Hybrid Intelligence Systems
Designing Interactions for Continuous Mutual Augmentation of Humans and AI

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

Christina Wiethof

2022

Hamburg, Germany

Date of disputation:

21.11.2022

Evaluators:

First Evaluator: Prof. Dr. Eva A. C. Bittner

Second Evaluator: Prof. Dr. Tilo Böhmann

Third Evaluator: Prof. Dr. Sofia Schöbel

Abstract

Motivation

Technologies enabled by artificial intelligence (AI) are increasingly deployed by organizations to substitute human workforce, enhancing the effectiveness and efficiency of business operations. However, AI is far away from resembling human intelligence, thus easily reaching its limits when facing complex tasks. To compensate for the limitations and allow for fast and flexible learning of AI, end users get involved in training AI on the job, employing the human-in-the-loop (HITL) approach. The indispensability and need for human intelligence call for human-centered socio-technical system design for user interaction with AI-enabled systems. This, in turn, enables the computer-in-the-loop (CITL) approach, which complements HITL and allows humans to get support and learn through interaction with AI. Combining HITL and CITL leverages the strengths of both artificial and human intelligence and enables interactions for mutual augmentation of human and AI to facilitate hybrid learning and enhanced task performance. This form of human-AI collaboration is coined hybrid intelligence (HI). Hybrid intelligence systems (HISs) are based on HI to allow for continuous mutual augmentation of human and AI and to enable their co-evolvement on the job. Research on AI-enabled technologies in text-based online customer service (OCS) shows great potential in and need for the implementation of HISs to facilitate the collaboration of service employees (SEs) and AI during customer service delivery. However, due to its novelty, research on HISs in OCS and HI in general is still scant and fragmentary. It lacks systematic and integrated human-centered design knowledge on the actual interactions of humans and AI combining HITL and CITL within HIS to facilitate the continuity of mutual augmentation. To address this research gap, the goal of this dissertation is to develop validated design knowledge for human-centered HISs that ensure interactions for continuous mutual augmentation of human and AI to facilitate hybrid learning and enhanced task performance in the context of OCS.

Research Design

This dissertation conducts design science research (DSR) to develop prescriptive design knowledge for HISs. Therefore, eight interconnected publications are included as a cumulative approach to achieve the overall research goal. Moreover, six overarching research questions have been formulated in accordance with the DSR cycles of relevance, rigor, and design to systematically integrate the included publications for the accumulation and evolution of design knowledge. To support this research endeavor, several research methods have been applied across the publications over the course of this dissertation. These include systematic literature reviews, semi-structured expert interviews, taxonomy development methods, standardized user tests, as well as qualitative and quantitative data collection and analysis methods. Regarding user tests, design artifacts were instantiated for demonstration and evaluation of design validity and utility as simulated proof-of-concept or prototypes, employing either the wizard of Oz technique or fully functioning AI capabilities.

Results

The main result of this dissertation is validated design knowledge toward a theory for design and action for designing and developing human-centered HISs. For this, prescriptive design knowledge in the form of several design principles (DPs) and instantiated design artifacts is developed within the included publications. These DPs are predominantly contextualized in the application domain of OCS. Based on the individual DPs, the dissertation presents six abstract DPs to design human-centered HIS, covering the form of AI appearance, augmentation strategies, and integration of motivational elements. Employing several evaluation methods on the instantiated design artifacts provides additional results on the validity and utility of the DPs as well as usage behavior, learning progress, continuance intention to use, perceived humanness, usefulness, and task performance. Underpinning the validated design knowledge, the dissertation contributes conceptualizations of hybrid learning, a framework for gamifying collaboration processes, and a taxonomy of AI integration into customer service.

Contribution

By designing and developing human-centered HISs to facilitate hybrid learning and enhanced task performance, this dissertation contributes to research on human-computer interaction, specifically user interaction with AI-enabled systems for human-AI collaboration within a socio-technical system. This entails contributions that address the (1) user, (2) system, (3) task and context, and (4) interaction and outcomes of HISs. Following a DSR approach, these contributions are twofold, that is, the dissertation contributes to both research and practice. First, the dissertation contributes patterns conceptualizing hybrid learning. Based on this, it introduces and evaluates a novel perspective of distinguishing between novice and expert users for enabling an implicit knowledge transfer through HISs. Practitioners can draw on this concept to facilitate learning and enhanced task performance through HISs. Second, a framework for gamifying collaboration processes is developed and successfully applied to both human-human collaboration as well as human-AI collaboration. Moreover, its application to an HIS discloses that humanizing AI can be considered a gamification element that increases team spirit. Practitioners can use the framework in a top-down manner to guide the implementation of gamification in HISs. Third, the dissertation contributes to research on AI in OCS by building and evaluating a taxonomy of AI integration into customer service. Based on the taxonomy, existing AI infusion archetypes are confirmed and new archetypes identified. The taxonomy serves both researchers and practitioners offering guidance to plan or analyze their AI integration endeavors along the dimensions in sequential order. Successfully applied to three organizational use cases, it constitutes the contextual foundation for the design knowledge developed. Fourth, the contributions to interaction and outcomes of HISs comprise the validated prescriptive design knowledge in the form of

DPs and instantiated design artifacts toward a theory for design and action. In this regard, the dissertation presents six abstract DPs to facilitate the causal coherences of design knowledge specified to the research goal. Based on this, a preliminary design theory for human-centered HIS is presented.

Limitations

The dissertation faces a few limitations in terms of the knowledge sources, focus on human-centered design knowledge, evaluation methodology, and contextual environment. First, the selection of stakeholders, experts, and participants for interviews and user tests, as well as the determination of the scientific databases and search strings for systematic literature reviews constrain the results of the respective research methods. Second, HISs combine HITL and CITL to enable interactions for mutual augmentation. Although the design knowledge developed throughout this dissertation includes HITL, the focus on human-centered design is aimed at facilitating learning and task performance through CITL. As such, the optimization or assessment of technical performance within the scope of HITL did not fall within the scope of this dissertation. Third, employing qualitative evaluation methods entails several limitations, for example, the number of instantiated design artifacts and the selected OCS use cases. Furthermore, effects on hybrid learning and task performance are constrained by simulating customers or AI within short-term semi-naturalistic evaluations. For developing a solid design theory, a more in-depth evaluation is necessary. Fourth, the results are predominantly constrained by the contextual environment of OCS.

Future Research

Along with the contributions, the dissertation discloses several avenues for future research on (1) interaction and outcomes, (2) system characteristics, (3) user characteristics, and (4) task and context concerning user interaction with AI-enabled systems for human-AI collaboration within HISs. First, long-term, quantitative, and naturalistic evaluations of the developed design knowledge could compensate for the limitations of the selected evaluation methodology. Additionally, in-depth investigations of HITL within HISs from a technical perspective could complement the results of this dissertation. Future research may also contribute descriptive research on hybrid learning as well as an evaluation framework for HISs. Second, the framework for gamifying collaboration processes demands further research to consider other solutions than gamification to motivate participation in collaboration processes and to facilitate a descriptive perspective of gamification elements and respective effects. Moreover, for system characteristics, an untapped potential for examining the relationship between human and AI within an HIS remains. Third, regarding user characteristics, future research might contribute descriptive knowledge of design choices and their effects on learning and task performance. Fourth, in terms of task and context, the dissertation encourages further research on AI integration into customer service. For instance, researchers could draw on the taxonomy of AI integration into customer service and similarly contribute descriptive knowledge of design

decisions with certain effects in specific contexts. Finally, the application of the developed design knowledge for HISs to other organizational contexts might support generalization.

Kurzfassung

Motivation

Unternehmen setzen zunehmend Technologien ein, die auf künstlicher Intelligenz (KI) basieren, um menschliche Arbeitskräfte zu ersetzen und die Effektivität und Effizienz von Geschäftsabläufen zu verbessern. Allerdings ist KI noch weit von der menschlichen Intelligenz entfernt und gelangt bei komplexen Aufgaben schnell an ihre Grenzen. Um dies zu kompensieren und ein schnelles und flexibles Lernen der KI zu ermöglichen, werden die Endnutzer in das Training der KI am Arbeitsplatz einbezogen, wobei der Human-in-the-Loop-Ansatz (HITL) zum Einsatz kommt. Die Unverzichtbarkeit und Involvierung menschlicher Intelligenz hierfür erfordert ein humanzentriertes soziotechnisches Systemdesign für die Benutzerinteraktion mit KI-gestützten Systemen. Dies wiederum ermöglicht den Computer-in-the-Loop-Ansatz (CITL), der HITL ergänzt, d. h. der Mensch erhält Unterstützung und lernt durch die Interaktion mit der KI. Die Kombination von HITL und CITL nutzt die Stärken sowohl der künstlichen als auch der menschlichen Intelligenz und ermöglicht Interaktionen zur gegenseitigen Augmentierung von Mensch und KI, um hybrides Lernen und eine verbesserte Aufgabenbearbeitung zu ermöglichen. Diese Form der Kollaboration zwischen Mensch und KI wird als hybride Intelligenz (HI) bezeichnet. In hybriden Intelligenzsystemen (HIS) wird HI eingesetzt, um eine kontinuierliche gegenseitige Augmentierung von Mensch und KI zu ermöglichen und ihre gemeinsame Entwicklung bei der Arbeit zu fördern. Forschungsarbeiten zu KI-gestützten Technologien im textbasierten Online-Kundenservice (OKS) zeigen das große Potenzial und die Notwendigkeit der Implementierung von HIS, um die Kollaboration von Servicemitarbeitenden (SE) und KI bei der Erbringung von Kundendienstleistungen zu ermöglichen. Aufgrund ihrer Neuartigkeit ist die Forschung zu HIS in OKS und HI im Allgemeinen jedoch noch spärlich und lückenhaft. Es fehlt an systematischem und integriertem Wissen über die konkreten Interaktionen von Mensch und KI, die HITL und CITL innerhalb von HIS kombinieren, um die Kontinuität der gegenseitigen Augmentierung zu unterstützen. Um diese Forschungslücke zu schließen, ist es das Ziel der Dissertation, validiertes Designwissen für humanzentrierte HIS zu entwickeln, die Interaktionen für kontinuierliche, gegenseitige Augmentierung von Mensch und KI gewährleisten, um hybrides Lernen und verbesserte Aufgabenbearbeitung im OKS zu ermöglichen.

Forschungsdesign

In dieser Dissertation wird der Design Science Research (DSR) Ansatz verfolgt, um präskriptives Gestaltungswissen für HIS zu entwickeln. Daher sind acht miteinander verbundene Publikationen kumulativ enthalten, die das übergeordnete Forschungsziel adressieren. Darüber hinaus wurden sechs übergreifende Forschungsfragen in Anlehnung an die DSR-Zyklen für Relevanz, Rigor und Design formuliert, um die einbezogenen Publikationen systematisch für den Aufbau und die Entwicklung von Designwissen zu

integrieren. Zur Umsetzung dieses Forschungsvorhabens wurden im Laufe dieser Dissertation verschiedene Forschungsmethoden in den Publikationen angewandt. Dazu gehören systematische Literaturrecherchen, semi-strukturierte ExpertInnen-Interviews, Methoden zur Entwicklung von Taxonomien, standardisierte Benutzertests sowie qualitative und quantitative Methoden zur Datenerhebung und –analyse. Für die Benutzertests wurden Designartefakte zur Demonstration und Evaluierung der Validität des Designs und dessen Nutzens als simuliertes Proof-of-Concept oder als Prototypen instanziiert, wobei entweder die Wizard-of-Oz-Methode oder voll funktionsfähige KI-Funktionen eingesetzt wurden.

Ergebnisse

Das zentrale Ergebnis dieser Dissertation ist validiertes Gestaltungswissen zur Entwicklung einer Gestaltungstheorie für das Design und die Entwicklung von humanzentrierten HIS. Daher wird in den einbezogenen Publikationen präskriptives Gestaltungswissen in Form von mehreren Designprinzipien (DPs) und instanziierten Designartefakten entwickelt. Diese DPs sind überwiegend in der OKS Anwendungsdomäne kontextualisiert. Basierend auf den einzelnen DPs werden in der Dissertation sechs abstrakte DPs zur Gestaltung von humanzentrierten HIS aufgeführt, die die Form der KI-Erscheinung, Augmentierungsstrategien und die Integration von motivierenden Elementen berücksichtigen. Die Anwendung verschiedener Methoden für die Evaluation der instanziierten Designartefakte liefert zusätzliche Ergebnisse in Bezug auf die Validität und Nützlichkeit der DPs sowie zum Nutzungsverhalten und Lernfortschritt, zur weiteren Nutzungsabsicht, wahrgenommenen Menschlichkeit, Nützlichkeit und zur Aufgabenbearbeitung. Zur Untermauerung des validierten Gestaltungswissens trägt die Dissertation Konzeptualisierungen hybriden Lernens, ein Framework für die Gamification von Kollaborationsprozessen und eine Taxonomie für die Integration von KI in den Kundenservice bei.

Beitrag

Mit dem Design und der Entwicklung von humanzentrierten HIS, die hybrides Lernen und verbesserte Aufgabenbearbeitung ermöglichen, leistet die Dissertation einen Beitrag zur Forschung im Bereich der Mensch-Computer-Interaktion, insbesondere für Benutzerinteraktion mit KI-gestützten Systemen für Mensch-KI-Kollaboration innerhalb eines soziotechnischen Systems. Dies beinhaltet Beiträge, die sich mit (1) dem Benutzer, (2) dem System, (3) der Aufgabe und dem Kontext und (4) der Interaktion und den Ergebnissen von HIS befassen. Dem DSR-Ansatz entsprechend sind diese Beiträge zweigeteilt, d. h. die Dissertation trägt sowohl zur Forschung als auch zur Praxis bei. Erstens leistet die Dissertation einen Beitrag zur Konzeptualisierung hybriden Lernens. Darauf aufbauend wird eine neue Perspektive zur Unterscheidung zwischen Novizen und Experten vorgestellt und evaluiert, um einen impliziten Wissenstransfer durch HIS zu ermöglichen. Anwender können sich dieses Konzept zunutze machen, um das Lernen zu unterstützen und die Aufgabenbearbeitung durch HIS zu verbessern. Zweitens wird ein

Framework für die Gamification von Kollaborationsprozessen entwickelt und erfolgreich sowohl auf die Mensch-Mensch-Kollaboration als auch auf die Mensch-KI-Kollaboration angewendet. Darüber hinaus zeigt die Verwendung des Frameworks für ein HIS, dass die Vermenschlichung von KI als ein Gamification-Element betrachtet werden kann, das den Teamgeist stärkt. Anwender können das Framework in einer Top-Down-Weise nutzen, um die Implementierung von Gamification in HIS zu unterstützen. Drittens trägt die Dissertation zur KI-Forschung im OKS bei, indem sie eine Taxonomie für die Integration von KI in den Kundenservice entwickelt und evaluiert. Auf der Grundlage dieser Taxonomie werden bestehende KI-Integrations-Archetypen bestätigt und neue Archetypen identifiziert. Sie dient sowohl Forschern als auch Anwendern als Unterstützung für die Planung oder Analyse ihrer KI-Integrationsvorhaben entlang der Dimensionen in sequentieller Reihenfolge. Die Taxonomie, die erfolgreich für drei Anwendungsfälle verwendet wurde, bildet die kontextuelle Grundlage für das entwickelte Gestaltungswissen. Viertens umfassen die Beiträge zur Interaktion und zu den Ergebnissen das validierte, präskriptive Gestaltungswissen in Form von DPs und instanziierten Designartefakten, die zu einer Gestaltungstheorie beitragen. In diesem Hinblick stellt die Dissertation sechs abstrakte DPs vor, um die kausalen Zusammenhänge des Gestaltungswissens im Sinne des Forschungsziels zu ermöglichen. Darauf aufbauend wird eine vorläufige Gestaltungstheorie für humanzentrierte HIS vorgestellt.

Limitationen

Die Dissertation unterliegt einigen Einschränkungen in Bezug auf die Wissensquellen, den Fokus auf humanzentriertes Gestaltungswissen, die Evaluierungsmethoden und die kontextuelle Einordnung. Erstens schränken die Auswahl der Stakeholder, ExpertInnen und TeilnehmerInnen für die Interviews und Nutzertests sowie die Festlegung der wissenschaftlichen Datenbanken und Suchbegriffe für systematische Literaturrecherchen die Ergebnisse der jeweiligen Forschungsmethode ein. Zweitens kombinieren HIS HITL und CITL, um Interaktionen zur gegenseitigen Augmentierung zu ermöglichen. Obwohl das in dieser Dissertation entwickelte Gestaltungswissen HITL beinhaltet, zielt der Fokus auf humanzentriertes Design darauf ab, das Lernen und die Aufgabenbearbeitung durch CITL zu unterstützen. Die Optimierung oder Bewertung der technischen Leistung im Rahmen von HITL war daher nicht Gegenstand dieser Dissertation. Drittens bringt der Einsatz qualitativer Evaluationsmethoden einige Einschränkungen mit sich, zum Beispiel die Anzahl der instanziierten Designartefakte und die ausgewählten Anwendungsfälle im OKS. Darüber hinaus sind die Auswirkungen auf das hybride Lernen und die Aufgabenbearbeitung durch die Simulation von Kunden oder KI im Rahmen von kurzzeitigen semi-naturalistischen Evaluationen eingeschränkt. Für die Entwicklung einer soliden Gestaltungstheorie ist eine tiefergehende Evaluation notwendig. Viertens sind die Ergebnisse überwiegend durch die kontextuelle Einordnung im OKS eingeschränkt.

