use during ideation and development.

One AI-specific challenge is the management of the AI lifecycle. Time and resources need to be invested even after deployment and system rollout. Unlike conventional systems, the monitoring of an AI-based system's performance during operation is of importance, whereby domain experts play a crucial role. As the head of IT innovations described: *"There is a huge difference, because suddenly it's no longer IT monitoring something, but it's the domain experts that have to monitor it."* So even after closely collaborating during the development, domain experts were especially in demand. This led INSURER to realize that AI expertise and understanding needed to be built outside the developer teams as well. Conveniently, this realization coincided with the business division's desire to better *"understand what AI entails and ask the right questions"* as INSURER's AI architect put it.

While for some companies setting up a monitoring framework for AI applications on internal processes might be uncharted territory, for INSURER this was nothing new. Due to the high regulatory requirements for health insurers set out by the German Social Security Code, INSURER already had many monitoring tools in place which surveil their processes and determine quality criteria for these processes. Adding AI into the mix therefore did not significantly raise complexity because as INSURER's AI architect remarked: "*[the domain experts] already monitor [their] processes anyway and AI is an automation component in this process."* Further, due to the slow-changing nature of the health insurance field, INSURER's AI-based applications were less prone to be confronted with abruptly changing environments or input variables, which is a typical challenge such systems might face in other industries. However, when attempting to design services that are customer facing, this regulatory-driven advantage turned into a disadvantage. The strict guidelines on the utilization and possible user applications set by the Social Security Code consequently explain why INSURER focused on leveraging AI for internal process optimization rather than customer value creation.

Data management is a field of activity that gained new priority when INSURER further pushed its AI endeavors. As INSURER's AI architect recounted: *"that's when we teased the topic for the first time: […] we need a data lake, because without fast data delivery and regulated but good data access, we won't pick up speed."* Aside from further improvements to its data infrastructure INSURER was faced with numerous new and old questions regarding data management. Like questions on how to manage the data preparation process or on what constitutes good data quality from the perspective of AI. All this goes to show that the topic of data management is a core field of activity which needs continuous attention even after a solid foundation for working with AI is set.

Intermediates with a focus on customer value creation

Intermediates with a focus on customer value creation aim at using AI to enrichen offered products or create new customer-facing services. Thus, use cases are focused on the customer experience and how AI can be a facilitator for improving it.

Case Example Intermediate with a focus on customer value creation

A media group and newspaper publishing house from our case study, anonymously referred to as PUBLISHER (Case ID 08), employs almost 5,000 employees and has been in business for more than five decades.

Because of technological advancements and changes in how information and media are presented and consumed today, PUBLISHER’s industry is rapidly changing and experiencing disruption. Following and leveraging this trend is a big challenge PUBLISHER is facing. Consequently, AI-supported efforts are majorly focused on its online publishing platforms and thus the content customers read and interact with online.

PUBLISHER has no overarching AI strategy. It follows a rather practical approach to AI. As PUBLISHER’s executive board member and head of digital research and development (R&D) explained: *“It's always concrete questions where—independently of a larger strategic context—attempts are being made at developing the best solution for a concrete problem.”* Many questions PUBLISHER faced were not necessarily unique, which is why most AI-based applications were bought from outside software providers. Third-party solutions that are successfully in use are text-to-speech and hate speech detection applications. The latter for example monitors PUBLISHER’s forums and immediately blocks posts containing hate speech, which in turn removes the need for constant supervision by an employee. These AI applications came pre-trained and merely needed to be fed with a minimal amount of PUBLISHER’s own data to be fully operational. This way, PUBLISHER only had to invest resources in intermittently checking and retraining the software’s algorithm.

Other AI-specific considerations and questions PUBLISHER had in conversations with third-party providers concerned the database used to train these models. GDPR requirements and potential future regulations could mean that the employment of these applications might not be possible in the future. With third-party cookies for example, INSURER had a lot of concerns regarding future-proofing. The potential risk was highlighted by INSURER's head of R&D as follows: *"We all assume that third-party cookies will disappear from the market within the next two to three years. [...] And then of course I ask: 'Do you use third-party cookies as your database? And if so, what will you do?' […] 'How is the data collected?', 'Will there be technological or legal changes that will make the use of this data basis impossible in the future?'"*

Despite the large amount of knowledge on and positive experiences with out-of-the-box AI solutions, PUBLISHER decided to additionally put efforts into building and encouraging their own expertise in developing AI. When considering what project to choose, PUBLISHER factored in two main considerations. First, whether the respective field was part of the core business and second, how company-specific the context and especially the data of the problem was. According to the head of digital R&D, the subject of churn prediction fits these criteria: *"This is know-how that we would like to have in the company because everything that has to do with […] generating subscriptions and avoiding subscription cancellation is a core business of ours. So that's one of those disciplines that we must master and […] it is all the better, the better we master it also technologically."* Hence, PUBLISHER initiated a churn prediction project in collaboration with an external technology partner. The goal of the project was to develop an AI-based solution which predicts if a customer is likely to cancel their subscription and proposes what needs to be done to prevent this. The team consisted of three PUBLISHER employees and two external team members. This decision enabled the company to outsource some parts of the development while still doing most of the work in-house. The mix of building and outsourcing allowed PUBLISHER to quickly get started on the project, yet ensured that AI know-how stayed in the company long term.

In terms of data and data management, a challenge PUBLISHER faced was how to work with multiple data sources. Thinking about this roadblock, the head of R&D recounted: *"What we were not able to do in the past is link data from different data sources because we knew it was going to [lead to] chaos."* To tackle this problem PUBLISHER introduced a new role namely a *"head of data"*. One main task of this role was to ensure data integrity which placed the foundation for PUBLISHER on achieving a more coherent and workable database and infrastructure.

12.4.3 Strategic Visionaries

In contrast to *Explorers* and *Intermediates, Strategic Visionaries* have numerous running AI-based applications. Questions surrounding data management, tool use, or basic AI functionality are not at the forefront as these are areas in which *Strategic Visionaries* have already built a high level of expertise. *Strategic Visionaries* go further than the other transformation types; they see AI as a key enabler and a competitive advantage. As a result, they explicitly define AI as a part of the company's business strategy. Extracting best practices, developing guidelines, setting up a comprehensive governance, and pipelining AI incubation are all fields *Strategic Visioners* are highly invested in. These investments in turn facilitate the scaling of AI solutions and thus make large-scale implementation of AI possible.

Case Example Strategic Visionary

A *Strategic Visionary* case example company, anonymously referred to as RETAIL (Case ID 01), employs well over 20,000 employees. The company has been in business for decades and is presently very active in the field of e-commerce.

RETAIL's AI vision is to be a leading AI company by 2030, and it has the best conditions to do so. Presently, a total of 40 ML-based products and services are part of its portfolio. Application scenarios range from general applications like e-mail classification and forecasting to e-commerce-specific ones, like dynamic pricing or image similarity-based outfit recommendations. In terms of AI expertise, RETAIL employs over 20 teams with roughly 100 developers that work on implementing ML as part of their output. Concerning AI governance RETAIL concerned itself with defining development standards for example regarding system architecture and coding best practices or understanding regulatory requirements. RETAIL also gave special attention to the development of ethical guidelines, which entailed ensuring fit with its corporate values and considerations on public image and impact.

One main pillar of RETAIL's strategy is the expansion and reinforcement of AI knowledge and know-how. At RETAIL this was approached *"in depth, in the form of a high level of excellence among all data scientists and engineers [and] in breadth, in the form of a solid basic knowledge among all roles that are indirectly involved"* (RETAIL internal documents). In other words, all levels starting from (top) management down to employees that are not in direct contact with AI use cases are educated on AI.

Starting with process automation RETAIL established a framework for streamlining AI ideation and incubation. In RETAIL's approach AI incubation teams were set as the heart of the operation. Their task was to first discover, prototype, and validate ideas. After validating an idea, the incubation team would get together with the product team to plan, develop, and then implement and automate the application. This centralized approach helped RETAIL deploy automated processes more quickly and increase its general expertise on the incubation of AI innovations.

A major bottleneck RETAIL repeatedly encountered was its engineering capacities. In view of this bottleneck, RETAIL set the goal to build its own cloud-native AI platform. This meant that establishing development standards and best practices gained an even greater priority, as the answers to this would be the foundation of the platform.

In contrast to *Explorers* and *Intermediates* RETAIL as a *Strategic Visionary* utilized partnerships and collaborations not only to develop use cases but also to enable the general exchange on knowledge, processes, and technologies. Potential partners were not limited to IT service providers but included actors in other industries, in education, and in research.

12.5 Recommendations for Action

The cases in our study illustrate different AI transformation stages of organizations. Before setting up a comprehensive AI strategy, it is advisable for companies to first assess which one of the four transformation types they are. The main factors in assessing this are AI expertise, the existence of AI enabling data infrastructure, the number of successful AI projects, and the scope of the organization's AI governance. Knowledge of the company's transformation type allows the company to determine its positioning among competitors, manage expectations, and set appropriate goals when initiating AI projects. For example, BROKER as an *Explorer* had to realize that one main objective for them is to generally experiment with AI and test the collaboration relationship with their new external IT partner. In contrast, new projects for a *Strategic Visionary* would not be focused on experimenting but would be very result-oriented with a focus on finding and defining best practices, like in RETAIL's case.

As a result, it is only after assessing the transformation type that an organization is ready to set up a strategy for enterprise-wide AI change.

Based on our research and the companies' experiences, we presented a 3D-Model for guiding AI transformation. As depicted in **Figure 1**, the three dimensions for strategic action are (1) *Core Capability Building*, (2) *Value Stream Embedding*, and (3) *Organizational Enabling*. These dimensions span the space of possible AI activities and configurations.

In the following section, we explain the dimensions and for each, present recommendations for possible actions practitioners setting up an AI transformation strategy can take.

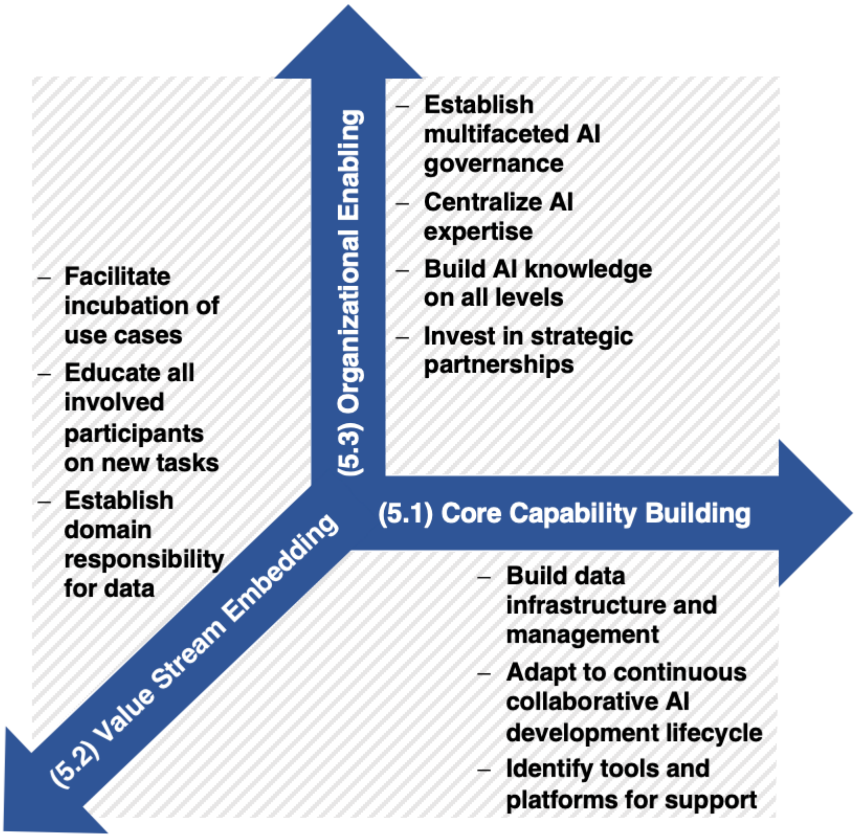

Figure 1. 3D-Model for guiding AI transformation

12.5.1 Recommendations on Core Capability Building

A company's AI core capability is comprised of the activities and infrastructure an organization must have to successfully initiate, implement, and maintain AI activities and applications. As some level of capability is necessary to get started on an AI project, *Explorers* must especially set this dimension as a top priority. Nevertheless, *Intermediates* and *Strategic Visionaries* must continually (re)evaluate their core capabilities to determine if existing and planned AI activities are still sufficiently supported. Consequently, companies must set aside long-term resources for managing and observing these new and changing requirements.

Build data infrastructure and establish data management

Concerning data management, AI introduces novel requirements. Structured and unstructured data must be centrally accessible and working with multiple actors requires regulated data access. Moreover, when consolidating multiple data sources, data quality and integrity become crucial to identify items across different data sources. To ensure this, companies should set up a data infrastructure that facilitates regulated and fast provisioning of different kinds of data (e.g., data lake and data hubs). Additionally, if absent, organizations should introduce a new role, such as head of data or data officer, whose main job is to manage the data infrastructure and keep track of ongoing and upcoming needs.

Adapt to AI's continuous and collaborative development lifecycle

The non-deterministic nature of many AI algorithms introduces a new volatile variable into the development and operation of AI-based systems. Unfavorable changes in continually learning AI applications or changes in the application's environment (e.g., shifting input data) require monitoring and regular maintenance. Companies must recognize this and set aside resources in terms of time and people during development and beyond.

Additionally, the development process calls for close interdisciplinary collaboration of product teams, AI developers, system engineers, and legal teams. To address this, organizations should establish fixed and flexible possibilities for exchange between those involved in the development lifecycle.

Identify tools and platforms that support and best fit AI ambitions and capabilities

In our research, we observed that most companies use cloud-based solutions to set up their infrastructure for AI development. This is advisable as it allows for more flexibility and supports the collaborative nature of AI development.

In terms of development tools, there is an abundance of tools to choose from when working with AI. For many case organizations, we noticed an uncontrolled development regarding AI tool use, as most developers are often free to select preferred tools. However, we suggest organizations, especially when moving from the *Explorer* to the *Intermediate* stage, to decide how much they want to control tools, and if tooling should be limited to a consolidated set. At the very least companies should ensure compatibility inside collaborating teams and with existing systems.

12.5.2 Recommendations on Value Stream Embedding

AI projects are initially highly experimental and often already fail as proof-of-concepts. A reason for this is that organizations focus on conceptual or technical aspects, leading to a disconnect from the concrete value stream or actual process environment. However, AI adoption and change require seamless integration of existing business processes, knowledge systems, user interfaces, and customer channels. Thus, value stream embedding is a key dimension as it describes the sum of all endeavors and measures taken to seamlessly optimize and automate specific services, internal processes, and workflows.

Actively facilitate incubation of use cases

During project incubation, different teams with diverse perspectives and responsibilities interact and communicate on specific needs relating to a (potential) AI solution. These exchanges not only act as a catalyst and accelerator for innovation but are also an enabler for value stream embedding and seamless integration—if all perspectives are considered. For this reason, companies should provide numerous touch points and opportunities for interdisciplinary and interdivisional collaboration and design during this process. Possible formats can be workshops or periodic brainstorming and discussions with product teams, AI and software developers, system engineers, or other affected parties.

Regarding resource management, it is advantageous for companies to separate the incubation process from general operations. In doing so, firms ensure that investments and efforts into new use cases do not interfere with daily business operations.

Educate all involved participants on new tasks and responsibilities

Uncertainty on how AI changes the workflow of the involved employees can be a reason for poor integration. To combat this, we recommend that companies educate the teams on how the introduction of AI or the AI development process itself changes their duties. For instance, affected departments must be educated on their role as a critical asset for validating and monitoring their AI applications. Where possible, companies should routinize these new tasks, as introducing them as a part of a daily, weekly, or monthly routine appropriately consolidates them.

Establish domain responsibility for data

Identifying the correct and necessary data and in part obtaining it, is a task that falls within the responsibility of the domain experts as it necessitates a deep understanding of the domain-specific processes. Our research suggests that especially for data-intensive or data-driven departments this task is reoccurring throughout the project duration which necessitates particular attention and sufficient resources. For such cases, it is advisable to appoint a department member that is responsible for this task. This role defined as data steward by one of our study's cases does not need to be filled by someone who is overly technically versed, as this role should mainly concern itself with data content and contextualization of the data. Technical considerations and issues still mainly fall within the responsibility of AI developers and members of the data management team.

12.5.3 Recommendations on Organizational Enabling

Organizational enabling describes the strategic and enterprise-wide integration and establishment of AI. Primarily initiated by and in the responsibility of the company's (top) management, activities in this dimension are not focused on the individual AI solution or process but are concerned with strategically enabling the organization.

Establish multifaceted AI governance

Firms are confronted with a multitude of challenges that arise when they leave the explorative stage to move on to implement truly embedded AI systems and scale their development. Establishing best practices for AI development as part of a company's technical governance is just one area of governance with which a company must concern itself. Other types of governance owing to the AI transformation are regulatory governance, dealing with legal requirements, organizational governance, entailing the business' structure, and ethical governance, which reflects company-specific ethical guidelines.

Centralize AI expertise & knowledge in the AI core team

On the one hand, we observed that organizations often have a hard time finding AI experts but on the other hand we also found that they struggle to leverage existing AI expertise. Considering these difficulties, we suggest companies concentrate their AI expertise on one AI core team. This core team works as a foundation for mutual education and knowledge transfer. Additionally, the team is valuable for managing AI efforts and can function as a point of contact for company-wide AI-related ideas and propositions. Thus, this centralizing knowledge approach helps streamline AI efforts, even for smaller companies. Larger companies with many AI experts can either continue with this centralized approach by allowing for multiple AI core teams with different specializations or can alternatively experiment with decentralized approaches where AI experts might be embedded into functional teams.

Build AI skills and knowledge on all employee levels

Knowledge in general and employee capability specifically are concerns for all transformation types. While some companies primarily focus on certain employees or specific teams, teaching AI to enable the entire organization must address all company levels. Consequently, we suggest that companies go beyond directly involved product and developer teams. Software developers, system engineers, and other actors directly involved in developing or using AI applications should at least have a basic understanding of AI methods and standards. Further, the (top) management must be educated on AI potential and needs, to ensure sufficient resources are allocated for developing and maintaining AI-based systems. We advise businesses to offer seminars and information workshops open to the general staff. Programs like this demystify AI, reduce hesitations, foster an open-minded innovation culture, and promote interdisciplinary interactions, constituting an ideal foundation for launching new AI projects.

Invest in strategic partnerships

AI change does not need to be an isolated process. Collaborating, outsourcing, and communicating with other players enable significant and competitive advantages. We suggest collaborating with external IT providers for companies that have little expertise but want to quickly get started on their AI transformation. This way, the entire development, or at least those parts the company has not mastered yet, can be outsourced.

Beyond outsourcing and collaborating on projects, we recommend organizations that want to advance their AI endeavors to look for fresh impulses. Having an exchange on AI activities with industry peers can be an impulse. Additionally, connecting with businesses from other industries or startups as well as players in education and research can also be beneficial for companies to gain new insights and keep pace with the continuously changing AI landscape.

12.6 Concluding Remarks

Owing to rapid advancements in AI, organizations today are presented with a myriad of exciting AI technologies and application scenarios. Thus, many companies are actively investing in utilizing and developing AI. Reducing cost, increasing productivity, and creating new services are just a few potential avenues (Alsheibani et al., 2020). However, despite great interest and initiated efforts, many companies fail at adopting and thus levering AI for their organizations (Jöhnk et al., 2021; Pumplun et al., 2019). In this vein, we conducted a multi-case study to gain insights from eleven organizations with differing AI profiles. Based on our research, we highlight four AI transformation types reflecting different transformation stages and journeys. Further, we develop a 3D-Model for AI transformation and present concrete recommendations for action on each dimension. Our insights on transformation types and dimensions for action equip practitioners with the necessary knowledge to assess their current practices and develop a roadmap for future AI endeavors. Hereby, becoming AI-savvy organizations that can unlock AI potential and retain an AI-enabled competitive position long term.

12.7 Acknowledgments

The research was financed by the German Federal Ministry of Education and Research and the European Social Fund under the "Future of Work" program (funding ref.: INSTANT - 02L18A111).

The research's data collection was supported and facilitated by the IT-Executive Club e.V..

12.8 Appendix – Table Research Method

ID	Industry	Employees	Position of Participants		Duration
01	Retail	> 20,000	CIO, Vice President Business Intelligence, Head of Data Science, Head of System Services, Corporate Responsibility Lead, Head of IT and Process Management, Technical IT Consultant	F O C U S G R O U P S	Σ= ~6h
02	Consumer Goods	> 20,000	Director Data & Analytics, Head of Data Science Hub		
03	Health Insurance	5,000-10,000	CDO, Head of AI, Product Owner		
04	Public Sector	1,000-5,000	AI Consultant, Head of Data Science & AI, Board Member (Digital Transformation)		
05	Health Insurance	10,000-20,000	Head of IT-Innovations, AI Architect, Scrum Master, IT Division Head	I N T E	45 min
06	Medical Technology	5,000-10,000	CIO, IT Demand Manager		30 min
07	Financial services & insurance	5,000-10,000	Head of Data & Data Analytics, Product Owner Data Analytics Platform		1h 5 min
08	Publishing/ Media	3,000	Head of Digital Research & Development		50 min

09	Public Sector	> 20,000	CDO, Advisor Digital Strategies	R V I E W S	1h 10 min
10	Publishing/ Media	1,000-5,000	Head of Data, Head of Data Intelligence		50 min
11	Insurance Brokerage	1,000-5,000	CDO, Consultant Digital Corporate Development, Digital Transformation Manager		50 min

12.9 References

Alsheibani, S. A., Cheung, Y., & Messom, C. (2019a). Factors inhibiting the adoption of artificial intelligence at organizational-level: A preliminary investigation. *Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Alsheibani, S. A., Cheung, Y., & Messom, C. (2019b). Towards an artificial intelligence maturity model: From science fiction to business facts. *Pacific Asia Conference on Information Systems (PACIS)*, China.

Alsheibani, S. A., Messom, C., Cheung, Y., & Alhosni, M. (2020). Reimagining the strategic management of artificial intelligence: Five recommendations for business leaders. *Americas Conference on Information Systems (AMCIS)*, Salt Lake City, United States.

Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS quarterly*, *45*(1).

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, *19*(4).

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS quarterly*, *45*(3).

Bock, D. E., Wolter, J. S., & Ferrell, O. C. (2020). Artificial intelligence: Disrupting what we know about services. *Journal of Services Marketing*, *34*(3).

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, *41*(13).

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*(1).

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, *13*(1).

Crowston, K., & Bolici, F. (2019). Impacts of machine learning on work. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Eitle, V., Zöll, A., & Buxmann, P. (2022). Organizational readiness concept for AI: A quantitative analysis of a multi-stage adoption process from the perspective of data scientists. *European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Fridgen, G., Hartwich, E., Rägo, V., Rieger, A., & Stohr, A. (2022). Artificial intelligence as a call for retail banking: Applying digital options thinking to artificial intelligence adoption. *European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Fukas, P., Rebstadt, J., Remark, F., & Thomas, O. (2021). Developing an artificial intelligence maturity model for auditing. *European Conference on Information Systems (ECIS)*, Virtual Conference.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. *Business & Information Systems Engineering*, *63*(1).

Jovanovic, B. (2022). *55 Fascinating AI statistics and trends for 2022*. Retrieved May 20, 2022, from https://dataprot.net/statistics/ai-statistics/

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Laut, P., Dumbach, P., & Eskofier, B. M. (2021). Integration of artificial intelligence in the organizational adoption – A configurational perspective. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and organization*, *17*(1), 2-26.

Nguyen, Q. N., Sidorova, A., & Torres, R. (2022). Artificial intelligence in business: A literature review and research agenda. *Communications of the Association for Information Systems (CAIS)*, *50*(7).

Pandl, K. D., Teigeler, H., Lins, S., Thiebes, S., & Sunyaev, A. (2021). Drivers and inhibitors for organizations' intention to adopt artificial intelligence as a service. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A new organizational chassis for artificial intelligence - Exploring organizational readiness factors. *European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Sagodi, A., Engel, C., Schniertshauer, J., & Van Giffen, B. (2022). Becoming certain about the uncertain: How AI changes proof-of-concept activities in manufacturing–Insights from a global automotive leader. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Yin, R. K. (2003). *Case study research: Design and methods* (3 ed., Vol. 5). Sage Publications.

13 Publication No. 5: Heuer et al. (2023)

Heuer, M., Lewandowski, T., Weglewski, J., Mayer, T., Kubicek, M., Lembke, P., Ortgiese, S. & Böhmann, T. (2023). Rethinking Interaction with Conversational Agents: How to Create a Positive User Experience Utilizing Dialog Patterns.

International Conference on Human-Computer Interaction 2023 (HCII 2023), Springer, Copenhagen (Denmark).

Abstract

Conversational agents (CAs) are increasingly used as an additional convenient and innovative customer service channel to relieve service employees, as in the studied organization. In the process of analyzing and maintaining the present AI-based agent, however, user satisfaction is low as the CA lacks understanding and offers unsatisfactory solutions to users. Nonetheless, solving requests and providing a positive user experience is crucial to relieve the service employees' workload permanently. For CAs' improvement, this study followed action design research (ADR) and used design thinking. We identified the central interaction problems (findability, welcome message, dialog control, and fallback issues) with a monitoring process and analysis. Afterward, we interviewed users about their expectations and requirements, and addressed these problems by creating user-centric mock-ups. Through a quantitative survey, the most popular solutions were implemented in a prototype. Finally, the resulting CA prototype was evaluated, showing a significantly improved user experience afterward, and design guidelines were discovered.

Keywords: Conversational agents · Chatbot user experience (UX) · Fallback strategy · Interaction design · Artificial intelligence (AI)

13.1 Introduction

CAs have become increasingly common in organizations as a central channel for customer contact (Castillo et al., 2020; Zierau et al., 2020b). Organizations introduce CAs in customer service to reduce the workload of customer service employees, leading to cognitive relief and increasing the productivity of entire service units, such as an internal or external information technology (IT) help desk (Corea et al., 2020). Unfortunately, dissatisfied users stand out during CA analysis and maintenance. This is mainly due to communication difficulties between the CA and the user, thus leading to unsatisfactory or no solutions. The challenges of today's CAs are that they are learning systems and, initially, are limited (Zierau et al., 2020c). Therefore, CAs often misunderstand user requests and fail to meet user expectations (Gnewuch et al., 2017; Janssen et al., 2021b). Additionally, CAs have problems when a user request cannot be assigned to a learned subject (e.g., intent/entity recognition). Frequently, formulations need to be more variable and holistically designed to reach a successful conclusion (Zierau et al., 2020c). Customers become particularly dissatisfied when the reformulation of their query does not lead to success. In such cases, a successful solution proposal from the CA to the customer may fail. In addition, CAs require continuous improvement, as new use cases are emerging from user queries and need to be included (Lewandowski et al., 2022b; Meyer von Wolff et al., 2022).

To tackle this issue, it is necessary to resolve the concerns and create a positive user experience. The goal is to ensure a lasting interaction with the user and always offer a solution. We characterize successful communication as ensuring a permanent interaction between the user and the CA and offering a solution in all cases. Customer satisfaction and the continued use of the CA depend on successful communication. Indeed, interaction problems are a common challenge (**Table 1**). This paper presents two novel concepts to help keep customer satisfaction and CA usage high.

Table 1. Potential Effort Matrix

Potential / Effort	Small	Medium	High
Small		Message length, Help category	Learning, Text completion, User control
Medium	Response time, sentiment analysis, small talk	Character, Forwarding, UX	Fallbacks
High	Callback, Complaint	Conversation starter, Fallback at greeting	Findability, Buttons, Message variance

The central research question was as follows: “How can a user’s CA experience be improved by analyzing, understanding, and optimizing interaction problems?” The sub-questions were “What interaction problems exist between user and CAs, why do they arise, and how are they solved?” and “To what extent can the user experience be improved?” The focus was on users and their perceived usage. The two-level category system and the restructured fallback scheme are explained later.

13.2 Background

The idea of communicating with computers using natural language (e.g., via voice or text messages) has existed for several decades (Lewandowski et al., 2021). Weizenbaum took initial steps toward a text interface between humans and information systems (IS) with *ELIZA* in 1966 (Weizenbaum, 1966). *ELIZA* is an IS that generates responses to text inputs, simulating a psychotherapist in a therapy session (Brandtzaeg & Følstad, 2017). Since then, numerous other CAs have been developed, such as *Parry* (Colby, 1975) or *ALICE* (Shawar & Atwell, 2007; Wallace, 2009), which mainly answered simple rule-based commands and questions, simulating conversations (Brandtzaeg & Følstad, 2017; Gnewuch et al., 2017). However, the speech comprehension of these systems was not particularly robust (Diederich et al., 2019a), which impaired interaction. They could not hold long conversations with their users or give more than simple and often rule-based responses to the input commands (Gnewuch et al., 2017). However, advances in technology in the field of artificial intelligence (AI), specifically in the context of NLP/NLG (Zierau et al., 2020c), are leading to a massive proliferation of these systems in numerous workplaces (Feng & Buxmann, 2020; Meyer von Wolff et al., 2020a) and support contexts (e.g., Herrera et al., 2019; Zierau et al., 2020b). In recent years, this massive technological progress has allowed the development of progressively more intelligent CAs, characterized as AI-based systems, which are (1) user-centric, (2) social and intuitive, (3) learning (Lewandowski et al., 2021; Zierau et al., 2020a), and as the ChatGPT and large language models (LLM) trend indicates, becoming more (4) aggregative and (5) generative, with the ability to access numerous sources of information in the background, and recognize and create new content for users and customers using machine learning (ML) technologies, which also enables interaction.

CAs in the scientific literature are often divided into two subclasses (in taxonomies described with the dimension “communication mode”) (Diederich et al., 2019b; Knote et al., 2019). The first class of text-based CAs is usually referred to as chatbots, chatterbots, talkbots, or dialogue systems (Gnewuch et al., 2017; Winkler & Söllner, 2018). The second class of speech-based CAs is commonly called virtual, digital, or intelligent assistants as well as digital companions (Gnewuch et al., 2017). In many academic publications, however, no distinction is made at all (Meyer von Wolff et al., 2019a). The distinction is also marginal from a technical perspective because speech-based input can easily be transferred to text-based input and vice versa (Diederich et al., 2019b). Further, combinations of text and speech-based forms of input and output approaches exist (Diederich et al., 2019a).

This paper defines CAs as text-based and AI-based representations, such as chatbots (see above) (Gnewuch et al., 2017; Io & Lee, 2017; Vaidyam et al., 2019; Winkler & Söllner, 2018). There are two main differences and novelties of this kind of information system (IS) (Knote et al., 2019). First, users interact with the system using natural languages like English, German, or Spanish. Second, they rely on "the assistant's knowledgeability and human-like behavior, often summarized as artificial intelligence" (Knote et al., 2019, p. 2025), which has great potential to assist, solve, augment, or automate user tasks intuitively (Diederich et al., 2019b; Lewandowski et al., 2022b; Semmann et al., 2018). A unique characteristic of CAs is their ability to learn and improve through naturalistic interactions (Lewandowski et al., 2022b). In this paper, the term CA includes all AI-based IS that communicate with users (employees or customers) through a text-based natural language interface enabled by NLP/NLU technologies such as CA frameworks (e.g., Microsoft LUIS, Google Dialogflow, SAP CAI, or RASA.ai), including an intelligent communication and built-in self-learning component. The actual prototype in this work was developed using SAP CAI.

Currently, CAs are especially adopted in interactive domains, such as customer service and support, marketing, sales and entertainment, teaching and education, and in different workplace applications (Diederich et al., 2019a, 2019b; Meyer von Wolff et al., 2019a). CAs have the potential to change the entire interaction and communication channel between customers and service employees (Følstad et al., 2018a; Gnewuch et al., 2017; Wilson, 2018; Xu et al., 2017). As a result, they can transform service provision and value (co)-creation, and thus, entire business models and service systems (Böhmann et al., 2014; Maglio et al., 2009) in the future (Dämon, 2017; Gnewuch et al., 2017; Tewes et al., 2020).