Ausblick

Zusätzlich zu den Beiträgen zeigt die Dissertation mehrere Möglichkeiten für zukünftige Forschung zu (1) Interaktion und Ergebnissen, (2) Systemcharakteristika, (3) Nutzereigenschaften und (4) Aufgabe und Kontext in Bezug auf die Nutzerinteraktion mit KI-gestützten Systemen für Mensch-KI-Kollaboration in HIS auf. Erstens können langfristige, quantitative und naturalistische Evaluationen des entwickelten Gestaltungswissens die Einschränkungen der Evaluationsmethodik ausgleichen. Weiterhin könnten vertiefende Untersuchungen von HITL innerhalb von HIS aus technischer Sicht die Ergebnisse dieser Dissertation ergänzen. Zukünftige Forschungsarbeiten könnten auch einen Beitrag zur deskriptiven Forschung über hybrides Lernen sowie zu einem Evaluierungsframework für HIS leisten. Zweitens bedarf das Framework für die Gamification von Kollaborationsprozessen weiterer Forschung, um andere Lösungen als Gamification in Betracht zu ziehen, um die Teilnahme an Kollaborationsprozessen zu motivieren, und um eine deskriptive Perspektive der Gamification-Elemente und ihrer jeweiligen Auswirkungen zu ermöglichen. Zudem gibt es bei den Systemcharakteristika noch ungenutztes Potenzial für die Untersuchung der Beziehung zwischen Mensch und KI innerhalb eines HIS. Drittens könnte die künftige Forschung im Hinblick auf die Benutzereigenschaften deskriptives Wissen über Gestaltungsentscheidungen und deren Auswirkungen auf das Lernen und die Aufgabenbearbeitung liefern. Viertens regt die Dissertation in Bezug auf Aufgabe und Kontext weitere Forschungen zur Integration von KI in den Kundenservice an. So könnten Forscher beispielsweise auf die Taxonomie der KI-Integration in den Kundenservice zurückgreifen und auch deskriptives Wissen über Designentscheidungen und deren Auswirkungen in spezifischen Kontexten beitragen. Schließlich könnte die Anwendung des erarbeiteten Gestaltungswissens für HIS in anderen Organisationskontexten die Verallgemeinerung unterstützen.

Table of Contents

I. List of Figures

II. List of Tables

III. List of Abbreviations

AI .... Artificial intelligence
CA .... Conversational agent
CASA .... Computers are social actors
CITL .... Computer-in-the-loop
CW .... Collaborative writing
DF .... Design feature
DP .... Design principle
DSR .... Design science research
HI .... Hybrid intelligence
HIS .... Hybrid intelligence system
HITL .... Human-in-the-loop
IS .... Information system
MDA .... Mechanics, dynamics, aesthetics
ML .... Machine learning
MR .... Meta-requirement
OCS .... Online customer service
RG .... Research goal
RQ .... Research question
SE .... Service employee
SLR .... Systematic literature review
UI .... User interface
WOz .... Wizard of Oz

1 Introduction

1.1 Motivation and Problem Statement

Organizations are increasingly leveraging the analytical capabilities of artificial intelligence (AI) to improve their business operations and services in terms of speed, efficiency, and accuracy (Ghosh et al. 2019). To achieve this, AI makes use of machine learning (ML) techniques to understand and react to its environment (Dellermann et al. 2019b; Russell and Norvig 2016). It processes information gathered as input to recognize patterns and make predictions directed toward actionable output (Dellermann et al. 2019b; Dellermann et al. 2019a; Kaplan and Haenlein 2019). In domains with a huge amount of available training data, AI has already achieved notable successes in improving the efficiency and effectiveness of algorithms and results (e.g., in speech recognition, recommender systems, or autonomous vehicles) (Holzinger 2016a, 2016b; Holzinger et al. 2017). With these developments, AI has become an extensive and influential research field in computer science. A major focus here is placed on advancing the underlying ML processes with the aim of creating artificial general intelligence that resembles human intelligence (Dellermann et al. 2019b; Dellermann et al. 2019a; Holzinger 2016a). Yet, the development of such AI is still a long way off (Dellermann et al. 2019b; Dellermann et al. 2019a). Therefore, the focus of AI is usually narrowed to specific application domains and requires intensive training (Ghahramani 2015). Such training is usually based on traditional automatic ML (Holzinger 2016a; Holzinger et al. 2016, 2017) and requires an ML expert as a mediator. This expert collects the data from the end users, processes it, and creates design parameters as inputs for learning. Thus, the development, training, and assessment of AI is led by the ML expert. For further iterations, the end users' involvement is limited to data provision and feedback, leading to asynchronous iterations as well as the scant influence of the end users on the resulting systems (Amershi et al. 2014). Hence, a new and alternative approach to automatic ML involves and focuses on the end user in the learning process, eliminating the need for asynchronous cycles with an ML expert. This, in turn, enables more flexible, focused, and faster learning cycles, as well as promotes higher trust and acceptance among end users. This approach is defined as human-in-the-loop (HITL) since it involves a human in the learning process of AI. It encompasses interactive ML, which builds on various ML concepts involving humans (Gaurav 2016; Holzinger 2016a, 2016b; Martínez et al. 2019). This is specifically valuable in domains with less available training data, where ML is stretched to its limits (i.e., starting with only a few training examples). In such domains, HITL learning is faster and leads to the early availability of high-quality datasets (Holzinger 2016a, 2016b; Holzinger et al. 2017; Yimam et al. 2016). It is evident that research emphasizes the indispensability of human knowledge and intelligence in system design (Holzinger 2016a, 2016b).

Information systems (ISs) research recognizes the need for human intelligence and investigates human-computer interactions to allow for human-centered socio-technical system design (Bittner et al. 2019a; Seeber et al. 2020). In such research, interdisciplinary research streams related to human-computer interaction within socio-technical systems need to be considered (Rzepka and Berger 2018), including studies on such systems, their users, or task characteristics (Amershi et al. 2014; Chevalier-Boisvert et al. 2019; Dellermann et al. 2019a; Holzinger 2016a; Holzinger et al. 2016, 2017; Martínez et al. 2019; Rzepka and Berger 2018; Yimam et al. 2016; Yimam and Biemann 2018). There are two sides to human-centered system design underlying user interaction with AI-enabled systems. First, there is HITL, which focuses on ML improvement through human involvement. Second, Shneiderman (2020) introduced the so-called computer-in-the-loop (CITL) approach, which addresses the learning and support that human users receive by interacting with AI (Dellermann et al. 2019b; Shneiderman 2020). With CITL, AI can provide human users with predictions, greater transparency, and feedback. Based on these AI outputs, the human user can make or adapt decisions leading to higher efficiency, speed, and accuracy (Abdel-Karim et al. 2020; Amershi et al. 2014). Additionally, the human can learn from AI, which, in turn, leads to optimized human input and the enhanced task performance of a socio-technical system. Finally, CITL and HITL are complementary counterparts (Abdel-Karim et al. 2020), leveraging the strengths of either humans or AI to augment the respective other in terms of learning and task performance.

Researchers in the field of human-computer interaction have previously investigated opportunities to complement the strengths of artificial and human intelligence by tackling the research stream of human-AI collaboration. It facilitates the combination of the intuitive capabilities of human intelligence (e.g., empathy, flexibility, and creativity) with the analytical capabilities of AI (e.g., consistency, speed, and efficiency) to achieve better results (Dellermann et al. 2019b). Thus, instead of replacing humans, AI joins forces with them to create a team (Wilson and Daugherty 2018). Initial studies already exist on human-AI teaming or hybrid teamwork, respectively, for example, on designing collaborative agents (Strohmann et al. 2019) and conversational agents (CAs) in collaborative work (Bittner et al. 2019b; Poser and Bittner 2020), machines in creative teams (Przybilla et al. 2019), machines as teammates within the scope of a research agenda (Seeber et al. 2020), as well as collaborative intelligence as a concept (Epstein 2015).

Recently, HITL, CITL, and the two in combination have been taken into closer consideration for the strengthening of human-AI collaboration through the mutual augmentation of artificial and human intelligence to facilitate hybrid learning and enhanced task performance. Based on this idea, the concept of hybrid intelligence (HI) emerged from human-AI collaboration research and was defined by Dellermann et al. (2019b, p. 640): "*[HI] is the ability to achieve complex goals by combining human and artificial intelligence, thereby reaching superior results to those each of them could have accomplished separately, and continuously improve by learning from each other.*" A hybrid intelligence system (HIS) is a socio-technical IS that includes humans and AI and enables their co-evolvement based on HI. Previous research on HISs has focused on human-AI collaboration and teaming related to HI, for example, a taxonomy of design knowledge for HIS (Dellermann et al. 2019a), use cases for human-AI teaming (Dubey et al. 2020), hybrid human-AI collaboration in clinical decision-making (Hun Lee et al. 2021), design principles for HI ML algorithms focusing on trust (Ostheimer et al. 2021), and a research agenda for HI (Akata et al. 2020). Nevertheless, due to the novelty of the HI concept, existing research is still scant and fragmentary, either addressing the broad concept, specific applications, or focusing exclusively on either HITL or CITL. However, with the shift to more human-centered approaches in human-AI interactions, research on HIS lacks systematic and integrated human-centered design knowledge on the actual interactions of humans and AI combining HITL and CITL within HIS to facilitate continuous mutual augmentation.

To address this research gap, this dissertation contributes prescriptive knowledge for designing and developing a socio-technical artifact – specifically, an HIS – by conducting design science research (DSR) (Gregor and Hevner 2013; Hevner et al. 2004; March and Smith 1995). Therefore, the associated research endeavors seek the relevance and rigor of the developed design knowledge, ensuring consideration of the needs of the environment and existing knowledge from the knowledge base. The environment is initially approached to conceptualize the problem space in terms of context and goals with an eye to projectability (Gregor and Hevner 2013; Hevner et al. 2004; Maedche et al. 2019; Venable 2006; vom Brocke et al. 2020). Following Maedche et al. (2019), to contextualize the problem space, stakeholders need to be identified that are representative of the environment and involved throughout the DSR project to ensure relevance. By means of accomplishing complex goals, problem-solving, and natural language processing (Dellermann et al. 2019b; Russell and Norvig 2016), organizations are making use of AI for several business operations (e.g., candidate selection, process planning, fraud detection, and customer interaction). Customer service, however, is one of the most prominent areas significantly exploited by AI research and technologies, particularly in the text-based online customer service (OCS) frontstage (e.g., self-service technology or chatbots taking over service encounters). Especially with the recent rise of the “service encounter 2.0” (Larivière et al. 2017) and the introduction of frontline service technology infusion archetypes (Keyser et al. 2019) involving augmentation scenarios that facilitate the collaboration of service employees (SEs) and AI, there is great potential in and need for the implementation of HISs in text-based OCS. This further informs the non-context-sensitive goal of the dissertation, which is to develop validated design knowledge for human-centered HISs that ensure interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and enhanced task performance.

1.2 Research Goal and Research Questions

Derived from the motivation and problem statement, the overall research goal (RG) is defined as follows:

> **RG: Develop validated design knowledge for human-centered HISs that ensure interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and enhanced task performance.**

The RG is approached within the scope of a cumulative dissertation representing an overarching DSR project. This includes eight interconnected publications with individual RGs, questions, and contributions. While each publication can stand on its own, they all contribute to the accumulation and evolution of design knowledge to achieve the overall RG of the DSR project.

To properly integrate the individual publications in the context of the overall DSR project and address the RG, six research questions (RQs) have been constructed and formulated, following Thuan et al. (2019). These are mapped to the three DSR cycles, thus ensuring appropriate activities for relevance, rigor, and design in building and evaluating design knowledge (Hevner et al. 2004; Hevner 2007) (Figure 1).

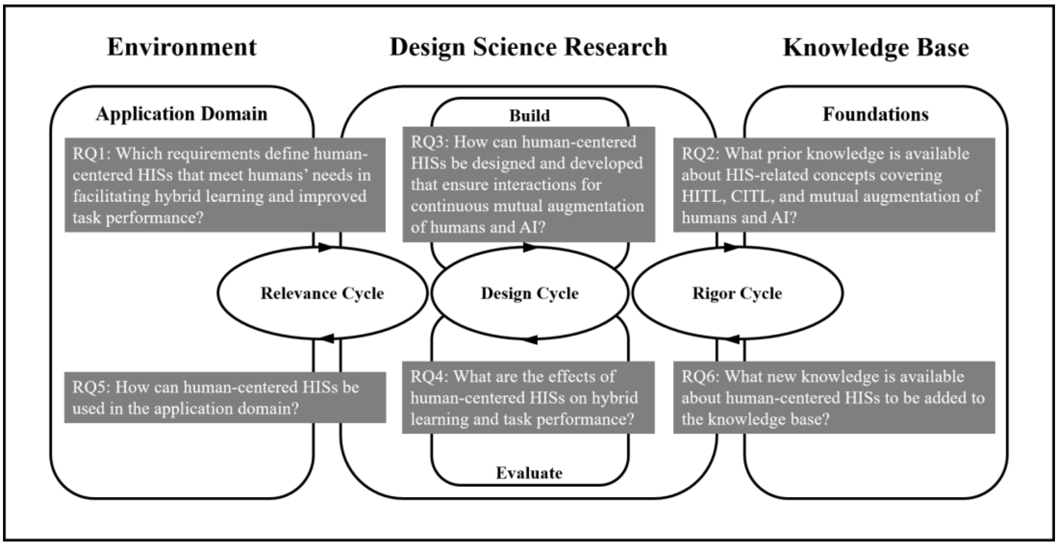

Figure 1. DSR cycles relevant to the RQs (Hevner et al. 2004; Hevner 2007; Thuan et al. 2019)

In DSR, design knowledge connects solutions from the solution space (i.e., in the form of artifacts) to the problem space. The elaboration of both the problem and solution spaces is essential to balance the trade-off between the projectability and fitness of design knowledge to any context (Maedche et al. 2019; Venable 2006; vom Brocke et al. 2020). To initiate the DSR project, the relevance cycle defines and examines the application context and related requirements for the solution space based on the problem space (Hevner 2007), as queried by RQ1:

<u>RQ1</u>: Which requirements define human-centered HISs that meet humans' needs in facilitating hybrid learning and improved task performance?

Three of the included publications address this RQ by identifying meta-requirements (MRs) as objectives (Peffers et al. 2007) or suggestions (Kuechler and Vaishnavi 2012) for a solution from humans in the application domain of customer service (i.e., SEs) (Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022). To investigate the OCS environment, the fourth publication (Poser et al. 2022b) applies an inductive approach to investigating real-world solutions (Lösser et al. 2019; Nickerson et al. 2013).

To prevent the development of design knowledge in terms of routine design that is only relevant to the contextual environment, the rigor cycle complements the relevance cycle. By grounding on appropriate foundations and applying methods from the knowledge base, rigor ensures actual contributions to the knowledge base. Accordingly, RQ2 addresses the study of prior knowledge:

<u>RQ2</u>: What prior knowledge is available about HIS-related concepts covering HITL, CITL, and mutual augmentation of humans and AI?

Analogous to the inductive approach, Poser et al. (2022b) adopt a deductive approach by conducting a systematic literature review (SLR) on existing scientific knowledge about the integration of AI into customer service (vom Brocke et al. 2015; Webster and Watson 2002). In addition to scientific knowledge on OCS, Wiethof and Bittner (2021) conduct an SLR on combining CITL and HITL to facilitate HI and hybrid learning. Elsewhere, Wiethof et al. (2021a) draw on research on motivation and meaningful engagement for gamifying collaboration processes, which serves as a foundation for gamifying HITL in HIS (Wiethof et al. 2022b). Moreover, further non-systematic literature reviews yield related work and conceptual and theoretical backgrounds for all the included publications.

By considering the requirements of the contextual environment and grounding on foundations and methods from the knowledge base, RQ3 addresses the design cycle building design knowledge and artifacts:

<u>RQ3</u>: How can human-centered HISs be designed and developed that ensure interactions for continuous mutual augmentation of humans and AI?

RQ3 is addressed by five of the included publications, following a DSR process (Kuechler and Vaishnavi 2012; Peffers et al. 2007) in terms of design, development, and demonstration (Poser et al. 2022a; Wiethof et al. 2021a, 2021b; Wiethof et al. 2022a; Wiethof and Bittner 2022). We derive action- and materiality-oriented design principles (DPs) to facilitate the design and development of an HIS that ensures interactions for continuous mutual augmentation of humans and AI (Chandra et al. 2015). The DPs are further instantiated within prototypes for demonstration.

After building the design knowledge and artifact in the design cycle, the next steps are evaluation and validation (Hevner et al. 2004; Hevner 2007; Thuan et al. 2019). To determine how and what to evaluate, the RG and requirements are reconsidered. HISs should enable interactions for mutual augmentation to ensure hybrid learning and optimized task performance, which is addressed by RQ4:

RQ4: What are the effects of human-centered HISs on hybrid learning and task performance?

To answer this RQ with evaluation confidence (vom Brocke et al. 2020), Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) conduct semi-naturalistic evaluations, deploying their HIS artifacts in the field. We investigate effects on hybrid learning and task performance by means of user test runs (Venable et al. 2012, 2016), qualitative expert interviews (Mayring 2014; Meuser and Nagel 1991; Myers and Newman 2007), and quantitative data gathering for triangulation (Mayring 2001).

As RQ4 closes the design cycle with evaluations, RQ5 closes the relevance cycle by applying the validated design knowledge to the contextual environment:

RQ5: How can human-centered HISs be used in the application domain?

Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) have deployed and evaluated an HIS artifact in OCS and provide design knowledge in the form of prescriptive statements and artifacts. This can be used by organizations to implement an HIS in their OCS to facilitate hybrid learning and enhance task performance. Furthermore, Poser et al. (2022b) contribute a taxonomy of AI integration into customer service to inform organizations' design decisions.

Analogous to RQ5, RQ6 closes the rigor cycle by adding the validated design knowledge to the knowledge base:

RQ6: What new knowledge is available about human-centered HISs to be added to the knowledge base?

The dissertation's DSR project contributes new design knowledge for HISs to the knowledge base by accumulating the individual publications' contributions. By following a DSR process (Kuechler and Vaishnavi 2012; Peffers et al. 2007), Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) contribute design knowledge in the form of prescriptive DPs and instantiated artifacts in the context of OCS. Furthermore, all the publications extend existing knowledge or add new knowledge to the knowledge base on user interaction with AI-enabled systems (Rzepka and Berger 2018) in terms of HISs.

1.3 Thesis Outline

This dissertation is structured as depicted in table 1. It encompasses a wrapper and eight articles published during the dissertation’s research journey.

The wrapper provides an overarching summary of the dissertation journey connecting the individual contributions of each article to address the overall RG and respective RQs. It starts with an introduction in section 1 to motivate the research, state the problem, and formulate the RQs. After the introduction, section 2 elaborates on theoretical concepts, which provide the scientific foundations for the research endeavors of this dissertation. Next, section 3 presents the research design, integrating the publications in the overall dissertation’s DSR project toward addressing the overarching goal. After this, section 4 provides an overview of each included publication. The two sections that follow outline the concluding contributions to theory in section 5 and to practice in section 6. Limitations are considered in section 7 and implications for further research are provided in section 8. The remaining eight sections (9 to 16) each present one publication.

Table 1. Thesis outline

Wrapper	1. Introduction	
	2. Theoretical Foundations	
	3. Research Design	
	4. Publications	
	5. Theoretical Contribution	
	6. Practical Contribution	
	7. Limitations	
	8. Implications for Further Research	
Publications	9. Paper 1	Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI
	10. Paper 2	Designing and Evaluating a Collaborative Writing Process with Gamification Elements: Toward a Framework for Gamifying Collaboration Processes
	11. Paper 3	Hybrid Intelligence – Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review
	12. Paper 4	Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service
	13. Paper 4	Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI
	14. Paper 6	Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions
	15. Paper 7	Gamifying the Human-in-the-Loop: Toward Increased Motivation for Training AI in Customer Service
	16. Paper 8	Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service

2 Theoretical Foundations

This section introduces the theoretical foundations upon which the dissertation and its included publications are built. It presents the current state of research and derives the research gaps motivating the RG to develop validated design knowledge for HISs to ensure interactions for continuous mutual augmentation of humans and AI.

Taking into account the definition of an HIS, the concept of HI derives from the area of human-AI interaction, specifically, human-AI collaboration (Dellermann et al. 2019a). Human-human collaboration is commonly and frequently applied in both private as well as business settings, as by working together, humans can achieve goals that one individual cannot achieve alone (i.e., together they can complement each other's resources, insights, and skills) (Leimeister 2014; Traumer et al. 2017). As such, collaboration constitutes the work of two or more individuals on the same task to achieve a common goal (Leimeister 2014). They complete the underlying task, which is central to their collaboration, with communication, coordination, and cooperation (Bittner et al. 2019a; Dellermann et al. 2019a; Leimeister 2014). By working together interdependently toward the same goal, such individuals form a team (Leimeister 2014), which is characterized by a functional structure, shared goals and values, mutual interactions between team members, and team spirit (Forster 1978).

Similarly, "*[t]he goal of hybrid intelligence is to create superior results through a collaboration between humans and machines*" (Dellermann et al. 2019a, pp. 277-278). Hence, HI incorporates the concept of collaboration and, in line with team characteristics (Forster 1978), addresses the complementation of humans and AI with their functional strengths as they work toward the same goals, as well as interact with each other through mutual augmentation (Akata et al. 2020; Dellermann et al. 2019b). Mutual augmentation, which leverages the combination of artificial and human intelligence, is central to HI (Akata et al. 2020; Dellermann et al. 2019b; Traumer et al. 2017). In their research endeavor investigating the future of work involving humans and machines, Traumer et al. (2017) present a taxonomy of interaction types (i.e., machine-machine, machine-human, and human-human) that accords with the level of task complexity. Although organizations increasingly implement AI within the scope of human workforce replacement to enhance efficiency, there are work tasks that greatly benefit from and even require putting the human in the loop of AI (i.e., HITL). Accordingly, in terms of mutual augmentation, the human can benefit from CITL within the scope of decision support, learning, and task performance (Akata et al. 2020; Dellermann et al. 2019b; Dellermann et al. 2019a; Traumer et al. 2017; Wiethof and Bittner 2021). The combination of HITL and CITL enables human-AI collaboration to achieve HI. Based on this, an HIS constitutes a socio-technical IS consisting of the technical sub-system (i.e., the AI), the social sub-system (i.e., the human), and their aligned interactions to achieve a certain goal of an underlying task (Koch and Gross 2006; Leimeister 2014). Although there is extant literature on human-computer interaction, including user interaction with AI-enabled systems (Rzepka and Berger 2018; Zhang and Li 2005), knowledge on human-centered HISs that combine HITL and CITL to facilitate interactions for mutual augmentation is lacking. As HI “*combine[s] the complementary strengths of heterogeneous intelligences (i.e., human and artificial agents) into a socio-technological ensemble*” (Dellermann et al. 2019a, p. 276), interdisciplinary research into human-centricity needs to be considered. To provide a systematic and structured theoretical foundation, the theoretical concepts of this dissertation build on the adapted framework of human-computer interaction for user interaction with AI-enabled systems (Rzepka and Berger 2018; Zhang and Li 2005) to facilitate the design of a socio-technical human-centered HIS (Figure 2).

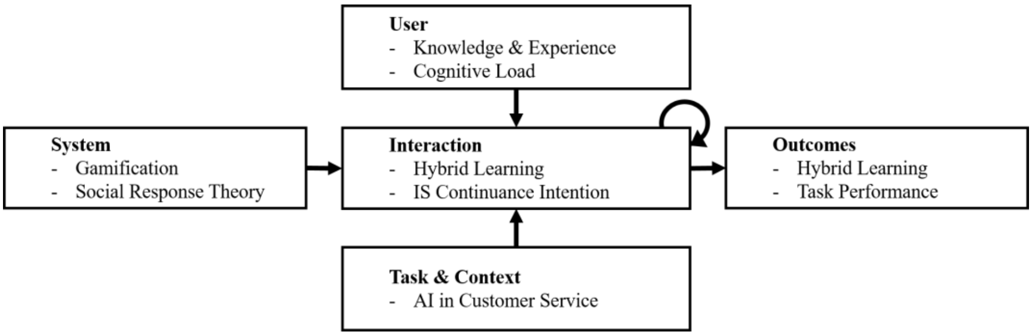

Figure 2. Framework of human-computer interaction (Rzepka and Berger 2018; Zhang and Li 2005), adapted

Considering the knowledge gap regarding the actual interactions of humans and AI that combine HITL and CITL within an HIS for continuous mutual augmentation, the **interaction** construct covers the core concepts of this dissertation, namely, hybrid learning (Dellermann et al. 2019a; Kulesza et al. 2015; Wiethof and Bittner 2021, 2022) and continuance intention to use (Bhattacherjee 2001; Bhattacherjee et al. 2008), underlying users’ perceptions, attitudes, intentions, and behavior (Rzepka and Berger 2018). In alignment with the definition of HI, the enhancement of artificial and human intelligence in terms of hybrid learning and superior results in terms of task performance determine the **outcomes** of the interactions (Dellermann et al. 2019a). To leverage system and user characteristics for successful interaction, the dissertation further draws on social response theory and gamification for the **system,** as well as the cognitive load, knowledge, and experience of the **user**. Within the scope of the contextual research environment, research on AI in customer service is taken into account for the construct of **task and context**.

2.1 Interaction and Outcomes

2.1.1 Hybrid Learning

The core interactions of an HIS are the mutual augmentation activities of human and AI through HITL and CITL (Dellermann et al. 2019b; Dellermann et al. 2019a; Wiethof and Bittner 2021) (Figure 3).

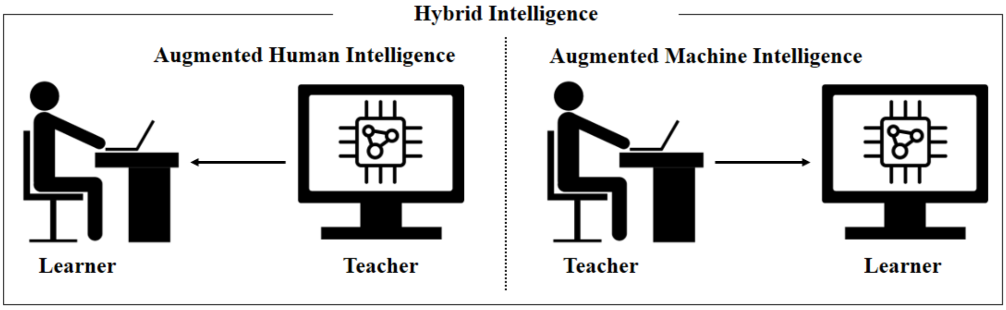

Figure 3. HI including human and machine intelligence (Dellermann et al. 2019b), adapted

Based on mutual augmentation, humans and AI can continuously learn from each other and improve (Dellermann et al. 2019a). On the one hand, HITL derives from the idea of improving AI in terms of efficiency and speed by involving the end user in the learning process (Amershi et al. 2014; Holzinger 2016a, 2016b; Holzinger et al. 2019; Martínez et al. 2019). On the other hand, CITL is based on the idea of putting humans at the center of AI (Shneiderman 2020) to accrue learning benefits to the human (Schneider 2020). Within the scope of augmenting human intelligence, learning can be regarded from different perspectives. For exmaple, some researchers specifically investigate agent tutors for human intelligence (Chen et al. 2022; Chhibber and Law 2019; Giraffa and Viccari 1998; Hjorth 2021; Wambsganss et al. 2021). However, within HISs, augmentation usually happens on the job during collaboration with AI. Thus, instead of direct teaching, it occurs in the form of decision support, exploration, or integration (Wiethof and Bittner 2021). First, regarding decision support, AI can provide a prediction with an accuracy rate to allow the human to adapt the decision (Luong et al. 2019), analyze data, and provide useful information and insights for the human to derive implications (Paschen et al. 2020), or give explanations about its reasoning and predictions (Kulesza et al. 2015). Second, regarding exploration, AI might provide results that do not have the highest accuracy or score but enable the human to gain new or different insights (McCamish et al. 2017; Oliveira et al. 2020; Salam et al. 2019; Smith et al. 2018). Third, some researchers integrate artificial and human intelligence, making HITL and CITL inseparable (e.g., by aggregating the predictions of the human and AI) (Dellermann et al. 2017). Eventually, for each form of augmentation, there are bidirectional interactions with respect to the input and output of either the AI (HITL) or the human (CITL) that facilitate hybrid collaborative learning (Wiethof and Bittner 2021). As such, hybrid learning resembles one outcome of the interactions (Dellermann et al. 2019b; Dellermann et al. 2019a; Rzepka and Berger 2018). In addition, one must not forget the outcome regarding task performance. In fact, task performance is the main driver for organizations to implement AI in their operations to achieve better results and efficiency (Dellermann et al. 2019a; Ghosh et al. 2019; Rzepka and Berger 2018). In line with hybrid learning, Schneider (2020) suggests that optimizing human inputs can also lead to improved task performance.

However, in the available body of research on HI and HISs, the integration of these central aspects is so far neglected (e.g., researchers focusing on enhanced task performance miss the aspect of continuous improvement through hybrid learning) (Dubey et al. 2020; Hun Lee et al. 2021; Ostheimer et al. 2021). This might stem from the scant and fragmentary research available, highlighting the need for a comprehensive understanding of HI, as well as for design knowledge to guide the design and development of HISs for hybrid learning and enhanced task performance (Poser et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2021, 2022).

2.1.2 IS Continuance Intention

Continuous bidirectional interactions for augmentation are essential for any HIS. To enable the continuity of iterative interactions, there is a need for users to exhibit a high level of intention to continue working with AI within HIS (Dellermann et al. 2019b). To address this, the dissertation builds on the IS continuance model (Bhattacherjee 2001; Bhattacherjee et al. 2008) based on the expectation-confirmation theory of Oliver (1980) (Figure 4).

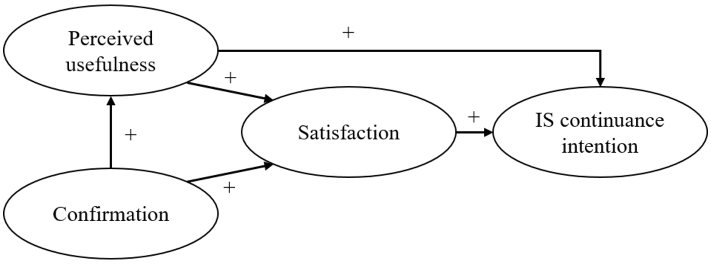

Figure 4. A post-acceptance model of IS continuance (Bhattacherjee 2001), adapted

According to the model in figure 4, a high IS continuance intention is achieved through high satisfaction with IS use and perceived usefulness. User satisfaction itself is further influenced by confirmation of expectation and perceived usefulness. The latter is also influenced by the confirmation level (Bhattacherjee 2001). Concerning the design and development of HISs, users' expectations need to be identified to facilitate the usefulness of an HIS. Eventually, if users intend to continue using the system for both exploiting (CITL) and teaching (HITL) AI, the iterative nature of a hybrid collaborative learning cycle within the scope of mutual augmentation can be ensured (Wiethof and Bittner 2021).

2.2 System Characteristics

2.2.1 Gamification

Building on the work of Bhattacherjee and Premkumar (2004), Lowry et al. (2015) established the multimotive IS continuance model, which integrates motivational constructs into the IS continuance model. Hence, users' satisfaction, performance, and continuance intention to use are influenced by underlying motivations (i.e., hedonic, intrinsic, and extrinsic motivations) (Lowry et al. 2015). One possible solution to trigger different kinds of motivation is gamification (Darejeh and Salim 2016; Lowry et al. 2015, 2015; Steffens et al. 2015). Deterding et al. (2011a, p. 10) define gamification as "*the use* *of game design elements in non-game contexts*." As such, in the context of collaboration, Marczak et al. (2015) and Steffens et al. (2015) investigated the use of gamification elements within software collaboration teams. They thereby created a framework matching gamification elements with desired behaviors and showed the potential of gamification to motivate a team in its collaboration endeavors. However, there are several gamification elements, such as points, leaderboards, levels, or badges, specifically designed for increasing extrinsic motivation (i.e., external rewarding) that are criticized based on the overjustification effect (Charms 1983), which states that increasing users' extrinsic motivation will decrease their intrinsic motivation (i.e., their satisfaction with the work itself) (Charms 1983; Deci and Ryan 2000; Meske et al. 2015). Consequently, users might focus on quantity to gain points and thus neglect quality (Meske et al. 2015). To avoid the occurrence of the overjustification effect, gamification needs to invite, challenge, and allow users to meaningfully engage in their work (Hunicke et al. 2004; Liu et al. 2017; Suh et al. 2017). After all, gamification supports achieving a goal or outcome in a non-game context (e.g., education, workplace, or healthcare) (Darejeh and Salim 2016; Deterding et al. 2011a). Meaningful engagement will then enable the user to feel "*a sense of meaning and more deeply understand the essence of the experienced events*" toward continuance intention to use (Suh et al. 2017, p. 270). To enable meaningful engagement through gamification, there are several theoretical concepts to consider, such as the mechanics, dynamics, and aesthetics (MDA) framework (Hunicke et al. 2004), the concept of aesthetic experience (Suh et al. 2017), the framework for designing and researching gamified systems (Liu et al. 2017), and the differentiation of gameful and playful interactions (Tseng and Sun 2017).