CAs facilitate a new form of flexibility, quality, speed, and personalization of the customer relationship (Lewandowski et al., 2021; Wilson, 2018). Moreover, they represent a highly scalable and cost-effective solution, saving money by replacing manually executed tasks performed by previously required service employees (Oracle, 2016; Wirtz et al., 2021; Wirtz et al., 2018). A transformation towards an innovative, convenient, automated, self-learning 24/7 customer service communication and interaction channel available to customers worldwide and multilingually is feasible and could take over initial routine tasks (Brandtzaeg & Følstad, 2017; Følstad et al., 2018a; Gnewuch et al., 2017). In addition, CAs offer shorter resolution times and high availability (Waizenegger et al., 2020).

Based on these findings, particularly the dissatisfaction of users with their interaction with CAs and consequent interaction problems, the need to rethink the design of and interaction with CAs is apparent. Accordingly, we present approaches utilizing user-centered mock-ups and prototypes, which are implemented and allow us to derive design guidelines.

13.3 Methodology

The goal of this article was to identify problems in interaction design. To achieve our research goal, we followed the ADR approach (Sein et al., 2011) by combining various sub-steps. Initially, we created user-centered CA solution alternatives (mock-ups) based on expert interviews to solve the identified interaction problems. Subsequently, we selected alternatives quantitatively through a survey, implemented them as prototypes, and finally evaluated them. An overview of the sub-steps of the methodology is shown in **Table 2**.

13.3.1 Action Design Research

As explained by Sein et al. (2011), ADR is a research approach for examining and co-creatively solving a specific constellation of problems within an organization. As the name suggests, the approach combines action research (AR, Susman & Evered, 1978) with design science research (DSR, Hevner et al., 2004) (Mullarkey & Hevner, 2019; Venable et al., 2017). This combination aims to create practice-relevant IT artifacts. The approach consists of four stages, starting with the problem formulation phase. The central concept within ADR is the “Building, Intervention, and Evaluation” (BIE) cycle (Sein et al., 2011). Generally, an ADR project starts in the first phase by identifying a research gap in practice or theory and formulating initial research questions (Sein et al., 2011; Venable et al., 2017), similar to DSR. The IT artifact idea resembles this BIE cycle, where the artifact is built, put into action within an organization (the AR part), and subsequently evaluated (Venable et al., 2017). Thereby, the artifact is improved and refined in a cycle-wise manner (Sein et al., 2011). A significant part of the ADR approach involves reflection and the subsequent formalized learning stage. The third and fourth stages lead to an improved IT artifact and aid in the development of design principles and the research outcome (Sein et al., 2011). These can take the form of new, initially researched designs or selective improvements of existing designs (Sein et al., 2011). Mullarkey and Hevner (2019) propose an “elaborated action design research process model” as a more flexible approach for immersive industry-based projects.

13.3.2 Data Collection, Analysis, and Procedure

Table 2. Procedure

Step	Action
1	Interaction problems defined Identified measures for improving the user experience
2	Developed monitoring guide Analyzed logs
3	Potential-effort-matrix defined Prioritizing
4	User-centric solution alternatives created
5	Expert interviews regarding user interaction problems
6	Developed survey for interaction problems and choosing the prototype implementation
7	Focus group setting for evaluating the prototype

As a first step, *interaction problems* were defined and the measures that would improve the user experience were explored. We did this at various levels, i.e., both in the basic design and in the direct interaction. We started to *identify* the CAs’ interaction problems so we could solve them and thus improve the user experience. Second, we developed a *monitoring guide* and analyzed the monitoring and chat logs of the current CA (Peras, 2018) and its 443 conversations, allowing us to identify the main problems in the current interaction design. Based on the identified problems, a *potential effort matrix* was derived (see **Table 1**). We focused on the four areas that offered high potential for improvement combined with comparatively low effort.

To address these four core problems, *user-centric solution alternatives* were created. To improve the user experience, six potential users were involved in the creation process, and an open *interview* was conducted to solicit initial approaches and ideas (Helfferich, 2019). These interviews gave us valuable insights on the expectations and requirements regarding the interaction problems of potential users. Next, we enhanced these findings (four interaction problems) with literature review results (vom Brocke et al., 2009).

As a sixth step, we developed a *survey* with questions about each core problem to elaborate on the identified interaction problems (Albers et al., 2009; Nardi, 2018). Eleven questions focused on the findability of the CA, and an additional 14 questions were related to the welcome message. Nine questions addressed issues regarding dialog control, and 11 questions examined the fallback issue. One hundred and twelve people participated in the survey, which gave us a representative view of the issues. With the help of a survey and design thinking techniques (Meinel & von Thienen, 2016), alternative solutions were discarded, while the most popular ones were implemented in a prototype. In the seventh step, we evaluated the prototype in a *focus group* setting (Krueger & Casey, 2002; Morgan, 1996). The final guideline was created based on the findings of the monitoring, the surveys, and the concrete implementation of improvements in the prototype with subsequent evaluation (Kushner, 1993).

13.3.3 Evaluation

For the evaluation, the prototype was first evaluated, followed by the developed guideline (Venable et al., 2016). The evaluation strategy ensured a structured and correct evaluation of the prototype. The goal was to compare the existing CA with the optimized version (i.e., the prototype). The focus was on comparing the change in the user experience and confirming the researched guideline aspects, which were established as explained in the previous subsection.

In order to compare the prototype with the existing CA, we used a between-subjects design (Charness et al., 2012), which is suitable for such situations. In this evaluation, the prototype and the existing CA were the independent variables, while the subject areas of the scenarios and questions explained below were the dependent variables. The System Usability Score (SUS) score can be used to compare the two variants, i.e., the existing CA and the prototype (Bangor et al., 2009). Different scenarios were developed so that the prototype could be evaluated in a stepwise manner. The scenarios were created based on the problems identified from the first two steps described above. The problems were selected based on the topics most frequently addressed in the monitoring process. We tackled the interaction problems in the prototype and evaluated its success. Four areas were addressed, and we compared the prototype with the existing CA. For this comparison, we created four scenarios (in connection with the four selected topics), which were to be executed by the test persons for the respective assigned CA. These four scenarios were related to the corresponding problem topic areas. In addition to these topic areas, which were optimized in the prototype, the results of the expert interviews, monitoring, and survey were also used in the implementation of the prototype. These addressed the user experience in general, the fallback scheme, reusability, findability, and the two-level category system.

The design guideline was also evaluated as a summarizing end product. The goal of the guideline was to create helpful knowledge for future projects. For this purpose, a focus group was formed as the first evaluation method to obtain expert feedback on the design guideline (Helfferich, 2019; Krueger & Casey, 2002). This was a formative evaluation, as it was part of the process of developing the design guideline (Scriven, 1972). Since the evaluation and discussion were carried out directly with those affected (i.e., a focus group) (Morgan, 1996), the procedure can be considered a "naturalistic evaluation" (Kushner, 1993; Venable et al., 2016). This step of the process aimed to gather suggestions for improving the design guideline. First, we presented the status of the design guideline to the focus group. Then, we discussed the particular content-related topics with the CA team.

Additionally, the structure was discussed. The second evaluation method was a summative "ex post evaluation" (Stockdale & Standing, 2006). A structured survey with predefined statements (Phellas et al., 2011) was administered to the project owners, who were asked to rate the quality of the guideline on a Likert scale with five values ranging from 1 (strongly disagree) to 5 (strongly agree) (Likert, 1932). Nine categories were defined and formulated in advance in a Google Forms questionnaire.

13.4 Results

We started by analyzing the current conversations already happening with the existing CA. It became clear that the CA failed to answer user queries in a satisfactory way. In particular, the CA failed to understand users and therefore asked them to rephrase their queries. Furthermore, the CA forwarded the user to other contact means (e.g., to a human service employee). As the misunderstanding of the CA and subsequent failure of the conversation were the most pressing issues, we analyzed all conversations with the CA itself. Of the 443 conversations, the CA failed to ask users to reformulate their queries 367 times. Additionally, 181 times the CA suggested categories or topics that the users could check themselves since the CA could not help. On 99 occasions, the CA did not understand the query and answered incorrectly. Further, the CA redirected the user to use other customer service methods 60 times.

In order to differentiate better to improve the user experience with the CA, we isolated the different contexts of the CA. We distinguished between individual queries (249 instances), company-related queries (80 instances), and general queries (85 instances). We regarded 38 further instances as non-relevant, as the users in those cases obviously entered misleading information to mess with the CA. The logs also showed the CA erred in communication by asking multiple queries in one (four instances) and connecting the greeting with a question (43 instances). Although the CA had a high percentage of dissatisfactory responses, users were tolerant regarding a one-time CA failure. In 96 instances, the CA had to fallback and rephrase one time, but there were only 24 instances with four or more fallbacks. Overall, the CA had 548 single fallbacks.

We focused on the fallback issues and classified them to their contexts. The greatest number of fallbacks were related to individual topics users asked about (255 instances). Another 105 cases resulted in fallbacks regarding general topics but there were only three instances of company-related topics. Forty-two questions were not classifiable. The main reason behind the fallbacks in 183 instances was that information on the queried topic existed in the CA but was not kept up-to-date. In 77 instances, information on the topic did not exist. Of 273 user reactions to a fallback, the conversation was ended in 74 cases, while in 187 cases the query was reformulated. Only 42 cases were solved after the first fallback, and an additional 145 instances led to another fallback. In total, 97 users left the conversation after the first fallback. Further, 74 left after the second fallback, an additional 30 quit after the third fallback, and 11 left after at least four fallbacks. The data showed that most users quit after the first or second fallback, and thus the design of a fallback handling strategy would be of paramount importance.

To offer help to the user in case of a fallback, topic/category suggestions were offered by the CA 125 times in the hope that they might cover the user's question and solve the problem. However, users took advantage of these suggestions in only 73 cases, of which no solution was found and the problem persisted in 57 cases.

Table 3. Interaction problems

Interaction problems
Findability
Welcome Message
Dialog control
Fallback issue

Once the core interaction problems of the CA were defined (see **Table 3**), the problem understanding and selection phase was considered complete. Then, the solution phase began. In order to find suitable and satisfactory solutions for the problems, alternative solutions had to be designed, from which suitable ones were then selected. Since the context was the design of the interaction and a system, it made sense to involve possible users in the process of generating alternatives as well as in the selection process. The first step in opening up the solution space was to conduct interviews with six potential users in order to gather initial suggestions for solutions and to open up the solution space. The interviewees were asked openly formulated questions about the central problems.

Regarding findability, the test subjects offered various suggestions. Most preferred an always visible icon on the side of a website. Out of habit, the bottom right was suggested. Placement in a navigation bar was also mentioned occasionally. However, the respondents were relatively divided on the design of the icon; in addition to a speech bubble as an icon, a CA icon was also mentioned several times. There was also disagreement about effects and animations. Some participants were negatively disposed towards them, while others expressed a contrary opinion, for example, related to effects that appear when the website is first loaded (so the CA draws attention to itself). In terms of color, the icon should be adapted to the corporate identity but still stand out somewhat from the rest of the CA and its environment.

The default message at the beginning of a conversation should explain that the chat is a conversation with a CA. After a greeting, a short but friendly and helpful message was expected. In some cases, subjects also mentioned topics or sample questions that the system could answer. A privacy notice in the form of a link was desired in order to not to make the message too long. Concrete language features or humor should be chosen depending on the company context. The respondents preferred a casual approach, but not a humorous one. The message should build trust and express helpfulness, as this is an important basis for the further course of the chat.

Most test subjects preferred free input of their requests in the form of free text. However, there was isolated criticism that many CAs cannot handle this type of input well, which can cause the system to take over and direct the conversation or suggest buttons. Especially when the CA needed feedback to answer the question, some of the control went to the bot. A mix of free text input and buttons thus seemed to make sense to many test subjects. According to the test subjects, buttons were particularly suitable for questions with few but clearly defined answer options (e.g., yes/no questions).

If the CA did not understand a question and made a fallback, there were different opinions from the respondents. Some wanted to be forwarded directly to a real service employee, while others were prepared to rephrase the question or expected similar suggestions from the CA. In such cases, the respondents indicated that they wanted to reformulate once or twice at most. If this still did not lead to a solution, a fallback was inevitable. Direct chat transfer or a contact form via e-mail were preferred as channels, and in some cases, a telephone call or direct callback were also mentioned. Without a fallback, direct forwarding should not be an option. For most of the respondents, the variability of the messages was not very important. Accordingly, slight variations in wording were sufficient, for example, in the fallback context. In general, however, this was a quality feature of a good implementation.

We then conducted a survey. Of the 112 survey participants, only 69 completed questionnaires were used for data analysis. Of these 69 participants, 32 were female, 35 were male, and one was diverse, with one of the participants also not providing gender information. On average, the participants were between 25 and 39 years of age (35 participants). Of the 69 participants, 15 were between 18 and 24 years of age, while 19 participants were between 40 and 59 years of age. None were older than 59. It should also be noted that 65 participants had interacted with a CA before, while only four participants had never interacted with a CA. Regarding the variability in CA responses, 13.04% of the 69 participants stated it is "very important," 39.13% said it was "important," 28.99% were "neutral," 13.04% indicated it was "not important," and 5.80% said it was "not important at all."

Based on the survey results on findability, a disagreement emerged regarding the pages of a website on which a CA should be found. About 45% of the respondents wanted access on every page, while about 41% indicated the contact or customer service page was sufficient. According to the respondents, the reference to the CA should be found on the right edge of the website, 46% preferred the middle, and 39% favored the bottom right side of the website. Regarding the size of the CA icon, a medium size was the most popular (about 59%). The survey results showed that in terms of scrolling behavior, the option of a fixed, ever-visible CA icon was desired by 84% of the respondents. With regard to the color design of a notice, 64% preferred a design matching the color of the website, which nevertheless stands out clearly so that the CA can be found quickly. Approximately 35% of the respondents preferred a design that matches the website and does not stand out strongly. Regarding the shape of the symbol, a square was preferred by 72% of the respondents. However, it was also noted in a comment that the CA design should orient to the design of the website and the corporate design of the company. Similar to the color scheme, the focus should always be on a design that matches the rest of the environment. The graphic symbol of a speech bubble was also popular (72%). Regarding the textual indication of the symbol, the opinions were mixed. About 30% of the respondents stated that no textual reference was necessary, 26% favored “chatbot” as a reference, and about 22% preferred “Ask your question here.” The respondents also noted that a textual hint should be as short as possible. Furthermore, 67% of the respondents did not want effects to attract attention. The effect of flying in the CA icon had the most supporters of all the effects shown (12%). Other comments mainly indicated that a CA icon should never overlay content on a website. Intrusive designs were not desired, as they look like advertising. One suggestion was to resize the design depending on which page it is on. Another suggestion was that the icon in the customer service area could be more prominent than on the homepage.

The survey results regarding the default message at the beginning of the conversation showed that the message that is currently used is best received by potential users: “Hello, I am Roberta, your digital assistant. Please ask me your question.” Above all, it was emphasized that the brevity of the message is crucial. Based on the survey results, however, a trend was discernible indicating that an optimal default message has not yet been determined. The ratings were mostly in the “like” range (63.77%), while the rest of the responses were “like very much” (21.74%) and “mediocre” (13.04%). If the default message had been optimal, a significant shift to “I like very much” would have been seen in the ratings. It could be concluded that the default message used currently (short, personal, (optional) emojis; welcome formula with introduction of the CA and offer of help) promotes the user experience adequately. With regard to the presentation of the General Data Protection Regulation (GDPR) information, as required within the EU, the button with text content variant was preferred. A total of 84.06% of the respondents found this variant to be at least “average” and predominantly responded “I like” and “I like very much.” The previous form of a text block was also rated as “good” by some of the respondents, although the proportion of “bad” ratings was 26.09%. The respondents thought a plain button above the default message seemed unobtrusive in contrast to a larger text block. The information on the GDPR could still be easily accessed. Comments like "one likes to overlook" illustrated the inconspicuousness.

Regarding the dialog design, the majority (57.97%) of the respondents preferred free input. However, some noted that while free input is desirable, it usually does not work well in current CAs, which is why guided CA input is then preferred. Buttons (11.59% of respondents) or quick replies (13.04% of respondents) were almost equally important to the users, which explains why the term button is used synonymously with quick replies in the following. In general, a mix of free text and buttons was preferred (31.88% of respondents for a mix of free text and button or free text and quick replies, i.e., a total of 63.76%), where the use of buttons is independent of the content. The respondents did not differ much in whether the clicked button or "Quick Reply" should be displayed in the user dialog as an answer (34.78% of respondents) or its label or title (37.68% of respondents). Accordingly, a further 17.39% of the respondents indicated that a mix could be used flexibly, and 10.14% of the respondents selected the option "no matter." Buttons should be used after the results, for topic suggestions (63.77% of the asked ones), fallbacks (52.17% of the asked ones), forwardings (68,12% of the asked ones), and yes–no answers (73.91% of the asked ones). In addition, unrestricted free text input was preferred by most of the respondents (82.61%). Text entry was considered the primary input standard (75.36%). Further, intelligent sentence completion was viewed as reasonable as a future feature of CA (13.04% of respondents). Sentence completion automatically displays canned questions for the user to select when entering individual words. As a result, customer satisfaction and comprehension should be improved by correct answers from the CA.

The survey was able to determine a new fallback scheme. Based on the responses, the following formulation was rated as the best: "Sorry, I don't have an answer to that yet. I will be happy to give you the opportunity to speak to a human employee. Your request will be dealt with as soon as possible." A total of 42.03% of the respondents said "I like it a lot," while an additional 40.58% stated "I like it." The subjects generally preferred a transparent response, with the CA disclosing non-knowledge. Compared to rephrasing, a redirect was preferred. In addition, the importance of making an effort to find a solution was conveyed at the end of the message. Based on the survey, the structure of the fallback scheme should be as follows: Make CA limits clear + inspire confidence in the CA + do not blame the user. For fallback 1: Clarify misunderstanding + suggest rewording + give option to forward. In case of fallback 2: Clarify misunderstanding again using a different wording + suggest forwarding. In the formulation of the first fallback, a reformulation option is suggested, while this is omitted in the second fallback. The wording of the second fallback consists of the respondent's request for forwarding, which is reflected in the evaluation of the variant: "Unfortunately, I don't have an answer to this either. My colleagues will take care of it and find a solution. How should the contact be made?" Of the respondents, 71.02% classified this option as "I like" and "I like very much." Compared to other formulations that stood for a second fallback, it was clear that the forwarding option was more appealing. The formulation was intended to produce an optimal scenario for a fallback process. The results of the survey clearly showed that forwarding is desired after only one fallback. This would make it impossible to achieve the goal of relieving the burden on customer service. If a reformulation was suggested to the respondent after one or even two fallbacks, it was rejected in most cases, especially after the second fallback. It can be concluded that the patience of the users is rather low and that a solution should be offered as soon as possible. Meanwhile, category suggestions were a good option. However, categories that are inappropriate and do not offer solutions but rather serve typical FAQs were problematic. Individual concerns characterized the conversations. Forwarding, often chosen in the scenarios, also received a clear vote in the survey regarding the two forms of implementation. While call, QR code, and e-mail were not desired, the respondents felt that a callback option and, above all, chat takeover should be used. Indeed, chat takeover received ratings of "like it a lot" (42.03%) and "like it" (43.48%). Callback received the second-highest rating, with 30.43% of the respondents saying they "like it a lot" and 40.58% saying they "like it." The sales of the other three options were somewhat unclear clear, so they were excluded. Only 5% of the respondents indicated "I like" or "I like very much" in response to the e-mail option.

Table 4. Interaction problems and dialog patterns

Interaction problems	Solution guideline
Findability	Fixed, visible CA icon
Welcome Message	Short, personal, (optional) emojis; welcome formula with introduction of the CA and offer of help. GDPR information as button
Dialog Control	Free text input + buttons
Fallback Issue	Make CA limits clear + inspire confidence in the CA + do not blame the user.

	For fallback 1: Clarify misunderstanding + suggest rewording + give option to forward. In case of fallback 2: Clarify misunderstanding again using a different wording + suggest forwarding

Following the evaluation of the survey, the prototype was created. The CA symbol was adapted to make it easier to find, so the image of a speech bubble was included on the website without a textual reference. Some of the aspects collected could only be integrated locally for the evaluation, such as resizing the icon and moving it from the bottom right of the web page edge to a middle position on the right edge of the page. According to the survey, only a small change was required in the welcome message (see **Table 4**). The final representation of the GDPR notice (see **Figure 1**) was removed due to the previous form of a text block.

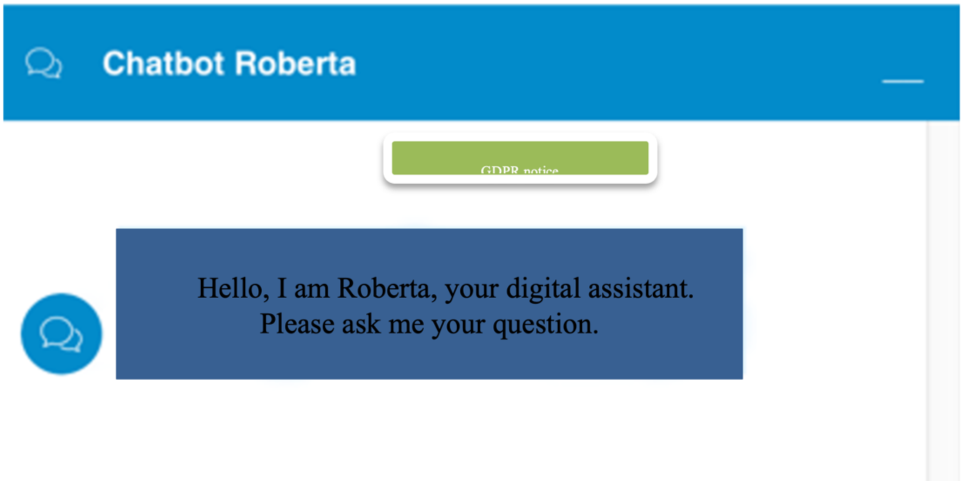

Figure 1. Welcome message with GDPR (translated to English, design slightly modified)

The topic of dialog design turned out to be appropriate when comparing the results, as many of the results were already used in the CA. The results of the monitoring showed that the topic suggestions were not relevant to the goal. However, this relevance was mainly related to the fallback area where a solution was developed. Intelligent sentence completion was another potential change that emerged from the results. The new fallback scheme was based on the following aspects. First, the wording that the CA sends during a fallback was taken from the survey. Further, the preparation for prototyping revealed a discrepancy between the desire for forwarding after only one fallback and the goal of user relief. This would not have been feasible by forwarding early. Due to this, the fallback was formulated so that a reformulation occurs first but at the same time there is the option to choose one of the two forwardings. In the case of a second fallback, the reference to a reformulation option is omitted. The forwarding options remain. The new prototype distinguishes the causes of fallbacks. The previously mentioned fallbacks occurred because the CA could not assign a single intent.

In order to exclude missing expressions as a cause, keyword recognition was implemented in the prototype. The feature was declared as a two-level category system. While it is technically supposed to run via the fallback path, visually it looks to the user like query recognition. The CA can recognize specific keywords. If the exact request cannot be assigned to an intent, suitable super-categories are suggested with the help of the recognition. After selecting these categories, questions are displayed in a second stage, which the user can choose. In the case of a wording that matches the query, the user has reached their goal despite the fallback. Unsuitable questions are followed by the option of a redirect. The resulting benefit can be primarily related to improving the UX. The fallback with no prospect of resolution is bypassed, and a second chance is given to process the query. This also prevents premature forwarding and partially relieves customer service. The rate of solved requests cannot yet reach 100% due to the individuality of the requests, which makes forwarding to an employee indispensable for some users. Following prototyping, the prototypes were evaluated. Ten people were interviewed for each variant. The results were positive for all interviewees.

Our results show a significantly improved user experience and a reduced error occurrence without finding a solution. Design guidelines were derived that enable replication of the results. The restructured fallback scheme avoids permanent requests for reformulation and inappropriate category suggestions. A two-level category system intercepts fallbacks and helps to find the user a solution through a topic-specific category set by keyword recognition. Based on the subjects surveyed (>9), the prototype increased the SUS score (interpretation by Bangor et al. (2009)) of the CA from 67 to 85.25. The first score is acceptable and represents an "OK" usability. The higher score of 85.25 indicates excellent usability and a significant improvement.

The shift in fallback to a new scheme that separates fallbacks according to their state of entry causes an alteration in thinking, where fallbacks are no longer primarily to be monitored. Rather, fallbacks do not lead to a solution. This commonly occurring problem (Brendel et al., 2020b) could be solved here. The callback option and chat pickup, identified as the most popular redirects, have been widely implemented and meet user expectations (see **Figure 2**). General satisfaction could be raised and the CA prototype use was positively evaluated. The core problems arise from frequent errors and a lack of solutions for the user. This has a direct impact on the user experience and is improved by using the prototype's implemented features. A fallback scheme that does not allow permanent rewording but offers specific suggestions minimizes errors and improves the UX. Further, the existing possibility of a redirection ensures recurring use of the CA without a loss of trust. Even so, CAs require constant interaction with the user and thus continuous monitoring and training. The design guideline was also evaluated positively.

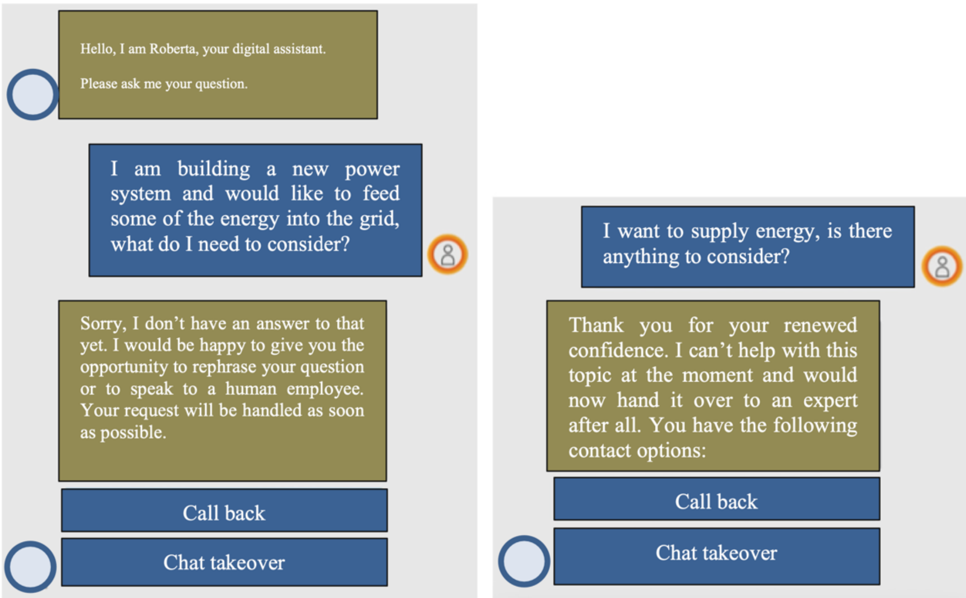

Figure 2. Twice restructured Fallback-scheme (translated to English)

13.5 Discussion

Despite the evaluation of the prototype and the design guideline, dissemination of the survey during the COVID-19 pandemic proved problematic but was possible digitally. Evaluation of the prototype was also possible. On-site interviews were extremely helpful for the "Thinking Aloud" method as part of the prototype evaluation. However, it was not possible to ensure that all subjects had the same conditions, so a laboratory experiment was not possible. In addition, the design guideline should have been tested more extensively in practice, which was not done here. The detailed collection of demographic data, both for the survey and the prototype, was not possible in the organization. In this regard, permission for a more detailed collection of demographic data would have been desirable. Other aspects that should be considered critically are the formulation of the prototype as well as a design guideline evaluation. During the implementation of both processes, there were queries regarding questions. It may be assumed that questions were interpreted differently, at least in part, which could have influenced the results. There were other organizational challenges in conducting the interviews in general due to the pandemic. Facilitation as well as conducting interviews via video conferencing systems, such as Microsoft Teams, could prove difficult and complicated. However, no such problems arose during the interviews.

Chronologically to the procedure, first the extent of the solution ideas is to be mentioned. The interviews with six test subjects, each with their own ideas as well as implementations from the literature, do not reflect all possible solutions. Thus, ideas that were not discussed could have resulted in a better UX and more successfully achieved the research objective. Regarding the conduct of the survey, demographic data were included but were not considered further in the comparison. Hence, no evaluation of different age groups or genders was made to determine differences in this regard. This approach was sufficient to obtain a general impression of the proposed solutions but could nevertheless lead to differential results. Studies involving diverse industries, different target groups, and a more detailed analysis would be useful. Moreover, no deep analysis of the demographic data was performed in the prototype evaluation.

Of the 69 survey participants, 65 had interacted with a CA before. It would be interesting to survey more people who had not interacted with a CA, which would lead to a better result on user experience. It is possible that this group would have new ideas about CAs, as they would not have a preconceived image. Furthermore, there might be reasons why this group of people has not yet used a CA. Such information is essential for improving the user experience of CAs. A balanced mix between these two groups (users and non-users) should be ensured in future evaluations so that similarities and differences can be identified, which could then be used to optimize the CA.

In addition, the prototype evaluation does not show perfect statistical significance (requiring 30 test subjects). Thus, the evaluation of the prototype is not sufficient to make a final assessment of the changes. Rather, the guideline aspects must be implemented in a productive environment and monitored over several weeks and then re-evaluated. Above all, the occurrence of fallbacks must be tracked. However, the keyword recognition with the new fallback scheme changes the perspective on fallbacks themselves. A fallback should no longer be regarded as an error. Resolution through the new categories could also lead to a solution, even if the CA technically executes a fallback. The importance of the unresolved request is now the new core aspect to monitor. The UX is significantly dependent on this, as every unresolved request leaves the user dissatisfied.

Overall, the results of this work including the wide range of surveys and user interviews suggest that there has been little research (e.g., Bouguelia et al., 2021) on so-called Dialog Patterns and therefore, they offer additional research value. By highlighting interaction problems, new patterns can be identified for the four core problems, which we refer to as *Dialog Patterns*. Little systematic research has been done in this area so far, so this work contributes to it despite its limitations.

13.6 Conclusion

In summary, optimizing the prioritized problem aspects of the investigated CA show a significant improvement in error occurrence without solution finding and in the user experience. The restructured fallback scheme avoids permanent requests for reformulation and inappropriate category suggestions are avoided. A two-tier category system intercepts fallbacks and gives the user another chance to find a solution via a topic-specific category set with the help of keyword recognition. The change in fallback to a new scheme that separates fallbacks according to their entry status is causing a shift in thinking that fallbacks are no longer primarily to be monitored. Instead, it is fallbacks that lead to no resolution. With the help of the research, the research question: “How can a user's CA experience be improved by analyzing, understanding, and optimizing interaction problems?” can be answered. This paper shows several approaches by means of user-centered mock-ups and prototypes to improve the pressing issue of user satisfaction with the CA interaction, to solve core interaction problems, and shows the need for further research on dialog patterns.

13.7 References

Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J. (2009). Methodik der empirischen Forschung (Vol. 3). Springer, Wiesbaden, Germany.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of usability studies, 4(3), 114-123.

Böhmann, T., Leimeister, J. M., & Möslein, K. (2014). Service Systems Engineering. Business & Information Systems Engineering, 6(2), 73-79.

Bouguelia, S., Brabra, H., Zamanirad, S., Benatallah, B., Baez, M., & Kheddouci, H. (2021). Reusable abstractions and patterns for recognising compositional conversational flows. Advanced Information Systems Engineering: 33rd International Conference, CAiSE 2021, LNCS, vol. 12751, pp. 161–176. Springer, Cham. Melbourne, VIC, Australia, June 28–July 2, 2021, Proceedings.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. International Conference on Internet Science (INSCI), Cham.