Regarding employing motivation and gamification models to increase continuance intention to use, there is a need for integrated design knowledge on gamifying IS in collaboration processes (Briggs et al. 2010; Richter et al. 2018; Seeber et al. 2020; Wiethof et al. 2021a). When developing HISs, such knowledge may be beneficial to facilitate humans' motivation and continuance intention to involve in the loop of AI learning within the scope of HITL (Wiethof et al. 2022b).

2.2.2 Social Response Theory

Apart from gamifying HISs, when considering system characteristics, social response theory, which builds on the "computers are social actors" (CASA) paradigm (Nass et al. 1994; Nass and Moon 2000), should not be neglected. It predicates that "*individuals are responding mindlessly to computers to the extent that they apply social scripts […] that are inappropriate for human-computer interaction. […] [I]ndividuals must be presented with an object that has enough cues to lead the person to categorize it as worthy of social responses, while also permitting individuals who are sensitive to the entire situation to note that social behaviors were clearly not appropriate*" (Nass and Moon 2000, p. 83). Therefore, rudimentary cues are already sufficient, for example, names, emotions (Frick

2015), or typing indicators (Gnewuch et al. 2018). Consequently, when technology is equipped with social cues to make it more human, humans will apply social behaviors and rules when dealing with the technology (Nass et al. 1994; Nass and Moon 2000). It can positively impact perceptions of social presence (Robb and Lok 2014) as well as relationship-building between humans and technology (Elshan and Ebel 2020; Nass and Moon 2000). As such, much extant research draws on this theory, striving for design knowledge to build acceptance of technology as well as human-AI teaming or collaboration. For instance, Elshan and Ebel (2020) developed design knowledge for CAs as teammates by identifying the need for the social orientation of a CA, as well as social conversations. Gnewuch et al. (2017) similarly drew on social response theory to design cooperative and social CAs for customer service. They advised implementing social cues that meet the expectations of a human SE (e.g., friendliness, expertise, appearance, and language). Additionally, it is important that the social cues match the context and the technological capabilities, that is, they should not raise inappropriate expectations. Otherwise, the human user will be frustrated as soon as these expectations are not met (Gnewuch et al. 2017). This phenomenon is known and has been investigated under the concept of the uncanny valley by Mori et al. (2012). It argues for a well-established balance of social cues and capabilities in creating human-likeness for maximum affinity (Gnewuch et al. 2017; Gnewuch et al. 2018; Mori et al. 2012).

Eventually, socio-emotional behavior, as well as human-like attributes imbued in AI, can positively influence the acceptance of and interaction with AI in terms of affection and relatedness (Bittner et al. 2019b; Elshan and Ebel 2020; Nass and Moon 2000; Wiethof et al. 2021b). Accordingly, when investigating HISs, it is necessary to consider social response theory for successful human-AI teamwork and collaboration.

2.3 User Characteristics

2.3.1 Knowledge and Experience

Similar to system characteristics, specific user characteristics have an impact on the interaction construct (Rzepka and Berger 2018), which covers hybrid learning and IS continuance intention toward continuous hybrid learning (Dellermann et al. 2019a). Specifically, in the context of learning, users' task experience plays a relevant role in system use (Rzepka and Berger 2018). In this area, initial research suggests considering different levels of experience and knowledge (i.e., distinguishing between expert and novice users) (Dellermann et al. 2019b; Hu et al. 2019; Liu et al. 2014; Oliveira et al. 2020; Rzepka and Berger 2018; Wiethof and Bittner 2021). In fact, research can leverage this differentiation to enable implicit knowledge transfer from experts to novices (Dellermann et al. 2019b). For example, Liu et al. (2014) developed a workflow-based HITL system that learns from the experience of experts and provides recommendations to novices. Moreover, the time- and location-independent availability of AI ensures constant real-time support for learning (Rao et al. 2020). As such, on-the-job learning within an HIS is analogous to professional development expanding knowledge and skills (Joy-Matthews et al. 2004; Mikołajczyk 2022; Sutherland and Canwell 2004). In the case of novices, this includes general knowledge expansion, learning from experience, and learning from other employees (Sutherland and Canwell 2004). In terms of professional development, even experts can take advantage of hybrid learning (i.e., rebuilding and improving skills, developing new attitudes, and also learning from other employees) (Sutherland and Canwell 2004). When measuring learning, several constructs should be taken into account, including objective learning progress, as well as subjective perceptions of enhanced knowledge, experience, and new insights (Samarasinghe and Tretiakov 2009; Wiethof and Bittner 2022). In the end, for both expert and novice users, hybrid learning can increase productivity, improve human input, and lead to better task performance (Liu et al. 2014; Rao et al. 2020; Schneider 2020).

Accordingly, when designing interactions for mutual augmentation in HISs, the knowledge and experience of human users should be considered to investigate the outcomes in terms of hybrid learning and task performance more effectively. Thus, there is a need to extend the concept of hybrid learning by including and leveraging the differentiation between experts and novices.

2.3.2 Cognitive Load

Expert users are differentiated from novices in that the former have more experience. Hence, they take more advantage of an HIS for purposes of task performance and efficiency (Rzepka and Berger 2018) instead of learning (Wiethof and Bittner 2022). As such, different expectations toward the design of HISs in terms of cognitive style need to be met to ensure a cognitive fit between users and the system. The better the cognitive fit, the more likely the users are to accept the AI's output (e.g., in the form of recommendations), thus enabling learning and enhancing efficiency and task performance (Rzepka and Berger 2018). This demands the investigation of human augmentation by AI (e.g., how information or recommendations are presented to the users). For instance, users often tend to choose the default option (Münscher et al. 2015; Weinmann et al. 2016) and do not base their decisions on rational thinking but instead overly rely on AI (Mele et al. 2021; Münscher et al. 2015; Rzepka and Berger 2018; Weinmann et al. 2016). This might lead to the opposite of intended effects in terms of hybrid learning and task performance (Amershi et al. 2014; Dellermann et al. 2019b; Rzepka and Berger 2018; Seeber et al. 2020). Therefore, users should feel encouraged to leverage their intelligence when collaborating with AI (i.e., by processing and questioning AI output to enable learning as well as enhance task performance).

For adequate system design, accordingly, cognitive load theory provides a valuable foundation as it integrates learning and task performance by means of humans' cognitive processing capacities (Sweller 1988; Sweller et al. 2011). Sweller et al. (2011) define cognitive load as working memory load that exploits humans' cognitive capacities. They divide it into intrinsic, extraneous, and germane cognitive load. Intrinsic cognitive load, on the one hand, "*is imposed by the basic structure of the information […] for achieving […] goals irrespective of the instructional procedures used*" (Sweller et al. 2011, p. 57). Extraneous cognitive load, on the other hand, "*is imposed solely because of the instructional procedures being used*" (Sweller et al. 2011, p. 57). Lastly, "*[u]nlike intrinsic and extraneous cognitive load […], germane cognitive load is not imposed by the learning materials […] [but] can be better understood as working memory resources that are devoted to information that is relevant or germane to learning*" (Sweller et al. 2011, p. 57). Cognitive load theory states that by reducing cognitive load to a manageable level, humans' information-processing capabilities can be enhanced to improve learning and task performance (Sweller 1988; Sweller et al. 2011). While intrinsic cognitive load occurs from the task itself and related requirements, extraneous cognitive load depends on how and how much information is presented. Hence, extraneous cognitive load can be reduced and managed by a suitable system design that frees up working memory capacities in terms of germane cognitive load to facilitate learning and task processing (Sweller 1988; Sweller et al. 2011).

Finally, to ensure a cognitive fit between users and the system, design considerations should include the cognitive load of experts as well as novices for enhanced learning capacities and improved task performance. On the one hand, users need information for learning and task performance; on the other hand, too much information will increase cognitive load and potentially result in overreliance.

2.4 Task and Context

To specify interactions and outcomes in terms of learning and task performance, task and context need to be defined. For this dissertation, the concrete environment is comprised by OCS (i.e., making customer service delivery more efficient and enabling on-the-job learning of SEs). Related work on AI in customer service provides contextual foundations.

With progress in digitalization and advances in technology, customer service has gone through several changes. In fact, "*[information technology] enables service, and its input and output, information, is central to service*" (Rust and Huang 2014, p. 208). Hence, technology plays an important role in the transformation of customer service. As such, it is a promising and popular application field for researchers and practitioners designing and developing technology to make business operations more effective and efficient (Bitner et al. 2000; Ghosh et al. 2019). For customer service, this means enhancing the overall service encounter and the associated customer satisfaction (Bitner et al. 2000). For instance, since service can be defined as "*any direct provision or co-creation of value between a provider and a customer*" (Rust and Huang 2014, p. 207), the provider does not necessarily have to be a physical customer contact personnel (Froehle 2006; Rust and Huang 2014; Verhagen et al. 2014). In fact, technological advances enable new forms of customer contact, independent of time and location (e.g., text-based OCS) (Froehle 2006; Froehle and Roth 2004; Rust and Huang 2014; Verhagen et al. 2014). This includes technology-mediated customer contact, such as chat or email (Froehle 2006; McLean and Wilson 2016; Turel and Connelly 2013), and technology-generated customer contact, such as self-service solutions (Froehle 2006; Meuter et al. 2000; Scherer et al. 2015; Verhagen et al. 2014). Moreover, organizations are increasingly deploying AI for technology-generated customer contact (Huang and Rust 2018; Robinson et al. 2020; Xu et al. 2020). They thereby exploit the advantages of technology and AI capabilities, such as consistency, speed, and efficiency in terms of permanent availability, information accessibility, and data processing (Dellermann et al. 2019b; Rust and Huang 2014). After all, service AI can be defined as "*the configuration of technology 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. 331). However, reducing contact with SEs leads to a decrease in personal touch and individualization, which has a negative impact on overall customer experience and satisfaction (Ameen et al. 2021; Verhagen et al. 2014). To mitigate that risk, organizations use CAs as SE substitutes, making customer service more human-like through conversational interactions and social cues (Følstad and Skjuve 2019; Gnewuch et al. 2017). However, research and practice are still far away from completely replacing human intelligence with AI, confirming the necessity of keeping domain experts in the loop of AI (Raisch and Krakowski 2021). Even domain-specific CAs often fail to handle certain complex customer requests demanding recovery strategies and solutions, such as CA-SE handovers (Benner et al. 2021; Følstad and Skjuve 2019; Poser et al. 2021). Still, handovers only allow for a sequential request escalation.

Similar to the conceptual archetypes of customer contact in relation to technology of Froehle and Roth (2004), new archetypes for technology infusion in frontline customer service, the "service encounter 2.0" (Larivière et al. 2017), have been defined by Keyser et al. (2019) as involving technology, the SE, and the customer (Figure 5). These archetypes can cover several roles of technology, including augmentation, substitution, and support of SEs (Larivière et al. 2017; Ostrom et al. 2019). In the end, the inability of AI to replace SEs, as well as the customers' need for personalization, prompt the involvement of SEs in real-time customer service delivery. Consequently, as depicted in figure 5, (1) the customer is in direct contact with the SE (i.e., human-human interaction), and (2) the SE is augmented by AI and vice versa, invisibly for the customer (i.e., human-AI interaction).

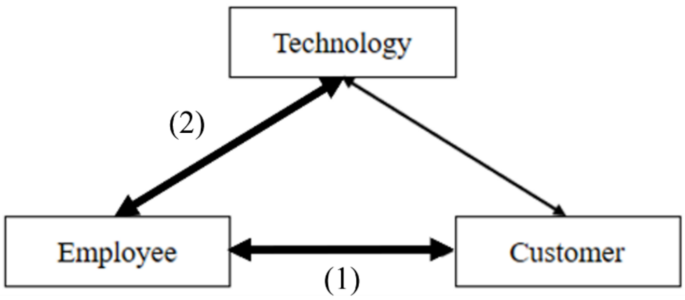

Figure 5. Frontline service technology infusion, augmentation scenario highlighted (Keyser et al. 2019), adapted

This concept is analogous to HI as it enables the combination of both artificial and human intelligence through interactions for mutual augmentation. However, extant research has mainly investigated customer interactions with AI and has only very recently arrived at developing concepts and requirements for augmentation scenarios in OCS. For example, Gao and Jiang (2021) and Hohenstein and Jung (2018) investigated the implementation of CAs augmenting humans with message suggestions. They call for further research to apply their findings to the customer service domain. Any design contributions in OCS so far are fragmentary or limited to specific constructs of hybrid intelligence. For instance, Feine et al. (2020b) focus on how to engage domain experts in CA development, and Graef et al. (2021) specifically investigate the feedback approach. Moreover, they neither implemented nor evaluated their designs in a real-world OCS setting. Similarly, Dubey et al. (2020) established a framework for developing human-AI teaming, which is applied to OCS but only evaluated in terms of applicability for developers.

Finally, the service encounter 2.0 and AI integration for SE augmentation make OCS a promising application domain for HI. Regarding the amount of currently available research throughout the transformations of customer service, there is still a lack of systematic design knowledge on how to integrate AI into customer service, considering various infusion types (Poser et al. 2022b). Also, design knowledge on how to design interactions for mutual augmentation within HISs for text-based OCS is developed in this dissertation. Additionally, different forms of human augmentation by AI are further investigated, such as augmentation through embedded AI within the scope of a dashboard (Dubey et al. 2020; Poser et al. 2022a; Wiethof and Bittner 2022) or augmentation through CAs (Hohenstein and Jung 2018; Meyer von Wolff et al. 2019a; Meyer von Wolff et al. 2019b; Wiethof et al. 2022a).

3 Research Design

This dissertation research project addresses the need for systematic and integrated human-centered design knowledge for interactions for continuous mutual augmentation of humans and AI within an HIS. This design knowledge constitutes a solution for the contextualized problem space, that is, by enhancing efficiency and effectiveness within the scope of hybrid learning and task performance in OCS (Venable 2006). DSR is an IS paradigm specifically dedicated to this class of problems. Unlike behavioral scientists in IS conducting descriptive research to construct theories of analysis, explanation, or prediction, design science researchers work toward a theory for design and action (Baskerville et al. 2018; Gregor 2006). A theory for design and action provides prescriptive design knowledge with utility character on how to address the specific problem, linking representations from the solution space in the form of solution technologies to the problem space (Gregor and Hevner 2013; Venable 2006; vom Brocke et al. 2020) as "*practically applicable means-end conclusions*" (Österle et al. 2011). As such, design is "*the search for an effective artifact […] utilizing available means to reach desired ends while satisfying laws in the problem environment*" (Hevner et al. 2004, p. 83). Artifacts produced by DSR can have the form of constructs, models, methods, or instantiations (Hevner et al. 2004; March and Smith 1995; Österle et al. 2011). Therefore, DSR not only contributes to the scientific knowledge base but also to finding a solution for prevalent problems in a real-world environment (Gregor and Hevner 2013; Hevner et al. 2004). Depending on the level of maturity, the developed design knowledge ranges from well-developed design theory, to nascent design theory, to a situated implementation of an artifact. Nascent design theory can include constructs, methods, models, DPs, and technological rules, which are already sufficient and valuable design knowledge contributions for research and practice (Gregor and Hevner 2013). The same applies to situated implementations (Baskerville et al. 2018). In the end, the relatedness to practice as well as research not only distinguishes DSR from other IS research (Gregor and Hevner 2013) but also makes it a highly suitable paradigm for the research endeavor at hand.

3.1 Design Science Research Context

"*[The] goal of DSR is to invent new artifacts where none exist and to improve existing artifacts to enhance organizational, group, and individual human productivities and effectiveness*" (Baskerville et al. 2018, p. 362).

By conducting DSR, this dissertation generates prescriptive design knowledge for HISs toward a theory for design and action (Gregor 2006) and applies it to the construction of an artifact in terms of building and evaluation (Baskerville et al. 2018; Gregor and Hevner 2013). This HIS artifact is a socio-technical system that includes people, organizations, and technology (Gregor and Hevner 2013; Österle et al. 2011). Thus, DSR activities are not limited to studying the knowledge base of scientific literature but also involve investigating experiences and knowledge in the application environment (Österle et al. 2011). In the context of the overarching DSR project, an HIS is implemented in OCS to improve on-the-job hybrid learning and task performance of customer service delivery.