Brendel, A. B., Kolbe, L. M., & Diederich, S. (2020). Designing Anthropomorphic Enterprise Conversational Agents. Business & Information Systems Engineering, 62(3), 193-209.

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. The Service Industries Journal, 40(13-14), pp. 900-925.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. Journal of economic behavior & organization, 81(1), 1-8.

Colby, K. M. (1975). Artificial paranoia: A computer simulation of paranoid processes. Pergamon Press, Oxford.

Corea, C., Delfmann, P., & Nagel, S. (2020). Towards intelligent chatbots for customer care-practice-based requirements for a research agenda. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Dämon, K. (2017, 10.07.2017). MEGATREND CHATBOT - Was Chatbots der Wirtschaft bringen. WirtschaftsWoche. Retrieved 06.02.2023 from https://www.wiwo.de/erfolg/trends/megatrend-chatbot-was-chatbots-der-wirtschaft-bringen/19997318.html

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019a). On conversational agents in information systems research: Analyzing the past to guide future work. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019b). Towards a taxonomy of platforms for conversational agent design. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. International Conference on Internet Science (INSCI), Springer, Cham. https://doi.org/10.1007/978-3-030-01437-7_16

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. International Conference on Information Systems (ICIS), Seoul, Korea.

Helfferich, C. (2019). Leitfaden- und Experteninterviews. In Handbuch Methoden der empirischen Sozialforschung (pp. 669-686). Springer Fachmedien.

Herrera, A., Yaguachi, L., & Piedra, N. (2019). Building Conversational Interface for Customer Support Applied to Open Campus an Open Online Course Provider. IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Maceió, Brazil.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in Information Systems research. Management Information Systems Quarterly (MISQ), 28(1), 75-105.

Io, H. N., & Lee, C. B. (2017). Chatbots and Conversational Agents: A Bibliometric Analysis. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore.

Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do Chatbots fail? A Critical Success Factors Analysis. International Conference on Information Systems, Austin, TX, USA.

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying smart personal assistants: an empirical cluster analysis. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Krueger, R., & Casey, M. A. (2002). Focus group interviewing on the telephone. State Health Access Data Assistance Center (SHADAC), Minneapolis.

Kushner, S. (1993). Naturalistic evaluation: Practical knowledge for policy development. Research Evaluation, 3(2), 83-94.

Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-art analysis of adopting AI-based conversational agents in organizations: A systematic literature review. Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022). Design knowledge for the lifecycle management of conversational agents. International Conference on Wirtschaftsinformatik (WI), Nürnberg, Germany.

Likert, R. (1932). A Technique for the Measurement of Attitudes. Archives of psychology, 140, 44-60.

Maglio, P. P., Vargo, S. L., Caswell, N., & Spohrer, J. (2009). The service system is the basic abstraction of service science. Information Systems and e-business Management, 7(4), 395-406.

Meinel, C., & von Thienen, J. (2016). Design Thinking. Informatik-Spektrum, 39, 310-314. https://doi.org/https://doi.org/10.1007/s00287-016-0977-2

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), 3. https://doi.org/10.17705/1pais.12203

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2019). How May I Help You?–State of the Art and Open Research Questions for Chatbots at the Digital Workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Meyer Von Wolff, R., Hobert, S., & Schumann, M. (2022). Chatbot Introduction and Operation in Enterprises–A Design Science Research-based Structured Procedure Model for Chatbot Projects. Hawaii International Conference on System Sciences (HICSS), Hawaii, HI, USA.

Morgan, D. L. (1996). Focus Groups. Annual review of sociology, 22(1), 129-152.

Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. European Journal of Information Systems, 28(1), 6-20.

Nardi, P. M. (2018). Doing Survey Research: A Guide to Quantitative Methods (4 ed.). Routledge, New York, NY, USA. https://doi.org/https://doi.org/10.4324/9781315172231

Oracle. (2016). Can Virtual Experiences Replace Reality? Oracle. Retrieved 02.03.2020 from https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc35/CXResearchVirtualExperiences.pdf

Peras, D. (2018, December). Chatbot evaluation metrics. Economic and Social Development: Book of Proceedings - Intrnational Scientific Conference on Economic and Social Development, Zagreb, Croatia.

Phellas, C. N., Bloch, A., & Seale, C. (2011). Structured methods: Interviews, questionnaires and observation. Researching society and culture, 3(1), 23-32.

Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Ed.), Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen (pp. 60-91). R. Piper & Co. Verlag.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. MIS quarterly, 37-56.

Semmann, M., Grotherr, C., Vogel, P., Bittner, E., Biemann, C., & Böhmann, T. (2018). Intelligent Collaboration of Humans and Language-Based Assistants (INSTANT). International Conference on Information Systems (ICIS), San Francisco, USA.

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they Really Useful? J.Lang.Technol.Comput.Linguist. (Former LDV-Forum) 29–49 (2007).

Stockdale, R., & Standing, C. (2006). An interpretive approach to evaluating information systems: A content, context, process framework. European journal of operational research, 173(3), 1090-1102.

Susman, G. I., & Evered, R. D. (1978). An Assessment of the Scientific Merits of Action Research. Administrative science quarterly, 23(4), 582-603.

Tewes, S., Niestroj, B., & Tewes, C. (2020). Geschäftsmodelle in die Zukunft denken. Springer Gabler, Wiesbaden, Germany.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. The Canadian Journal of Psychiatry, 64(7), 456-464.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. European Journal of Information Systems, 25(1), 77-89. https://doi.org/10.1057/ejis.2014.36

Venable, J. R., Pries-Heje, J., & Baskerville, R. L. (2017). Choosing a Design Science Research Methodology. Australasian Conference on Information Systems, Hobart, Australia.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. European Conference on Information Systems (ECIS), Verona, Italy.

Waizenegger, L., Seeber, I., Dawson, G., & Desouza, K. (2020). Conversational agents-exploring generative mechanisms and second-hand effects of actualized technology affordances. Proceedings of the 53rd Hawaii International Conference on System Sciences.

Wallace, R. S. (2009). The anatomy of ALICE. In: Epstein,R., Roberts,G., Beber,G. Parsing the Turing Test, pp. 181–210. Springer, Dordrecht (2009).

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

Wilson, J. H. D., Paul R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. Havard Business Review. Retrieved 18.07.2020 from https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. Academy of Management Annual Meeting (AOM), Chicago, USA.

Wirtz, J., Kunz, W., & Paluch, S. (2021). The Service Revolution, Intelligent Automation and Service Robots. European Business Review.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. Journal of Service Management, 29(5), 907-931. https://doi.org/10.1108/josm-04-2018-0119

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. Conference on Human Factors in Computing Systems (CHI), New York, NY, United States. https://doi.org/10.1145/3025453.3025496

Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020a). A review of the empirical literature on conversational agents and future research directions. International Conference on Information Systems (ICIS), India.

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020b). Towards developing trust-supporting design features for AI-based chatbots in customer service. International Conference on Information Systems (ICIS), India.

Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., & Leimeister, J. M. (2020c). The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. International Conference on Information Systems (ICIS), India.

14 Publication No. 6: Lewandowski et al. (2023b)

Lewandowski, T., Poser, M., Kučević, E., Heuer, M., Hellmich, J., Raykhlin, M., Blum S. & Böhmann, T. (2023). Leveraging the Potential of Conversational Agents: Quality Criteria for the Continuous Evaluation and Improvement.

Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS), Hawaii (USA).

Abstract

Contemporary organizations are increasingly adopting conversational agents (CAs) as intelligent and natural language-based solutions for providing services and information. CAs promote new forms of personalization, speed, cost-effectiveness, and automation. However, despite their hype in research and practice, organizations fail to sustain CAs in operations. They struggle to leverage CAs' potential because they lack knowledge on how to evaluate and improve the quality of CAs throughout their lifecycle. We built on this research gap by conducting a design science research (DSR) project, aggregating insights from the literature and practice to derive a validated set of quality criteria for CAs. Our study contributes to CA research and guides practitioners by providing a blueprint to structure the evaluation of CAs to discover areas for systematic improvement.

Keywords: Artificial intelligence assistants · Conversational agents · Chatbots · Quality criteria set · Design science research (DSR)

14.1 Introduction

Due to ongoing developments in artificial intelligence (AI) and improvements in underlying machine learning (ML) algorithms, CAs are becoming increasingly relevant in organizations as essential gateways to digital services and information (Følstad et al., 2021; Gnewuch et al., 2018). Primarily operating in external or internal organizational environments, CAs can conveniently provide users (e.g., customers and employees) access to information from several connected systems and data sources. In addition, CAs are able to execute standardizable processes and tasks that have conventionally been performed by employees (Meyer von Wolff et al., 2020a). Equipped with these capabilities, organizations can deploy CAs in various work contexts to efficiently and cost-effectively automate routine tasks or assist users in performing tasks (Meyer von Wolff et al., 2020a). Due to their massive economic potential and capability to deliver personalized services, much research has been conducted on these AI-based systems (Cui et al., 2017; Zierau et al., 2020b). More specifically, previous research has focused on aspects that are technical (e.g., technology selection and NLP improvements), behavioral (e.g., user trust), and conceptual or design-oriented (Diederich et al., 2019a; Meyer von Wolff et al., 2021; Zierau et al., 2020a).

Despite its promising potential, the adoption of CAs in organizational environments does not always have a positive impact because the technology is still error-prone and fails in interactions (Gnewuch et al., 2017; Janssen et al., 2021b). Therefore, recent research has adopted a management perspective to identify the reasons for the moderate success of CAs. In this vein, factors for success and failure, as well as a continuous evaluation (e.g., monitoring) and improvement process, have been proposed to ensure the successful operation of CAs (Janssen et al., 2021b; Meyer von Wolff et al., 2021). Thus far, however, there is a lack of knowledge on how CAs can be evaluated with criteria to test and improve their quality throughout their lifecycle (Lewandowski et al., 2022b). Therefore, this paper explores the following research question:

RQ: *What are relevant criteria for continuously evaluating the quality of CAs, and how can they be applied?*

In this paper, a set of relevant criteria was developed to evaluate the quality of CAs, and a procedure model to apply the criteria was derived and evaluated. Since, in practice, many CAs fail due to a lack of knowledge concerning evaluation, a criteria-based approach can close this gap in CA research on lifecycle topics and support the CAs' operation phase (Lewandowski et al., 2022b). From a practical lens, the quality criteria and procedure model can serve as an initial overview for organizations to systematically structure CA evaluation to discover areas for improvement. Following DSR activities, we present insights from the literature and practice to derive a validated set of quality criteria for CAs. Hence, the remainder of our paper is structured as follows. Section 2 presents the related CA research and delineates the research gap. In Section 3, we describe the DSR approach to developing our artifact. In Section 4, we present the findings of our study, including an overview of our final quality criteria set. Subsequently, Section 5 outlines the instantiation of the quality criteria set using a real-life case in an IT organization. We discuss our findings and conclude with our limitations and contributions in Section 6.

14.2 Related Research

The vision of communicating with information systems (IS) has been around for nearly 50 years. An early example is ELIZA, which allowed initial natural language-based interactions with a computer (Weizenbaum, 1966). However, technical limitations restricted early attempts at CAs (Diederich et al., 2019a). Nevertheless, in recent decades, massive technological progress has allowed the development of progressively more intelligent CAs. Consequently, CAs, known under numerous designations, such as chatbots, chatterbots, or dialog systems, have gained interest, leading to discussions in the literature about a delimitation of the terms. We use the term "conversational agent" in this paper to refer to all AI- and text-based representations, such as chatbots (cf. Gnewuch et al., 2017), since the CAs investigated in the real-life DSR project were text-based.

Today, CAs are increasingly adopted and have attained popularity in various commercial and private application domains (Meyer von Wolff et al., 2020a). Integrated into various front and back-end systems, such as websites or messaging applications (e.g., MS Teams), CAs support organizations' ongoing digitization and automation by e.g., filtering information or efficiently assisting employees in daily work tasks (Zierau et al., 2020a). Hence, with their scalability and 24/7 availability (Gnewuch et al., 2017; Xu et al., 2017), CAs can make a transformative contribution by providing a convenient way for more individual interactions, such as acting as a central service platform and first point of contact for customers before they reach out to actual employees (Zierau et al., 2020a). Thereby, users' high load of information is reduced (Xu et al., 2017). Moreover, employees can concentrate on their core and non-routine tasks.

Nevertheless, developed CAs still have a high failure rate (Janssen et al., 2021b). Many fail in real-world environments due to, among other things, frustrating user experiences (Følstad et al., 2018a).

As a result, multiple organizations take their CAs offline because they lack knowledge of quality criteria and aspects relevant to continuous evaluation and improvement, resulting in an uncoordinated and highly explorative development process (Janssen et al., 2021b). Moreover, CAs represent a novel form of learning, unfinished, user-centric, and socially interactive IS that has introduced, so far, unsolved challenges (Lewandowski et al., 2021; Zierau et al., 2020a). A distinctive feature of CAs is their capability to learn and improve via naturalistic interactions. Accordingly, CAs' learning progress is highly context-driven and thus dependent on actual application and usage (Clark et al., 2019; Zierau et al., 2020c). Because of this unfinished and learning nature of CAs, novel approaches to handle their implementation and improvement in their lifecycle are required since they initially possess limited functions and require several interdisciplinary design activities (Lewandowski et al., 2022b; Meyer von Wolff et al., 2021).

Consequently, the highest effort occurs in operations, where CAs require continuous evaluation and later training and improvement in a real-world context, often characterized by rapid changes and high dynamics in which it is generally impossible to predict how users will interact and what information will be retrieved long-term (Janssen et al., 2021b). Although CAs have gained a great deal of research attention from specific conceptual, usability, or technical design perspectives, the operation in general, and continuous improvement process, specifically, lack of detailed theoretical and practice-based knowledge is an issue. (Lewandowski et al., 2022b; Meyer von Wolff et al., 2021). Hence, a clear criteria-based approach to continuously evaluate CAs' quality in further development is needed to sustain them. First researchers have already investigated success and failure factors for CA implementations from an organizational perspective (e.g., Janssen et al., 2021b; Lewandowski et al., 2021; Meyer von Wolff et al., 2021). However, they tend to address the managerial perspective and do not focus on the continuous improvement process. Other authors have studied the different effects of CAs on an individual level, either on perceived trust, enjoyment, or affordance theory (Stoeckli et al., 2019; Zierau et al., 2020b) or in the wider context of IS acceptance theories, such as in the "Technology Adoption Model" (e.g., Pillai & Sivathanu, 2020). However, there is little research on concrete quality criteria that can be applied to ensure systematic CA improvement. Initial contributions exist in evaluating CA design. Nevertheless, current research is (1) confined to technical measurements (e.g., Alonso et al., 2009), (2) other agent classes (e.g., Kuligowska, 2015), and (3) individual design aspects (e.g., Seeger et al., 2021), while (4) being segregated. Further, research (5) focused on human behavior or ethical aspects (e.g., Radziwill & Benton, 2017) and (6) initial classifications and typologies for only a high-level analysis and guidance on interaction design (Følstad et al., 2018b), which for CA teams only play a superordinate role in CA development. Thus far, a holistic overview of criteria for researchers and practitioners for constant evaluation and sustainability throughout the CA lifecycle is lacking.

14.3 Research Approach

This article aims to provide CA quality criteria that will allow organizations to continuously evaluate and improve CAs during their lifecycle, as proposed by Lewandowski et al. (2022b). To achieve this goal, we adopted the DSR paradigm and applied the three-cycle view by Hevner (2007). Overall, we conducted seven research activities (see **Figure 1**).

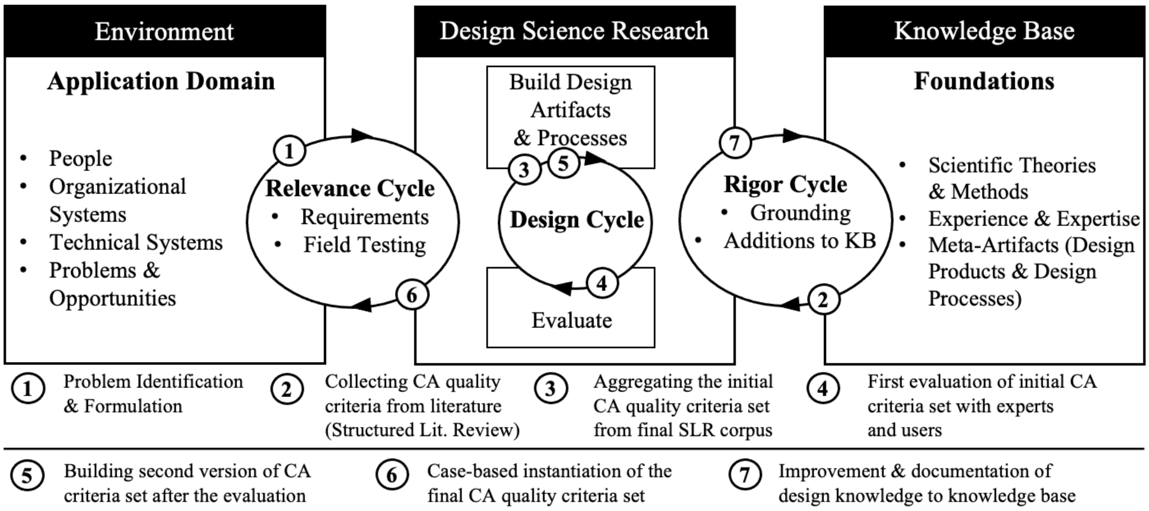

Figure 1. Design cycles and research activities, according to Hevner (2007).

The first step of the DSR approach is the identification of a pervasive real-world problem. In accordance with the Introduction and Related Research, we are building our research on the current lack of an overview in organizations concerning how a criteria-based approach could sustain the operation and continuous improvement of CAs to ensure their success throughout their lifecycle (see **Section 13.2**).

Based on this problem, in Step 2 we conducted a structured literature review (SLR) according to the five-step process of vom Brocke et al. (2009) in the databases of AISeL, ACM DL, IEEE Xplore, EBSCO, and ProQuest ABI/INFORM to derive initial criteria for evaluating CA quality. In this process, we based the subphases on established methods. For instance, we followed the taxonomy proposed by Cooper (1988) to define our SLR scope and Brink (2013) for the well-founded creation of a synonym list to structure the search process. We identified and verified suitable keywords via an initial database search to create the following search string: (“chatbot” OR “dialogue system” OR “conversational agent” OR “virtual assistant” OR “cognitive assistant”) AND (“qualit*” OR “design”

OR “criteria” OR “effectiveness” OR “evaluation” OR “usability”). Applying the search string to the aforementioned databases, we obtained 1,895 articles. We selected 180 of these for an in-depth analysis by screening each article’s title, abstract, and keywords. Utilizing deselection criteria ((1) technical or architectural aspects, (2) physical machines/robots, and (3) lack of CA application case) and deleting duplicates, we arrived at 94 articles. In a final full-text analysis, we classified 67 articles as relevant.

As part of Step 3, we initiated the design cycle to create an initial quality criteria set. To do so, we followed a multistep procedure. First, two researchers independently analyzed the full texts of the final 67 articles from Step 2 to identify suitable criteria. Second, the resulting criteria set (containing 221 potential criteria) was revised and condensed by (1) filtering out non-CA-specific criteria (e.g., related to the design of the messenger front end), (2) synthesizing similar and redundant criteria, (3) weighting aspects that multiple authors have addressed, and (4) deleting aspects irrelevant for evaluating text-based CAs (e.g., only relevant for speech-based assistance systems).

In Step 4, the initial literature-based criteria set was evaluated and expanded by interviewing seven CA users and experts. These interviews were conducted in December 2021 and lasted an average of 37 minutes. To ensure a systematic procedure, we developed a semi-structured guide, following the instructions of Gläser and Laudel (2009). Experts were asked (1) about their CA experience and possible quality criteria, and afterward (2) we presented the quality criteria from the literature to let them rate existing criteria and point out missing aspects.

Building on these insights, as part of Step 5, we created the final criteria set consisting of meta-criteria, criteria, and sub-criteria (see **Section 13.4**). Utilizing the insights from Step 2, as well as the statements from the expert interviews, we decided whether (1) a criterion had to be retained, revised, or added to the criteria set and (2) whether the criteria set was comprehensible or needed to be restructured overall.

In Step 6, we conducted a naturalistic evaluation of the final quality criteria set by supervising its instantiation in an IT organization (see **Section 13.5**). The goal was to verify whether the criteria set can be utilized to evaluate CA quality and whether it has the potential to help organizations continuously improve CAs in a structured way. Finally, we incorporated the findings from the naturalistic instantiation into the criteria set and improved and communicated them.

14.4 Quality Criteria Set for CAs

Based on the DSR research activities, we have derived a final criteria set consisting of 6 meta-criteria and 15 criteria with 33 sub-criteria to evaluate and improve CAs' quality throughout their lifecycle. The criteria set supports a cyclical evaluation process carried out at specified intervals in CAs' lifecycle which is performed based on previously collected data (e.g., monitoring, performance, or user data). In **Table 1**, we list the criteria along with example references that provide corresponding sources and interview insights.

14.4.1 Input

Input comprises criteria that focus on creating and submitting requests to the CA. In this context, the diverse **interaction abilities** of CAs can be evaluated (e.g., Kowald & Bruns, 2020). Many CA teams employ existing *communication channels* (e.g., messenger front ends, such as MS Teams, or websites), ensuring users are comfortable and familiar with their basic functions (Feng & Buxmann, 2020). However, reflecting, exchanging, or expanding channels with progressive development is essential. Moreover, various input *control elements* can be evaluated and integrated to facilitate dialog flow. For example, it may be helpful to allow users to interact with CA responses via buttons (Kowald & Bruns, 2020). Especially in the interviews, the need to continuously refine the selection and functionality of control elements was emphasized (e.g., text, buttons, reactions, and carousel selections). In addition, the **context awareness** of CAs should be evaluated. The ability to grasp *dialog-oriented context* allows CAs to incorporate previous user utterances to conduct a conversation with users. These conversations should be evaluated to ensure that users do not have to enter input repetitively (Saenz et al., 2017). Connected to this, resumption and return points in the dialog tree are fundamental aspects for evaluation. A well-structured conversation flow helps users provide the correct input, achieve their goals, and avoid deadlocks (Diederich et al., 2020). In addition, the *technical context* needs to be established to enable unrestricted usage, especially in complex use cases. From the first to the last user touchpoint, background systems should be conveniently accessed to provide correct data for the user's input (e.g., one-time user identification to address background systems to resolve requests).

14.4.2 Output

Regarding **output**, the **format** of CA responses should be reflected. The responses require the appropriate selection of a *suitable output format* in terms of a user- and content-oriented presentation (e.g., with texts, images, and tiles). An *appealing output formatting or visual representation of CA responses* is recommended (Kowald & Bruns, 2020). Especially in the CA context, users prefer short and manageable CA answers (Edirisooriya et al., 2019). In terms of **content**, the CA should transparently present its *capabilities* and *limitations* to evoke an appropriate user expectation that is consistent with the nature of the CA as an unfinished and learning IS (Diederich et al., 2020). Furthermore, CA answers should be reviewed to evaluate whether users' information needs have been fulfilled. The relevance and meaningfulness of presented information and the "up-to-dateness" of the knowledge base for *information retrieval* should be checked to decide whether background knowledge must be updated or expanded (Diederich et al., 2020). Apart from recognizing the user's intent and presenting the correct output, Feng and Buxmann (2020) emphasized the evaluation of different representations and levels of *detail of the knowledge.* Especially for more complex CAs (e.g., those that combine numerous background systems as a central platform), it is challenging to present solutions that are often complex in an abstract and *convergent* way that provide users with appropriate answers to their concerns. The interview experts highlighted that solutions sorted by relevance and *justification* of the CAs' answers could increase user trust in these answers. For example, a CA could refer to the background system/source to make transparent from where the knowledge was obtained (e.g., clickable link below the answer). Closely related, the CAs' **calibration** of *response appropriateness* should be evaluated to provide concise and manageable CA answers. In this context, CAs' *response accuracy* (also referred to as response quality, e.g., Jiang & Ahuja, 2020) needs to be evaluated to present knowledge correctly (e.g., length, tonality, fluency) to the target group. Regarding the **timing of responses**, on the one hand, the *technical response time* is considered a relevant factor for CAs. For example, Edirisooriya et al. (2019) identified quick responses—within two to five seconds of the user's request—as essential. On the other hand, the criterion *balance between proactivity and interruption* refers to the fact that CAs' proactive utterances may interrupt users. This behavior and its effects on users should be evaluated.

14.4.3 Anthropomorphism

Anthropomorphism refers to human characteristics, such as emotions, applied to nonhuman objects (Schuetzler et al., 2021). Anthropomorphism can positively affect the use of CAs and can be divided into three aspects: humanlike identity, verbal cues, and non-verbal cues (Seeger et al., 2021). First, evaluable criteria in the context of **humanlike identity** represent aspects that strengthen CA *identity* (e.g., profile pictures or avatars), and other *characteristics*, such as demographic information, including gender, age, or name (Seeger et al., 2021). In addition, the general *visual representation* was also highlighted during several interviews. A CA team should reflect on how the CA can be easily detected as the first contact point with the user, including, for example, its integration into a website, such as position, size, attractive [humanlike] appearance, and colors. Furthermore, CAs' **verbal cues** should be reviewed. In addition to the ability to engage in social dialogues, called *"chitchat," emotional expressions* (e.g., apologizing by the CA), *verbal style*, and *self-reference* (e.g., the chatbot referring to itself as "I" or "me"), or context-sensitive responses, *tailored personality, and lexical alignment* (e.g., by the CA adapting its responses to the users' utterances; Saenz et al., 2017) can also be used to make CAs seem more humanlike (Schuetzler et al., 2021; Seeger et al., 2021). In particular, chitchat and character definition were emphasized in the interviews, since many users first check the CA for its social capabilities and quickly lose interest if it fails, even at slight initial social interactions. Further possibilities of humanlike design are **non-verbal cues**, such as *emoticons,* or artificially induced *typing delays and indicators*, such as typing dots (Gnewuch et al., 2018). Continuously improving social skills has already had a short-term impact on the success of a CA. However, researchers (e.g., Grudin & Jacques, 2019) have also noted that a humanlike CA can be repellant to users. Seeger et al. (2021) indicated that the different anthropomorphism criteria must be combined and evaluated practically.

14.4.4 Dialog Control

For successful **Dialog Control,** CAs' understanding of users' requests, along with their intentions and goals, should be evaluated (Clark et al., 2019). However, CAs are learning IS and are, therefore, initially error-prone. In particular, user input in long and complex sentences poses a challenge for CAs (Michaud, 2018). Thus, proactive dialog handling in regular operations and reactive handling in failure operations should be evaluated to ensure that CAs avoid, reduce, or recover from failures. In **regular operations**, organizations should continuously reflect on whether the CA proactively avoids error scenarios by, for example, asking the user to *reformulate the request* (Diederich et al., 2020) or prompting the user for more information (Chaves & Gerosa, 2021). If no appropriate answer was elicited, the CA could proactively refer to misunderstandings or reintroduce skills (interviews). Afterward, the CA could provide *alternative responses* to keep the conversation alive (Chaves & Gerosa, 2021). Another way is to provide *conversational prompts*. Through the use of prompts, the CA provides *suggestions* for prospective requests in addition to its response (e.g., in the case of a long response time by the user). The aim is to predict the user's intentions (e.g., by suggestions on text buttons) and to proactively avoid error cases when processing a user's free text input (Li et al., 2020). In **failure operations,** it is crucial to define and evaluate *(e.g., proactive and resilient) repair strategies* to overcome conversational breakdowns, since their existence can result in a negative experience for users and impair future CA success (Benner et al., 2021). In the case of a breakdown, the CA should fail gracefully to maintain user trust (Feng & Buxmann, 2020). For instance, the CA can apologize and propose new solutions (Benner et al., 2021). However, if repair attempts fail repeatedly and the CA's capabilities are exceeded, the CA should encourage *fallbacks* or a *handover* to a service representative (Poser et al., 2021; Poser et al., 2022b).

14.4.5 Performance

A holistic evaluation of CA performance represents a strong predictor for CA success (Peras, 2018). By combining design-and technically-oriented principles, the CAs' **performance** is directly related to user satisfaction (Liao et al., 2016). The performance demonstrates the effective and efficient completion of executed tasks between the user and the CA (Peras, 2018). Regarding CAs' **effectiveness**, *the task (success) rate* and the *task failure rate* could be used to collect the average number of (successful) tasks and the average number of default fallback intents to trigger appropriate countermeasures (Peras, 2018). In the interviews, the *retention and feedback rates* were mentioned regarding recording returning users and continuously evaluating users' average ratings to uncover weaknesses to derive improvement potential. Furthermore, it is necessary to consider CAs' **efficiency** because the effective performance of tasks explicates only a few insights into whether the CA also performs the tasks with a resource-based approach. Given this perspective, evaluating the average time used to complete a task (*task completion time*) and the average number of rounds of dialogue required (*average number of turns*) is essential to capture efficiency (Holmes et al., 2019; Peras, 2018). In addition, the *human handover rate* is significant in evaluating at which points the CA cannot complete a task (Wintersberger et al., 2020).

14.4.6 Data Privacy

Data privacy includes criteria related to the implementation and communication of data protection. In the **implementation of data protection**, a relevant criterion is that the conversations with the CA should be kept as *private* and *anonymous* as possible, especially if the CA's context is confidential and personal data are processed (Feng & Buxmann, 2020). During the interviews, it was emphasized that as little data as possible should be stored during a conversation, and anonymized data should be stored if conversational data is obligatory to improve a CAs' performance. The **communication of data protection** contains the criterion of *transparency* toward users, meaning the disclosure of which user data is processed. In this context, it is helpful to provide data protection policies for users (Rajaobelina et al., 2021).