On the way to accumulating design knowledge for HIS, several artifacts have been built, which are interconnected through the iterative, complementary, and evolving nature of DSR but also stand on their own. Hence, each artifact required an individual evaluation to validate the design knowledge in terms of its utility (i.e., whether it achieves its goal) (Hevner et al. 2004; March and Smith 1995). The different knowledge contributions generated and validated across the publications included in this dissertation were able to build on, enhance, and interconnect with each other for the accumulation and evolution of design knowledge to address the overarching RG (Rothe et al. 2020; vom Brocke et al. 2020) as follows.

First, Wiethof et al. (2021b) motivate the research endeavors by showing the potential of human-AI teaming for building a synergy between humans and AI. Next, Wiethof and Bittner (2021) elaborate on the concept of HI in terms of hybrid learning. We provide a foundation for Wiethof and Bittner (2022), deriving first DPs for an HIS in OCS focused on hybrid learning. This represents the first cycle of a multicyclic DSR sub-project. In the second cycle, Poser et al. (2022a) extend the design knowledge for HIS in OCS with additional DPs to facilitate improved task performance. In addition, we deploy a fully functioning HIS prototype that integrates ML techniques. While Wiethof and Bittner (2022) and Poser et al. (2022a) design and develop an HIS with embedded AI, Wiethof et al. (2022a) contribute design knowledge for an HIS with a CA as user interface (UI). Due to the human-centered approach, the DPs have a strong focus on CITL to facilitate human learning and enhance utility for task performance. Still, we always ensure and implicate HITL with feedback and implicit learning mechanisms. To investigate motivational elements for humans to train AI during work, Wiethof et al. (2022b) apply gamification to HITL. Therefore, we make use of the framework for gamifying collaboration processes by Wiethof et al. (2021a). Finally, Poser et al. (2022b) provide the contextual environment for Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) in terms of interactions for mutual augmentation of SEs and AI in the OCS frontstage.

3.2 Design Science Research Strategy

To enable the accumulation and evolution of design knowledge in one overarching DSR project, the research endeavors are guided by the three DSR cycles emphasizing relevance, rigor, and design (Hevner et al. 2004; Hevner 2007; Thuan et al. 2019). Figure 6 depicts the accumulation of design knowledge mapped to the RQs and DSR cycles of the dissertation.

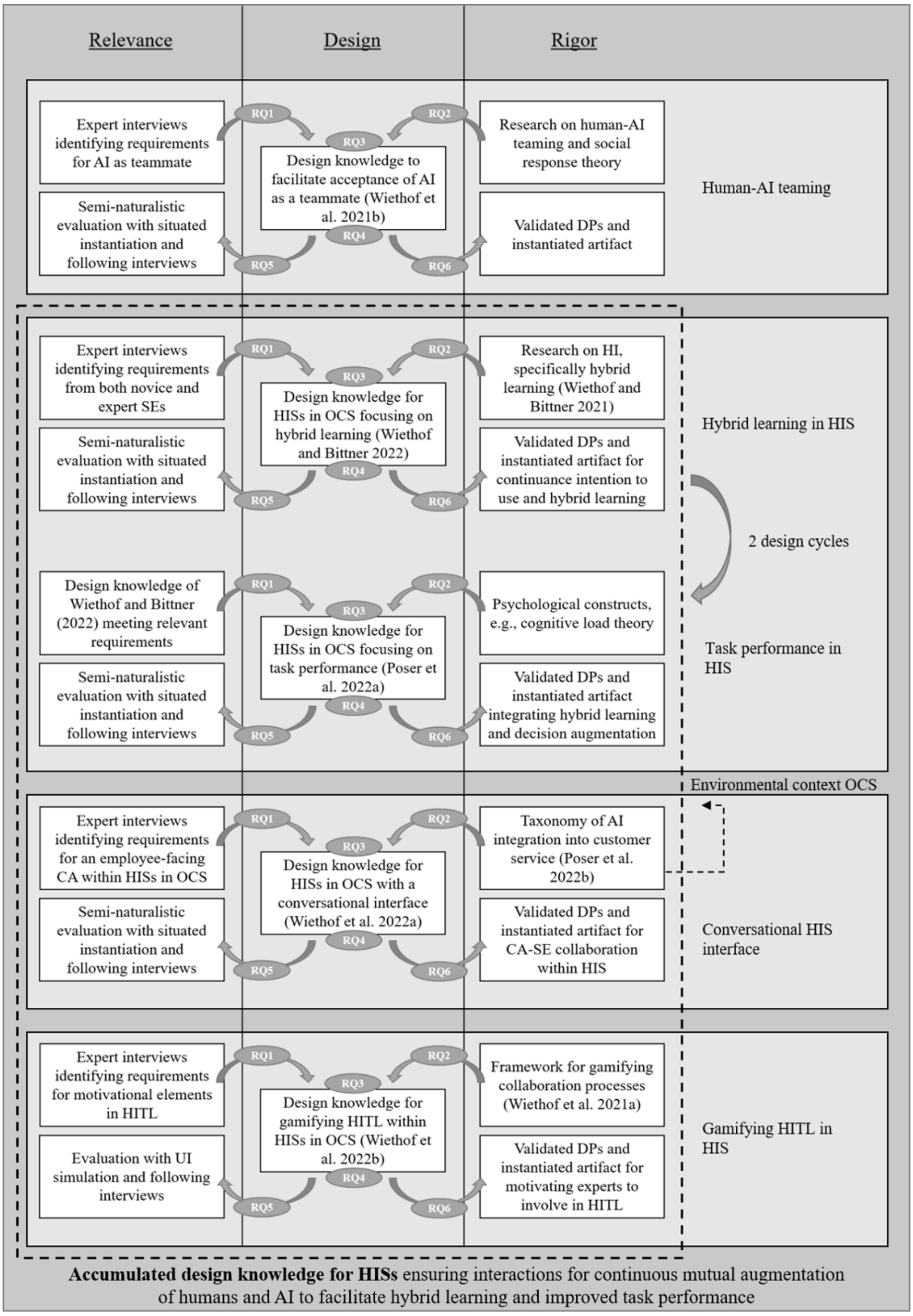

Figure 6. DSR strategy

3.2.1 Relevance

As DSR contributions should be relevant to the real-world environment (Hevner et al. 2004), the relevance cycle connects design knowledge to the specific application domain. It informs design with contextual knowledge and applies newly generated design to the environment (Gregor and Hevner 2013; Hevner et al. 2004). The contextual knowledge, which is inductively drawn from the field, practice, and experience (Gregor and Hevner 2013; Koppenhagen et al. 2012), usually motivates and initializes any DSR project with issues and problems identified (Koppenhagen et al. 2012). To close the relevance cycle, a design artifact is applied in the environment and validated in terms of its solution fitness to the problem space (Koppenhagen et al. 2012; vom Brocke et al. 2020).

RQ1. Wiethof and Bittner (2022), Wiethof et al. (2022b), and Wiethof et al. (2022a) conduct expert interviews (Mayring 2014; Meuser and Nagel 1991; Myers and Newman 2007) to identify MRs as objectives (Peffers et al. 2006; Peffers et al. 2007) or suggestions (Kuechler and Vaishnavi 2012) for a solution. The experts were selected SEs working in customer service. Wiethof and Bittner (2022) further grouped the experts into novice and experienced SEs to better address the hybrid learning part of the RG. In the context of a multicyclic DSR sub-project, Wiethof and Bittner (2022) inform Poser et al. (2022a) with design knowledge meeting relevant requirements. To identify requirements for motivational elements in HITL, Wiethof et al. (2022b) specifically involve experienced SEs. Apart from that, in the course of another publication, we investigate the integration of AI into customer service by applying both empirical-to-conceptual induction as well as conceptual-to-empirical deduction (Poser et al. 2022b). For the inductive approach, we considered various real-world solutions (Lösser et al. 2019; Nickerson et al. 2013).

RQ5. To apply the generated design knowledge in the application domain, Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) have deployed and evaluated an HIS artifact in OCS. They provide prescriptive statements and artifacts, which can be used by organizations to implement HISs in OCS. Furthermore, while Wiethof and Bittner (2022), Poser et al. (2022a), and Wiethof et al. (2022b) deploy an HIS within the same cooperating organization, the design knowledge developed by Wiethof et al. (2022a) is implemented in another cooperating organization. Both use cases can also be applied to the taxonomy of AI integration into customer service (Poser et al. 2022b), specifically, the ex post evaluation of the taxonomy itself considered the use case from Wiethof and Bittner (2022), Poser et al. (2022a), and Wiethof et al. (2022b), while the use case from Wiethof et al. (2022a) used the taxonomy to identify MRs. This implies well-balanced projectability and fitness of the accumulated design knowledge for the contextual environment (vom Brocke et al. 2020). Moreover, the taxonomy can be used to inform organizations' design decisions in their augmentation endeavors integrating AI into OCS. Finally, organizations receive guidance on how to enable mutual augmentation of SEs and AI to facilitate hybrid learning and enhanced task performance by implementing HISs in OCS (Poser et al. 2022a; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022).

3.2.2 Rigor

Though relevance to practice is the essential characteristic that differentiates DSR from other IS research, rigor must not be neglected. Consequently, research should be grounded on scientific knowledge, as well as make contributions to the knowledge base (Gregor and Hevner 2013). While practice usually initiates a DSR project with inductively identified requirements (Gregor and Hevner 2013; Koppenhagen et al. 2012), rigor requires further grounding on existing scientific knowledge deductively derived from the knowledge base (Hevner et al. 2004; Koppenhagen et al. 2012). Eventually, "*[t]he combined result of the practical induction and the theoretical deduction forms a class of problems and respective solutions as basis of design science research projects*" (Koppenhagen et al. 2012, p. 4). To close the rigor cycle, the generated design knowledge enhances or adds to the existing knowledge base (Koppenhagen et al. 2012).

RQ2. Complementing the inductive empirical-to-conceptual approach, Poser et al. (2022b) adopt a deductive conceptual-to-empirical approach by conducting an SLR on existing scientific knowledge about the integration of AI into customer service (vom Brocke et al. 2015; Webster and Watson 2002). The final taxonomy of Poser et al. (2022b) serves as a foundation for the knowledge base that informs Wiethof et al. (2022a) in identifying MRs. In addition to the SLR investigating the knowledge base of the contextual environment, Wiethof and Bittner (2021) conduct an SLR on HI combining HITL and CITL in hybrid learning processes. Wiethof and Bittner (2022) elaborate on the idea of hybrid learning for designing HIS in OCS. Apart from that, all publications build on non-systematic literature reviews disclosing related work and research gaps, as well as conceptual and theoretical backgrounds, such as HIS research and hybrid learning (Wiethof and Bittner 2021, 2022), cognitive load theory (Poser et al. 2022a), team research and social response theory (Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof et al. 2022b), IS continuance intention model (Wiethof et al. 2021a; Wiethof and Bittner 2022), AI infusion into customer service (Poser et al. 2022b), and a framework for gamifying collaboration processes (Wiethof et al. 2021a; Wiethof et al. 2022b). Finally, Poser et al. (2022b) and Wiethof and Bittner (2021) provide conceptual foundations and guidance in the form of a taxonomy for integrating AI into OCS (Poser et al. 2022b), as well as hybrid learning patterns on how to combine HITL with CITL in developing HI (Wiethof and Bittner 2021).

RQ6. Regarding contributions to the knowledge base, all publications included in this dissertation contribute new knowledge. First of all, Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) develop, evolve, and contribute design knowledge in the form of prescriptive DPs and instantiated artifacts in the context of OCS, which can be accumulated in the scope of the overarching DSR project of this dissertation. Apart from that, by building on and discussing the results with existing conceptual and theoretical knowledge, several contributions extend existing knowledge on user interaction with AI-enabled systems (Rzepka and Berger 2018) in terms of HISs. Furthermore, Wiethof and Bittner (2021) and Wiethof and Bittner (2022) contribute concepts for hybrid learning within HISs, whereby Wiethof and Bittner (2022) specifically focus on differentiating between novice and experienced human users. To support motivation and continuance intention in collaboration processes, Wiethof et al. (2021a) contribute a framework for gamifying such processes. Wiethof et al. (2022b) draw on this framework to contribute to the knowledge base by linking gamification with HITL. Lastly, Poser et al. (2022b) contribute to research on OCS. First, they create a taxonomy encompassing conceptual and empirical knowledge to provide design decision guidance on the integration of AI into customer service. Second, they identify additional infusion archetypes on how to integrate AI in the front- and backstage of OCS in terms of augmentation.

3.2.3 Design

In searching for an effective artifact, the core of any DSR project lies in the design cycle (Hevner et al. 2004). It entails the execution of the essential activities of building and evaluating (Hevner et al. 2004; Koppenhagen et al. 2012). Still, it must not disregard the connections to the other two cycles (Hevner et al. 2004). Hence, both evaluation and building activities should equally consider and draw on the requirements of the environment from the relevance cycle as well as the knowledge base from the rigor cycle (Hevner et al. 2004). Thus, a rigorous methodology with appropriate methods must be selected for design and evaluation (Gregor and Hevner 2013). Therefore, the design science research methodology of Peffers et al. (2007) or the high-level design research process of Kuechler and Vaishnavi (2012) are commonly applied. The objectives for a solution (Peffers et al. 2007) or suggestions (Kuechler and Vaishnavi 2012) usually cover requirements to address the problem space. Combined with knowledge from the knowledge base, they inform the development of the design artifact. In this way, DPs are derived, which are "*high level responses to the identified key requirements and, therefore, are instantiations of the class of potential solutions*" (Koppenhagen et al. 2012). They come in the form of prescriptive statements (Gregor et al. 2020) at different abstraction levels (Wache et al. 2022). For example, action- and materiality-oriented DPs prescribe characteristics of the system to be built and what the users can and are supposed to do with it (Chandra et al. 2015). They contribute to the scientific knowledge base toward a theory for design and action and provide utility for goal achievement (Gregor 2006; Gregor et al. 2020; Jones and Gregor 2007; Venable 2006). For demonstration and evaluation, DPs are implemented as design features (DFs) in an instantiated artifact (Koppenhagen et al. 2012). The form of such an artifact can range from a proof-of-concept (Gregor and Hevner 2013; Wilde and Hess 2006) to a fully functioning solution technology (Baskerville et al. 2018; Koppenhagen et al. 2012). The generated design knowledge in the form of artifacts and DPs can then be evaluated based on criteria such as validity, efficacy, utility, or quality (Gregor and Hevner 2013). Venable et al. (2012) provide a framework for evaluation methods in DSR based on strategies for DSR evaluation (Pries-Heje et al. 2008). Depending on when (ex ante vs. ex post) and how (naturalistic vs. artificial) the designed artifact is to be evaluated, appropriate methods can be selected (Gregor and Hevner 2013; Pries-Heje et al. 2008; Venable et al. 2012).

RQ3. Five of the included publications follow a DSR process (Kuechler and Vaishnavi 2012; Peffers et al. 2006; Peffers et al. 2007) and build design artifacts in accordance with the steps of design, development, and demonstration (Poser et al. 2022a; Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022). We identified MRs from literature and practice and then derived DPs providing prescriptive knowledge (Chandra et al. 2015; Gregor 2006; Gregor et al. 2020) for designing and developing HISs that ensure interactions for continuous mutual augmentation of humans and AI. Although each of the five publications constitutes individual design knowledge, they are interdependently connected and cumulative within the overall DSR project of this dissertation as follows. Wiethof et al. (2021b) provide initial DPs to facilitate the general acceptance of AI as a teammate in hybrid work. Based on the potential of hybrid work, a multicyclic DSR sub-project derives DPs for the design and development of HIS in OCS. The first cycle (Wiethof and Bittner 2022) focuses on hybrid learning, differentiating between novice and expert SEs. The second cycle (Poser et al. 2022a) extends the design knowledge of the first cycle with DPs toward improved task performance based on psychological constructs. To ensure projectability and validity, Wiethof et al. (2022a) build on this idea and investigate a different UI for HIS. Instead of embedding AI into an HIS (Poser et al. 2022a; Wiethof and Bittner 2022), Wiethof et al. (2022a) develop DPs for an HIS in OCS using a conversational interface. Lastly, Wiethof et al. (2022b) derive DPs for motivational elements that support the continuous involvement of expert users in HITL. All five of these publications instantiate the DPs within a prototype for demonstration. While Wiethof et al. (2021b) provide a fully functioning prototype that enables collaborative writing (CW) of humans and AI, Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) develop and evolve one HIS prototype for OCS, starting with a prototype enabled by the wizard of Oz (WOz) technique (Wiethof and Bittner 2022) and resulting in a fully functioning prototype based on real data and AI (Poser et al. 2022a; Wiethof et al. 2022a).