Table 1. Final CA Quality Criteria Set

<table>
<tr><th>Meta-criteria</th><th>Criteria</th><th>Sub-criteria</th><th>Example References</th></tr>
<tr><td rowspan="4">Input</td><td rowspan="2">Interaction abilities</td><td>Communication channel</td><td>(Feng & Buxmann, 2020), Interviews</td></tr>
<tr><td>Control elements</td><td>(Kowald & Bruns, 2020; Li et al., 2020), Interviews</td></tr>
<tr><td rowspan="2">Context awareness</td><td>Dialog-oriented context</td><td>(Diederich et al., 2020; Michaud, 2018; Saenz et al., 2017)</td></tr>
<tr><td>Technical context</td><td>Interviews</td></tr>
<tr><td rowspan="12">Output</td><td rowspan="2">Format</td><td>Suitable format</td><td rowspan="2">(Edirisooriya et al., 2019; Feng & Buxmann, 2020; Kowald & Bruns, 2020), Interviews</td></tr>
<tr><td>Appealing formatting and visualization</td></tr>
<tr><td rowspan="4">Content</td><td>Transparent capabilities and limitations</td><td>(Diederich et al., 2020; Saenz et al., 2017)</td></tr>
<tr><td>Information retrieval</td><td>(Diederich et al., 2020; Edirisooriya et al., 2019), Interviews</td></tr>
<tr><td>Detail of knowledge</td><td>Interviews</td></tr>
<tr><td>Solution convergence and justification</td><td>Interviews</td></tr>
<tr><td rowspan="2">Calibration</td><td>Response appropriateness</td><td rowspan="2">(Hu et al., 2018; Jiang & Ahuja, 2020)</td></tr>
<tr><td>Response accuracy</td></tr>
<tr><td rowspan="2">Time</td><td>Technical response time</td><td>(Edirisooriya et al., 2019; Meyer-Waarden et al., 2020), Interviews</td></tr>
<tr><td>Balance between proactivity and interruption</td><td>(Feng & Buxmann, 2020)</td></tr>
<tr><td rowspan="7">Anthropo-morphism</td><td rowspan="2">Humanlike identity</td><td>Identity and characteristics</td><td>(Schuetzler et al., 2021; Seeger et al., 2021)</td></tr>
<tr><td>(Humanlike) visual representation</td><td>Interviews</td></tr>
<tr><td rowspan="3">Verbal cues</td><td>Emotional expressions</td><td>(Saenz et al., 2017; Seeger et al., 2021)</td></tr>
<tr><td>Chitchat / Small talk</td><td rowspan="2">(Grudin & Jacques, 2019; Huiyang & Min, 2022; Schuetzler et al., 2021)</td></tr>
<tr><td>Tailored personality and lexical alignment</td></tr>
<tr><td rowspan="2">Non-verbal cues</td><td>Typing delay and indicator</td><td rowspan="2">(Gnewuch et al., 2018; Schuetzler et al., 2021; Seeger et al., 2021), Interviews</td></tr>
<tr><td>Emoticons</td></tr>
<tr><td rowspan="4">Dialog control</td><td rowspan="2">Regular operation</td><td>Reformulate requests and alternative responses</td><td>(Diederich et al., 2020; Saenz et al., 2017), Interviews</td></tr>
<tr><td>Conversational prompts and suggestions</td><td>(Kowald & Bruns, 2020; Li et al., 2020)</td></tr>
<tr><td rowspan="2">Failure operation</td><td>(Proactive & Resilient) repair strategies</td><td>(Benner et al., 2021; Diederich et al., 2020; Feng & Buxmann, 2020), Interviews</td></tr>
<tr><td>Fallbacks and handover</td><td>(Poser et al., 2021; Poser et al., 2022b; Wintersberger et al., 2020)</td></tr>
<tr><td rowspan="6">Perfor-mance</td><td rowspan="3">Effectiveness</td><td>Task (success) rate</td><td rowspan="2">(Peras, 2018), Interviews</td></tr>
<tr><td>Task failure rate</td></tr>
<tr><td>Retention and feedback rate</td><td>Interviews</td></tr>
<tr><td rowspan="3">Efficiency</td><td>Task completion time</td><td rowspan="2">(Holmes et al., 2019; Peras, 2018), Interviews</td></tr>
<tr><td>Average number of turns</td></tr>
<tr><td>Human-handover rate</td><td>(Wintersberger et al., 2020), Interviews</td></tr>
<tr><td rowspan="2">Data privacy</td><td rowspan="2">Implementation and communication</td><td>Privacy and anonymity</td><td rowspan="2">(Feng & Buxmann, 2020; Janssen et al., 2021b; Lewandowski et al., 2021; Rajaobelina et al., 2021), Interviews</td></tr>
<tr><td>Transparency</td></tr>
</table>

14.5 Case-Based Instantiation

After the rigorous derivation process, the final quality criteria set including all meta-criteria was instantiated in an IT organization to evaluate its applicability and feasibility. To this end, an existing CA *(ExpertBot)* was evaluated and improved along the criteria set by using various evaluation methods. The *ExpertBot* operates within organizational boundaries, is integrated into a messenger, and facilitates employees' search for internal experts (and their skills) to help employees and staff projects. Therefore, the CA participates in chat-based conversations involving multiple employees to suggest suitable experts by accessing diverse data sources (e.g., skill database, document management systems, internal chat forums).

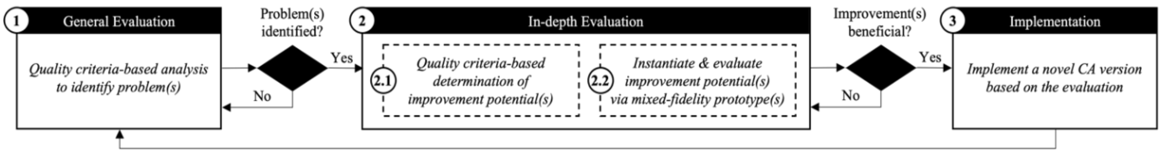

Figure 2. Procedure Model for the Evaluation and Improvement of CAs

Based on the criteria, an interdisciplinary team comprising experts from research and the IT organization conducted the evaluation of the *ExpertBot*'s quality. As there is limited substantiated knowledge on the procedure and selection of CA evaluation methods and required experts, an *explorative and iterative process* was initiated in cooperation with the IT organization. As a result, a procedure comprising three phases was completed (see **Figure 2**): In phase 1, the **general evaluation**, we performed a quality criteria-based analysis to identify problems of the current CA version in operations. The general evaluation revealed that the overall performance of *ExpertBot* was insufficient. Significant problem indicators, such as user retention and feedback rate, were considered throughout the criteria-based analysis to start an in-depth evaluation. Consequently, we initiated an improvement project to increase its performance.

As part of phase 2, the **in-depth evaluation**, we first conducted phase 2.1, an evaluation in cooperation with the organization to assess CAs' quality and determine improvement potentials based on our criteria set. In this context, 13 interdisciplinary participants (e.g., CA developers, employees responsible for staffing, and employees outside the subject area) were recruited to conduct an in-depth evaluation of the other meta-criteria to gain insights into the current *ExpertBot* quality and possible criteria interdependencies. This procedure allowed a multi-perspective in-depth evaluation of the *ExpertBot* due to the participants' varying experience levels regarding CAs and the broad discussion of criteria and weaknesses of the *ExpertBot*. Thereby, phase 2.1 was performed using a mixed-method approach. Semi-structured interviews were conducted with seven of the 13 participants in the beginning. We presented the current *ExpertBot* version to ask the participants about the general implementation and relevance for improvement along with the individual criteria from our set. Based on the analyzed qualitative data, a survey was conducted to ask participants to rank previously determined potentials according to their relevance for improving the *ExpertBot*.

In phase 2.2, we instantiated prototypes illustrating the highest ranked improvement potentials uncovered with the criteria set to test their feasibility and demonstrate an CA improvement. In this context, the prototypes provide a well-founded comparison of the current and novel CA version(s) and allow the involvement of real users by providing a basis for decision-making on whether the identified improvement potentials are beneficial when they are implemented or need to be revised. Therefore, the results of phase 2.1 were used to create mixed-fidelity prototypes, present them to participants, and compare them to the current CA version. Prototypes were designed based on the prioritized improvement potentials corresponding to the analyzed *ExpertBot*. For this purpose, the Figma (2022) design tool was used. The improvements were arranged into several scenarios, assembling suitable criteria (e.g., one scenario focused on the meta-criterion output with selected sub-criteria) to visualize and evaluate them with individual prototypes. Thereby, we presented all participants two prototypes for each scenario during semi-structured interviews. The first prototype contained the assumed improvements, while the other prototype represented the current CA state. During the presentation, three questions were asked regarding each scenario. First, participants were asked to evaluate which of the two prototypes was more effective at first glance and which aspects were crucial to this impression. Second, scenario criteria were individually addressed, and the participants were asked to determine which criterion is conceivable for increasing CA quality. Third, we asked which implemented criteria were the most important in elevating CA quality.

Finally, in phase 3, the improvement potentials identified in the evaluation to be particularly effective for increasing CAs' quality were implemented in a novel CA version by the CA team. After phase 3, the procedure should again start with phase 1 to examine the quality of the new CA version, which, however, was not part of the instantiation.

14.6 Discussion and Conclusion

Contemporary CAs have attracted considerable attention in organizations and academic research, introducing a paradigm shift in how users interact with IS (Zierau et al., 2020c). However, CAs have a high discontinuation rate (Gnewuch et al., 2017; Janssen et al., 2021b). In this context, a holistic overview of how to evaluate and improve CA quality throughout its lifecycle is lacking. From a research perspective, primary contributions exist in evaluating CA design (e.g., Seeger et al., 2021). However, existing scientific knowledge is segregated and does not yet address how CAs can be continuously evaluated and improved. Moreover, structured knowledge on how to conduct an improvement process is so far lacking. This is unsatisfactory, since CA development comprises several novel and effortful activities that should be systematically orchestrated.

To close this knowledge delta, we conducted a rigorous DSR project by aggregating insights from the literature supplemented by experiences from the practice-based, real-life environment to derive a systemized and synthesized set of CA quality criteria. In addition, we developed a procedure model in the context of the instantiation of the criteria set.

We contribute the presented criteria set, serving organizations as an overview of relevant aspects to evaluate and improve the quality of CAs as part of their operations. In combination with the application of the prototype method, the instantiation of the criteria set can pave the way to systematically evaluate and improve CAs by comparing different CA versions. First, the application of the criteria set enables organizations and CA teams to check whether a new CA version possesses better quality than the current version. Consequently, it can be ensured that a new CA version will be deployed only if its quality is at least as high as the previous version. Second, comparing the quality of the two CA versions may reveal improvement potentials before going live. Third, against the backdrop of moderate CA success, determining proper criteria can help CA teams (even beforehand) to design better CAs and evaluate them with users to confirm their intended use.

In addition to the criteria set, we contribute a procedure model, serving as a blueprint to apply the criteria set. This allows to structure the evaluation of CAs and discover areas for systematic improvement. Regarding required experts, we discovered that the involvement of people from different departments is beneficial for the evaluation process, as CA development is highly interdisciplinary and demands the combination of technical and design-oriented aspects (e.g., intent recognition, dialog design, CAs' front channel). For the instantiation, people from the IT, business and data protection departments were involved. Furthermore, people outside the subject area can significantly contribute to CA evaluation and improvement. In general, CAs' quality criteria evaluation should be conducted as naturally and quickly as possible to identify actual user behavior.

Overall, the combination of the applied criteria set and procedure model in the IT organization helps to address the knowledge gap on how to reduce the discontinuation rate of CAs and evaluate and improve CAs' quality throughout their lifecycle to sustain their operation.

However, the instantiation of the criteria set also revealed three challenges and aspects that need further research. First, as there is limited substantiated knowledge on the procedure and selection of CA evaluation methods and required experts in general, an explorative process was selected which was time-consuming in terms of both the actual activities as well as the application of the methods. Further research is needed to explore alternative or faster ways of performing activities and methods for criteria-based evaluation of CAs' quality. There may also be automation potentials with tool support. Closely related, a guideline is needed to determine when such an evaluation should be performed in general and who must be involved during the process. Second, in phase 1 of our instantiation (see **Chapter 5**), we determined the need for a quality criteria-based in-depth evaluation of CAs' performance in their natural context. The performance criterion proved to be a valid indicator for the improvement of *ExpertBot*. Nevertheless, further investigation is required to determine whether there are additional indicators to start the in-depth evaluation. For example, with increasing CA progress and more team expertise, other aspects of the criteria set could trigger an in-depth evaluation. Furthermore, there may be criteria that need to be more or less frequently evaluated, designed, or technically improved. Third, concerning the criteria set, we observed that different criteria have varying levels of influence on CAs' quality. Additionally, we discovered that specific criteria from our set differed in their importance depending on the expertise of our interview partners. For instance, anthropomorphism was less significant in the interviews and ranked low in phase 2.1 compared to the findings in the literature. Although the cue design of the CA was dominant in the reviewed literature, several experts stated that CA humanization was not as relevant for the use case considered in the project, which can be attributed to the fact that many employees in the IT organization have an IT-versed background. A classification or assessment ranking of the criteria's influence and importance combined with a more in-depth procedure offers additional potential for future research.

Despite these valuable insights, there are a few methodological limitations which provide further avenues for future research. First, concerning DSR, one objective was to apply our final quality criteria set in a naturalistic evaluation setting to verify whether the set could serve CA teams in evaluating and revealing potential improvements in a procedural and structured way. Although our set was applicable and could meet those objectives, the instantiation referred only to a single CA team in an IT organization. Further studies need to identify whether the criteria set can be applied to other organizations or if it needs to be extended or reorganized based on further perspectives. Second, the experts in this study and their domain-specific experiences influenced the study's external validity. In particular, our derived knowledge is dependent on their experiences. Finally, we recognized that the results depend on the authors' literature selection, aggregation, and judgment. Further studies could modify or prioritize the quality criteria set and reveal significant interdependencies for CA teams.

14.7 Acknowledgments

The research was financed by the German Federal Ministry of Education and Research and the European Social Fund (funding ref.: INSTANT, 02L18A111).

14.8 References

Alonso, A. F., Fuertes Castro, J. L., Martínez Normand, L., & Soza, H. (2009). Towards a set of measures for evaluating Software agent autonomy. Mexican International Conference on Artificial Intelligence (MICAI), Guanajuato, México.

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A review on recovery strategies to overcome conversational breakdowns of conversational agents. International Conference on Information Systems (ICIS), Austin, TX, United States.

Brink, A. (2013). Anfertigung wissenschaftlicher Arbeiten (5th ed.). Springer Gabler.

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. International Journal of HCI, 37(8), 729–758.

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., & Munteanu, C. (2019). What makes a good conversation? Challenges in designing truly conversational agents. Conference on Human Factors in Computing Systems, New York, NY, United States.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104–126.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A customer service chatbot for e-commerce websites. Meeting of the Association for Computational Linguistics-System Demonstrations, Vancouver, Canada.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019). On conversational agents in information systems research: Analyzing the past to guide future work. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. Business & Information Systems Engineering, 62(3), 193–209.

Edirisooriya, M., Mahakalanda, I., & Yapa, T. (2019). Generalized framework for automated conversational agent design via QFD. Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka.

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Figma Inc. (2022). Retrieved May 31, 2022, from https://www.figma.com

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., & Ischen, C. (2021). Future directions for chatbot research: An interdisciplinary research agenda. Computing, 103(12), 2915–2942.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. International Conference on Internet Science (INSCI).

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2018). Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. International Conference on Internet Science (INSCI).

Gläser, J., & Laudel, G. (2009). Experteninterviews und qualitative Inhaltsanalyse: Als Instrumente rekonstruierender Untersuchungen (3rd ed.). VS Verlag für Sozialwissenschaften.

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. European Conference on Information Systems (ECIS), Portsmouth, United Kingdom.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. International Conference on Information Systems (ICIS), Seoul, Korea.

Grudin, J., & Jacques, R. (2019). Chatbots, humbots, and the quest for artificial general intelligence. Conference on Human Factors in Computing Systems (CHI), Glasgow, Scotland.

Hevner, A. R. (2007). A three cycle view of design science research. Scandinavian journal of information systems, 19(2), 4.

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? European Conference on Cognitive Ergonomics.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. Conference on Human Factors in Computing Systems (CHI), Montréal, Canada.

Huiyang, S., & Min, W. (2022). Improving Interaction Experience through Lexical Convergence: The Prosocial Effect of Lexical Alignment in Human-Human and Human-Computer Interactions. International Journal of Human–Computer Interaction, 38(1), 28-41.

Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do chatbots fail? A critical success factors analysis. International Conference on Information Systems (ICIS), Austin, TX, United States.

Jiang, J., & Ahuja, N. (2020). Response quality in Human-Chatbot Collaborative Systems. Conference on Research and Development in Information Retrieval (SIGIR), Virtual Event, China.

Kowald, C., & Bruns, B. (2020). Chatbot Kim: A digital tutor on AI. How advanced dialog design creates better conversational learning experiences. International Journal of Advanced Corporate Learning, 13(3), 26.

Kuligowska, K. (2015). Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. Professionals Center for Business Research, 2.

Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-art analysis of adopting AI-based conversational agents in organizations: A systematic literature review. Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022). Design knowledge for the lifecycle management of conversational agents. International Conference on Wirtschaftsinformatik (WI), Nürnberg, Germany.

Li, C.-H., Yeh, S.-F., Chang, T.-J., Tsai, M.-H., Chen, K., & Chang, Y.-J. (2020). A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. Conference on Human Factors in Computing Systems (CHI).

Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What can you do? Studying social-agent orientation and agent proactive interactions with an agent for employees. ACM Conference on Designing Interactive Systems (DIS), Brisbane, Australia.

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), 3.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2021). Sorry, I can't understand you! – Influencing factors and challenges of chatbots at digital workplaces. International Conference on Wirtschaftsinformatik (WI), Essen, Germany.

Meyer-Waarden, L., Pavone, G., Poocharoentou, T., Prayatsup, P., Ratinaud, M., Tison, A., & Torné, S. (2020). How service quality influences customer acceptance and usage of chatbots. Journal of Service Management Research, 4(1), 35–51.

Michaud, L. N. (2018). Observations of a new chatbot: Drawing conclusions from early interactions with users. IT Professional, 20(5), 40–47.

Peras, D. (2018). Chatbot evaluation metrics. Economic and Social Development: Book of Proceedings, 89–97.

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. International Journal of Contemporary Hospitality Management, 32(10).

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Poser, M., Wiethof, C., & Bittner, E. A. (2022). Integration of AI into customer service: A taxonomy to inform design decisions. European Conference on Information Systems (ECIS), Timisoara, Romania.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint.

Rajaobelina, L., Prom Tep, S., Arcand, M., & Ricard, L. (2021). Creepiness: Its antecedents and impact on loyalty when interacting with a chatbot. Psychology & Marketing, 38(12), 2339–2356.

Saenz, J., Burgess, W., Gustitis, E., Mena, A., & Sasangohar, F. (2017). The Usability Analysis of Chatbot Technologies for Internal Personnel Communications. IIE Annual Conference, Norcross.

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding whether and how to deploy chatbots. MIS Quarterly Executive, 20(1), 4.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. Journal of the Association for Information systems, 22(4), 8.

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2019). How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electronic Markets, 30, 1–35.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. European Conference on Information Systems (ECIS).

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

Wintersberger, P., Klotz, T., & Riener, A. (2020). Tell me more: Transparency and time-fillers to optimize chatbots' waiting time experience. Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, Tallinn, Estonia.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. Conference on Human Factors in Computing Systems (CHI), New York, NY, United States.

Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020). A review of the empirical literature on conversational agents and future research directions. International Conference on Information Systems (ICIS), India.

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020). Towards developing trust-supporting design features for AI-based chatbots in customer service. International Conference on Information Systems (ICIS), India.

Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., & Leimeister, J. M. (2020). The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. International Conference on Information Systems (ICIS), India.

15 Publication No. 7: Lewandowski et al. (2023a)

Lewandowski, T., Kučević, E., Leible, S., Poser, M., & Böhmann, T. (2023). Enhancing conversational agents for successful operation: A multi-perspective evaluation approach for continuous improvement. Electronic Markets, 33(1), 39.

Abstract

Contemporary organizations increasingly adopt conversational agents (CAs) as intelligent and natural language-based solutions for providing services and information. CAs offer new forms of personalization, speed, (cost-)effectiveness, and automation. However, despite their hype in research and practice, many organizations still fail to seize CAs' potential because they lack knowledge of how to evaluate and improve the quality of CAs to sustain them in organizational operations. We aim to fill this knowledge gap by conducting a design science research project in which we aggregate insights from the literature and practice to derive an applicable set of quality criteria for CAs. Our article contributes to CA research and guides practitioners by providing a blueprint to structure the evaluation of CAs and to discover areas for systematic improvement.

Keywords: Artificial intelligence assistants · Conversational agents · Chatbots · Quality criteria set · Design science research (DSR)

15.1 Introduction

Recent technological advancements in intelligent and natural language-based information systems (IS) transform everyday life, work, and interactions (Brynjolfsson & McAfee, 2014; Davenport & Kirby, 2016; Diederich et al., 2020). As a result of ongoing developments in artificial intelligence (AI) and improvements in the underlying machine learning (ML) and natural language processing (NLP) algorithms, conversational agents (CAs) are becoming increasingly relevant in organizations as essential gateways to digital services and information (Følstad et al., 2021; Gnewuch et al., 2018). In this context, a recent analysis valued the global market for CAs at $3.49 billion in 2021 and expects it to grow to $22.9 billion by 2030, indicating their increasing importance (Research and Markets, 2022). Primarily operating in external or internal organizational environments (Patel et al., 2021), CAs interact with users (e.g., customers and employees) via natural language to provide convenient access to information from multiple connected systems and data sources. Moreover, CAs can perform standardizable processes and (cost-)effectively automate or assist tasks conventionally performed by employees (Meyer von Wolff et al., 2020a). In terms of automation, Gartner predicts that CAs will automate one in ten agent interactions by 2026 (Rimol, 2022). Consequently, CAs are expected to deliver significant economic value in existing and future applications, businesses, and digital ecosystems (Seeger et al., 2021; Seiffer et al., 2021).

Due to their potential, an extensive stream of research has focused on these AI-based systems (Cui et al., 2017; Zierau et al., 2020b). Since 2016, known as the "year of the chatbot" (Dale, 2016, p. 811), interdisciplinary research has explored various aspects related to CAs (Diederich et al., 2019a; Janssen et al., 2020), leading to a significant increase in both scientific and practical knowledge (Zierau et al., 2020a). More specifically, previous research has examined technical aspects (e.g., NLP improvements) as well as framework and platform selection (Diederich et al., 2019a, 2019b; Følstad et al., 2021). Scholars have also investigated user attitudes toward CAs, including acceptance, motivation, and behavioral implications, such as user trust (e.g., Brandtzaeg & Følstad, 2017; Go & Sundar, 2019; Seeger et al., 2017). In addition, prior studies have also focused on interaction design (e.g., Bittner et al., 2019; Gnewuch et al., 2018) and user preferences for visual cues and conversational design of CAs (e.g., Feine et al., 2019a; Schuetzler et al., 2021). Furthermore, social, ethical, and privacy challenges associated with CAs' implementation and use have been explored (e.g., Ischen et al., 2020; Ruane et al., 2019; Wambsganss et al., 2021).

Despite the steep increase in CA research and their vast opportunities, adopting CAs in organizational environments does not always have a positive impact because the technology is still error-prone and fails in interactions (Gnewuch et al., 2017; Riquel et al., 2021). These deficiencies concern CAs of varying maturity levels and regularly result in incorrect responses and conversational breakdowns (Weiler et al., 2022). Attributable to inadequate CAs, employees have developed negative feelings toward CAs and their providers in recent years (Diederich et al., 2020; Feine et al., 2019b; Schuetzler et al., 2021). To date, several potential reasons for the shortcomings and moderate success of CAs have been identified.

First, a primary reason for the limited success of CAs is their premature deployment, often driven by high expectations and management pressure, and usually combined with little knowledge of the CA development process in general and of CA quality in particular. This practice often leads to non-use, dissent, or complete failure, as highlighted by Janssen et al. (2021b) and Lewandowski et al. (2022b). Unsatisfactory CA design and limited capabilities can result in a frustrating user experience that triggers resistance and a loss of trust in the CA, further hindering its successful adoption in real-world environments (Weiler et al., 2022). The failure of CAs is frustrating not only for employees but also for the CA vendor, who has invested significant effort, time, and money into developing the CA (Janssen et al., 2021b; van der Goot et al., 2021).

Second, CAs are only marginally or not continuously evaluated to ensure their improvement, successful operation, and overall progress in organizations (Janssen et al., 2021b; Meyer von Wolff et al., 2021). Therefore, previous research has proposed continuous evaluation (e.g., via monitoring (Corea et al., 2020) or chatlog data (Kvale et al., 2019)) and operation and improvement processes (Lewandowski et al., 2022b; Meyer von Wolff et al., 2022) to regularly assess their use, quality, and added value (Brandtzaeg & Følstad, 2018; Meyer von Wolff et al., 2022). However, little is known about how to systematically organize this operation and improvement process to improve CAs. Previous studies have focused on single perspectives, such as continuous technical adaptations (e.g., retraining the NLP algorithm (Meyer von Wolff et al., 2022)), adjusting the knowledge base (Janssen et al., 2021b; Jonke & Volkwein, 2018), and improving individual CA functionalities and the dialog flow based on previous failures identified by chatlogs (e.g., Kvale et al., 2019). Despite these insights, there is a lack of knowledge on how CAs can be evaluated with criteria to test and improve their quality throughout their lifecycle (Lewandowski et al., 2022b). In addition, the current findings are often relatively fragmented across disciplines and application domains, and they therefore lack a cohesive axis of transferability for sustained practical usage (Elshan et al., 2022a; Følstad et al., 2021; Li & Suh, 2022). In this regard, experts in the field urge for more collaboration and aggregation in interdisciplinary research on CAs, and encourage further research on the topics around measurement, modeling, and evaluation approaches for CAs, as outlined, for example, in the CA research agenda by Følstad et al. (2021). To the best of our knowledge, there is no holistic overview of criteria for researchers and practitioners and approaches to continuously evaluate, improve, and sustain CAs that would facilitate organizations in this problem context. Therefore, this article explores the following research question:

RQ: *What are relevant criteria for continuously evaluating the quality of CAs, and how can they be applied?*

By addressing the research question, this article aims to systematize the continuous evaluation and improvement of CAs to counteract CA failure in organizational environments. To successfully operate a CA, measurements or criteria are needed for orientation to adapt CAs to user needs (Følstad et al., 2021; Meyer von Wolff et al., 2022). Therefore, we pursued a twofold contribution: (1) a set of relevant criteria to evaluate the quality of CAs and (2) a procedure model as part of the instantiation of the quality criteria set in an IT organization, prescribing its application and evaluation activities. The criteria set and procedure model define a cyclical criteria-based evaluation process that can be triggered by different impulses. These results address the identified research gap and present an approach for practice. Specifically, our proposed quality criteria set addresses this lack of knowledge about the successful operation of CAs. In this context, the evaluated set of quality criteria and the procedure model can serve as an initial overview for organizations to structure CA evaluations systematically and discover areas for improvement. Following design science research (DSR) activities mapped to the three-cycle view by Hevner (2007), we approach the derivation of these results with the following structure: First, we present the related CA research and delineate the research gap in more detail. Next, we describe our research approach to developing our artifact. We then present the findings of our work, including an overview of our final quality criteria set. Subsequently, we outline the instantiation of the quality criteria set using a real-life case in an IT organization. Finally, we discuss our findings as well as their implications and conclude with our limitations and potential future research.

15.2 Related Research

15.2.1 Text-based conversational agents as specific AI-based IS

Research on AI-based IS has attracted substantial attention (Elshan et al., 2022b; Felderer & Ramler, 2021) and transformed from a technical trend to a pervasive phenomenon in our daily lives (Maedche et al., 2019). AI-based systems proliferate in various application domains and contribute to multiple innovations (L. Wang et al., 2020). One application area that has seen renewed interest and increasingly utilizes AI is communication with computers via natural language, which has been a topic of research and practice for several decades (Gnewuch et al., 2017). Since the 1960s, researchers have worked on text-based and later speech-based CAs to automate procedures and assist users with various tasks (Følstad et al., 2021). An early example was ELIZA, which allowed initial natural language-based interactions with a computer (Weizenbaum, 1966). However, technical limitations (e.g., computational power and storage capacity) and overly simplistic capabilities (e.g., non-learning algorithms) restricted early attempts at CAs as they could not meet the high expectations (Diederich et al., 2019a; Gnewuch et al., 2017). According to Dale (2016) and Klopfenstein et al. (2017), ELIZA and other previously developed CAs used simple rule-based mechanisms to generate responses.

Nevertheless, in recent decades, technological progress has allowed the development of more sophisticated CAs that utilize novel AI, ML, and NLP algorithms and models (Gnewuch et al., 2017). In this context, the CA attempts to understand the user's intention behind the input prompt to provide an adequate response output. In particular, the techniques of supervised learning, unsupervised learning, and human-in-the-loop (where humans are involved in the training process) lead to increasingly better CAs (Radziwill & Benton, 2017; Wiethof & Bittner, 2021). As a result, they have gained widespread adoption and can now better address the needs of the general public and the mass market (Maedche et al., 2019).

CAs support the ongoing digitalization and automation of organizations by performing various activities, such as filtering information or efficiently assisting employees in their daily tasks (Zierau et al., 2020a). Hence, with their scalability and 24/7 availability (Gnewuch et al., 2017; Xu et al., 2017), CAs can have a transformative impact on business operations by acting as a central service platform and first point of contact for customers, providing a convenient way to handle service requests more individually before human intervention (Zierau et al., 2020a), and reducing information overload for users (Xu et al., 2017). Accordingly, employees can concentrate on more complex, creative, and non-routine tasks.

The widespread use of CAs has generated significant research interest, with a rapidly growing body of contributions. However, CA research has a strong interdisciplinary character and is fragmented into several research streams (Følstad et al., 2021): Multiple perspectives and disciplines, including "informatics, management and marketing, media and communication science, linguistics and philosophy, psychology and sociology, engineering, design, and human-computer interaction" are employed to study CAs (Følstad et al., 2021, p. 2916). This interdisciplinary research has introduced numerous designations, such as chatbots (e.g., Dale, 2016), conversational (user) interfaces (e.g., Herrera et al., 2019), or dialog systems (e.g., McTear, 2021), leading to debates in the literature about their terminology and classifications. In this context, the authors Gnewuch et al. (2017), for example, have divided these AI-based IS into two subclasses: text-based CAs (e.g., chatbots or natural dialog systems) and speech-based CAs (e.g., smart speakers or virtual assistants).

In this article, we use the term “conversational agent” to refer to all AI and text-based representations such as chatbots. Although some research indicates that the distinction between text and speech-based CAs is marginal since speech-based input can be transferred to text-based input and vice versa from a technical viewpoint (Diederich et al., 2019b), research has also revealed that evaluating speech-based CAs requires distinct criteria compared to text-based CAs. For instance, evaluating the quality of a smart speaker involves design elements, such as overall (hardware) appearance (including styling elements and imagery), as discussed in Su and Hsia (2022). In addition, privacy handling is an important issue, for instance, when referring to the proactive (i.e., listening continuously to react) or reactive (i.e., reacting restricted to specific keywords) activation of speech-based CA, as discussed by Burbach et al. (2019). Furthermore, the ability of smart speakers to process audio speech and handle different dialects, tonalities, and noise in different input environments is crucial (e.g., Bisio et al., 2018), as is robust output generation (e.g., text-to-speech generation and perception of understandability and naturalness (Schmitt et al., 2021)). In summary, while text and speech-based CAs share some commonalities, evaluating their quality requires considering specific criteria unique to each modality.

15.2.2 Continuous evaluation and improvement of conversational agents

While the development of a CA has become much more accessible, the underlying IS is complex by nature (Maroengsit et al., 2019). Besides the described possibilities and applications of CAs, the management, evaluation, and improvement of these AI-based systems pose new challenges for organizations. These activities are essential because disregarding them can result in high failure and discontinuation rates (Diederich et al., 2020; Janssen et al., 2021b). Many CAs have failed in real-world environments due to, among other reasons, frustrating user experiences (Følstad et al., 2018a). As a result, multiple organizations have taken their CAs offline since they lack knowledge of how to ensure continuous evaluation and improvement, leading to an uncoordinated and highly exploratory development process (Janssen et al., 2021b). As CAs represent a novel form of IS with distinct characteristics that differentiate them from traditional IS and other AI-based systems, they require new approaches for design, evaluation, and improvement.

One unique characteristic of CAs is their sociability. As *social* IS, they are capable of interacting with users via natural language, representing a new sociotechnical application class (Maedche et al., 2019). These AI-based systems impact traditional service delivery and enable new individualized and convenient sociotechnical interactions (Klaus & Zaichkowsky, 2020), requiring humanlike, user-centered, and socially interactive IS design (Lewandowski et al., 2022a). Contrary to the classification of AI-based CAs as IS from a technological perspective, the existing literature shows that the organizational adoption and practical use of CAs must be viewed in fundamentally different ways (e.g., Corea et al., 2020; Lewandowski et al., 2021). As a result, CA teams are designing chatbots differently from traditional IS and from multiple new perspectives, having equipped them with social features, names, avatars, and communicative behaviors to attract users' attention and simulate natural conversation (McTear et al., 2016). Nonetheless, enhancing the user experience of CAs remains a crucial challenge owing to the absence of a comprehensive overview to determine whether they are well-designed and useful, and because of the lack of widely applied approaches to evaluate and improve them, as described in the interdisciplinary chatbot agenda by Følstad et al. (2021).

Another unique characteristic of current CAs is their level of intelligence and ability to learn and improve via naturalistic interactions. As such, they can be classified as a form of *learning and intelligent IS,* depending on the ongoing development and introduction of, so far, unsolved challenges (Lewandowski et al., 2021; Zierau et al., 2020a). CAs often have limited skills initially, and learning progress depends on the application area and the actors' engagement in training these systems. Accordingly, CAs' learning progress is highly context-driven and thus dependent on actual application and usage (Clark et al., 2019; Zierau et al., 2020c). The learning nature of CAs indicates the necessity for novel approaches to their evaluation and improvement (Lewandowski et al., 2022b; Meyer von Wolff et al., 2021).

Consequently, the highest effort needs to be invested in operations where CAs require continuous evaluation and later training and improvement in a real-world context. This endeavor is complicated by rapid changes and high dynamics, in which it is generally impossible to predict how users will interact and what information will be retrieved long-term (Janssen et al., 2021b). CAs have gained a great deal of research attention, with perspectives ranging from specific conceptual or usability-related aspects to technical design. However, detailed theoretical and practical knowledge is lacking for the operation in general and the continuous improvement process of CAs in particular (Lewandowski et al., 2022b; Meyer von Wolff et al., 2021). Hence, a comprehensive and systemized criteria-based approach to continuously evaluate CAs' quality can help to improve and sustain them.

15.2.3 Evaluation criteria for conversational agents

In recent years, the overall user experience and improvement of CAs have been prominent topics in research endeavors. There is a growing body of knowledge on methods and measures to evaluate the overall user experience with CAs, resulting in initial factors contributing to a positive or negative user experience (Følstad et al., 2021; Zarouali et al., 2018). In addition, authors have examined various effects of CAs at the individual level, either on perceived human likeness, trust, perceived social support, enjoyment, affordance theory (Lee & Choi, 2017; Stoeckli et al., 2019; Zierau et al., 2020b), or in the broader context of IS acceptance theories, such as in the "Technology Adoption Model" (e.g., Pillai & Sivathanu, 2020). However, there is little research on concrete quality criteria that can be applied to ensure systematic CA evaluation and improvement. Thereby, scholars call to establish convergence in interdisciplinary CA research in measurements, models, and approaches for evaluating CAs (Følstad et al., 2021).

Contributions referring to the design and evaluation of CAs are beginning to emerge. According to Følstad et al. (2021), there is a rapidly growing body of work on CA interaction design (e.g., Ashktorab et al., 2019), CA personalization (e.g., Laban & Araujo, 2020; Shumanov & Johnson, 2021), use of interaction elements (e.g., Jain et al., 2018), social cues (e.g., Feine et al., 2019a; Seeger et al., 2021), and capability representation. However, current research is often confined to (1) single design issues or the effects of dedicated design elements (e.g., Seeger et al., 2021), (2) technical measurements or technical performance (e.g., Alonso et al., 2009; Goh et al., 2007), (3) other agent classes, such as embodied or speech-based CAs (e.g., Kuligowska, 2015; Meira & Canuto, 2015), and (4) individual design aspects (e.g., Seeger et al., 2021), while (5) being segregated through the interdisciplinary CA research landscape. In addition, CA-oriented research has (6) focused on satisfaction issues, such as human behavior or ethical aspects (e.g., Neff & Nagy, 2016; Radziwill & Benton, 2017), affect and emotions, such as mood adjustment, entertainment, and authenticity (e.g., Meira & Canuto, 2015; Pauletto et al., 2013; Radziwill & Benton, 2017) and (7) initial classifications and typologies for high-level analysis and guidance on interaction design (Følstad et al., 2018b), which only play an overarching role for development.

Important preliminary work includes ISO 9241-oriented CA evaluation criteria sets, such as those of Radziwill and Benton (2017), Casas et al. (2020), and Johari and Nohuddin (2021), representing first CA quality criteria sets and approaches. However, they tend to focus on improvements at a high meta-level, such as those regarding efficiency (e.g., robustness to manipulation or unexpected input), effectiveness (e.g., if the CA passes the Turing test), impact and accessibility (e.g., meets neurodiverse needs), trustworthiness and transparency (e.g., security and intrusiveness), and humanity and empathy (e.g., the realness of a CA or personalization). While these criteria may provide valuable guidance in the initial evaluation and improvement of CAs by addressing technical concerns, such as increasing the accuracy of NLP components or conducting user surveys to gauge initial perceptions, they have limited utility for CA teams in organizations seeking to ensure the long-term success of CAs within an application environment. This limitation necessitates a comprehensive system-wide perspective, for example, with respect to the overall input processing, the output presentation, representation elements, or the design of the dialog flow. To fill this research gap, it is essential to develop a more detailed, multi-perspective, and comprehensive set of quality criteria for researchers and practitioners that addresses a broader range of requirements for the long-term success of CAs.

15.3 Research Approach

Our objective is to create a set of quality criteria for CAs as a central artifact that allows organizations to continuously evaluate and improve their CAs. To achieve this objective, we used the DSR paradigm and applied the three-cycle view presented by Hevner (2007). DSR is well-established in IS research and appropriate for our research because we aim to create an artifact that addresses a real-world problem and enables the continuous improvement of CAs to counteract their failure (Gregor & Hevner, 2013). Following the classification of contribution types in the DSR of Gregor and Hevner (2013), this research contributes knowledge at different levels. We contribute to level two by creating an operational artifact in the form of a set of quality criteria, including a procedure model (design knowledge). We also contribute to level one (artifact instantiation) by applying the quality criteria in a real-world context. We aim to derive and generate prescriptive knowledge from the descriptive knowledge extracted and evaluated from the knowledge base (Drechsler & Hevner, 2018). This knowledge will serve as a normative blueprint for practitioners and a starting point for further research. To structure our research endeavor according to the established ground rules of DSR, we conducted seven research steps, as illustrated in **Figure 1.**

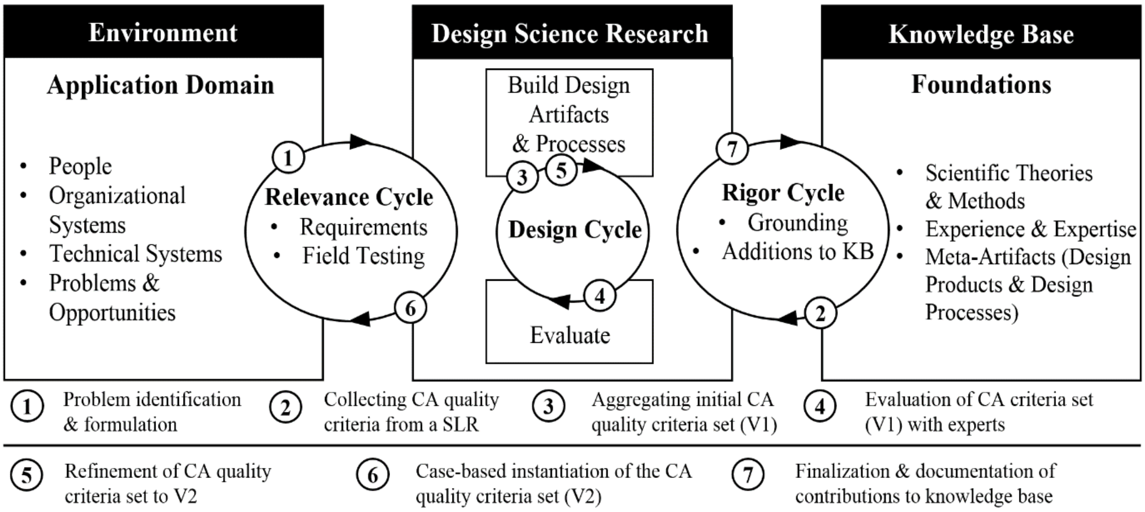

Figure 1. DSR three-cycle view and our research steps based on Hevner (2007)

Step 1 of the DSR approach refers to the identification and formulation of a pervasive real-world problem. The initial situation was investigated through two semi-structured interviews following a prepared interview guide (see **Table 1**), revealing that the overall quality and usage rate of their used CA (*ExpertBot*) was insufficient, and at the same time, the IT organization lacked concrete criteria and an improvement process for it. Supplementing these insights, we examined the successful and failed use cases of organizational CAs in the current body of literature to highlight the practical relevance of the problem beyond our specific case. This status quo demonstrated the need for a solution approach that defined the addressed overarching problem class. Therefore, our research was based on the current knowledge gap regarding how a criteria-based endeavor could sustain the operation and continuous improvement of CAs to ensure their long-term success. This knowledge gap was grounded and described in the Introduction and Related research sections. As a result, we adopted a problem-centered perspective at the beginning of our research, based on Peffers et al. (2007).

Table 1. List of interviewees of Step 1 and the first relevance cycle.

ID	Interviewee	Duration
Exp1.1	AI expert & senior project manager	56 minutes
Exp2.1	Software engineer & project manager	57 minutes

Based on the formulated problem, in *Step 2*, we conducted a structured literature review (SLR) to derive the initial criteria for evaluating CA quality. We followed the five-step process of vom Brocke et al. (2009) in the databases of AISeL, ACM DL, IEEE Xplore, EBSCO, and ProQuest ABI/INFORM. We defined the scope of our SLR using the taxonomy proposed by Cooper (1988), as shown in Table 2. We focus on research outcomes, practices, and applications of quality criteria for CAs. Our goal is to address the lack of anchor points that allow continuous evaluation and improvement of CAs by synthesizing the relevant literature. We adopted a neutral perspective by paying attention to different existing (interdisciplinary) criteria sets of CAs and methods of measuring their effectiveness. Our coverage strategy followed a representative nature, focusing specifically on essential and influential literature to answer our research question. A conceptual organization was chosen to cluster the existing research contributions. The results of our literature review are intended for IS researchers and interdisciplinary researchers concerned with CAs. Furthermore, practitioners can apply the derived quality criteria and procedures to improve their CAs.

Table 2. Applied taxonomy of literature reviews by Cooper (1988)

Characteristic	Categories			
Focus	Research Outcomes	Research methods	Theories	Applications
Goal	Integration	Criticism	Central issues	
Perspective	Neutral representation	Espousal of position		
Coverage	Exhaustive	Exhaustive (selective citation)	Representative	Central or pivotal
Organization	Historical	Conceptual	Methodological	
Audience	Specialized scholars	General scholars	Practitioners or policy makers	General public

We first identified the central terms in our research question and decomposed them into related concepts to construct a search term (Brink, 2013; Xiao & Watson, 2019). Next, we used the resulting terms to conduct an initial unstructured literature search of the databases: "conversational agent," "evaluation," "criteria," and "qualit*." We extracted keywords, synonyms, and homonyms from the relevant papers found (Rowley & Slack, 2004; vom Brocke et al., 2009) and used them to form the following search string: ("chatbot" OR "dialogue system" OR "conversational agent" OR "virtual assistant" OR "cognitive assistant") AND ("qualit*" OR "design" OR "criteria" OR "effectiveness" OR "evaluation" OR "usability"). We applied the search string to the aforementioned databases, resulting in 1895 articles. After screening the titles, abstracts, and keywords of each article, we selected 180 articles for in-depth analysis. To further filter the literature corpus, we established exclusion criteria to ensure that only relevant articles were included in the dataset. Two researchers independently used these criteria to screen the articles and reduce potential selection biases. Subsequently, we removed articles that addressed (1) technical or architectural aspects, (2) physical machines or robotics and their interfaces, or (3) no specific use cases for CAs. We also removed duplicates. During this process, we reduced our literature dataset to 94 articles by examining their research questions and results sections. In a final rigorous full-text analysis, we identified 67 articles as relevant, consisting primarily of journal and conference articles. **Figure 2** illustrates the literature review process.

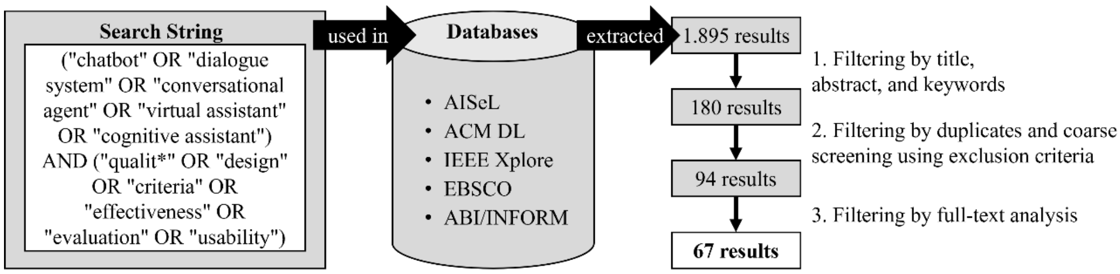

Figure 2. Literature review process according to vom Brocke et al. (2009)

In *Step 3*, we embarked on the first design cycle to establish a quality criteria set, version 1 (V1). To do so, we followed a multi-step procedure. Initially, we independently extracted appropriate quality criteria by conducting a full-text analysis of the final 67 articles from *Step 2*. Next, we integrated the extracted criteria into a shared document containing 221 criteria, with brief descriptions and references. We then refined and streamlined the criteria based on three aspects. First, we sorted all criteria by topic and removed non-relevant criteria for our research scope (*Step 2*). Therefore, we excluded, for example, non-CA-specific criteria, irrelevant to the evaluation of text-based CAs (e.g., those relevant only to speech-based or embodied assistants). Second, we combined criteria that were indistinguishable and removed redundant criteria. Third, we weighted the criteria based on their frequency in the reviewed literature. Due to the quantity and complexity of the collated quality criteria set, we developed a multi-level model consisting of three levels: meta-criteria, criteria, and sub-criteria. This hierarchical arrangement allows for the holistic or selective application of the quality criteria set, enabling the evaluation of specific (topic-based) areas without needing to use the entire set.

In *Step 4*, we evaluated the initial literature-based quality criteria set (V1) through semi-structured interviews to expand the set in a second design cycle. We used Venable et al.'s (2016) Framework for Evaluation in Design Science (FEDS) throughout this process to define the overarching evaluation strategy. Our primary goal was to review and improve the quality criteria set developed for evaluating and improving CAs. Therefore, we chose a formative ex-ante approach to evaluate the quality criteria set for this design cycle. To prepare for the interviews, a semi-structured interview guide with questions about all quality criteria was created to ensure a systematic procedure and comparably gathered data. We then conducted and recorded seven interviews with experts from an IT organization with professional experience in CA projects and external researchers, following the methods of Gläser and Laudel (2009), and Meuser and Nagel (2009b). We discussed the possible quality criteria of CAs with the interviewees based on their expertise before individually presenting and assessing our identified ones, and having them extend our existing quality criteria set and point out missing aspects. **Table 3** presents the list of interviewees of this second design cycle.

Table 3. List of interviewees of Step 4 and the second design cycle

ID	Interviewee	Duration
Exp1.2	AI expert & senior project manager	42 minutes
Exp2.2	Software engineer & project manager	42 minutes
Exp3	Principal data manager	36 minutes
Exp4	Branch manager	35 minutes
Exp5.1	CA developer/engineer	29 minutes
Exp6	CA researcher	40 minutes
Exp7	CA researcher	34 minutes

Building on the insights gained from the interviews conducted in *Step 4*, we developed V2 of our quality criteria set in *Step 5*. During the interviews, we received feedback from experts and gathered valuable input on the initial criteria set (V1). We decided whether a criterion had to be retained, revised, or added to the criteria set. In this context, we considered the experts' suggestions on the wording and structural arrangement of the criteria for reasons of comprehensibility, leading to design adjustments in V2. Furthermore, the experts' experience with CAs in real-world contexts led to the identification of additional quality criteria, which were integrated into V2 in a complementary manner where appropriate.

In *Step 6*, we conducted a summative naturalistic ex-post evaluation of the quality criteria set (V2) by supervising its case-based instantiation in an IT organization using the FEDS (Venable et al., 2016). Our goal was to verify if the criteria set could be used to evaluate CA quality and whether it could help organizations improve their CAs in a structured and normative way by emphasizing its usefulness and relevance. To achieve this, we developed a procedure model for the application and instantiation of the quality criteria set and conducted two interview rounds. The first round included seven experts, three of them from *Step 4* and four new participants; in the second round, one of the new participants was not available, so 13 interviews were conducted in total, as shown in **Table 4**. During the first round, we asked the experts about the current state version of *ExpertBot,* and its problems and potential for improvement before transitioning to the individual criteria from our set to create suitable scenarios. We used scenarios as flexible containers that included a certain number of our quality criteria that matched a collective evaluation. In this context, mockups were created with Figma (2022) as prototypes to simulate each scenario with a current state version and a modified version of *ExpertBot,* incorporating altered criteria aligned with our criteria set (V2). In the second round, we used the created prototypes to simulate each scenario previously defined in A/B tests related to Young (2014). This allowed us to determine which criteria were considered highly influential and most important to the experts. In addition, we paid attention to whether the experts mentioned new criteria in the instantiation that were not yet included in our quality criteria set. The procedure is more detailed in the "Case-based instantiation of the quality criteria set" section below.

Table 4. List of interviewees of Step 6 and the second relevance cycle

ID	Interviewee	Duration – round 1	Duration – round 2
Exp1.3	AI expert & senior project manager	34 minutes	41 minutes
Exp2.3	Software engineer & project manager	39 minutes	28 minutes
Exp5.2	CA developer/engineer	41 minutes	38 minutes
Exp8	Product owner	33 minutes	28 minutes
Exp9	Management assistance	35 minutes	34 minutes
Exp10	Senior software architect	38 minutes	32 minutes
Exp11	Senior software engineer	34 minutes	-

Finally, in *Step 7*, we incorporated the evaluation results from *Step 6*, the naturalistic case-based instantiation, into the quality criteria set and developed the final version presented and documented in the next section. Following Gregory and Muntermann's (2014) theorizing framework, we iteratively developed and improved an abstracted artifact version that met the larger problem class derived in *Step 1*. We communicated the quality criteria set for CAs as a rigorously elaborated prescriptive artifact, providing applicable knowledge that contributes to the knowledge base as a solution design entity for practitioners with an adaptable framework for situational instantiations to improve their CAs by applying the derived quality criteria (Drechsler & Hevner, 2018). In addition, our set provides descriptive knowledge as an observation and classification concept for researchers, with new insights and starting points for further research on evaluating, understanding, and improving CAs for their long-term success.

15.4 Quality Criteria Set for Conversational Agents

Based on the DSR research activities, we derived the final criteria set for evaluating and improving the quality of CAs. This set incorporates a hierarchical structure consisting of 6 metacriteria, 14 criteria, and 33 sub-criteria, enabling a systematic and rigorous evaluation process. The *meta-criteria* are the highest level of abstraction, representing the overarching evaluation areas of a CA. The *criteria* at the second level break them down. These can be used, for example, to create responsibilities in a CA team for (meta-)criteria areas, ensuring that accountability is clearly defined and understood. This structure also supports informed decision-making (e.g., prioritizing specific criteria of the CA). Although (meta-)criteria provide logical and structural clarity and classification, they are not sufficiently granular for evaluation purposes. Therefore, at the third level, *sub-criteria* have been defined as specific elements that can be evaluated using qualitative or quantitative methods. Overall, this approach allows for a detailed and comprehensive evaluation of CAs. The following section presents the quality criteria along the six meta-criteria and their hierarchical structures depicted in **Table 5**.

15.4.1 Input

Input comprises criteria that focus on creating and submitting requests to the CA. In this context, the diverse *interaction abilities* of CAs can be evaluated (e.g., Kowald & Bruns, 2020). Many CA teams employ existing *communication channels* (e.g., messenger front ends, such as Microsoft Teams or websites), ensuring that users are comfortable and familiar with their basic functions (Feng & Buxmann, 2020). However, reflecting, exchanging, or expanding channels with progressive development is essential. Moreover, various input *control elements* can be evaluated and integrated to facilitate dialog flow. For example, it may be helpful to allow users to interact with CA responses via buttons (Kowald & Bruns, 2020). The interviews emphasized the need to continuously evaluate and refine the selection and functionality of control elements (e.g., text, buttons, reactions, and carousel selections). In addition, the *context awareness* of CAs should be evaluated. The ability to grasp *dialog-oriented contexts* allows CAs to incorporate previous user utterances to conduct sophisticated conversations with users. These conversations should be evaluated to ensure that users do not have to enter input repetitively (Saenz et al., 2017). Connected to this, resumption and return points in the dialog tree are fundamental aspects of evaluation. A well-structured dialog flow helps users provide the correct input, achieve their goals, and avoid deadlocks (Diederich et al., 2020). Moreover, the *technical environment* needs to be established to enable unrestricted usage, especially in complex use cases. From the first to the last user touchpoint, background systems should be conveniently accessed (e.g., single sign-on) to address background systems that resolve requests and provide information.

15.4.2 Output

Output refers to criteria related to the CA-generated response provided in return to the user request. Regarding output, the *format* of the CA responses should be reflected. The responses require an appropriate selection of suitable *visual elements* in terms of a user and content-oriented presentation (e.g., with texts, images, and tiles), as well as high *readability* (Kowald & Bruns, 2020). Especially in the context of CAs, *consistency* in language and terminology is important for avoiding complexity and confusion for users (Edirisooriya et al., 2019). In terms of *content*, the CA should transparently disclose its *capabilities* and *limitations* to evoke appropriate user expectations that are consistent with the nature of the CA as a learning IS (Diederich et al., 2020). Furthermore, CA answers should be reviewed to evaluate whether users' (information) needs have been fulfilled. The relevance and meaningfulness of the presented information and the "up-to-dateness" of the knowledge base for *information retrieval* should be checked to determine whether background knowledge must be updated or expanded (Diederich et al., 2020). Apart from recognizing the user's intent and presenting the correct output, Feng and Buxmann (2020) emphasized the evaluation of different representations and levels of *detail of the knowledge*. Especially for more complex CAs (e.g., those that combine numerous background systems as a central platform), it is challenging to present the often multifaceted solutions in an abstract and *convergent* way that provides users with appropriate answers to their concerns. The interview experts highlighted that solutions sorted by the relevance and *justification* of the CAs' answers could increase user trust in these outputs. For example, a CA could refer to the background system or source to make it transparent from where the knowledge was obtained (e.g., clickable link below the answer). Closely related, the CAs' *calibration* of *response appropriateness* should be evaluated to provide concise and manageable CA answers. In this context, CAs' *response accuracy* (e.g., also referred to as "response quality" (Jiang & Ahuja, 2020)) needs to be evaluated to present knowledge correctly (e.g., length, tonality, fluency) to the target audience. Regarding the *timing of responses*, *technical response time* is considered a relevant factor for CAs. For example, Edirisooriya et al. (2019) identified quick responses—within two to five seconds of the user's request—as essential. However, the criterion *balance between proactivity and interruption*, which refers to the fact that CAs' proactive utterances may interrupt users, indicates that this behavior and its effects on users should be evaluated.

15.4.3 Anthropomorphism

Anthropomorphism relates to human characteristics, such as emotions, applied to nonhuman objects (Schuetzler et al., 2021). Anthropomorphism can positively affect the use of CAs and can be divided into three aspects: humanlike identity, verbal cues, and non-verbal cues (Seeger et al., 2021). First, evaluable criteria in the context of *humanlike identity* represent aspects that strengthen CA *identity* (e.g., profile pictures or avatars) and other *characteristics*, such as demographic information, including gender, age, or name (Seeger et al., 2021). In addition, the general *visual representation* was highlighted during several interviews. A CA team should reflect on how the CA can be easily detected as the first contact point with the user, including, for example, its integration into a website, such as its position, size, responsive (humanlike) appearance, and colors. Furthermore, CAs' *verbal cues* should be reviewed. Besides the ability to engage in social dialogs, called "*chitchat*," *emotional expressions* (e.g., apologizing by the CA), verbal style, and self-reference (e.g., the CA referring to itself as "I" or "me"), or context-sensitive responses, *tailored personality and lexical alignment* (e.g., by the CA adapting its responses to the users' utterances (Saenz et al., 2017)) can also be used to make CAs seem more humanlike (Schuetzler et al., 2021; Seeger et al., 2021). In particular, chitchat and character definition were emphasized in the interviews, since many users first check the CA for its social capabilities and quickly lose interest if it fails, even at slight initial social interactions. Further possibilities of humanlike design are *non-verbal cues*, such as *emoticons*, or artificially induced *typing delays and indicators*, such as typing dots (Gnewuch et al., 2018). However, researchers have also noted that a humanlike CA can be repellent to users (e.g., Grudin & Jacques, 2019). Seeger et al. (2021) indicated that the different anthropomorphism criteria must be combined and evaluated practically.

15.4.4 Dialog control

For successful *dialog control*, CAs' understanding of users' requests, along with their intentions and goals, should be evaluated (Clark et al., 2019). However, CAs are learning IS and, therefore, are initially error-prone. In particular, user input in lengthy and complex sentences poses a challenge for CAs (Michaud, 2018). Thus, proactive dialog handling in regular operations and reactive handling in failure operations should be evaluated to ensure that CAs avoid, reduce, or recover from failures. In *regular operations*, organizations should continuously reflect on whether the CA proactively avoids error scenarios by, for example, asking the user *to reformulate the request* (Diederich et al., 2020) or prompting the user for more information (Chaves & Gerosa, 2021). Further, the interviews revealed the expectation that if no appropriate answer was elicited, the CA should proactively refer to misunderstandings or reintroduce its skills. Afterward, the CA could provide *alternative responses* to keep the conversation alive (Chaves & Gerosa, 2021). Another way is to provide *conversational prompts*. Through the use of prompts, the CA provides *suggestions* for prospective requests in addition to their responses (e.g., in the case of a long response time by the user). The aim is to predict the user's intentions (e.g., by offering suggestions on text buttons) and proactively avoid error cases when processing a user's text input (Li et al., 2020). In *failure operations*, it is crucial to define and evaluate (e.g., *proactive and resilient) repair strategies* to overcome conversational breakdowns, since their existence can result in a negative experience for users and impair future CA success (Benner et al., 2021). In the case of a breakdown, the CA should fail gracefully in order to maintain user trust (Feng & Buxmann, 2020). For instance, the CA can apologize and propose new solutions (Benner et al., 2021). However, if repair attempts fail repeatedly and the CA's capabilities are exceeded, the CA should encourage *fallbacks or a handover* to a service representative (Poser et al., 2021; Poser et al., 2022b).

15.4.5 Performance

A holistic evaluation of CA *performance* represents a strong predictor of CA success (Peras, 2018). By combining design and technically-oriented principles, CAs' performance relates directly to user satisfaction (Liao et al., 2016). The performance demonstrates the effective and efficient completion of tasks between the user and the CA (Peras, 2018). Regarding CAs' *effectiveness*, the *task success rate* and the *task failure rate* could be used to collect the number of successful tasks and the number of default fallback intents to trigger appropriate countermeasures (Peras, 2018). In the interviews, the *retention and feedback* rates were mentioned regarding the recordings of returning users and continuously evaluating users' average ratings to uncover weaknesses and derive improvement potential. Furthermore, it is necessary to consider CAs' *efficiency* because the adequate performance of tasks explicates only a few insights into whether the CA also performs the tasks with a resource-friendly approach. Given this perspective, evaluating the time required to complete a task (*task completion time*) and the (average) number of rounds of dialog required (average *number of turns*) is essential to capture efficiency (Holmes et al., 2019; Peras, 2018). In addition, the *human handover rate* is significant in evaluating at which points the CA cannot complete a task (Wintersberger et al., 2020).

15.4.6 Data Privacy

Data privacy includes criteria related to the *realization and communication of data protection* endeavors. One important aspect is ensuring that conversations with the CA are kept as *private* and *anonymous* as possible, particularly when the CA deals with confidential or personal data (Feng & Buxmann, 2020). During the interviews, we received feedback emphasizing the importance of minimizing the storage of conversational data and ensuring that any stored data is anonymized to the greatest extent feasible, especially when such data is necessary to improve a CA's performance. The communication of data protection contains the criterion of *transparency* toward users, meaning the disclosure of which user data is processed. In this context, it is helpful to provide data protection policies (Rajaobelina et al., 2021).

Table 5. Final CA quality criteria set

Meta-criteria	Criteria	Sub-criteria	Example references
Input	Interaction abilities	Communication channel	(Feng & Buxmann, 2020), Interviews
		Control elements	(Kowald & Bruns, 2020; Li et al., 2020), Interviews
	Context awareness	Dialog-oriented context	(Diederich et al., 2020; Michaud, 2018; Saenz et al., 2017)
		Technical environment	Interviews
Output	Format	Visual elements	(Edirisooriya et al., 2019; Feng & Buxmann, 2020; Kowald & Bruns, 2020), Interviews
		Readability and consistency	
	Content	Transparent capabilities and limitations	(Diederich et al., 2020; Saenz et al., 2017)
		Information retrieval	(Diederich et al., 2020; Edirisooriya et al., 2019), Interviews
		Detail of knowledge	Interviews
		Solution convergence and justification	Interviews
	Calibration	Response appropriateness	(Hu et al., 2018; Jiang & Ahuja, 2020)
		Response accuracy	
	Time	Technical response time	(Edirisooriya et al., 2019; Meyer-Waarden et al., 2020), Interviews
		Balance between proactivity and interruption	(Feng & Buxmann, 2020)
Anthropo-morphism	Humanlike identity	Identity and characteristics	(Schuetzler et al., 2021; Seeger et al., 2021)
		(Humanlike) visual representation	Interviews
	Verbal cues	Emotional expressions	(Saenz et al., 2017; Seeger et al., 2021)
		Chitchat / small talk	(Grudin & Jacques, 2019; Huiyang & Min, 2022; Schuetzler et al., 2021)
		Tailored personality and lexical alignment	
	Non-verbal cues	Emoticons	(Gnewuch et al., 2018; Schuetzler et al., 2021; Seeger et al., 2021), Interviews
		Typing delay and indicator	
Dialog control	Regular operation	Reformulate requests and alternative responses	(Diederich et al., 2020; Saenz et al., 2017), Interviews
		Conversational prompts and suggestions	(Kowald & Bruns, 2020; Li et al., 2020)
	Failure operation	(Proactive & resilient) repair strategies	(Benner et al., 2021; Diederich et al., 2020; Feng & Buxmann, 2020), Interviews
		Fallbacks and handover	(Poser et al., 2021; Poser et al., 2022b; Wintersberger et al., 2020)
Performance	Effectiveness	Task success rate	(Peras, 2018), Interviews
		Task failure rate	
		Retention and feedback rate	Interviews
	Efficiency	Task completion time	(Holmes et al., 2019; Peras, 2018), Interviews
		Number of turns	
		Human handover rate	(Wintersberger et al., 2020), Interviews
Data privacy	Realization and communication	Privacy and anonymity	(Feng & Buxmann, 2020; Janssen et al., 2021b; Lewandowski et al., 2021; Rajaobelina et al., 2021), Interviews
		Transparency	

15.5 Case-based Instantiation of the Quality Criteria Set

After the research activities of the DSR project in *Steps 1* to *5*, the final quality criteria set was instantiated in *Step 6* in an IT organization to investigate, evaluate, and improve the quality of an existing AI and text-based CA. Due to the organization's limited in-depth knowledge of a systematic CA evaluation procedure, including methods, a systemized procedure model was initiated and documented. It comprises three main phases and was applied to utilize the final CA quality criteria set throughout each phase (see **Figure 3**).