RQ4. For evaluation, Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), and Wiethof et al. (2022b) apply rigorous methods from the knowledge base to evaluate the design knowledge and artifact. Accordingly, we deploy the HIS artifact in the field and conduct semi-naturalistic evaluations by means of user test runs or UI simulations (Venable et al. 2012, 2016), as well as qualitative expert interviews (Mayring 2014; Meuser and Nagel 1991; Myers and Newman 2007). To investigate the effects on hybrid learning, Wiethof and Bittner (2022) additionally conduct a quantitative evaluation in terms of triangulation (Mayring 2001) and measure SEs' learning, their perception of AI teachability, and continuance intention to use. They thereby specifically differentiate between novice and expert employees. Poser et al. (2022a) and Wiethof et al. (2022a) also apply a mixed methods approach and gather additional data on usage behavior (Poser et al. 2022a; Wiethof et al. 2022a) and perceived humanness, usefulness, and continuance intention to use (Wiethof et al. 2022a) to measure effects on task performance.

3.3 Research Methods

Across the publications created along the research journey of this dissertation, several research methods have been selected and utilized for optimal consideration of the contextual environment and knowledge base, as well as to build and evaluate design knowledge. The variety and combination of different research methods support the achievement of better research results than one method could have produced alone (Mingers 2001).

First, to ground DSR on existing scientific knowledge (Gregor and Hevner 2013; vom Brocke et al. 2020), SLRs are conducted to investigate previous literature. This includes any prior knowledge related to the problem space, as well as existing knowledge from the solution space, to potentially address the problem (Gregor and Hevner 2013). A literature search can have different goals, such as summarizing and aggregating the results or progress of a research project, disclosing knowledge gaps, or deriving conceptual and theoretical foundations prior to a research project (vom Brocke et al. 2009; vom Brocke et al. 2015; Webster and Watson 2002). Although an unstructured search of the literature might provide some initial insights into the subject and research field to be investigated, Webster and Watson (2002), vom Brocke et al. (2015), and vom Brocke et al. (2009) provide guidelines for conducting an SLR. A rigorous search process guarantees high confidence in the grounding (vom Brocke et al. 2020). In the course of this dissertation, two SLRs were performed and documented accordingly. Wiethof and Bittner (2021) review prior work on HI and how it enables hybrid learning through HITL and CITL. Based on the results, we conceptualize hybrid learning. Poser et al. (2022b) conduct an SLR as part of our taxonomy-development endeavors. It serves as the first iteration, identifying existing work on AI integration into customer service.

Second, stakeholders related to the application domain are often involved in identifying requirements for as well as evaluating the design artifact. Semi-structured expert interviews are therefore commonly conducted (Österle et al. 2011). By making use of guidelines, rigor and qualitative analyses of the interviews can be ensured (Mayring 2014; Meuser and Nagel 1991; Myers and Newman 2007). Wiethof and Bittner (2022), Wiethof et al. (2022a), and Wiethof et al. (2022b) selected SEs from cooperating organizations as experts for the identification of MRs. For evaluation, semi-structured expert interviews followed standardized user tests with the respective users (Poser et al. 2022a; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022).

Third, the standardized user tests were conducted as pilot applications or simulation procedures in semi-naturalistic settings ex post (i.e., after artifact instantiation) (Österle et al. 2011; Pries-Heje et al. 2008; Venable et al. 2012, 2016). The instantiation of the artifact design into a prototype for demonstration follows the prototyping method (Koppenhagen et al. 2012; Österle et al. 2011; Wilde and Hess 2006). This allows for the rapid development of a system (Wilde and Hess 2006) that can be evaluated in terms of goal achievement and successful implementation of the design knowledge (Koppenhagen et al. 2012; Wilde and Hess 2006). In the research endeavors of this dissertation project, several artifacts have been developed based on accumulated and evolved design knowledge. Wiethof and Bittner (2022) instantiate an artifact without real AI capabilities but apply the WOz technique to simulate AI functionalities with a hidden wizard (Böttcher and Nüttgens 2013; Krannich 2010; Salber and Coutaz 1993). We selected this method to ensure consistent and controllable behavior of the system for each user as the focal research object was the social and not the technical sub-system (i.e., our goal was to investigate human learning and continuance intention to use and not to measure AI performance). Wiethof et al. (2022b) similarly used UI simulation procedures to investigate human users' perceptions of gamifying HITL in HIS. For a more naturalistic evaluation of design knowledge utility, Poser et al. (2022a) and Wiethof et al. (2022a) developed and instantiated fully functioning prototypes.

Fourth, apart from qualitative data collection and analysis, triangulation is ensured by also measuring and analyzing quantitative data (Mayring 2001). This includes data on human learning (Wiethof and Bittner 2022), usage behavior (Poser et al. 2022a; Wiethof et al. 2022a), perceived humanness, usefulness (Wiethof et al. 2022a), and continuance intention to use (Wiethof et al. 2022a; Wiethof and Bittner 2022).

Finally, starting with the SLR conducted in the first iteration, Poser et al. (2022b) build a taxonomy relying on DSR for its construction and evaluation (Hevner et al. 2004; Kundisch et al. 2021; Nickerson et al. 2013; Sonnenberg and vom Brocke 2012). For its development, we additionally investigate various real-world solutions for the integration of AI into customer service in three further iterations (Lösser et al. 2019; Nickerson et al. 2013). For its evaluation, we conduct formative ex ante evaluations assessing objective and subjective ending conditions and a summative ex post evaluation applying the taxonomy to illustrative scenarios (Kundisch et al. 2021; Venable et al. 2016).

4 Publications

Eight peer-reviewed articles published in internationally recognized outlets are included in the cumulative dissertation. They are listed in table 2 in chronological order according to their publication dates. Tables 3 to 10 provide an overview of one publication each. For all publications (Sections 9 to 16), the content and references are adopted verbatim. Only the format was adjusted to ensure a coherent appearance of the dissertation.

Table 2. Included publications

No.	Publication	Section
1	Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021b. **"Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI"** 54th Hawaii International Conference on System Sciences (HICSS). Virtual Event.	9
2	Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021a. **"Designing and Evaluating a Collaborative Writing Process with Gamification Elements: Toward a Framework for Gamifying Collaboration Processes"** AIS Transactions on Human-Computer Interaction (13:1).	10
3	Wiethof, C., and Bittner, E. A. C. 2021. **"Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review"** 42nd International Conference on Information Systems (ICIS). Austin, Texas, USA.	11
4	Poser, M., Wiethof, C., Banerjee, D., Subramanian, V. S., Paucar, R., and Bittner, E. A. C. 2022a. **"Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service"** The Transdisciplinary Reach of Design Science Research. DESRIST 2022. Lecture Notes in Computer Science, vol 13229.	12
5	Wiethof, C., and Bittner, E. A. C. 2022. **"Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI"** 30th European Conference on Information Systems (ECIS). Timișoara, Romania.	13
6	Poser, M., Wiethof, C., and Bittner, E. A. C. 2022b. **"Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions"** 30th European Conference on Information Systems (ECIS). Timișoara, Romania.	14

7	Wiethof, C., Roocks, T., and Bittner, E. A. C. 2022b. **“Gamifying the Human-in-the-Loop: Toward Increased Motivation for Training AI in Customer Service”** Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science, vol 13336.	15
8	Wiethof, C., Poser, M., and Bittner, E. A. C. 2022a. **“Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service”** Pacific Asia Conference on Information Systems (PACIS). Virtual Event.	16

Table 3. First publication of the cumulative dissertation

Citation	Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021b. "Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI," in *54th Hawaii International Conference on System Sciences (HICSS)*, Virtual Event, pp. 400-409 (doi: 10.24251/HICSS.2021.047).
Ranking	WKWI: B VHB-JOURQUAL3: C CORE2018: A
Type	Completed Research Paper
Track	Collaboration with Intelligent Systems: Machines as Teammates
Methodology	DSR. Qualitative evaluation with test runs and semi-structured interviews.
Research questions	RQ1: What are the requirements to ensure acceptance of an agent as teammate in CW? RQ2: How can an agent be designed and implemented as teammate contributing to the goal of the CW process? RQ3: How do the human teammates perceive and accept the contributions of the agent?
Research contribution	This paper develops prescriptive design knowledge in the form of MRs, DPs, and an instantiated design artifact for acceptance of an AI as teammate. It contributes to research on human-AI teaming and collaboration and provides a starting point for further research on human-AI interaction in text-based collaboration.
Co-authors' contribution	Navid Tavanapour and Eva Bittner co-authored this publication. Navid Tavanapour supported on the conceptual design of the paper. Navid Tavanapour and Eva Bittner provided overall feedback.

Table 4. Second publication of the cumulative dissertation

Citation	Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021a. "Designing and Evaluating a Collaborative Writing Process with Gamification Elements: Toward a Framework for Gamifying Collaboration Processes," *AIS Transactions on Human-Computer Interaction* (13:1), pp. 38-61 (doi: 10.17705/1thci.00141).
Ranking	ABDC: A AJG2021: 2
Type	Completed Research Paper
Methodology	Action Design Research. Collaboration Engineering for the design of a collaboration process.
Research questions	RQ1: How can one gamify digital collaboration processes in order to encourage users to intend to continue to use them? RQ2: How can one supplement a collaborative writing process with gamification elements to increase users' motivation to use, help them meaningfully engage with, and help them successfully accomplish such processes?
Research contribution	This paper provides a preliminary framework for gamifying collaboration processes by integrating relevant motivation and gamification models and principles. The paper also applies the framework to a CW process and provides prescriptive design knowledge in the form of MRs, DPs, a collaboration process, and an instantiated design artifact. It contributes to research on gamification and human-human collaboration and provides a starting point for further research on gamifying text-based collaboration.
Co-authors' contribution	Navid Tavanapour and Eva Bittner co-authored this publication. Navid Tavanapour supported on the conceptual design of the paper. Navid Tavanapour and Eva Bittner provided overall feedback.

Table 5. Third publication of the cumulative dissertation

Citation	Wiethof, C., and Bittner, E. A. C. 2021. "Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review," in *42nd International Conference on Information Systems (ICIS)*, Austin, Texas, USA.
Ranking	WKWI: A VHB-JOURQUAL3: A CORE2018: A*
Type	Completed Research Paper
Track	AI in Business and Society
Methodology	SLR
Research questions	RQ1: How do humans and computers learn from each other through HITL and CITL learning processes? RQ2: How can the learning effects encouraged by HITL and CITL be measured?
Research contribution	This paper provides conceptual foundations of collaborative learning processes in HIS. It contributes to research on human-AI collaboration and HI by identifying patterns combining HITL and CITL. The paper also presents measurements for evaluating learning effects of each pattern.
Co-authors' contribution	Eva Bittner co-authored this publication. She provided overall feedback.

Table 6. Fourth publication of the cumulative dissertation

Citation	Poser, M., Wiethof, C., Banerjee, D., Subramanian, V. S., Paucar, R., and Bittner, E. A. C. 2022a. “Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service,” in *The Transdisciplinary Reach of Design Science Research. DESRIST 2022. Lecture Notes in Computer Science, vol 13229*, A. Drechsler, A. Gerber and A. Hevner (eds.), Springer, Cham, pp. 142-153 (doi: 10.1007/978-3-031-06516-3_11).
Ranking	WKWI: B VHB-JOURQUAL3: C CORE2018: A
Type	Completed Research Paper
Track	Intelligent Systems and HCI
Methodology	Second cycle of a larger DSR project with Wiethof and Bittner (2022). Online field study within a semi-naturalistic evaluation. Expert interviews for qualitative analysis. Descriptive statistical methods for quantitative assessment of usage data.
Research question	RQ: How should a HIS be designed in a human-centered way to augment real-time decision-making for online customer service encounters?
Research contribution	This paper develops prescriptive design knowledge in the form of MRs, DPs, and an instantiated design artifact for human-centered HIS in OCS. It contributes to research on HI and OCS providing a solution for mutual augmentation of SE and AI. Drawing on the results of Wiethof and Bittner (2022), the paper extends the design knowledge by grounding on kernel theories to integrate real-time decision augmentation and hybrid learning within HIS.
Co-authors’ contribution	Mathis Poser, Debayan Banerjee, Varun Subramanian, Richard Paucar, and Eva Bittner co-authored this publication. Mathis Poser defined the research design and was responsible for the derivation of the design and data analysis. Debayan Banerjee and Varun Subramanian developed the backend of the prototype. Richard Paucar developed the frontend of the prototype. Eva Bittner provided overall feedback. The following activities were managed together with Mathis Poser: idea and concept of the publication, conception and execution of the evaluation, project management and conceptual guidance of the development team.
Best Student Paper Award	

Table 7. Fifth publication of the cumulative dissertation

Citation	Wiethof, C., and Bittner, E. A. C. 2022. "Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI," in *30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.
Ranking	WKWI: A VHB-JOURQUAL3: B CORE2018: A
Type	Completed Research Paper
Track	Artificial Intelligence and Digital Work
Methodology	First cycle of a larger DSR project with Poser et al. (2022a). Test runs using the WOz technique. Expert interviews for qualitative analysis. Descriptive statistical methods for quantitative assessment of learning and continuance intention to use.
Research questions	RQ1: How can continuous collaborative learning of customer service employees and AI be designed and implemented in a HIS? RQ2: How do the learning effects and continuance intention to use differ between novice and expert employees when working with a HIS in customer service?
Research contribution	This paper develops prescriptive design knowledge in the form of MRs, DPs, and an instantiated design artifact for HIS in OCS enabling collaborative learning of human and AI. It contributes to research on HI establishing a collaborative learning cycle between humans and AI differentiating between experts and novices. The paper also provides a starting point for the work of Poser et al. (2022a).
Co-authors' contribution	Eva Bittner co-authored this publication. She provided overall feedback.

Table 8. Sixth publication of the cumulative dissertation

Citation	Poser, M., Wiethof, C., and Bittner, E. A. C. 2022b. “Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions,” in *30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.
Ranking	WKWI: A VHB-JOURQUAL3: B CORE2018: A
Type	Completed Research Paper
Track	Artificial Intelligence in IS Research and Practice
Methodology	Taxonomy development based on an SLR and a qualitative analysis of market solutions and companies’ contact channels. Ex-ante evaluation of subjective ending conditions and ex-post evaluation by applying the taxonomy to two illustrative scenarios.
Research question	RQ: How can conceptual and empirical knowledge on the integration of AI in customer service be classified to provide design decision guidance?
Research contribution	The paper develops a taxonomy, which represents a DSR artifact of the type model, for the integration of AI into customer service. It contributes to research on AI-infused customer service summarizing and integrating scientific insights and the status quo in pratice. The paper also identifies additional archetypes for AI integration into customer service.
Co-authors’ contribution	Mathis Poser and Eva Bittner co-authored this publication. Mathis Poser defined the research design and conducted the quantitative data analysis of the ex-ante evaluation. Eva Bittner provided overall feedback. The following activities were managed together with Mathis Poser: idea and concept of the publication, development and analysis of the database for developing the taxonomy, conception and execution of the evaluation.

Table 9. Seventh publication of the cumulative dissertation

Citation	Wiethof, C., Roocks, T., and Bittner, E. A. C. 2022b. "Gamifying the Human-in-the-Loop: Toward Increased Motivation for Training AI in Customer Service," in *Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science, vol 13336*, H. Degen and S. Ntoa (eds.), Springer, Cham, pp. 100-117 (doi: 10.1007/978-3-031-05643-7_7).
Ranking	WKWI: B VHB-JOURQUAL3: C
Type	Completed Research Paper
Track	Artificial Intelligence in HCI
Methodology	DSR. Application of the framework for gamifying collaboration processes (Wiethof et al. 2021a). Qualitative evaluation with user test simulations and semi-structured expert interviews.
Research questions	RQ1: What effects do gamification elements have on expert users in customer service training an ML system? RQ2: How can gamification elements be integrated into HITL learning in customer service toward higher user motivation? RQ3: Which gamification elements are suitable for motivating expert users in customer service to train an ML system?
Research contribution	This paper develops prescriptive design knowledge in the form of MRs, DPs, and an instantiated design artifact for gamifying HITL in HIS. It contributes to interdisciplinary research combining gamification and ML. The paper also confirms the applicability of the framework for gamifying collaboration processes (Wiethof et al. 2021a) to human-AI collaboration processes.
Co-authors' contribution	Tim Roocks and Eva Bittner co-authored this publication. Tim Roocks conducted the expert interviews for identifying MRs. Eva Bittner provided overall feedback.

Table 10. Eighth publication of the cumulative dissertation

Citation	Wiethof, C., Poser, M., and Bittner, E. A. C. 2022a. "Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service," in *Pacific Asia Conference on Information Systems (PACIS)*, Virtual Event.
Ranking	WKWI: B VHB-JOURQUAL3: C CORE2018: A
Type	Completed Research Paper
Track	Service Science and IS
Methodology	DSR. User tests within a semi-naturalistic evaluation. Semi-structured expert interviews for qualitative analysis. Descriptive statistical methods for quantitative assessment of usage data and perceived humanness, usefulness, and continuance intention to use.
Research question	RQ: How can an employee-facing CA be designed and developed to augment SEs during customer interaction within HIS?
Research contribution	This paper develops prescriptive design knowledge in the form of MRs, DPs, and an instantiated design artifact for an employee-facing CA in HIS. It contributes to research on HI, CAs, and OCS providing a solution for a conversational HIS interface enabling mutual augmentation of SE and AI. Drawing on team research, the paper also provides a starting point for further research on human-AI teaming in OCS.
Co-authors' contribution	Mathis Poser and Eva Bittner co-authored this publication. Mathis Poser supported the derivation of design knowledge and scientific contextualization. Eva Bittner provided overall feedback. The following activities were managed together with Mathis Poser: idea and concept of the publication, definition of the research design, conception of the evaluation to create the database consisting of qualitative and quantitative data, project management including the conceptual responsibility of the prototype development.