15.5.1 Case setting for applying the procedure model

The DSR project considered the following case setting to apply the procedure model, evaluate CAs' quality, and address an existing real-world problem: (1) The procedure model requires a suitable use case to evaluate the applicability and feasibility to indicate a CAs' quality. To this end, an existing AI and text-based CA (*ExpertBot*) was investigated, evaluated, and improved in an IT organization. Based on our interview analysis (as outlined in *Step 1* of our DSR project), *ExpertBot* was deemed to be a suitable case for a root cause analysis, since the overall quality and usage rate were insufficient. The IT organization uses *ExpertBot* within organizational boundaries to identify, prioritize, and select needed experts. Therefore, *ExpertBot* participates in chat conversations and accesses various data sources, such as skill databases, document management systems, and internal chat forums, to provide fitting recommendations for experts and their skills. *ExpertBot* is integrated into an existing text-based communication channel in Microsoft Teams and works intent-based, using Microsoft Language Understanding (LUIS) and Azure Cognitive Services in the background. (2) Furthermore, forming an expert team with varying experience levels and backgrounds regarding CAs and their application field is crucial to provide a multi-perspective view enabling a broad discussion of the quality criteria and shortcomings of CAs. In our case, to implement the procedure model for evaluating the quality of *ExpertBot* through all phases, the existing CA team formed an interdisciplinary team of experts from the IT organization (e.g., CA developers, product owners, management assistance responsible for staffing, and employees from other departments, as outlined in *Step 6* of our DSR project). (3) Finally, an expert team requires an appropriate data basis to evaluate CAs. For this purpose, prepared data, such as the user retention rate or other criteria, can be applied. In this context, the newly formed expert team evaluated *ExpertBot* based on our quality criteria set and derived improvement potentials. Overall, this case setting served as the starting point for instantiating the CA quality criteria set through the procedure model, as shown in **Figure 3**.

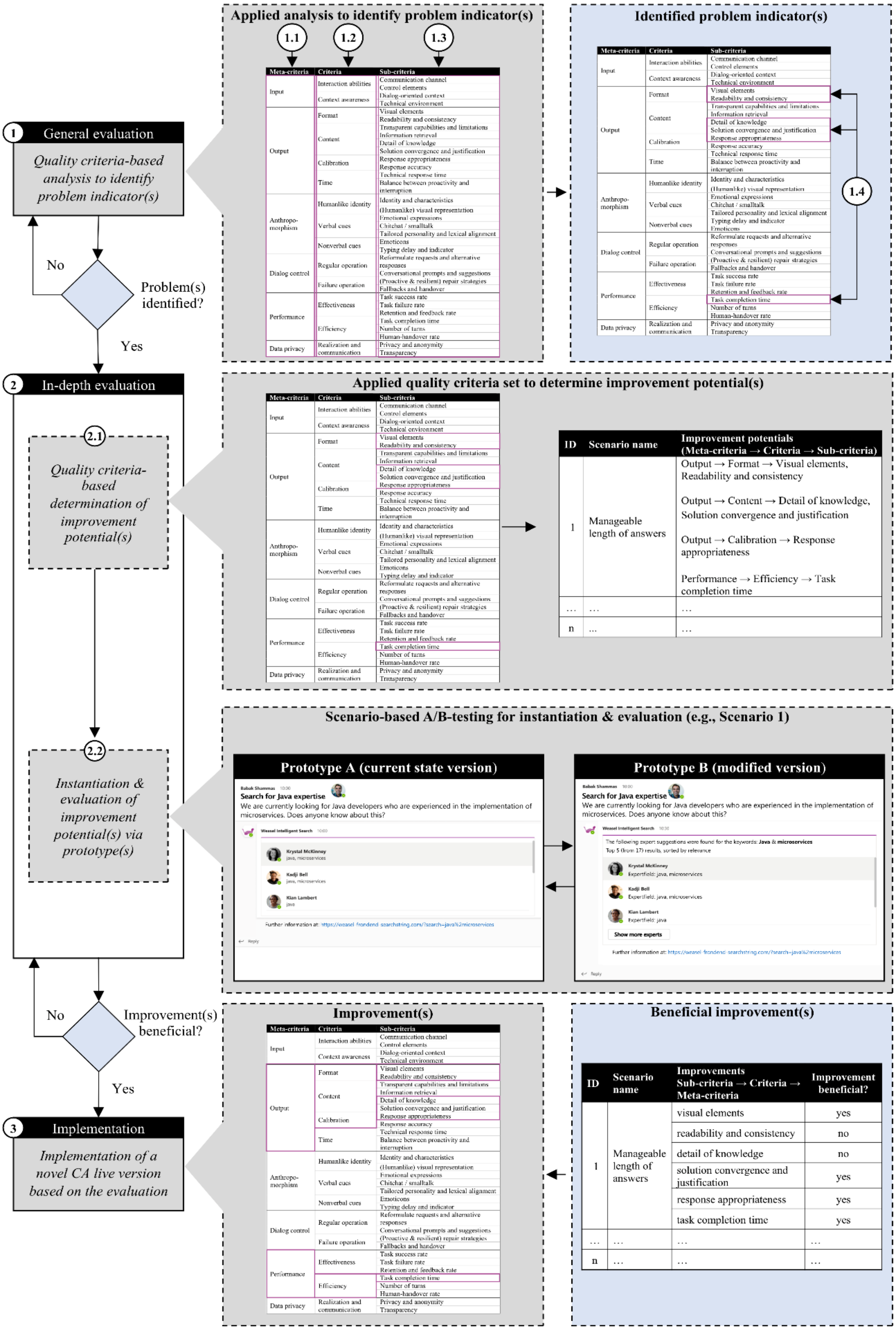

Figure 3. Procedure model for the evaluation and improvement of CAs using our quality criteria set

15.5.2 Utilization of the procedure model

Our procedure model is designed with three main phases and several sub-phases to provide a fine-grained approach that fosters comprehensiveness and traceability. The sub-phases enable us to (1) create progressive guidance for each phase of the procedure model, facilitating the evaluation of CAs; (2) ensure that every aspect of the procedure is thoroughly documented, which is crucial for properly evaluating CAs; and (3) create a more detailed and extensive procedure that helps the expert team to ensure a systematic CA evaluation.

Phase 1: General evaluation

In Phase 1 (general evaluation), we performed a quality criteria-based analysis to identify problems with the current CA version. More specifically, in Sub-phase 1.1, the derived meta-criteria were used to provide a starting point for the CA team's initial evaluation of the *ExpertBot* and to identify possible problem areas (see **Figure 3**). In Sub-phase 1.2, the corresponding criteria of these problem areas served as a more detailed level to narrow the scope of analysis. Thereby, in Sub-phase 1.3, the sub-criteria belonging to the criteria could be used as indicators of potential problems. Based on these phases and the analysis of appropriate data related to the corresponding sub-criteria, specific problem indicators of the *ExpertBot* were identified in Sub-phase 1.4.

In our illustrated example from our instantiation (see **Figure 3**), the general evaluation revealed that the overarching meta-criteria "output" and "performance" of the *ExpertBot* needed to be improved. Six problem indicators, such as "detail of knowledge," "solution convergence and justification," and "task completion time," were considered throughout the criteria-based analysis to start an in-depth evaluation. As a result, we initiated an improvement project to address the identified indicators.

Phase 2: In-depth evaluation

As part of Phase 2 (in-depth evaluation), we first conducted Sub-phase 2.1. In cooperation with the IT organization, the CAs' quality was evaluated, and the potentials for improvement were determined based on the identified problem indicators from Sub-phase 1.4. Using appropriate evaluation methods, the consideration of these improvement potentials was found to be beneficial for the expert team (comprising members from the CA team; see "case setting").

To gain this insight, we conducted seven semi-structured interviews with the expert team members. We presented the live version of the *ExpertBot* and asked the participants about the general implementation, problems, and relevance for improvement, along with the corresponding criteria from our set. The resulting evaluated improvement potentials of the *ExpertBot* were then transformed into coherent scenarios in an aggregation process. Thereby, a collective evaluation of multiple quality criteria in each scenario could be conducted. In the single scenario we outlined, as shown in Sub-phase 2.1 of **Figure 3**, all six specific problem indicators were identified as improvement potentials during the interviews. Specifically, the scenario was called "manageable length of answers" and included the improvement potentials "visual elements," "readability and consistency," "detail of knowledge," "solution convergence and justification," "response appropriateness," and "task completion time".

In Sub-phase 2.2, we created mockup prototypes for the transformed scenarios to demonstrate, investigate, and evaluate the identified improvement potentials. In this context, the prototypes enabled a well-founded comparison between the current state version of the CA and the proposed modified CA version(s). The expert team provided valuable feedback to verify whether the identified improvement potentials would be beneficial if implemented or needed to be revised or discarded.

For the creation of prototypes, we employed the Figma (2022) design tool in combination with the Microsoft Teams UI Kit (2023) to ensure a familiar and consistent visual representation during the demonstration. Furthermore, the prototypes were designed based on the previously evaluated improvement potentials corresponding to the analyzed *ExpertBot*. Subsequently, we conducted A/B tests involving six participants by presenting them with two prototypes for each scenario during semi-structured interviews to achieve a data basis for deciding whether to implement the proposed changes. One prototype contained the current CA version, while the other represented the assumed improvements (modified version, as depicted in **Figure 3**). For each scenario, questions were asked in three areas during the interviews. First, we asked participants to evaluate which of the two prototypes was more effective at first glance and which aspects were crucial to this impression (e.g., perceptions of the prototype features and differences). Second, the improvement potentials were addressed individually, and the participants were asked to determine which sub-criteria were conceivable for increasing CA quality. Third, we asked which of the addressed sub-criteria was rated the most important in improving CA quality to prioritize the highest-ranked improvement potentials (e.g., number of mentions) in preparation for the last phase.

Phase 3: Implementation

Finally, in Phase 3 (implementation), the improvement potentials, identified in Sub-phase 2.1 and evaluated as beneficial in Sub-phase 2.2 for increasing the quality of the CA, were implemented in a revised CA live version. These improvements were communicated to the users to ensure their visibility in the organization. After Phase 3, the procedure should be repeated to improve the CA on a long-term basis, for instance, if problems are identified based on existing data, or as part of a general cyclical evaluation to examine the quality of the new CA live version as a whole or in defined segments, which, however, was not part of the instantiation.

15.6 Discussion

Organizations strive to implement CAs due to their potential to increase business value with their ability to assist or automate processes, tasks, and activities (Lewandowski et al., 2021). However, despite their strengths in improving organizational efficiency (Zierau et al., 2020c), many CAs across industries are still error-prone and fail during interactions (Gnewuch et al., 2017), leading to a high discontinuation rate (Janssen et al., 2021b). To strengthen the management of CAs in an organizational context and improve their success, we have developed a quality criteria set and procedure model for conducting holistic evaluations and improvements of CAs. In a multi-step DSR project, criteria were identified, aggregated, and evaluated ex-ante for applicability and operationality in real-world environments. In addition, a procedure model for the application of the quality criteria set was determined as part of a naturalistic ex-post evaluation. The conducted evaluation activities demonstrate that the incorporated criteria provide an integrated view of a CA evaluation. Regarding the procedure model, the results indicate that a systematic analysis of the problems, requirements, and status quo of a specific CA is supported to identify and improve its most relevant aspects. In combination, these findings have implications for research and practice.

15.6.1 Theoretical implications

First, our quality criteria set and procedure model contribute to CA research by providing a synthesized and systematized approach to improving the success of contemporary CAs. To achieve this, we contributed the quality criteria set derived from strongly dispersed CA research streams (Følstad et al., 2021). A large share of this research has focused on specific design and technical issues to elevate the user experience (e.g., Seeger et al., 2021), as these issues were considered the main challenges in the implementation of CAs (Følstad et al., 2018a; Janssen et al., 2021b; van der Goot et al., 2021). However, CAs are inherently complex IS (Maroengsit et al., 2019) with distinctive characteristics that require a comprehensive view and analysis, as failures can arise from multiple (interrelated) factors (Janssen et al., 2021b; Meyer von Wolff et al., 2021). Therefore, we extended the focus of current CA research to a consolidated set of essential quality criteria that should be considered to support the prevention of CA failure. We also provided a starting point for a more structured CA evaluation with our procedure model, as recommended by Følstad et al. (2021). The quality criteria set addresses the type of AI and text-based Cas in general domains, as classified by Gnewuch et al. (2017). Nevertheless, the results may also apply to other types of CAs. Additionally, our work complements other preliminary efforts, such as the evaluation criteria sets of Radziwill and Benton (2017), and Casas et al. (2020), to provide a better understanding of CAs in the improvement process with a system-wide view.

Recent technical advancements in the field of NLP and ML applications should also be highlighted, especially the emergence of large language models (LLMs). These models are pre-trained on billions of text samples from specific data sources on the Internet and can generate diverse types of content (Brown et al., 2020; Jiang et al., 2022). In particular, these models become widely available to (non-technical) users via the release of intuitive and conversational interfaces, such as OpenAI's ChatGPT or Google's Bard (Jiang et al., 2022; Teubner et al., 2023). These releases implicate a remarkable movement in CA research exploring new application scenarios and their potential, which is also referred to as a new "AI wave" by Schöbel et al. (2023). Consequently, the question arises of which quality criteria are affected in this new wave and what requirements result for CAs and their (further) development. We, therefore, expect a paradigm shift in the perception and utilization of the different criteria from our set in regard to the novel LLM applications, which could lead to high dynamics and flexibility in their adaptation and use.

Second, this article contributes to management research on CAs, which encompasses various aspects of the CA lifecycle, such as critical phases, factors, or tasks within CA development (Lewandowski et al., 2022b; Meyer von Wolff et al., 2022). However, these initial studies neither provide deeper insights into CAs' evaluation and improvement nor explain how a cyclical evaluation process can be executed. In this regard, our research provides an approach for evaluating and improving CAs, which can serve as a meta-model for other researchers using different qualitative and quantitative methods within the lifecycle of a CA. While researchers focus on the design of CAs by targeting specific aspects, such as increasing user trust or anthropomorphism (e.g., Seeger et al., 2017), they often disregard the importance of evaluating CAs on an ongoing basis, as elaborated in the "Related Research" section. Thereby, we provide knowledge regarding a structured and continuous CA evaluation to ensure the improvement of CAs during their operation in organizations (Janssen et al., 2021b; Meyer von Wolff et al., 2021). In addition, the quality criteria set and procedure can assist in other lifecycle phases, for instance, by providing an overview of initial design issues in the initiation phase. Furthermore, the quality criteria set can support a multi-perspective and comprehensive development process and the detection of problems before going live in the integration phase to avoid direct failure. Our article aggregates design knowledge, supplemented by practical insights, and introduces a structured approach that provides initial insights into activities, people, and data, which can foster operations and enhance the performance of CAs.

Third, from the DSR lens, we contribute prescriptive design knowledge with the quality criteria set and procedure model for their application. Both form our developed artifact. This artifact provides a foundation that can be applied in the identified higher-level problem class in other solution spaces (Hevner et al., 2004). The application of the artifact can be utilized to address and explore the problem class in more depth and further improve the ability to apply it in a generalized manner or to design more sophisticated artifacts as tools for similar problems.

15.6.2 Practical implications

In addition to its theoretical contributions, our article has practical implications for organizations. By providing a systematic approach to evaluation and improvement, our artifact can guide CA teams in various ways to support the successful development, operation, and evolution of CAs. First, the combination of the quality criteria set and procedure model allows practitioners to obtain a comprehensive overview of relevant criteria and to narrow down the evaluation of their CA to identify specific problems and improve the overall quality of CAs. Second, the procedure model can serve as a blueprint for CA teams to systematize the evaluation process. The delineation of content and the sequence of relevant steps provides a feasible approach for practitioners to structure their evaluation and improvement activities of existing CAs. In addition, the criteria set can serve as a basis for CA teams to decide whether a CA project should be established and whether requirements are present (e.g., prepared data, an interdisciplinary team) to enable a comprehensive and multi-perspective evaluation of the quality of CAs. Thereby, the execution of evaluation and improvement tasks could be accelerated. Apart from the description of relevant criteria and the evaluation steps, the artifact's application may positively affect organizations. For instance, following the systematized procedure to improve CAs, the perceived user satisfaction could increase, thus resulting in an improved acceptance and usage rate and consequently counteracting the discontinuation rate of CAs. In addition, the evolution of CAs and related positive effects could not be limited to the CA domain, as their success could foster the overall AI transformation of an organization, so that the increased quality and use of CAs can influence other learning and AI-based IS.

15.6.3 Limitations and future research

Our research is not without limitations that have implications for further research. The developed artifact comprises a comprehensive set of quality criteria. However, its application does not inevitably guarantee success in the deployment and continuous improvement of CAs. To achieve this broad goal, additional aspects, such as technical requirements (e.g., AI, ML, and NLP algorithms and tools), a fit of the technology to the use case (e.g., using a CA for complex tasks or in emotionally-sensitive environments), design (e.g., human-computer interface), and organizational communication to users (e.g., tutorials, highlighting benefits and restrictions) have to be considered. All factors in interaction lay the foundation for a successful CA operation. In pursuit of this goal, the quality criteria set and procedure model can be considered one piece of the greater puzzle.

The instantiation of the quality criteria set revealed several challenges and aspects that need further research. First, in Phase 1 of our instantiation, we determined the need for a quality criteria-based, in-depth evaluation of CAs' output and performance. These overarching meta-criteria proved to be valid starting points for exploring the improvement potentials of the *ExpertBot*. Nevertheless, further investigation is required to identify additional triggers that warrant in-depth evaluation. Broadening the perspective, triggers from outside the organizational boundaries, such as feedback from customers, are possible. However, we did not identify any of these triggers in our project due to the inward-facing use case of the *ExpertBot*.

Second, further research on how organizations can generate a CA evaluation strategy, including aspects such as evaluation intervals, criteria selection, and suitable evaluation methods, is needed. In our real-world instantiation, we have resorted to semi-structured interviews and A/B testing as qualitative evaluation methods, which are not necessarily suitable for all criteria. Overall, a general framework could assist organizations and researchers in the selection of suitable evaluation and data analysis methods for relevant areas (such as our meta-criteria). In particular, longitudinal studies that explore application of the quality criteria set and procedure model in real-world environments can provide deeper insights into evaluation strategies and the impact of their use.

Third, we observed that different quality criteria of our set have varying levels of impact on CAs' quality. The criteria exhibit an indeterminate degree of interdependence as they influence each other. In addition, we found that the skills of the participants (e.g., CA team) can influence this factor. We have derived three directions for further research: (1) Conception, design, and evaluation of a modular, context-adaptive procedure model that can be tailored to arbitrary CA application environments and their essential quality criteria, including the individual conditions; (2) Investigations to determine the needs of AI and data literacy experts for designing, continuously evaluating, and improving CAs, as well as the needs of (non-expert) users for utilizing and validating the information output to counteract their failure; (3) Identification of criteria in our set that may need to be evaluated more or less frequently. A combined classification or ranking of the influence and importance of the criteria (e.g., by an empirical research approach) offers additional potential.

Moreover, we expect further technical progress and research in the context of customizable CAs, as also described in a study by Schöbel et al. (2023). Aspects such as social presence and anthropomorphism, as well as personalization and empathy of human-AI interactions, are to be considered. Especially the new wave of AI technologies and LLM could lead to improvements in terms of better customization and contextualization. By exploring these areas, further research could counteract the skepticism of users toward conventional CAs, perceiving them as unnatural, impersonal, or deceptive (Schöbel et al., 2023), and reduce the overall failure of CAs (Gnewuch et al., 2017). In this regard, our set of criteria contains anthropomorphism criteria, such as identity, visual representation, and tailored personalization. However, these criteria need to be further explored in the wake of the recent customization and contextualization capabilities of LLMs that could make CAs more adaptive to users' emotional states, for example, by tailoring responses to individual needs and preferences, fostering a wider acceptance in the future.

In addition to information retrieval scenarios, CAs augmented with LLM capabilities could act in a broader spectrum of possible use cases. In conjunction with our demonstrated procedure model, new approaches for evaluating and improving CAs extended by LLMs may become mandatory. For example, generative activities (e.g., content created based on statistical methods and available data) should be handled differently from information retrieval activities (e.g., content extracted unchanged from a connected data source). The question whether the generated content corresponds to the truth or contains false and misleading facts arises. The range of application scenarios in practice and the exploration of these technologies are in their infancy and will be an engaging field of research (Schöbel et al., 2023).

While our article provides insights and an approach to the evaluation and improvement of CAs, it has methodological limitations. Although our artifact proved to be applicable by instantiating it in an IT organization, its transferability to other application environments with CAs of different use cases, or other CA teams and conditions remain to be proven to further address the overarching problem class. In this vein, our set of quality criteria could be a building block for adaptation. Overall, several foundations are laid for research on the design, validation, and adaptation of the quality criteria set and procedure model.

15.7 Conclusion

CAs have become increasingly relevant in facilitating convenient access to information and services, representing essential gateways for organizations to interact with customers or employees (Følstad et al., 2021). However, due to their frequent premature deployment and varying maturity levels, CAs can be error-prone and fail to meet the requirements of their intended use cases, ultimately leading to their abandonment. To address this challenge, we conducted a DSR project that demonstrates how organizations can leverage a systematized procedure model based on criteria-based analysis to foster continuous evaluation and improvement of CAs. Our article provides guidance for organizations to better understand and evaluate the quality of their CAs, thereby laying the foundation for their long-term success. As a result, this article expands the knowledge base on CAs and emphasizes that evaluating and improving them is an ongoing challenge due to their complex and unique nature. Additional research is needed to further explore how organizations can conduct criteria-based evaluations of their CAs and develop effective evaluation strategies.

15.8 References

Alonso, A. F., Fuertes Castro, J. L., Martínez Normand, L., & Soza, H. (2009). Towards a set of measures for evaluating Software agent autonomy. Mexican International Conference on Artificial Intelligence (MICAI), Guanajuato, México. https://doi.org/10.1109/MICAI.2009.15

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. Conference on human factors in computing systems (CHI), Glasgow, Scotland, UK. https://doi.org/10.1145/3290605.3300484

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A review on recovery strategies to overcome conversational breakdowns of conversational agents. International Conference on Information Systems (ICIS), Austin, TX, United States.

Bisio, I., Garibotto, C., Grattarola, A., Lavagetto, F., & Sciarrone, A. (2018). Smart and robust speaker recognition for context-aware in-vehicle applications. IEEE Transactions on Vehicular Technology, 67(9), 8808-8821. https://doi.org/10.1109/TVT.2018.2849577

Bittner, E. A. C., Oeste-Reiß, S., & Leimeister, J. M. (2019). Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. International Conference on Internet Science (INSCI), Cham.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. interactions, 25(5), 38-43.

Brink, A. (2013). Anfertigung wissenschaftlicher Arbeiten: Ein prozessorientierter Leitfaden zur Erstellung von Bachelor-, Master-und Diplomarbeiten (5th ed.). Springer Gabler, Wiesbaden.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. Conference on Neural Information Processing Systems (NeurIPS), Vancouver, Canada.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company, New York.

Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Valdez, A. C. (2019). "Hey, Siri","Ok, Google","Alexa". Acceptance-Relevant Factors of Virtual Voice-Assistants. International Conference on Professional Communication (IPCC), Aachen, Germany. https://doi.org/10.1109/ProComm.2019.00025

Casas, J., Tricot, M.-O., Abou Khaled, O., Mugellini, E., & Cudré-Mauroux, P. (2020). Trends & methods in chatbot evaluation. International Conference on Multimodal Interaction (ICMI), Virtual event, Netherlands. https://doi.org/10.1145/3395035.3425319

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. International Journal of Human–Computer Interaction, 37(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., & Munteanu, C. (2019). What makes a good conversation? Challenges in designing truly conversational agents. Conference on Human Factors in Computing Systems (CHI), New York, NY, United States. https://doi.org/10.1145/3290605.3300705

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. Knowledge in society, 1(1), 104–126.

Corea, C., Delfmann, P., & Nagel, S. (2020). Towards intelligent chatbots for customer care-practice-based requirements for a research agenda. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A customer service chatbot for e-commerce websites. Meeting of the Association for Computational Linguistics-System Demonstrations, Vancouver, Canada. https://doi.org/10.18653/v1/P17-4017

Dale, R. (2016). The return of the chatbots. Natural Language Engineering, 22(5), 811-817. https://www.doi.org/10.1017/S1351324916000243

Davenport, T. H., & Kirby, J. (2016). Just how smart are smart machines? MIT Sloan Management Review, 57(3), 21-25.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019a). On conversational agents in information systems research: Analyzing the past to guide future work. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019b). Towards a taxonomy of platforms for conversational agent design. International Conference on Wirtschaftsinformatik (WI), Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. Business & Information Systems Engineering, 62(3), 193–209. https://doi.org/10.1007/s12599-020-00639-y

Drechsler, A., & Hevner, A. R. (2018). Utilizing, producing, and contributing design knowledge in DSR projects. International Conference on Design Science Research in Information Systems and Technology (DESRIST), Chennai, India. https://doi.org/10.1007/978-3-319-91800-6_6

Edirisooriya, M., Mahakalanda, I., & Yapa, T. (2019). Generalized framework for automated conversational agent design via QFD. Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka. https://doi.org/10.1109/MERCon.2019.8818945

Elshan, E., Engel, C., Ebel, P., & Siemon, D. (2022a). Assessing the reusability of design principles in the realm of conversational agents. International Conference on Design Science Research in Information Systems and Technology (DESRIST), St. Petersburg, FL, USA. https://doi.org/10.1007/978-3-031-06516-3_10

Elshan, E., Siemon, D., De Vreede, T., De Vreede, G.-J., Oeste-Reiß, S., & Ebel, P. (2022b). Requirements for AI-based Teammates: A Qualitative Inquiry in the Context of Creative Workshops. Hawaii Conference on System Sciences (HICSS), Hawaii, HI, USA.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019a). A taxonomy of social cues for conversational agents. International Journal of Human-Computer Studies, 132, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Feine, J., Morana, S., & Maedche, A. (2019b). Designing a chatbot social cue configuration system. International Conference on Information Systems (ICIS), Munich, Germany.

Felderer, M., & Ramler, R. (2021). Quality assurance for AI-based systems: Overview and challenges (Introduction to interactive session). Software Quality: Future Perspectives on Software Engineering Quality (SWQD), Springer, Cham. https://doi.org/10.1007/978-3-030-65854-0_3

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Figma. (2022). Retrieved May 31, 2022, from https://www.figma.com

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., & Ischen, C. (2021). Future directions for chatbot research: An interdisciplinary research agenda. Computing, 103(12), 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018a). What makes users trust a chatbot for customer service? An exploratory interview study. International Conference on Internet Science (INSCI), Springer, Cham. https://doi.org/10.1007/978-3-030-01437-7_16

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2018b). Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. International Conference on Internet Science (INSCI), Springer, Cham. https://doi.org/10.1007/978-3-030-17705-8_13

Gläser, J., & Laudel, G. (2009). Experteninterviews und qualitative Inhaltsanalyse: Als Instrumente rekonstruierender Untersuchungen (3rd ed.). VS Verlag für Sozialwissenschaften, Wiesbaden.

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. European Conference on Information Systems (ECIS), Portsmouth, United Kingdom.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. International Conference on Information Systems (ICIS), Seoul, Korea.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in human behavior, 97, 304-316. https://doi.org/10.1016/j.chb.2019.01.020

Goh, O. S., Ardil, C., Wong, W., & Fung, C. C. (2007). A black-box approach for response quality evaluation of conversational agent systems. International Journal of Computational Intelligence, 3(3), 195-203.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS quarterly, 37(2), 337-355. https://doi.org/10.25300/misq/2013/37.2.01

Gregory, R. W., & Muntermann, J. (2014). Research note—Heuristic theorizing: Proactively generating design theories. Information Systems Research, 25(3), 639-653. https://doi.org/10.1287/isre.2014.0533

Grudin, J., & Jacques, R. (2019). Chatbots, humbots, and the quest for artificial general intelligence. Conference on Human Factors in Computing Systems (CHI), Glasgow, Scotland. https://doi.org/10.1145/3290605.3300439

Herrera, A., Yaguachi, L., & Piedra, N. (2019). Building conversational interface for customer support applied to open campus an open online course provider. International Conference on Advanced Learning Technologies (ICALT), Maceio, Brazil. https://doi.org/10.1109/ICALT.2019.00011

Hevner, A. R. (2007). A three cycle view of design science research. Scandinavian journal of information systems, 19(2), 4.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in Information Systems research. Management Information Systems Quarterly (MISQ), 28(1), 75-105.

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? European Conference on Cognitive Ergonomics (ECCE), Belfast, UK. https://doi.org/10.1145/3335082.3335094

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. Conference on Human Factors in Computing Systems (CHI), Montréal, Canada. http://dx.doi.org/10.1145/3173574.3173989

Huiyang, S., & Min, W. (2022). Improving Interaction Experience through Lexical Convergence: The Prosocial Effect of Lexical Alignment in Human-Human and Human-Computer Interactions. International Journal of Human–Computer Interaction, 38(1), 28-41.

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. Chatbot Research and Design. CONVERSATIONS 2019, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-030-39540-7_3

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. Designing Interactive Systems Conference (DIS), New York, NY, USA. https://doi.org/10.1145/3196709.3196735

Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do chatbots fail? A critical success factors analysis. International Conference on Information Systems (ICIS), Austin, TX, United States.

Janssen, A., Passlick, J., Cardona, D. R., & Breitner, M. H. (2020). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. Business & Information Systems Engineering, 62(3), 211–225. https://doi.org/10.1007/s12599-020-00644-1

Jiang, E., Olson, K., Toh, E., Molina, A., Donsbach, A., Terry, M., & Cai, C. J. (2022). PromptMaker: Prompt-based prototyping with large language models. Conference on Human Factors in Computing Systems (CHI), New Orleans, LA, United States. https://doi.org/10.1145/3491101.3503564

Jiang, J., & Ahuja, N. (2020). Response quality in Human-Chatbot Collaborative Systems. Conference on Research and Development in Information Retrieval (SIGIR), Virtual Event, China. https://doi.org/10.1145/3397271.3401234

Johari, N. M., & Nohuddin, P. (2021). Quality attributes for a good chatbot: a literature review. International Journal of Electrical Engineering and Technology (IJEET), 12(7), 109-119. https://doi.org/10.34218/IJEET.12.7.2021.012

Jonke, A. W., & Volkwein, J. B. (2018). From Tweet to Chatbot–Content Management as a Core Competency for the Digital Evolution. Digital Marketplaces Unleashed, 275-285. https://doi.org/10.1007/978-3-662-49275-8_28

Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: a services marketing research agenda. Journal of Services Marketing, 34(3), 389-398. https://doi.org/10.1108/jsm-01-2019-0043

Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The Rise of Bots. Conference on Designing Interactive Systems (DIS), Edinburgh, United Kingdom. https://doi.org/10.1145/3064663.3064672

Kowald, C., & Bruns, B. (2020). Chatbot Kim: A digital tutor on AI. How advanced dialog design creates better conversational learning experiences. International Journal of Advanced Corporate Learning (iJAC), 13(3), 26-34. https://doi.org/10.3991/ijac.v13i3.17017

Kuligowska, K. (2015). Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. Professionals Center for Business Research, 2.

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2019). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. International Workshop on Chatbot Research and Design, Springer, Cham. https://doi.org/10.1007/978-3-030-39540-7_13

Laban, G., & Araujo, T. (2020). The effect of personalization techniques in users' perceptions of conversational recommender systems. ACM International Conference on Intelligent Virtual Agents (IVA), New York, NY, USA. https://doi.org/10.1145/3383652.3423890

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies, 103, 95-105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-art analysis of adopting AI-based conversational agents in organizations: A systematic literature review. Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

Lewandowski, T., Grotherr, C., & Böhmann, T. (2022a). Managing Artificial Intelligence Systems for Value Co-creation: The Case of Conversational Agents and Natural Language Assistants. In B. Edvardsson & B. Tronvoll (Eds.), The Palgrave Handbook of Service Management (pp. 945-966). Springer International Publishing. https://doi.org/10.1007/978-3-030-91828-6_45

Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022b). Design knowledge for the lifecycle management of conversational agents. International Conference on Wirtschaftsinformatik (WI), Nürnberg, Germany.