5 Theoretical Contribution

This section presents the contributions of the dissertation's research project to the scientific knowledge base. According to the overarching RG, the overall theoretical contribution is constituted by validated design knowledge for human-centered HISs that ensure interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and improved task performance. This design knowledge is developed toward a theory for design and action that provides prescriptive design knowledge with utility character linking the problem with the solution space (Gregor 2006; Gregor and Hevner 2013; Venable 2006; vom Brocke et al. 2020). It contributes design propositions (D), namely, DPs, which are supposed to guide and design mechanisms (M) within the problem situation (P) that facilitate the desired outcomes (O), while being constrained by the contextual environment (C) (Carlsson et al. 2011). Referring to the adapted framework of human-computer interaction (Rzepka and Berger 2018; Zhang and Li 2005) (Figure 2), the newly generated design knowledge exerts the most impact on interaction and the resulting outcomes. This is depicted in figure 7, which integrates the causal coherences of design knowledge (Carlsson et al. 2011) with the adapted framework of human-computer interaction (Rzepka and Berger 2018; Zhang and Li 2005).

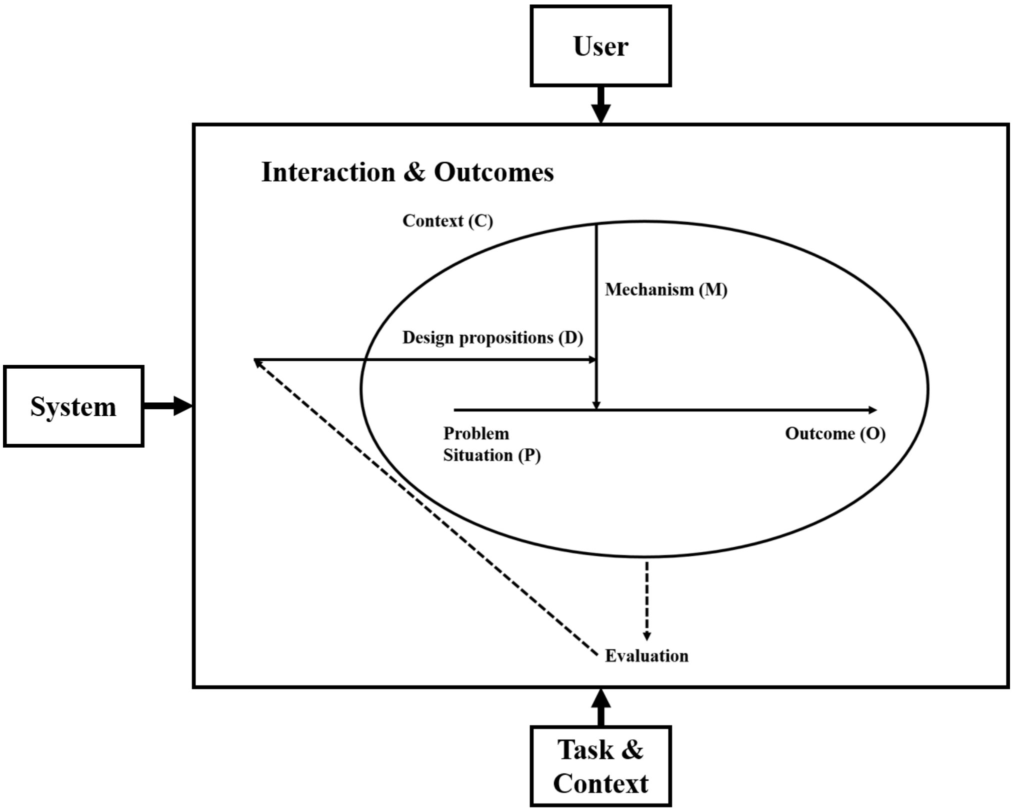

Figure 7. Integration of causal coherences of design knowledge with the framework of human-computer interaction (Carlsson et al. 2011; Rzepka and Berger 2018; Zhang and Li 2005)

Still, the theoretical contributions are not limited to design knowledge for interactions for mutual augmentation within HISs but also include contributions to the user, system, and task and context constructs. These extend and supplement the theoretical foundations for developing design knowledge for HIS (Figure 7). The contributions of this dissertation are analogously structured as follows.

First, regarding the **user** construct and its interactions with the system, hybrid learning is conceptualized (Wiethof and Bittner 2021, 2022). Second, for the **system** construct, a framework for gamifying collaboration processes is demonstrated and applied to HITL in HIS (Wiethof et al. 2021a; Wiethof et al. 2022b). Third, for the construct of **task and context**, a taxonomy drawing from scientific literature and practice is established and presented for AI integration into customer service (Poser et al. 2022b). Lastly, for the **interactions and outcome** constructs, design knowledge is accumulated toward a theory for design and action for designing and developing HIS (Poser et al. 2022a; Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022).

5.1 Hybrid Learning

Reviewing extant literature on how humans and AI work together by means of HITL and CITL, Wiethof and Bittner (2021) contribute to the overall research stream of human-AI collaboration by demonstrating how HITL and CITL can be combined for HI. Deriving from the research area of human-AI collaboration and teaming (Bittner et al. 2019b; Norman 2017; Seeber et al. 2020; Strohmann et al. 2019; Yu et al. 2019), HI constitutes improvement of both humans and AI through mutual augmentation (i.e., HITL and CITL) (Dellermann et al. 2019b; Dellermann et al. 2019a). However, as the concept of HI is rather new, existing literature is fragmentary, with different underlying terms and definitions for HI and hybrid learning. To address and include a wide scope of literature, Wiethof and Bittner (2021) covered several wordings in the search string of the SLR. By structuring and systemizing the available literature, we were able to create a common understanding of HI that combines HITL and CITL in a hybrid learning process (Wiethof and Bittner 2021) (Figure 8).

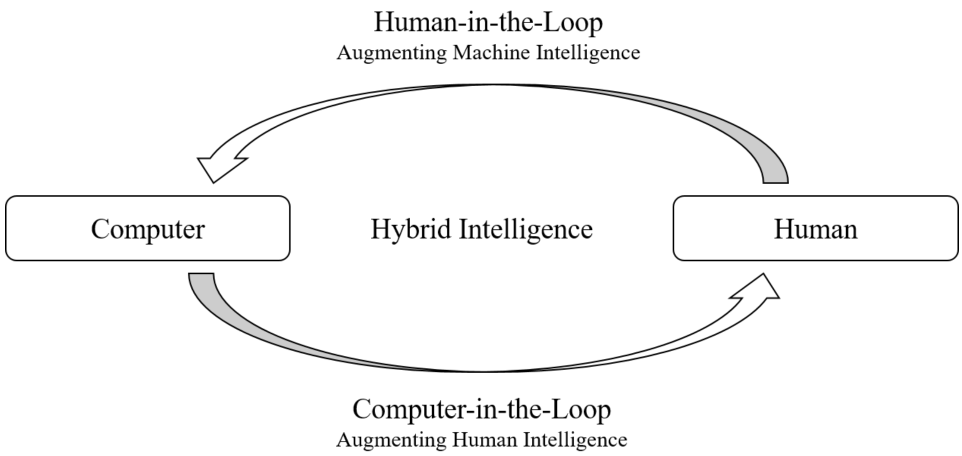

Figure 8. Generic hybrid learning process of HI (Wiethof and Bittner 2021)

Figure 8 depicts a generic hybrid learning process of HI, aggregating and representing the mutuality and iterative nature of the interactions for augmentation of humans and AI. Hence, all combinations of HITL and CITL identified in the available literature fit into this process (Wiethof and Bittner 2021). Moreover, through sense-making of the combinations, Wiethof and Bittner (2021) contribute descriptive knowledge by defining and presenting patterns of hybrid learning (Gregor and Hevner 2013): decision support, exploration, and integration. Decision support encompasses three sub-patterns, namely assimilation, exploitation, and explanation. These patterns provide a conceptual foundation for research on hybrid learning in HI. First, assimilation in decision support allows for iterative adjustments based on the comparison of outputs according to inputs (Amershi et al. 2014; Berger et al. 2021; Erbe 2001; Hu et al. 2019; Luong et al. 2019). Second, exploitation in decision support ensures that the human and AI iteratively exploit and validate each other’s knowledge (Cai et al. 2019; Hanika et al. 2019; Lees et al. 2011; Lindvall et al. 2021; Mullins et al. 2020; Paschen et al. 2020; Pereira and Paulovich 2020; Rundo et al. 2020; Steenwinckel et al. 2021; Traumer et al. 2017; Verdenius 1995; Zeni et al. 2019). Third, explanation in decision support enables AI to implicitly and explicitly collect data from the human, as well as provide explanations for teaching the human (Dellermann et al. 2019b; Holzinger et al. 2021; Hudec et al. 2021; Hun Lee et al. 2021; Kiefer 2022; Kulesza et al. 2015; Liu et al. 2014; Schneider and Handali 2019). Fourth, unlike exploitation, exploration enables AI and the human to explore and identify new insights (McCamish et al. 2017; Oliveira et al. 2020; Salam et al. 2019; Smith et al. 2018). Fifth and last, integration does away with the separation of artificial and human intelligence and instead directly integrates them (Bassano et al. 2020; Dellermann et al. 2017; Gavriushenko et al. 2020; Hekler et al. 2019). As these patterns all fit into the generic hybrid learning process, it is also possible to combine them, for example, exploring until finding an aspect or a way to exploit (McCamish et al. 2017) or providing explanations for exploratory results (Smith et al. 2018).

Overall, for the dissertation, the patterns of hybrid learning serve as a valuable foundation that represents and illustrates how interactions for mutual augmentation of humans and AI in terms of HITL and CITL are processed. Referring back to Carlsson et al. (2011), the mechanisms are represented within a problem situation leading to the desired outcomes. As such, further publications included in this dissertation draw on this conceptualization to develop design knowledge that aims to guide and design the mechanisms toward hybrid learning and improved task performance (Poser et al. 2022a; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022). For instance, Wiethof and Bittner (2022) investigate hybrid learning within an HIS. We contribute a novel perspective of distinguishing humans by knowledge and experience within HISs and leverage the different knowledge levels. Based on the mutuality and iterative nature of the interactions for augmentation of humans and AI within HISs, Wiethof and Bittner (2022) extend the work of Wiethof and Bittner (2021) and propose a conceptualization of hybrid learning that differentiates between experts and novices (Figure 9).

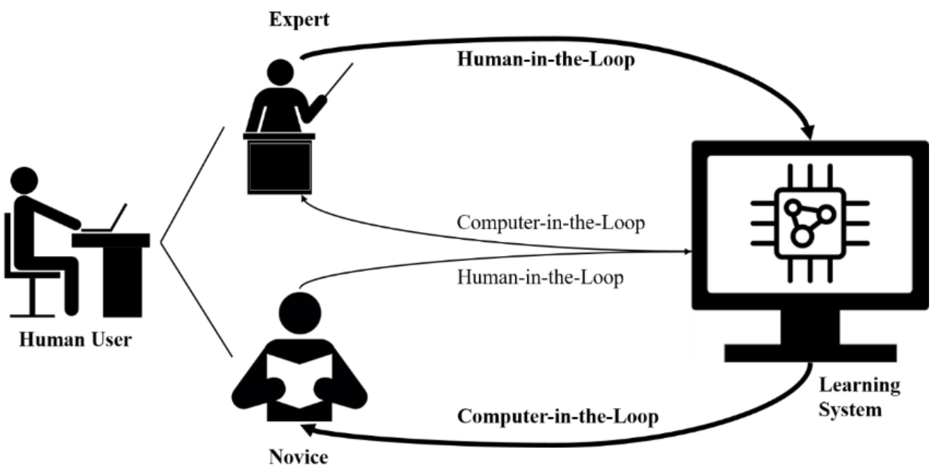

Figure 9. Hybrid learning differentiating between experts and novices (Wiethof and Bittner 2022)

The bold arrows in figure 9 depict an implicit knowledge transfer from experts to novices through AI from a long-term perspective. To ensure hybrid learning in terms of continuous improvement by learning from each other (Dellermann et al. 2019a), a high continuance intention to use is required from experts to teach the AI (HITL) and from novices to learn from the AI (CITL). Accordingly, drawing on the IS continuance intention model (Bhattacherjee 2001; Bhattacherjee et al. 2008), Wiethof and Bittner (2022) develop validated design knowledge for HISs that enable a high continuance intention to use for learning and teaching.

5.2 Gamifying Collaboration Processes

Building on the concept of hybrid learning that differentiates between experts and novices (Wiethof and Bittner 2022), Wiethof et al. (2022b) investigate expert users' motivation to involve in HITL. In domains with a limited amount of data, the involvement of end users for training through HITL leads to better results due to faster and more tightly coupled learning cycles (Amershi et al. 2014; Holzinger 2016a; Holzinger et al. 2016, 2017). However, this initially requires more training efforts from the experts. Although experts confirm their willingness to teach and continuance intention to use (Wiethof and Bittner 2022), interactively training AI on the job inevitably causes an addition to the workload.

As such, Wiethof et al. (2021a) and Wiethof et al. (2022b) contribute motivational affordances to increase humans' IS continuance intention and thus motivation to involve in HITL. First, addressing the research field of digital collaboration (Curtis and Lawson 2001; Johnson and Johnson R. T. 2004; Leimeister 2014; Richter et al. 2018), Wiethof et al. (2021a) develop and contribute a preliminary framework for gamifying collaboration processes (Figure 10) and apply it to text-based collaboration by means of CW.

<table>
<tr><th colspan="3">Gamification elements in collaboration processes</th></tr>
<tr><td rowspan="3">Mechanics</td><td>Gamification affordances</td><td>Status, competition, self-expression, etc.</td></tr>
<tr><td>Gamification objects</td><td>Items, characters, visual assets, etc.</td></tr>
<tr><td>Gamification mechanics</td><td>Rules</td></tr>
<tr><th colspan="3">Gamification principles for collaboration process interactions</th></tr>
<tr><td rowspan="3">Dynamics</td><td>User-system-interactions</td><td>User-to-system, system-to-user, user-to-user</td></tr>
<tr><td>Gameful interactions</td><td>Competition, cooperation</td></tr>
<tr><td>Playful interactions</td><td>Exploration, creation, pretending</td></tr>
<tr><th colspan="3">Intended outcomes of the gamified collaboration processes</th></tr>
<tr><td rowspan="3">User engagement (aesthetic/flow experience)</td><td rowspan="2">Meaningful engagement (aesthetic experience)</td><td>Experiential outcomes: sensory and cognitive experiences (sensation, fantasy, narrative, challenge, fellowship, discovery, expression, submission, meaning, self-expansion), attachment to outcome, attachment to system</td></tr>
<tr><td>Instrumental outcomes: functional, related to work context, prolonged use, increased use, increased learning</td></tr>
<tr><td>Deep engagement (flow experience)</td><td>Hedonic motivation</td></tr>
<tr><th colspan="3">Continuance intention to use gamified digital collaboration processes</th></tr>
</table>

Figure 10. Framework for gamifying collaboration processes (Wiethof et al. 2021a), adapted

Building on the MDA framework (Hunicke et al. 2004), this framework constitutes three segments – mechanics, dynamics, and user engagement – covering both aesthetic and flow experience. Each segment is aligned to the domain of digital collaboration processes (i.e., gamification elements in collaboration processes for mechanics, gamification principles for collaboration process interactions for dynamics, and intended outcomes of the gamified collaboration processes for user engagement). The overall goal of the framework is to increase continuance intention to use gamified collaboration processes. Integrating the work of Liu et al. (2017), Suh et al. (2017), and Tseng and Sun (2017), the framework provides elements to consider for each segment (e.g., gamification objects for mechanics, user-system-interactions for dynamics, and meaningful engagement for user engagement). Lastly, examples are assigned to each element.

Overall, Wiethof et al. (2021a) contribute the integration of motivation and gamification models and principles for gamifying digital collaboration processes to a preliminary framework. With this framework, we provide prescriptive knowledge toward a theory for design and action that provides guidance on how to design IS for increased continuance intention to use and the production of promising outcomes underlying enhanced satisfaction and creativity in a virtual environment (Gregor 2006; Gregor and Hevner 2013). We apply gamification to a text-based digital collaboration process, using and confirming the applicability of the framework (Wiethof et al. 2021a).