Li, C.-H., Yeh, S.-F., Chang, T.-J., Tsai, M.-H., Chen, K., & Chang, Y.-J. (2020). A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. Conference on Human Factors in Computing Systems (CHI), New York, NY, USA. https://doi.org/10.1145/3313831.3376209

Li, M., & Suh, A. (2022). Anthropomorphism in AI-enabled technology: A literature review. Electronic Markets, 32, 2245–2275. https://doi.org/10.1007/s12525-022-00591-7

Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What can you do? Studying social-agent orientation and agent proactive interactions with an agent for employees. ACM Conference on Designing Interactive Systems (DIS), Brisbane, Australia. https://doi.org/10.1145/2901790.2901842

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. Business & Information Systems Engineering, 61(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. International conference on information and education technology, Aizu-Wakamatsu, Japan. https://doi.org/10.1145/3323771.3323824

McTear, M. (2021). Introducing Dialogue Systems. In Conversational AI. Synthesis Lectures on Human Language Technologies (pp. 11-42). Springer, Cham. https://doi.org/10.1007/978-3-031-02176-3_1

McTear, M. F., Callejas, Z., & Griol, D. (2016). The conversational interface (Vol. 6). Springer, Cham, https://doi.org/10.1007/978-3-319-32967-3

Meira, M. d. O., & Canuto, A. d. P. (2015). Evaluation of Emotional Agents' Architectures: an Approach Based on Quality Metrics and the Influence of Emotions on Users. World Congress on Engineering (WCE), London, U.K.

Meuser, M., & Nagel, U. (2009). The expert interview and changes in knowledge production. Interviewing experts, 17-42.

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), 3. https://doi.org/10.17705/1pais.12203

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2021). Sorry, I can’t understand you! – Influencing factors and challenges of chatbots at digital workplaces. International Conference on Wirtschaftsinformatik (WI), Essen, Germany.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2022). Chatbot Introduction and Operation in Enterprises–A Design Science Research-based Structured Procedure Model for Chatbot Projects. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Meyer-Waarden, L., Pavone, G., Poocharoentou, T., Prayatsup, P., Ratinaud, M., Tison, A., & Torné, S. (2020). How service quality influences customer acceptance and usage of chatbots. Journal of Service Management Research, 4(1), 35–51. https://doi.org/10.15358/2511-8676-2020-1-35

Michaud, L. N. (2018). Observations of a new chatbot: Drawing conclusions from early interactions with users. IT Professional, 20(5), 40–47. https://doi.org/10.1109/MITP.2018.053891336

Microsoft Teams UI Kit. (2023). Retrieved 28.03.2023 from https://www.figma.com/community/file/916836509871353159/Microsoft-Teams-UI-Kit

Neff, G., & Nagy, P. (2016). Automation, Algorithms, and Politics| Talking to Bots: Symbiotic Agency and the Case of Tay. International Journal of Communication, 10(17), 4915–4931.

Patel, S., Chiu, Y.-T., Khan, M. S., Bernard, J.-G., & Ekandjo, T. A. (2021). Conversational Agents in Organisations: Strategic Applications and Implementation Considerations. Journal of Global Information Management (JGIM), 29(6), 1-25. http://doi.org/10.4018/JGIM.20211101.oa53

Pauletto, S., Balentine, B., Pidcock, C., Jones, K., Bottaci, L., Aretoulaki, M., Wells, J., Mundy, D. P., & Balentine, J. (2013). Exploring expressivity and emotion with artificial voice and speech technologies. Logopedics Phoniatrics Vocology, 38(3), 115-125. https://doi.org/10.3109/14015439.2013.810303

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77. https://doi.org/10.2753/MIS0742-1222240302

Peras, D. (2018). Chatbot evaluation metrics. Economic and Social Development: Book of Proceedings, 89–97.

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. International Journal of Contemporary Hospitality Management, 32(10). https://doi.org/10.1108/IJCHM-04-2020-0259

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Poser, M., Wiethof, C., & Bittner, E. A. (2022). Integration of AI into customer service: A taxonomy to inform design decisions. European Conference on Information Systems (ECIS), Timisoara, Romania.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint. https://doi.org/10.48550/arXiv.1704.04579

Rajaobelina, L., Prom Tep, S., Arcand, M., & Ricard, L. (2021). Creepiness: Its antecedents and impact on loyalty when interacting with a chatbot. Psychology & Marketing, 38(12), 2339–2356. https://doi.org/10.1002/mar.21548

Research and Markets. (2022). Chatbot Market. https://www.researchandmarkets.com/reports/5648691/chatbot-market-by-product-type-by-application?utm_source=BW&utm_medium=PressRelease&utm_code=f7b65b&utm_campaign=1751095+-+The+Worldwide+Chatbot+Industry+is+Expected+to+Reach+%2422.9+Billion+by+2030&utm_exec=jamu273prd#

Rimol, M. (2022). Gartner Predicts Conversational AI Will Reduce Contact Center Agent Labor Costs by $80 Billion in 2026. Gartner. Retrieved March 21, 2023 from https://www.gartner.com/en/newsroom/press-releases/2022-08-31-gartner-predicts-conversational-ai-will-reduce-contac

Riquel, J., Brendel, A. B., Hildebrandt, F., Greve, M., & Kolbe, L. M. (2021). "Even the Wisest Machine Makes Errors" – An Experimental Investigation of Human-like Designed and Flawed Conversational Agents. International Conference on Information Systems (ICIS), Austin, TX, United States.

Rowley, J., & Slack, F. (2004). Conducting a literature review. Management research news, 27(6), 31-39. https://doi.org/10.1108/01409170410784185

Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and Ethical Considerations. AICS.

Saenz, J., Burgess, W., Gustitis, E., Mena, A., & Sasangohar, F. (2017). The Usability Analysis of Chatbot Technologies for Internal Personnel Communications. IIE Annual Conference, Norcross.

Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a contemporary frontier of interaction design. European Conference on Information Systems (ECIS), Virtual Conference.

Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A., & Leimeister, J. M. (2023). Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. Information Systems Frontiers, 1-26. https://doi.org/10.1007/s10796-023-10375-9

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding whether and how to deploy chatbots. MIS Quarterly Executive, 20(1), 4. https://doi.org/10.17705/2msqe.00039

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2017). When do we need a human? Anthropomorphic design and trustworthiness of conversational agents. Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS Seoul, Korea.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. Journal of the Association for Information systems, 22(4), 8. https://doi.org/10.17705/1jais.00685

Seiffer, A., Gnewuch, U., & Maedche, A. (2021). Understanding employee responses to software robots: a systematic literature review. International Conference on Information Systems, Austin, US.

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. Computers in human behavior, 117, 106627. https://doi.org/10.1016/j.chb.2020.106627

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2019). How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electronic Markets, 30 (2), 369–403. https://doi.org/10.1007/s12525-019-00359-6

Su, Y.-S., & Hsia, J.-H. (2022). An Evaluation Model of Smart Speaker Design. In The Routledge Companion to Technology Management (pp. 141-156). Routledge.

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of chatgpt et al. the prospects of large language models. Business & Information Systems Engineering, 65(2), 95-101. https://doi.org/10.1007/s12599-023-00795-x

van der Goot, M. J., Hafkamp, L., & Dankfort, Z. (2021). Customer service chatbots: A qualitative interview study into customers' communication journey. International Workshop on Chatbot Research, CONVERSATIONS 2020, Amsterdam, the Netherlands. https://doi.org/10.1007/978-3-030-68288-0_13

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. European Journal of Information Systems, 25(1), 77-89. https://doi.org/10.1057/ejis.2014.36

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. European Conference on Information Systems (ECIS), Verona, Italy.

Wambsganss, T., Höch, A., Zierau, N., & Söllner, M. (2021). Ethical design of conversational agents: towards principles for a value-sensitive design. International Conference on Wirtschaftsinformatik (WI), Essen, Germany.

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2020). Effects of voice-based AI in customer service: Evidence from a natural experiment. International Conference on Information Systems (ICIS), India.

Weiler, S., Matt, C., & Hess, T. (2022). Immunizing with information–Inoculation messages against conversational agents' response failures. Electronic Markets, 32(1), 239-258. https://doi.org/10.1007/s12525-021-00509-9

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

Wiethof, C., & Bittner, E. (2021). Hybrid intelligence-combining the human in the loop with the computer in the loop: A systematic literature review. International Conference on Information Systems (ICIS), Austin, TX, United States.

Wintersberger, P., Klotz, T., & Riener, A. (2020). Tell me more: Transparency and time-fillers to optimize chatbots' waiting time experience. Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, Tallinn, Estonia.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. Journal of planning education and research, 39(1), 93-112. https://doi.org/10.1177/0739456X17723971

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. Conference on Human Factors in Computing Systems (CHI), New York, NY, United States. https://doi.org/10.1145/3025453.3025496

Young, S. W. (2014). Improving library user experience with A/B testing: Principles and process. Weave: Journal of Library User Experience, 1(1). https://doi.org/10.3998/weave.12535642.0001.101

Zarouali, B., Van den Broeck, E., Walrave, M., & Poels, K. (2018). Predicting consumer responses to a chatbot on Facebook. Cyberpsychology, Behavior, and Social Networking, 21(8), 491-497. https://doi.org/10.1089/CYBER.2017.0518

Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020a). A review of the empirical literature on conversational agents and future research directions. International Conference on Information Systems (ICIS), India.

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020b). Towards developing trust-supporting design features for AI-based chatbots in customer service. International Conference on Information Systems (ICIS), India.

Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., & Leimeister, J. M. (2020c). The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. International Conference on Information Systems (ICIS), India.

16 References

Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020). Challenges in chatbot development: A study of stack overflow posts. *International Conference on Mining Software Repositories (MSR)*, New York, NY, United States. https://doi.org/10.1145/3379597.3387472

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *Artificial Intelligence Applications and Innovations (AIAI)*, Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31

Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J. (2009). *Methodik der empirischen Forschung* (Vol. 3). Springer, Wiesbaden, Germany. https://doi.org/10.1007/978-3-322-96406-9

Alonso, A. F., Fuertes Castro, J. L., Martínez Normand, L., & Soza, H. (2009). Towards a set of measures for evaluating Software agent autonomy. *Mexican International Conference on Artificial Intelligence (MICAI)*, Guanajuato, México. https://doi.org/10.1109/MICAI.2009.15

Alsheibani, S. A., Cheung, Y., & Messom, C. (2019a). Factors inhibiting the adoption of artificial intelligence at organizational-level: A preliminary investigation. *Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Alsheibani, S. A., Cheung, Y., & Messom, C. (2019b). Towards an artificial intelligence maturity model: From science fiction to business facts. *Pacific Asia Conference on Information Systems (PACIS)*, X'ian, China.

Alsheibani, S. A., Messom, C., Cheung, Y., & Alhosni, M. (2020). Reimagining the strategic management of artificial intelligence: Five recommendations for business leaders. *Americas Conference on Information Systems (AMCIS)*, Salt Lake City, UT, United States.

Alter, S. (2001). Which Life Cycle--Work System, Information System, or Software? *Communications of the Association for Information Systems*, *7*(1), 17. https://doi.org/10.17705/1CAIS.00717

Alter, S. (2004). Desperately seeking systems thinking in the information systems discipline. *International Conference on Information Systems (ICIS)*, Washington DC, United States.

Alter, S. (2013). Work system theory: Overview of core concepts, extensions, and challenges for the future. *Journal of the Association for Information Systems*, *14*(2), 72-121. https://doi.org/10.17705/1jais.00323

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Conference on Human Factors in Computing Systems (CHI)*, Glasgow, Scotland, United Kingdom. https://doi.org/10.1145/3290605.3300484

Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, *45*(1), 315-341. https://doi.org/10.25300/MISQ/2021/15882

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114-123.

Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. (2015). Service innovation in the digital age: Key contributions and future directions. *MIS Quarterly*, *39*(1), 135-154. https://doi.org/10.25300/MISQ/2015/39:1.03

Baskerville, R. (2008). What design science is not. *European Journal of Information Systems*, *17*(5), 441-443. https://doi.org/10.1057/ejis.2008.45

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), 358-376. http://doi.org/10.17705/1jais.00495

Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., Vanzin, M., Junior, S. C., Silva, E., & Moreira, C. (2020). Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review*, *36*. https://doi.org/10.1016/j.cosrev.2020.100239

Bawack, R. E., Fosso Wamba, S., & Carillo, K. D. A. (2019). Artificial intelligence in practice: Implications for IS research. *Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, *19*(4). http://dx.doi.org/10.2139/ssrn.3741983

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, *22*(2), 10. https://doi.org/10.17705/1jais.00662

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A review on recovery strategies to overcome conversational breakdowns of conversational agents. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433-1450. https://doi.org/10.25300/MISQ/2021/16274

Berg, M. (2013). Natürlichsprachlichkeit in Dialogsystemen: Stand der Technik in industrie und Forschung. *Informatik-Spektrum*, *36*(4), 371-381. https://doi.org/10.1007/s00287-012-0650-3

Berg, M. (2015). NADIA: A simplified approach towards the development of natural dialogue systems. *International Conference on Applications of Natural Language to Information Systems*, https://doi.org/10.1007/978-3-319-19581-0_12

Bisio, I., Garibotto, C., Grattarola, A., Lavagetto, F., & Sciarrone, A. (2018). Smart and robust speaker recognition for context-aware in-vehicle applications. *IEEE Transactions on Vehicular Technology*, *67*(9), 8808-8821. https://doi.org/10.1109/TVT.2018.2849577

Bitner, M. J., & Wang, H. S. (2014). Service encounters in service marketing research. In *Handbook of service marketing research* (pp. 221-243). Edward Elgar Publishing. http://hdl.handle.net/10722/215942

Bittner, E. A. C., Oeste-Reiß, S., & Leimeister, J. M. (2019). Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Bock, A., Kaczmarek, M., Overbeek, S., & Heß, M. (2014). A comparative analysis of selected enterprise modeling approaches. *The Practice of Enterprise Modeling (PoEM)*, https://doi.org/10.1007/978-3-662-45501-2_11

Bock, D. E., Wolter, J. S., & Ferrell, O. C. (2020). Artificial intelligence: Disrupting what we know about services. *Journal of Services Marketing*, *34*(3), 317-334. https://doi.org/10.1108/JSM-01-2019-0047

Böhmann, T., Bittner, E., Biemann, C., Lewandowski, T., Poser, M., Heuer, M., Uba, C., Banerjee, D., Blum, S., Goertz, F., Pereira, M., Pietrowski, R. d. v., & Klein, H. (2023). *Das 3x3 der Chatbots*. https://www.inf.uni-hamburg.de/de/inst/ab/itmc/material/instant

Böhmann, T., Leimeister, J. M., & Möslein, K. (2014). Service Systems Engineering. *Business & Information Systems Engineering*, *6*(2), 73-79. https://doi.org/10.1007/s12599-014-0314-8

Bohus, D., & Rudnicky, A. I. (2009). The RavenClaw dialog management framework: Architecture and systems. *Computer Speech & Language*, *23*(3), 332-361. https://doi.org/10.1016/j.csl.2008.10.001

Bouguelia, S., Brabra, H., Zamanirad, S., Benatallah, B., Baez, M., & Kheddouci, H. (2021). Reusable abstractions and patterns for recognising compositional conversational flows. *International Conference on Advanced Information Systems Engineering (CAiSE)*, Melbourne, Australia. https://doi.org/10.1007/978-3-030-79382-1_10

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. *International Conference on Internet Science (INSCI)*, Cham.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions*, *25*(5), 38-43.

Breidbach, C., Brodie, R., & Hollebeek, L. (2014). Beyond virtuality: from engagement platforms to engagement ecosystems. *Managing Service Quality*, *24*(6), 592-611. https://doi.org/10.1108/MSQ-08-2013-0158

Brendel, A. B., Greve, M., Diederich, S., Bührke, J., & Kolbe, L. M. (2020a). You are an idiot! – How conversational agent communication patterns influence frustration and harassment. *Americas Conference on Information Systems (AMCIS)*, Salt Lake City, UT, United States.

Brendel, A. B., Kolbe, L. M., & Diederich, S. (2020b). Designing Anthropomorphic Enterprise Conversational Agents. *Business & Information Systems Engineering*, *62*(3), 193-209. https://doi.org/10.1007/s12599-020-00639-y

Briggs, R. O., Böhmann, T., Schwabe, G., & Tuunanen, T. (2019). Advancing Design Science Research with Solution-based Probing. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Brink, A. (2013). *Anfertigung wissenschaftlicher Arbeiten: Ein prozessorientierter Leitfaden zur Erstellung von Bachelor-, Master-und Diplomarbeiten* (5th ed.). Springer Gabler, Wiesbaden. https://doi.org/10.1007/978-3-658-02511-3

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. *Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, New York.

Brynjolfsson, E., & McAfee, A. (2017). *The business of artificial intelligence: What it can – and cannot – do for your organization*. Harvard Business Review. Retrieved May 03, 2023 from https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence

Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Valdez, A. C. (2019). "Hey, Siri","Ok, Google","Alexa". Acceptance-Relevant Factors of Virtual Voice-Assistants. *International Conference on Professional Communication (IPCC)*, Aachen, Germany. https://doi.org/10.1109/ProComm.2019.00025

Burns, M. B., & Igou, A. (2019). “Alexa, write an audit opinion”: Adopting intelligent virtual assistants in accounting workplaces. *Journal of Emerging Technologies in Accounting*, *16*(1), 81-92. https://doi.org/10.2308/jeta-52424

Casas, J., Tricot, M.-O., Abou Khaled, O., Mugellini, E., & Cudré-Mauroux, P. (2020). Trends & methods in chatbot evaluation. *International Conference on Multimodal Interaction (ICMI)*, Virtual event, The Netherlands. https://doi.org/10.1145/3395035.3425319

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, *41*(13-14). https://doi.org/10.1080/02642069.2020.1787993

Chandler, J. D., Danatzis, I., Wernicke, C., Akaka, M. A., & Reynolds, D. (2019). How does innovation emerge in a service ecosystem? *Journal of Service Research*, *22*(1), 75-89. https://journals.sagepub.com/doi/abs/10.1177/1094670518797479

Chandler, J. D., & Lusch, R. F. (2014). Service systems: A broadened framework and research agenda on value propositions, engagement, and service experience. *Journal of Service Research*, *18*(1), 6-22. https://doi.org/10.1177/1094670514537709

Chandler, J. D., & Vargo, S. L. (2011). Contextualization and value-in-context: How context frames exchange. *Marketing Theory*, *11*(1), 35-49. https://doi.org/10.1177/1470593110393713

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1-8. https://doi.org/10.1016/j.jebo.2011.08.009

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Checkland, P. (1999). Systems thinking. *Rethinking management information systems*, 45-56.

Chung, H., Iorga, M., Voas, J., & Lee, S. (2017). Alexa, Can I Trust You? *Computer (Long Beach Calif)*, *50*(9), 100-104. https://doi.org/10.1109/MC.2017.3571053

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., & Munteanu, C. (2019). What makes a good conversation? Challenges in designing truly conversational agents. *Conference on Human Factors in Computing Systems (CHI)*, New York, NY, United States. https://doi.org/10.1145/3290605.3300705

Colby, K. M. (1975). *Artificial paranoia: a computer simulation of paranoid process* (Vol. 49). Pergamon Press, https://doi.org/10.1016/j.ijinfomgt.2021.102383

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*(1), 1-17. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, *1*(1), 104–126.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1).

Corea, C., Delfmann, P., & Nagel, S. (2020). Towards intelligent chatbots for customer care-practice-based requirements for a research agenda. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5 ed.). Sage publications,

Cronholm, S., & Göbel, H. (2018). Guidelines supporting the formulation of design principles. *Australasian Conference on Information Systems (ACIS)*, Sydney, Australia.

Cronholm, S., & Göbel, H. (2019). Design science research constructs: A conceptual model. *Pacific Asia Conference on Information Systems (PACIS)*, X'ian, China.

Crowston, K., & Bolici, F. (2019). Impacts of machine learning on work. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A customer service chatbot for e-commerce websites. *Meeting of the Association for Computational Linguistics-System Demonstrations*, Vancouver, Canada. https://doi.org/10.18653/v1/P17-4017

Cui, T., Peng, X., & Wang, X. (2020). Understanding the effect of anthropomorphic design: Towards more persuasive conversational agents. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Dahl, D. A. (2013). Natural language processing: past, present and future. In *Mobile speech and advanced natural language solutions* (pp. 49-73). Springer. https://doi.org/10.1007/978-1-4614-6018-3_4

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811-817. https://www.doi.org/10.1017/S1351324916000243

Dämon, K. (2017, 10.07.2017). *MEGATREND CHATBOT - Was Chatbots der Wirtschaft bringen*. WirtschaftsWoche. Retrieved 06.02.2023 from https://www.wiwo.de/erfolg/trends/megatrend-chatbot-was-chatbots-der-wirtschaft-bringen/19997318.html

Davenport, T. H., & Kirby, J. (2016). Just how smart are smart machines? *MIT Sloan Management Review*, *57*(3), 21-25.

de Lacerda, A. R. T., & Aguiar, C. S. R. (2019). FLOSS FAQ chatbot project reuse - how to allow nonexperts to develop a chatbot. *International Symposium on Open Collaboration*, Skövde, Sweden.

Depietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. *The processes of technological innovation*, *199*(0), 151-175.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019a). On conversational agents in information systems research: Analyzing the past to guide future work. *International Conference on Wirtschaftsinformatik (WI)*, Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019b). Towards a taxonomy of platforms for conversational agent design. *International Conference on Wirtschaftsinformatik (WI)*, Siegen, Germany.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. *Business & Information Systems Engineering*, *62*(3), 193–209. https://doi.org/10.1007/s12599-020-00639-y

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). 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*, *23*(1), 96-138. https://doi.org/10.17705/1jais.00724

Drechsler, A., & Hevner, A. R. (2018). Utilizing, producing, and contributing design knowledge in DSR projects. *International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, Chennai, India. https://doi.org/10.1007/978-3-319-91800-6_6

Echeverri, P., & Skålén, P. (2011). Co-creation and co-destruction: A practice-theory based study of interactive value formation. *Marketing Theory*, *11*(3), 351-373. https://doi.org/10.1177/1470593111408181

Edirisooriya, M., Mahakalanda, I., & Yapa, T. (2019). Generalized framework for automated conversational agent design via QFD. *Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lanka. https://doi.org/10.1109/MERCon.2019.8818945

Eitle, V., Zöll, A., & Buxmann, P. (2022). Organizational readiness concept for AI: A quantitative analysis of a multi-stage adoption process from the perspective of data scientists. *European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Elshan, E., & Ebel, P. (2020). Let's team up: Designing conversational agents as teammates. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Elshan, E., Engel, C., Ebel, P., & Siemon, D. (2022a). Assessing the reusability of design principles in the realm of conversational agents. *International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, St. Petersburg, FL, USA. https://doi.org/10.1007/978-3-031-06516-3_10

Elshan, E., Siemon, D., De Vreede, T., De Vreede, G.-J., Oeste-Reiß, S., & Ebel, P. (2022b). Requirements for AI-based teammates: A qualitative inquiry in the context of creative workshops. Hawaii International Conference on System Sciences (HICSS), Hawaii, United States.

Espig, A., Klimpel, N., Rödenbeck, F., & Auth, G. (2019). Bewertung des Kundennutzens von Chatbots für den Einsatz im Servicedesk. *International Conference on Wirtschaftsinformatik (WI)*, Siegen, Germany.

Essaied, K., Marsan, J., Bazomanza, W. N., Tewamba, H. J. N., & Yombia, S. M. K. (2020). Organizational decision to adopt chatbot technology: The role of organizing vision and technological frame. *Americas Conference on Information Systems (AMCIS)*, Salt Lake City, UT, United States.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019a). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Feine, J., Morana, S., & Maedche, A. (2019b). Designing a chatbot social cue configuration system. *International Conference on Information Systems (ICIS)*, Munich, Germany.

Felderer, M., & Ramler, R. (2021). Quality assurance for AI-based systems: Overview and challenges (Introduction to interactive session). *Software Quality: Future Perspectives on Software*

Engineering Quality (SWQD), Springer, Cham. https://doi.org/10.1007/978-3-030-65854-0_3

Feng, S., & Buxmann, P. (2020). My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Figma. (2022). Retrieved May 31, 2022, from https://www.figma.com

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., & Ischen, C. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing*, *103*(12), 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38-42. https://doi.org/10.1145/3085558

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018a). What makes users trust a chatbot for customer service? An exploratory interview study. *International Conference on Internet Science (INSCI)*, Springer, Cham. https://doi.org/10.1007/978-3-030-01437-7_16

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2018b). Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. *International Conference on Internet Science (INSCI)*, Springer, Cham. https://doi.org/10.1007/978-3-030-17705-8_13

Frei, F. X. (2006). Breaking the trade-off between efficiency and service. *Harvard business review*, *84*(11), 92.

Fridgen, G., Hartwich, E., Rägo, V., Rieger, A., & Stohr, A. (2022). Artifificial intelligence as a call for retail banking: Applying digital options thinking to artifificial intelligence adoption. *European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Fu, K. K., Yang, M. C., & Wood, K. L. (2015). Design principles: The foundation of design. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, https://doi.org/10.1115/DETC2015-46157

Fukas, P., Rebstadt, J., Remark, F., & Thomas, O. (2021). Developing an artifificial intelligence maturity model for auditing. *European Conference on Information Systems (ECIS)*, Virtual Conference.

Glaser, B. G., & Strauss, A. L. (2006). The discovery of grounded theory: Strategies for qualitative research (Reprinted ed.). *New Brunswick London: Aldine Transaction*.

Gläser, J., & Laudel, G. (2009). *Experteninterviews und qualitative Inhaltsanalyse: Als Instrumente rekonstruierender Untersuchungen* (3rd ed.). VS Verlag für Sozialwissenschaften, Wiesbaden.

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. *European Conference on Information Systems (ECIS)*, Portsmouth, United Kingdom.

Gnewuch, U., Morana, S., Hinz, O., Kellner, R., & Maedche, A. (2023). More than a Bot? The impact of disclosing human involvement on customer interactions with hybrid service agents. *Information Systems Research*. https://doi.org/10.1287/isre.2022.0152

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *International Conference on Information Systems (ICIS)*, Seoul, Korea.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304-316. https://doi.org/10.1016/j.chb.2019.01.020

Goh, O. S., Ardil, C., Wong, W., & Fung, C. C. (2007). A black-box approach for response quality evaluation of conversational agent systems. *International Journal of Computational Intelligence*, *3*(3), 195-203.

Gregor, S. (2002). Design theory in information systems. *Australasian Journal of Information Systems*, *10*(1), 1. https://doi.org/10.3127/ajis.v10i1.439

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems*, *21*(6), 2. https://doi.org/10.17705/1jais.00649

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337-355. https://doi.org/10.25300/misq/2013/37.2.01

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 312-335.

Gregory, R. W., & Muntermann, J. (2014). Research note—Heuristic theorizing: Proactively generating design theories. *Information Systems Research*, *25*(3), 639-653. https://doi.org/10.1287/isre.2014.0533

Gronroos, C. (1988). Service quality: The six criteria of good perceived service. *Review of business*, *9*(3), 10.

Grotherr, C., Semmann, M., & Böhmann, T. (2018). Using Microfoundations of Value Co-Creation to Guide Service Systems Design – A Multilevel Design Framework. International Conference on Information Systems (ICIS), San Francisco, CA, USA.

Grudin, J., & Jacques, R. (2019). Chatbots, humbots, and the quest for artificial general intelligence. *Conference on Human Factors in Computing Systems (CHI)*, Glasgow, Scotland. https://doi.org/10.1145/3290605.3300439

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods*, *18*(1), 59-82. https://doi.org/10.1177/1525822X05279903

Helfferich, C. (2019). Leitfaden- und Experteninterviews. In *Handbuch Methoden der empirischen Sozialforschung* (pp. 669-686). Springer Fachmedien. https://doi.org/10.1007/978-3-658-21308-4_44

Helkkula, A., Kowalkowski, C., & Tronvoll, B. (2018). Archetypes of Service Innovation: Implications for Value Cocreation. *Journal of Service Research*, *21*(3), 284-301. https://doi.org/10.1177/1094670517746776

Herrera, A., Yaguachi, L., & Piedra, N. (2019). Building conversational interface for customer support applied to open campus an open online course provider. *International Conference on Advanced Learning Technologies (ICALT)*, Maceio, Brazil. https://doi.org/10.1109/ICALT.2019.00011

Heuer, M., Lewandowski, T., Weglewski, J., Mayer, T., Kubicek, M., Lembke, P., Ortgiese, S., & Böhmann, T. (2023). Rethinking Interaction with Conversational Agents: How to Create a Positive User Experience Utilizing Dialog Patterns. *International Conference on Human-Computer Interaction (HCII)*, Copenhagen, Denmark.

Hevner, A., & Chatterjee, S. (2010). *Design research in information systems: theory and practice* (Vol. 22). Springer, Bosten, MA, United States. https://doi.org/10.1007/978-1-4419-5653-8_2

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, *19*(2), 87-92.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75-105. https://doi.org/10.2307/25148625

Hildebrand, C., & Bergner, A. (2020). Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*, *49*, 659-676. https://doi.org/10.1007/s11747-020-00753-z

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *European Conference on Cognitive Ergonomics (ECCE)*, Belfast, United Kingdom. https://doi.org/10.1145/3335082.3335094

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. *Conference on Human Factors in Computing Systems (CHI)*, Montréal, Canada. http://dx.doi.org/10.1145/3173574.3173989

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, *21*(2), 155-172. https://doi.org/10.1177/1094670517752459

Huiyang, S., & Min, W. (2022). Improving interaction experience through lexical convergence: The prosocial effect of lexical alignment in Human-Human and Human-Computer interactions. *International Journal of Human–Computer Interaction*, *38*(1), 28-41. https://doi.org/10.1080/10447318.2021.1921367

Iivari, J. (2007). A paradigmatic analysis of information systems as a design science. *Scandinavian Journal of Information Systems*, *19*(2), 39-64.

Io, H. N., & Lee, C. B. (2017). Chatbots and Conversational Agents: A Bibliometric Analysis. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore.

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. *Chatbot Research and Design. CONVERSATIONS 2019*, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-030-39540-7_3

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. *Designing Interactive Systems Conference (DIS)*, New York, NY, USA. https://doi.org/10.1145/3196709.3196735

Janssen, A., Cardona, D. R., & Breitner, M. H. (2021a). More than FAQ! Chatbot Taxonomy for Business-to-Business Customer Services. *Chatbot Research and Design. CONVERSATIONS 2020*, https://doi.org/10.1007/978-3-030-68288-0_12

Janssen, A., Grützner, L., & Breitner, M. H. (2021b). Why do chatbots fail? A critical success factors analysis. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

Janssen, A., Passlick, J., Cardona, D. R., & Breitner, M. H. (2020). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. *Business & Information Systems Engineering*, *62*(3), 211–225. https://doi.org/10.1007/s12599-020-00644-1

Jiang, E., Olson, K., Toh, E., Molina, A., Donsbach, A., Terry, M., & Cai, C. J. (2022). PromptMaker: Prompt-based prototyping with large language models. *Conference on Human Factors in Computing Systems (CHI)*, New Orleans, LA, United States. https://doi.org/10.1145/3491101.3503564

Jiang, J., & Ahuja, N. (2020). Response quality in Human-Chatbot Collaborative Systems. *Conference on Research and Development in Information Retrieval (SIGIR)*, Virtual Event, China. https://doi.org/10.1145/3397271.3401234

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Johari, N. M., & Nohuddin, P. (2021). Quality attributes for a good chatbot: a literature review. *International Journal of Electrical Engineering and Technology*, *12*(7), 109-119. https://doi.org/10.34218/IJEET.12.7.2021.012

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. *Business & Information Systems Engineering*, *63*(1), 5–20. https://doi.org/10.1007/s12599-020-00676-7

Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. *Clinical and translational science*, *14*(1), 86-93. https://doi.org/10.1111/cts.12884

Johnston, R. (1995). The determinants of service quality: satisfiers and dissatisfiers. *International journal of service industry management*, *6*(5), 53-71. https://doi.org/10.1108/09564239510101536

Jonke, A. W., & Volkwein, J. B. (2018). From tweet to chatbot–content management as a core competency for the digital evolution. *Digital Marketplaces Unleashed*, 275-285. https://doi.org/10.1007/978-3-662-49275-8_28

Jovanovic, B. (2023, April 7, 2023). *55 Fascinating AI statistics and trends for 2022*. Retrieved May 2, 2023 from https://dataprot.net/statistics/ai-statistics/

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

Kaplan, B., & Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. In *Evaluating the organizational impact of healthcare information systems* (pp. 30-55). Springer. https://doi.org/10.1007/0-387-30329-4_2

Khan, R., & Das, A. (2018). *Build Better Chatbots: A complete guide to getting started with chatbots* (1 ed.). Apress Berkeley, CA, https://doi.org/10.1007/978-1-4842-3111-1

Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: a services marketing research agenda. *Journal of Services Marketing*, *34*(3), 389-398. https://doi.org/10.1108/jsm-01-2019-0043

Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The Rise of Bots. *Conference on Designing Interactive Systems (DIS)*, Edinburgh, United Kingdom. https://doi.org/10.1145/3064663.3064672

Knijnenburg, B. P., & Willemsen, M. C. (2016). Inferring capabilities of intelligent agents from their external traits. *ACM Transactions on Interactive Intelligent Systems*, *6*(4), 1-25. https://doi.org/10.1145/2963106

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying smart personal assistants: an empirical cluster analysis. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Kohne, A., Kleinmanns, P., Rolf, C., & Beck, M. (2020). *Chatbots: Aufbau und Anwendungsmöglichkeiten von autonomen Sprachassistenten*. Springer Fachmedien Wiesbaden GmbH,

Kowald, C., & Bruns, B. (2020). Chatbot Kim: A digital tutor on AI. How advanced dialog design creates better conversational learning experiences. *International Journal of Advanced Corporate Learning (iJAC)*, *13*(3), 26-34. https://doi.org/10.3991/ijac.v13i3.17017

Krueger, R., & Casey, M. A. (2002). *Focus group interviewing on the telephone*. https://www.shadac.org/sites/default/files/Old_files/FocGrp_KruegerCasey_Aug02.pdf

Kruse, L., Wunderlich, N., & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489-504. https://doi.org/10.1057/ejis.2008.40

Kuechler, W., & Vaishnavi, V. (2012). A framework for theory development in design science research: multiple perspectives. *Journal of the Association for Information Systems*, *13*(6), 3. https://doi.org/10.17705/1jais.00300

Kühl, N., Schemmer, M., Goutier, M., & Satzger, G. (2022). Artificial intelligence and machine learning. *Electronic Markets*, *32*, 2235–2244. https://doi.org/10.1007/s12525-022-00598-0

Kuligowska, K. (2015). Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professionals Center for Business Research*, *2*.

Kushner, S. (1993). Naturalistic evaluation: Practical knowledge for policy development. *Research Evaluation*, *3*(2), 83-94.

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2019). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. *International Workshop on Chatbot Research and Design*, Springer, Cham. https://doi.org/10.1007/978-3-030-39540-7_13

Laban, G., & Araujo, T. (2020). The effect of personalization techniques in users' perceptions of conversational recommender systems. *ACM International Conference on Intelligent Virtual Agents (IVA)*, New York, NY, USA. https://doi.org/10.1145/3383652.3423890

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wünderlich, N. V., & De Keyser, A. (2017). 'Service Encounter 2.0': An Investigation into the Roles of Technology, Employees and Customers. *Journal of Business Research*, *79*, 238-246. https://doi.org/10.1016/j.jbusres.2017.03.008

Laumer, S., Gubler, F. T., Racheva, A., & Maier, C. (2019a). Use Cases for Conversational Agents: An Interview-based Study. *Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Laumer, S., Maier, C., & Gubler, F. T. (2019b). Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis. European Conference on Information Systems (ECIS).

Laut, P., Dumbach, P., & Eskofier, B. M. (2021). Integration of artificial intelligence in the organizational adoption – A configurational perspective. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

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. *European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Lee, P. (2016). *Learning from Tay's introduction*. Lee, Peter. Retrieved 20.08.2020 from https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, *103*, 95-105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Lester, J., Branting, K., Mott, B. (2004). Conversational Agents. *The Practical Handbook of Internet Computing*. https://doi.org/10.1201/9780203507223.ch10

Lewandowski, T., Delling, J., Grotherr, C., & Böhmann, T. (2021). State-of-the-art analysis of adopting AI-based conversational agents in organizations: A systematic literature review. *Pacific Asia Conference on Information Systems (PACIS)*, Dubai, UAE.

Lewandowski, T., Grotherr, C., & Böhmann, T. (2022a). Managing Artificial Intelligence Systems for Value Co-creation: The Case of Conversational Agents and Natural Language Assistants. In B. Edvardsson & B. Tronvoll (Eds.), *The Palgrave Handbook of Service Management* (pp. 945-966). Springer International Publishing. https://doi.org/10.1007/978-3-030-91828-6_45

Lewandowski, T., Heuer, M., Vogel, P., & Böhmann, T. (2022b). Design knowledge for the lifecycle management of conversational agents. *International Conference on Wirtschaftsinformatik (WI)*, Nürnberg, Germany.

Lewandowski, T., Kučević, E., Leible, S., Poser, M., & Böhmann, T. (2023a). Enhancing conversational agents for successful operation: A multi-perspective evaluation approach for continuous improvement. *Electronic Markets*, *33*(1), 39. https://doi.org/10.1007/s12525-023-00662-3

Lewandowski, T., Poser, M., Kučević, E., Heuer, M., Hellmich, J., Raykhlin, M., Blum, S., & Böhmann, T. (2023b). Leveraging the Potential of Conversational Agents: Quality Criteria for the Continuous Evaluation and Improvement. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Li, C.-H., Yeh, S.-F., Chang, T.-J., Tsai, M.-H., Chen, K., & Chang, Y.-J. (2020). A Conversation Analysis of Non-Progress and Coping Strategies with a Banking Task-Oriented Chatbot. *Conference on Human Factors in Computing Systems (CHI)*, New York, NY, USA. https://doi.org/10.1145/3313831.3376209

Li, L. P., Juric, B., & Brodie, R. J. (2018). Actor engagement valence: conceptual foundations, propositions and research directions. *Journal of Service Management*, *29*(3), 491-516. https://doi.org/10.1108/JOSM-08-2016-0235

Li, M., & Suh, A. (2022). Anthropomorphism in AI-enabled technology: A literature review. *Electronic Markets*, *32*, 2245–2275. https://doi.org/10.1007/s12525-022-00591-7

Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What can you do? Studying social-agent orientation and agent proactive interactions with an agent for employees. *ACM Conference on Designing Interactive Systems (DIS)*, Brisbane, Australia. https://doi.org/10.1145/2901790.2901842

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, *140*, 44-60.

Lindner, M. (2014). *Dissertation: Privatheit im Informationszeitalter. Ethische Grundlagen von Privatheit und Anwendungsfragen in der Informationstechnologie* Universitätsbibliothek Tübingen]. http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-72166

López, G., Quesada, L., & Guerrero, L. A. (2018). Alexa vs. Siri vs. Cortana vs. Google Assistant: A comparison of speech-based natural user interfaces. *International Conference on Applied Human Factors and Ergonomics*, https://doi.org/10.1007/978-3-319-60366-7_23

Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? *Journal of Service Theory and Practice*, *30*(3). https://doi.org/10.1108/JSTP-04-2019-0088

Lusch, R. F., & Nambisan, S. (2015). Service Innovation: A Service-Dominant Logic Perspective. *MIS Quarterly*, *39*(1), 155-175. https://www.jstor.org/stable/26628345

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Maglio, P. P., Vargo, S. L., Caswell, N., & Spohrer, J. (2009). The service system is the basic abstraction of service science. *Information Systems and e-business Management*, *7*(4), 395-406. https://doi.org/10.1007/s10257-008-0105-1

Manseau, J. (2020). AI in the workplace: A qualitative analysis of intelligent employee assistants. *Americas Conference on Information Systems (AMCIS)*, Salt Lake City, UT, United States.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, *15*(4), 251-266. https://doi.org/10.1016/0167-9236(94)00041-2

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. *International conference on information and education technology*, Aizu-Wakamatsu, Japan. https://doi.org/10.1145/3323771.3323824

Mayring, P. (2014). Qualitative content analysis. Theoretical foundation, basic procedures and software solution.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence - August 31, 1955. *AI Magazine*, *27*(4), 12-14. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904

McTear, M. (2018). Conversation modelling for chatbots: current approaches and future directions. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*, 175-185.

McTear, M. (2020). Conversational ai: Dialogue systems, conversational agents, and chatbots. *Synthesis Lectures on Human Language Technologies*, *13*(3), 1-251.

McTear, M. (2021). Introducing Dialogue Systems. In *Conversational AI. Synthesis Lectures on Human Language Technologies* (pp. 11-42). Springer, Cham. https://doi.org/10.1007/978-3-031-02176-3_1

McTear, M. F., Callejas, Z., & Griol, D. (2016). *The conversational interface* (Vol. 6). Springer, Cham, https://doi.org/10.1007/978-3-319-32967-3

Meinel, C., & von Thienen, J. (2016). Design Thinking. *Informatik-Spektrum*, *39*, 310-314. https://doi.org/10.1007/s00287-016-0977-2

Meira, M. d. O., & Canuto, A. d. P. (2015). Evaluation of emotional agents' architectures: An approach based on quality metrics and the influence of emotions on users. *World Congress on Engineering (WCE)*, London, U.K.

Meske, C., Amojo, I., & Thapa, D. (2020). Understanding the affordances of conversational agents in mental mobile health sservices. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Meuser, M., & Nagel, U. (2009a). Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage. In S. Pickel, G. Pickel, H.-J. Lauth, & D. Jahn (Eds.), *Methoden der vergleichenden Politik- und Sozialwissenschaft* (pp. 465-479). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91826-6_23

Meuser, M., & Nagel, U. (2009b). The expert interview and changes in knowledge production. *Interviewing experts*, 17-42. https://doi.org/10.1057/9780230244276_2

Meyer von Wolff, R., Hobert, S., Masuch, K., & Schumann, M. (2020a). 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), 3. https://doi.org/10.17705/1pais.12203

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2019a). How May I Help You?–State of the Art and Open Research Questions for Chatbots at the Digital Workplace. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2020b). Einsatz von Chatbots am digitalen Büroarbeitsplatz–Eine praxisorientierte Betrachtung von Einsatzbereichen, Wirkungen und Handlungsempfehlungen. *HMD Praxis der Wirtschaftsinformatik*(57), 413-431. https://doi.org/https://doi.org/10.1365/s40702-020-00602-1

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2021). Sorry, I can’t understand you! – Influencing factors and challenges of chatbots at digital workplaces. *International Conference on Wirtschaftsinformatik (WI)*, Essen, Germany.

Meyer von Wolff, R., Hobert, S., & Schumann, M. (2022). Chatbot Introduction and Operation in Enterprises–A Design Science Research-based Structured Procedure Model for Chatbot Projects. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Meyer von Wolff, R., Masuch, K., Hobert, S., & Schumann, M. (2019b). What Do You Need Today?-An Empirical Systematization of Application Areas for Chatbots at Digital Workplaces. *Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Meyer-Waarden, L., Pavone, G., Poocharoentou, T., Prayatsup, P., Ratinaud, M., Tison, A., & Torné, S. (2020). How service quality influences customer acceptance and usage of chatbots. *Journal of Service Management Research*, *4*(1), 35–51. https://doi.org/10.15358/2511-8676-2020-1-35

Michaud, L. N. (2018). Observations of a new chatbot: Drawing conclusions from early interactions with users. *IT Professional*, *20*(5), 40–47. https://doi.org/10.1109/MITP.2018.053891336

Microsoft Teams UI Kit. (2023). Retrieved 28.03.2023 from https://www.figma.com/community/file/916836509871353159/Microsoft-Teams-UI-Kit

Mimoun, M. S. B., Poncin, I., & Garnier, M. (2012). Case study—Embodied virtual agents: An analysis on reasons for failure. *Journal of Retailing and Consumer services*, *19*(6), 605-612. https://doi.org/10.1016/j.jretconser.2012.07.006

Möller, F., Guggenberger, T. M., & Otto, B. (2020). Towards a Method for Design Principle Development in Information Systems. *International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, Kristiansand, Norway.

Morana, S., vom Brocke, J., Maedche, A., Seidel, S., Adam, M. T., Bub, U., Fettke, P., Gau, M., Herwix, A., & Mullarkey, M. T. (2018). Tool support for design science research—Towards a software ecosystem: A report from a DESRIST 2017 workshop. *Communications of the Association for Information Systems*, *43*(1), 17.

Morgan, D. L. (1996). Focus groups. *Annual review of sociology*, *22*(1), 129-152.

Morgan, H. (2022). Conducting a qualitative document analysis. *The Qualitative Report*, *27*(1), 64-77.

Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. *European Journal of Information Systems*, *28*(1), 6-20. https://doi.org/10.1080/0960085X.2018.1451811

Murphy, P. (2007). 'You Are Wasting My Time': Why Limits on Connectivity Are Essential for Economies of Creativity. *University of Auckland Business Review*, *9*(2), 16.

Myers, M. D. (2019). *Qualitative research in business and management*. Sage, Thousand Oaks, CA, United States.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and organization*, *17*(1), 2-26.

Nardi, P. M. (2018). *Doing Survey Research: A Guide to Quantitative Methods* (4 ed.). Routledge, New York, NY, USA. https://doi.org/10.4324/9781315172231

Nascimento, A. M., da Cunha, M. A. V. C., de Souza Meirelles, F., Scornavacca Jr, E., & de Melo, V. V. (2018). A Literature Analysis of Research on Artificial Intelligence in Management Information System (MIS). *Americas Conference on Information Systems (AMCIS)*, New Orleans, LA, United States.

Neff, G., & Nagy, P. (2016). Automation, Algorithms, and Politics| Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication*, *10*(17), 4915–4931.

Nguyen, Q. N., Sidorova, A., & Torres, R. (2022). Artificial intelligence in business: A literature review and research agenda. *Communications of the Association for Information Systems*, *50*(7), 175-207. https://doi.org/10.17705/1CAIS.05007

Niemann, M., Appel, M., Repp, N., & Steinmetz, R. (2009). Towards a consistent service lifecycle model in service governance. *Americas Conference on Information Systems (AMCIS)*, San Francisco, CA, United States.

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1990). Systems Development in Information Systems Research. *Journal of Management Information Systems*, *7*(3), 89-106. https://doi.org/10.1080/07421222.1990.11517898

Oracle. (2016). *Can Virtual Experiences Replace Reality?* Oracle. Retrieved 02.03.2020 from https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc35/CXResearchVirtualExperiences.pdf

Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer acceptance of AI in service encounters: understanding antecedents and consequences. In P. P. Maglio, Kieliszewski, C.A., Spohrer, J.C., Lyons, K., Patrício, L., Sawatani, Y. (Ed.), *Handbook of Service Science, Volume II* (pp. 77-103). Springer. https://doi.org/10.1007/978-3-319-98512-1_5

Pandl, K. D., Teigeler, H., Lins, S., Thiebes, S., & Sunyaev, A. (2021). Drivers and inhibitors for organizations' intention to adopt artificial intelligence as a service. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of marketing*, *49*(4), 41-50. https://doi.org/10.1177/002224298504900403

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183-199. https://doi.org/10.1016/j.im.2014.08.008

Patel, S., Chiu, Y.-T., Khan, M. S., Bernard, J.-G., & Ekandjo, T. A. (2021). Conversational agents in organisations: Strategic applications and implementation considerations. *Journal of Global Information Management*, *29*(6), 1-25. http://doi.org/10.4018/JGIM.20211101.oa53

Pauletto, S., Balentine, B., Pidcock, C., Jones, K., Bottaci, L., Aretoulaki, M., Wells, J., Mundy, D. P., & Balentine, J. (2013). Exploring expressivity and emotion with artificial voice and speech technologies. *Logopedics Phoniatrics Vocology*, *38*(3), 115-125. https://doi.org/10.3109/14015439.2013.810303

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45-77. https://doi.org/10.2753/MIS0742-1222240302

Peras, D. (2018). Chatbot evaluation metrics. *Economic and Social Development: Book of Proceedings*, 89–97.

Pereira, J., & Díaz, Ó. (2018). Chatbot dimensions that matter: Lessons from the trenches. *International Conference on Web Engineering (ICWE)*, Cáceres, Spain.

Peters, C., Maglio, P., Badinelli, R., Harmon, R. R., Maull, R., Spohrer, J. C., Tuunanen, T., Vargo, S. L., Welser, J. J., & Demirkan, H. (2016). Emerging digital frontiers for service innovation. *Communications of the Association for Information Systems*, *1*(39). https://doi.org/10.17705/1CAIS.03908

Pfeiffer, J. (2020). Interview with Omer Biran and Michael Halfmann on "Conversational Agents at SAP". *Business & Information Systems Engineering*, *62*(3), 249-252. https://doi.org/10.1007/s12599-020-00647-y

Phellas, C. N., Bloch, A., & Seale, C. (2011). Structured methods: Interviews, questionnaires and observation. *Researching Society and Culture*, *3*(1), 23-32.

Pillai, R., & Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. *International Journal of Contemporary Hospitality Management*, *32*(10), 3199-3226. https://doi.org/10.1108/IJCHM-04-2020-0259

Poser, M., Hackbarth, T., & Bittner, E. A. (2022a). Don't throw it over the fence! toward effective handover from conversational agents to service employees. *International Conference on Human-Computer Interaction (HCII)*, https://doi.org/10.1007/978-3-031-05412-9_36

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid service recovery: Design for seamless inquiry handovers between conversational agents and human service agents. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Poser, M., Wiethof, C., & Bittner, E. A. (2022b). Integration of AI into customer service: A taxonomy to inform design decisions. *European Conference on Information Systems (ECIS)*, Timisoara, Romania.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A taxonomy of evaluation methods for information systems artifacts. *Journal of Management Information Systems*, *32*(3), 229-267. https://doi.org/10.1080/07421222.2015.1099390

Premathilake, G. W., Li, H., Liu, Y., & Helander, N. (2021). A review of the empirical literature on service robots in information systems literature. *Pacific Asia Conference on Information Systems (PACIS)*, Dubai, UAE.

Pumplun, L., Tauchert, C., & Heidt, M. (2019). A new organizational chassis for artificial intelligence - Exploring organizational readiness factors. *European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Rädiker, S., & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA* (1 ed.). Springer, Wiesbaden. https://doi.org/10.1007/978-3-658-22095-2

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint*. https://doi.org/10.48550/arXiv.1704.04579

Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly*, *43*(1), iii-ix.

Rajaobelina, L., Prom Tep, S., Arcand, M., & Ricard, L. (2021). Creepiness: Its antecedents and impact on loyalty when interacting with a chatbot. *Psychology & Marketing*, *38*(12), 2339–2356. https://doi.org/10.1002/mar.21548

Research and Markets. (2022). *Chatbot Market*. https://www.researchandmarkets.com/reports/5648691

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Publishing Group, New York.

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing*, *36*(6).

Rimol, M. (2022). *Gartner Predicts Conversational AI Will Reduce Contact Center Agent Labor Costs by $80 Billion in 2026*. Gartner. Retrieved March 21, 2023 from https://www.gartner.com/en/newsroom/press-releases/2022-08-31-gartner-predicts-conversational-ai-will-reduce-contac

Riquel, J., Brendel, A. B., Hildebrandt, F., Greve, M., & Kolbe, L. M. (2021). "Even the Wisest Machine Makes Errors" – An Experimental Investigation of Human-like Designed and

Flawed Conversational Agents. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

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. *Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management research news*, *27*(6), 31-39. https://doi.org/10.1108/01409170410784185

Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and ethical considerations. *Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*, Galway, Ireland.

Rzepka, C., & Berger, B. (2018). User Interaction with AI-enabled Systems: A Systematic Review of IS Research. *International Conference on Information Systems (ICIS)*, San Francisco, CA, USA.

Saenz, J., Burgess, W., Gustitis, E., Mena, A., & Sasangohar, F. (2017). The usability analysis of chatbot technologies for internal personnel communications. *Industrial and Systems Engineering Conference*, Norcross, United States.

Sagodi, A., Engel, C., Schniertshauer, J., & Van Giffen, B. (2022). Becoming certain about the uncertain: How AI changes proof-of-concept activities in manufacturing–Insights from a global automotive leader. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Schmelzer, R. (2022). *The one practice that separates the AI successes from the failures*. Fobers Magazine. Retrieved 22.04.2024 from https://www.forbes.com/sites/cognitiveworld/2022/08/14/the-one-practice-that-is-separating-the-ai-successes-from-the-failures/?sh=62a7f2717cb5

Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a contemporary frontier of interaction design. *European Conference on Information Systems (ECIS)*, Virtual Conference.

Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A., & Leimeister, J. M. (2023). Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. *Information Systems Frontiers*, 1-26. https://doi.org/10.1007/s10796-023-10375-9

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding whether and how to deploy chatbots. *MIS Quarterly Executive*, *20*(1), 4. https://doi.org/10.17705/2msqe.00039

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. *Information and organization*, *21*(1), 1-16. https://doi.org/10.1016/j.infoandorg.2010.11.001

Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Ed.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (pp. 60-91). R. Piper & Co. Verlag. https://doi.org/10.25656/01:1423

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. B. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research*, *30*(1), 1-18. https://doi.org/10.1108/INTR-12-2019-0503

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2017). When do we need a human? Anthropomorphic design and trustworthiness of conversational agents. *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS* Seoul, Korea.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, *22*(4), 8. https://doi.org/10.17705/1jais.00685

Seiffer, A., Gnewuch, U., & Maedche, A. (2021). Understanding employee responses to software robots: a systematic literature review. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, *35*(1), 37-56. https://doi.org/10.2307/23043488

Semmann, M., Grotherr, C., Vogel, P., Bittner, E., Biemann, C., & Böhmann, T. (2018). Intelligent Collaboration of Humans and Language-Based Assistants (INSTANT). *International Conference on Information Systems (ICIS)*, San Francisco, CA, United States.

Shah, H., Warwick, K., Vallverdú, J., & Wu, D. (2016). Can machines talk? Comparison of Eliza with modern dialogue systems. *Computers in Human Behavior*, *58*, 278-295. https://doi.org/10.1016/j.chb.2016.01.004

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, *22*(1), 29-49. https://jlcl.org/article/download/88/86

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, *117*, 106627. https://doi.org/10.1016/j.chb.2020.106627

Siddike, M. A. K., & Kohda, Y. (2018). Towards a Framework of Trust Determinants in People and Cognitive Assistants Interactions. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Simon, H. A. (1996). *The sciences of the artificial* (3 ed.). MIT press, Boston, United States.

Sousa, D. N., Brito, M. A., & Argainha, C. (2019). Virtual Customer Service: Building your Chatbot. *International Conference on Business and Information Management (ICBIM)*, Paris, France. https://doi.org/10.1145/3361785.3361805

Stieglitz, S., Brachten, F., & Kissmer, T. (2018). Defining Bots in an Enterprise Context. *International Conference on Information Systems (ICIS)*, San Francisco, CA, United States.

Stockdale, R., & Standing, C. (2006). An interpretive approach to evaluating information systems: A content, context, process framework. *European Journal of Operational Research*, *173*(3), 1090-1102. https://doi.org/10.1016/j.ejor.2005.07.006

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2019). How affordances of chatbots cross the chasm between social and traditional enterprise systems. *Electronic Markets*, *30*(2), 369–403. https://doi.org/10.1007/s12525-019-00359-6

Storbacka, K., Brodie, R. J., Böhmann, T., Maglio, P. P., & Nenonen, S. (2016). Actor engagement as a microfoundation for value co-creation. *Journal of Business Research*, *69*(8), 3008-3017. https://doi.org/10.1016/j.jbusres.2016.02.034

Su, Y.-S., & Hsia, J.-H. (2022). An Evaluation Model of Smart Speaker Design. In *The Routledge Companion to Technology Management* (1 ed., pp. 141-156). Routledge.

Subramaniam, N. (2023). Digital transformation and artificial intelligence in organizations. *Journal of Financial Transformation*, *58*, 90-97.

Susman, G. I., & Evered, R. D. (1978). An Assessment of the Scientific Merits of Action Research. *Administrative Science Quarterly*, *23*(4), 582-603.

Symons, V. J. (1991). A review of information systems evaluation: content, context and process. *European Journal of Information Systems*, *1*(3), 205-212. https://doi.org/10.1057/ejis.1991.35

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of chatgpt et al. the prospects of large language models. *Business & Information Systems Engineering*, *65*(2), 95-101. https://doi.org/10.1007/s12599-023-00795-x

Tewes, S., Niestroj, B., & Tewes, C. (2020). *Geschäftsmodelle in die Zukunft denken*. Springer Gabler, Wiesbaden, Germany.

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). The Use of Focus Groups in Design Science Research. In *Design Research in Information Systems: Theory and Practice* (pp. 121-143). Springer US. https://doi.org/10.1007/978-1-4419-5653-8_10

Uba, C., Lewandowski, T., & Böhmann, T. (2023). The AI-based Transformation of organizations: The 3D-Model for guiding enterprise-wide AI change. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456-464. https://doi.org/10.1177/0706743719828977

van Aken, J. E. (2004). Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules. *Journal of management studies*, *41*(2), 219-246. https://doi.org/10.1111/j.1467-6486.2004.00430.x

van der Goot, M. J., Hafkamp, L., & Dankfort, Z. (2021). Customer service chatbots: A qualitative interview study into customers' communication journey. *International Workshop on Chatbot Research, CONVERSATIONS 2020*, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-030-68288-0_13

van Giffen, B., Borth, D., & Brenner, W. (2020). Management von Künstlicher Intelligenz in Unternehmen. *HMD Praxis der Wirtschaftsinformatik*, *57*(1), 4-20. https://doi.org/10.1365/s40702-020-00584-0

Van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. A. M. (2020). Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management*, *31*(2), 203-225. https://doi.org/10.1108/josm-06-2019-0175

Vargo, S. L., & Lusch, R. F. (2016). Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science*, *44*(1), 5-23. https://doi.org/10.1007/s11747-015-0456-3

Vargo, S. L., Wieland, H., & Akaka, M. A. (2015). Innovation through institutionalization: A service ecosystems perspective. *Industrial Marketing Management*, *44*, 63-72. https://doi.org/10.1016/j.indmarman.2014.10.008

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77-89. https://doi.org/10.1057/ejis.2014.36

Venable, J. R., Pries-Heje, J., & Baskerville, R. L. (2017). Choosing a design science research methodology. *Australasian Conference on Information Systems (ACIS)*, Hobart, Australia.

Venkatesh, V. (2022). Adoption and use of AI tools: a research agenda grounded in UTAUT. *Annals of Operations Research*, *308*, 641–652. https://doi.org/10.1007/s10479-020-03918-9

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21-54. https://www.jstor.org/stable/43825936

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *European Conference on Information Systems (ECIS)*, Verona, Italy.

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*(9), 205-224. https://doi.org/10.17705/1CAIS.03709

vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). 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), 520-544. https://doi.org/10.17705/1jais.00611

Wache, H., Möller, F., Schoormann, T., Strobel, G., & Petrik, D. (2022). Exploring the abstraction levels of design principles: the case of chatbots. *International Conference on Wirtschaftsinformatik (WI)*, Nürnberg, Germany.

Waizenegger, L., Seeber, I., Dawson, G., & Desouza, K. (2020). Conversational agents-exploring generative mechanisms and second-hand effects of actualized technology affordances. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Wallace, R. S. (2009). The anatomy of A.L.I.C.E. In *Parsing the Turing Test* (pp. 181-210). Springer. https://doi.org/10.1007/978-1-4020-6710-5_13

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, *3*(1), 36-59. https://doi.org/10.1287/isre.3.1.36

Wambsganss, T., Höch, A., Zierau, N., & Söllner, M. (2021). Ethical design of conversational agents: towards principles for a value-sensitive design. International Conference on Wirtschaftsinformatik (WI), Essen, Germany.

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2020). Effects of voice-based AI in customer service: Evidence from a natural experiment. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Wang, Y., Xiong, M., & Olya, H. (2020). Toward an understanding of responsible artificial intelligence practices. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), 8-23. https://www.jstor.org/stable/4132319

Weiler, S., Matt, C., & Hess, T. (2022). Immunizing with information–Inoculation messages against conversational agents' response failures. *Electronic Markets*, *32*(1), 239-258. https://doi.org/10.1007/s12525-021-00509-9

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

Wiethof, C., & Bittner, E. (2021). Hybrid intelligence-combining the human in the loop with the computer in the loop: A systematic literature review. *International Conference on Information Systems (ICIS)*, Austin, TX, United States.

Wilde, T., & Hess, T. (2006). *Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung*. L.-M.-U. t. M. n. Working Paper, Institut für Wirtschaftsinformatik und Neue Medien (WIM), München.

Wilde, T., & Hess, T. (2007). Forschungsmethoden der Wirtschaftsinformatik. *Wirtschaftsinformatik, 49*(4), 280-287.

Wilson, J. H. D., Paul R. (2018). *Collaborative Intelligence: Humans and AI Are Joining Forces*. Havard Business Review. Retrieved 18.07.2020 from https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

Winkler, R., Elshan, E., Söllner, M., & Leimeister, J. M. (2020). SPAM–A process model for developing smart personal assistants. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States.

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Annual Meeting (AOM)*, Chicago, USA. https://doi.org/10.5465/AMBPP.2018.15903abstract

Wintersberger, P., Klotz, T., & Riener, A. (2020). Tell me more: Transparency and time-fillers to optimize chatbots' waiting time experience. *Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, Tallinn, Estonia.

Wirtz, J., Kunz, W., & Paluch, S. (2021). The Service Revolution, Intelligent Automation and Service Robots. *European Business Review*.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management, 29*(5), 907-931. https://doi.org/10.1108/josm-04-2018-0119

Xiao, L., & Kumar, V. (2019). Robotics for customer service: A useful complement or an ultimate substitute? *Journal of Service Research, 24*(1), 9–29. https://doi.org/10.1177/1094670519878881

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of planning education and research, 39*(1), 93-112. https://doi.org/10.1177/0739456X17723971

Xiong, Y., Tang, Y., Kim, S., & Rosen, D. W. (2023). Human-machine collaborative additive manufacturing. *Journal of Manufacturing Systems, 66*, 82-91. https://doi.org/10.1016/j.jmsy.2022.12.004

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. *Conference on Human Factors in Computing Systems (CHI)*, New York, NY, United States. https://doi.org/10.1145/3025453.3025496

Xu, Y., Shieh, C.-H., van Esch, P., & Ling, I.-L. (2020). AI customer service: Task complexity, problem-solving ability, and usage intention. *Australasian Marketing Journal, 28*(4), 189-199. https://doi.org/10.1016/j.ausmj.2020.03.005

Yin, R. K. (2003). *Case study research: Design and methods* (3 ed., Vol. 5). Sage Publications, Young, S. W. (2014). Improving library user experience with A/B testing: Principles and process. *Weave: Journal of Library User Experience*, *1*(1). https://doi.org/10.3998/weave.12535642.0001.101

Zarouali, B., Van den Broeck, E., Walrave, M., & Poels, K. (2018). Predicting consumer responses to a chatbot on Facebook. *Cyberpsychology, Behavior, and Social Networking*, *21*(8), 491-497. https://doi.org/10.1089/CYBER.2017.0518

Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020a). A review of the empirical literature on conversational agents and future research directions. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020b). Towards developing trust-supporting design features for AI-based chatbots in customer service. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Zierau, N., Wambsganss, T., Janson, A., Schöbel, S., & Leimeister, J. M. (2020c). The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. *International Conference on Information Systems (ICIS)*, Hyderabad, India.

Appendix A: Eidesstattliche Versicherung

Hiermit versichere ich,

Tom Lewandowski, geboren am 05. September 1994 in Jesteburg,

an Eides statt, dass ich die vorliegende Dissertationsschrift mit dem Titel

„Artificial Intelligence in Organizations: Managing the Lifecycle of Conversational Agents"

selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Tostedt, 17.05.2024 Lewandowski

Ort, Datum Unterschrift