As Wiethof et al. (2021a) apply the framework to a collaboration setting involving only humans, Wiethof et al. (2022b) use the framework for gamifying HITL within an HIS. Thus, we confirm the applicability of the framework for gamifying human-AI collaboration. Moreover, we contribute knowledge on how to apply the framework in a systematic fashion to design and develop an HIS gamifying HITL within the scope of DSR. For this, we follow a top-down approach along the segments of the framework. First, intended outcomes for user engagement are defined as MRs. Second, gamification principles for dynamics are determined as DPs and derived from the MRs. Third, appropriate gamification elements are identified accordingly.

Finally, as gamification is supposed to optimize human-computer interaction (Khakpour and Colomo-Palacios 2021), Wiethof et al. (2022b) contribute to research on human-AI collaboration, specifically HISs, by combining gamification with HITL learning. Additionally, the transferability of the framework from human-human to human-AI collaboration can be reflected by drawing on social response theory (Nass et al. 1994; Nass and Moon 2000). This postulates that humans demonstrate social behavior toward computers if the latter are assigned certain social cues. As cooperation and fellowship are concepts commonly applied in gamification (Wiethof et al. 2021a; Wiethof et al. 2022b), there are two conclusions to be inferred. First, gamification can support and encourage team feeling in human-AI collaboration. Second, humanizing AI is itself a gamification element that increases team spirit (Wiethof et al. 2022b).

5.3 AI Integration into OCS

The environmental context constitutes an important construct for the causal coherences of design knowledge for socio-technical systems (Carlsson et al. 2011; Rzepka and Berger 2018; Zhang and Li 2005) (Figure 7). Although researchers intend to create general and abstract design knowledge to ensure projectability into new research contexts (Carlsson et al. 2011; vom Brocke et al. 2020), any socio-technical interactions are limited by contextual constraints (Carlsson et al. 2011). Hence, the contextual environment should be taken into account when considering a trade-off between the projectability and fitness of the solution with respect to the problem space (Carlsson et al. 2011; vom Brocke et al. 2020).

OCS represents a highly matured application domain. As such, the accumulated design knowledge for HISs in OCS developed in the course of this dissertation constitutes an improvement (i.e., "*new solutions for known problems*") (Gregor and Hevner 2013, p. 345) and extends the OCS knowledge base. For scientific contextual grounding, Poser et al.

(2022b) contribute the first taxonomy summarizing scientific insights and the status quo in practice for a better understanding of how to integrate AI into customer service (Figure 11).

The taxonomy "*help[s] researchers to organize knowledge by representing relevant dimensions and corresponding characteristics (i.e., constructs) and thereby reflect relationships among dimensions and characteristics*" (Kundisch et al. 2021, p. 4). It is a DSR artifact of the type model (Gregor and Hevner 2013; Kundisch et al. 2021; March and Smith 1995), providing prescriptive design knowledge toward a theory for design and action on AI integration into OCS from a socio-technical perspective (Gregor 2006; Gregor and Hevner 2013). Consequently, researchers can draw on the taxonomy to systematically derive design decisions along the (meta-)dimensions for designing AI-infused OCS.

MD	Dimensions		Characteristics			
Service Context	D_1: Service Stages	NE	Frontstage		Backstage	
	D_2: Service Process Continuity	NE	Disconnected		Connected	
Capabilities	D_3: AI Role	NE	Support	Augmentation		Performance
	D_4: Task Type	NE	Mechanical	Analytical	Intuitive	Empathetic
Deliverables	D_5: Knowledge and Data Insights	NE	Inquiry-related	Process-focused	Customer-related	Socio-emotional
	D_6: Performance Monitoring	NE	Human Agent Monitoring		AI Monitoring	
Integration	D_7: Hybrid Inquiry Handling	ME	Simultaneous	Consecutive - toward human	Consecutive - toward AI	Consecutive - alternating
	D_8: Level of AI Activity	NE	Reactive		Proactive	
	D_9: Form of AI Appearance	ME	AI-enabled agent		Embedded AI	
	D_{10}: AI Transparency to Customers	ME	Unknown		Known	
Intelligence	D_{11}: Data and Knowledge Processing	NE	Machine Learning		Rule-based Reasoning	
	D_{12}: Data and Knowledge Source	NE	Input before Interaction	Input during Interaction		Input after Interaction

Note: MD = meta-dimension; ME = mutually exclusive; NE = non-exclusive

Figure 11. Taxonomy of AI integration into customer service (Poser et al. 2022b)

Overall, the taxonomy contributes to the scientific knowledge base of OCS by systematically representing extant knowledge in science and practice. Furthermore, based on the taxonomy, Poser et al. (2022b) contribute a representation of AI infusion archetypes covering the front- and backstage (Figure 12). With this, they confirm existing archetypes from Keyser et al. (2019) and Ostrom et al. (2019), namely, substitution (I) and synchronous augmentation (IV & V) in the frontstage, and identify new archetypes, namely, asynchronous augmentation (II & III) in the frontstage and augmentation in the backstage (VI).

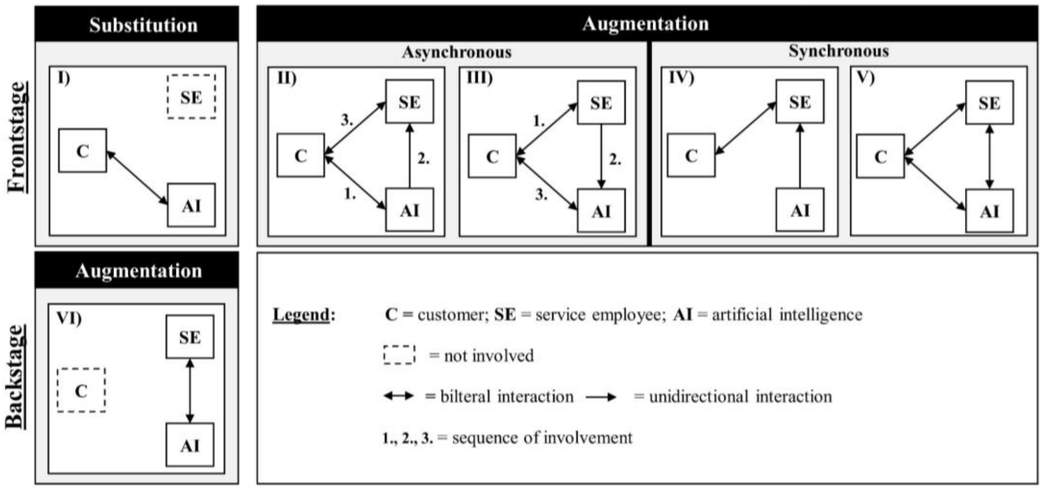

Figure 12. AI infusion archetypes covering front- and backstage (Poser et al. 2022b)

Finally, the taxonomy and depiction of augmentation scenarios serve as a relevant theoretical grounding for the dissertation's DSR project. In fact, Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022b), and Wiethof et al. (2022a) draw on use cases, in which AI infusion into OCS for synchronous SE augmentation in the frontstage is planned. Based on the overarching RG of designing a human-centered HIS to ensure interactions for continuous mutual augmentation of humans and AI, the AI infusion archetype to be designed in this dissertation is synchronous augmentation in the frontstage (Figure 12, V).

5.4 Design Knowledge for HISs in OCS

Grounded on extant theoretical foundations and informed by new theoretical contributions to user interaction with AI-enabled systems (RQ2 & RQ6) (Rzepka and Berger 2018), five of the included publications conduct DSR following a systematic process (Kuechler and Vaishnavi 2012; Peffers et al. 2006; Peffers et al. 2007) for designing, developing, and evaluating interactions for mutual augmentation in HISs (RQ3 & RQ4). These contribute validated design knowledge toward a theory for design and action in the form of prescriptive DPs and instantiated design artifacts (Chandra et al. 2015; Gregor 2006; Gregor and Hevner 2013; Koppenhagen et al. 2012). In terms of relevance, the design knowledge meets requirements identified from practice (RQ1) and is evaluated in the field by instantiating the design artifact in the application domain (RQ5). The presentation of the overall design knowledge is divided into the following two sub-sections. First, the publications' individual design contributions are accumulated in the form of abstract DPs for designing interactions for mutual augmentation in HISs to facilitate hybrid learning and improved task performance. To ensure a balance of projectability and fitness, the DPs are further specified in the OCS environment (vom Brocke et al. 2020; Wache et al. 2022). Subsequently, instantiated design artifacts in the form of HISs in OCS are demonstrated underlying the DPs. Next, serving as principles of form and function (Chandra et al. 2015;

Jones and Gregor 2007), the DPs contribute toward a design theory for HISs that enable interactions for continuous mutual augmentation of humans and AI (Gregor 2006; Jones and Gregor 2007). The second sub-section therefore presents the contribution to the development of such a design theory.

5.4.1 Interaction and Outcomes in HIS

In accordance with figure 7, the overarching RG of designing and developing interactions for continuous mutual augmentation within HISs for hybrid learning and improved task performance addresses the interaction and outcomes constructs. The causal coherences of design knowledge (Carlsson et al. 2011) specific to the RG of the dissertation are depicted in figure 13.

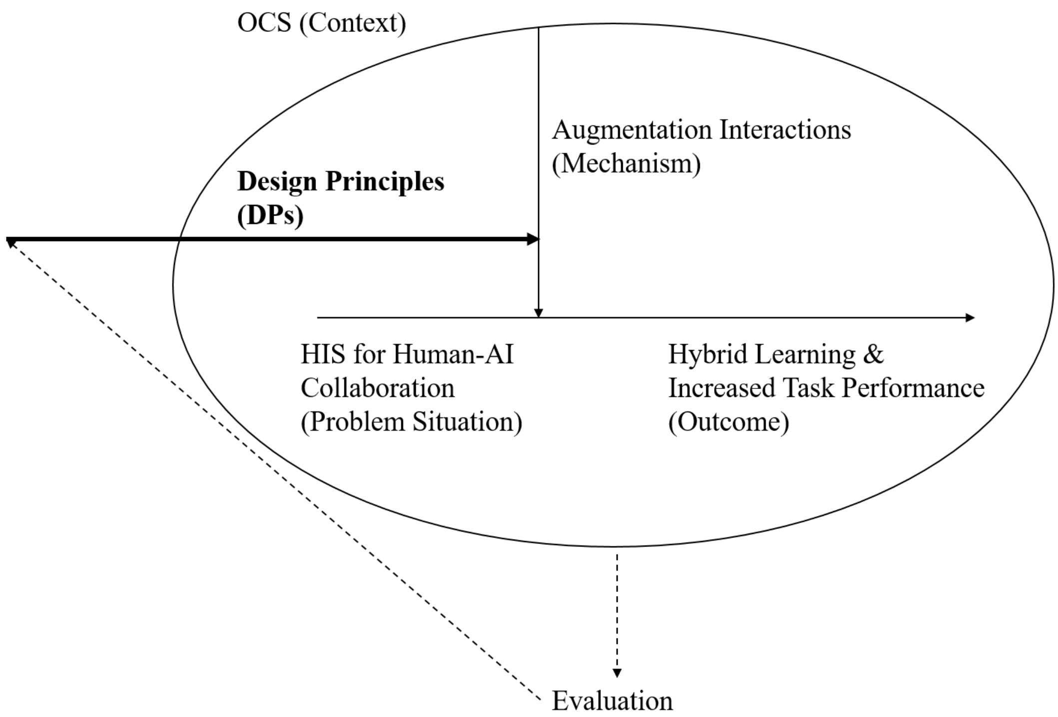

Figure 13. Causal coherences of design knowledge specified to the RG of the dissertation (Carlsson et al. 2011), adapted

As illustrated in figure 13 (Carlsson et al. 2011), in a contextualized problem situation, to achieve the desired outcome, DPs are to be executed on according mechanisms. The RG addresses the problem situation by calling for human-centered design knowledge for HISs to facilitate the outcome of hybrid learning and improved task performance by interactions for mutual augmentation as mechanisms, contextualized in OCS. The dissertation contributes validated design knowledge in the form of DPs and instantiated design artifacts, allowing for the causal coherences depicted in figure 13.

Five of the included publications contribute prescriptive design knowledge, in the form of DPs and instantiated design artifacts, to the overarching RG of the dissertation (Poser et al. 2022a; Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and

Bittner 2022). By accumulating design knowledge for projectability, six abstract DPs are identified. For design knowledge fitness, the DPs are further specified in the context of OCS (Chandra et al. 2015; vom Brocke et al. 2020; Wache et al. 2022).

> **DP1: Provide the HIS with AI capable of understanding domain-specific language in order for the AI to process task-related input to augment the human with its output, given a shared contextual understanding.**

Contextualizing DP1 within the synchronous augmentation scenario in the OCS frontstage (Keyser et al. 2019; Poser et al. 2022b), the AI needs to be able to understand and process customer requests (Wiethof et al. 2021b; Wiethof et al. 2022a). These capabilities allow the AI to augment the SE with context-specific output (e.g., response suggestions) in sync with the dynamic, real-time service delivery (Poser et al. 2022a). Wiethof and Bittner (2022), Poser et al. (2022a), and Wiethof et al. (2022a) therefore use a database of FAQ-based questions.

> **DP2: Provide the HIS with controllable AI in order for the human to involve in HITL to ensure continuous augmentation of the AI, given that the human uses, adapts, and/or provides feedback on the AI's output.**

Contextualizing DP2 within the synchronous augmentation scenario in the OCS frontstage (Keyser et al. 2019; Poser et al. 2022b), the SE is the main responsible within the socio-technical HIS (Wiethof et al. 2021b; Wiethof et al. 2022a). First, the SE represents the point of contact for customers and is therefore in charge of customer service delivery (Wiethof et al. 2022a). Although AI can process customer requests in synch with real-time service delivery (DP1), its output is only presented to the SE, remaining invisible to the customer. Second, by using, adapting, and/or providing feedback on the AI's output (e.g., in the form of response suggestions), the SE involves in HITL to improve the AI (Wiethof et al. 2022a; Wiethof and Bittner 2022). Third, the SE can augment the AI with information on the progress of the customer service process and with customer insights and information to facilitate shared understanding between the human and AI (Wiethof and Bittner 2022). Fourth, it is beneficial for AI to differentiate between novice and experienced SEs as it might learn more from experts. Lastly, an HIS should provide SEs with the option to prevent AI from learning (Wiethof and Bittner 2022).

> **DP3: Provide the HIS with motivational elements in order to encourage expert users to involve in HITL.**

DP3 builds upon DP2 providing implications to improve HITL. In the context of OCS, motivational elements are to be deployed in HIS to encourage expert SEs to involve in HITL. For the dissertation, gamification is considered to provide appropriate motivational elements. First, gamification elements include but are not limited to rewards and recognition for feedback. To ensure qualitative feedback, rewards are based on the value of the feedback. Second, working in teams not only gamifies HITL but also contributes to meaningful engagement (i.e., all SEs are involved in HITL and work toward the same goal). This goal is to improve AI for better work efficiency and learning of novices. Third, visualizing the progress of provided and needed feedback is another supportive gamification element. Lastly, imbuing the AI with its own personality (DP6) asking for and reacting to feedback has motivational effects in terms of gamification (Wiethof et al. 2022b).

DP4: Provide the HIS with dynamic UI elements in order for the AI to involve in CITL to ensure continuous augmentation of the human, given that the AI displays suitable output according to the context and the human's needs.

Contextualizing DP4 within the synchronous augmentation scenario in the OCS frontstage (Keyser et al. 2019; Poser et al. 2022b), the AI is able to understand and process customer requests and augment the SE with its output. First, such output can come in the form of response suggestions that address customers' requests (Wiethof et al. 2022a; Wiethof and Bittner 2022). This leads to enhanced task performance in terms of efficiency (Poser et al. 2022a; Wiethof et al. 2022b) and human learning, especially when differentiating between expert and novice SEs (Wiethof et al. 2022b; Wiethof and Bittner 2022). Additionally, response suggestions can support SEs in creating a personalized experience for the customer by providing a selection of suggestions with varying levels of sentiment (Wiethof et al. 2022a). Second, in fostering a shared understanding between the human and AI, dynamic UI elements allow for the AI to react to the SE's contributions (Wiethof et al. 2021b) and augment the SE with customer insights and information, as well as with the progress of the customer service process (Wiethof and Bittner 2022).

DP5: Provide the HIS with augmentation strategies toward the human in order for the human to learn and enhance task performance through CITL.

DP5 builds upon DP4 providing implications to improve CITL. First, in the context of OCS, AI's response suggestions should be easily applicable and/or adaptable to increase work efficiency (Poser et al. 2022a; Wiethof et al. 2022a). Second, in sync with the dynamic interaction, the number of response suggestions provided should be manageable and not overwhelm the SE with information (Poser et al. 2022a; Wiethof and Bittner 2022). Third, upon the SE's request, the AI elaborates on its response suggestions to verify the applicability and simplify selection (Poser et al. 2022a). Fourth, visualizing the utility of the response suggestions can further facilitate response selection. Therefore, the suggestions can be assigned a numerical utility value calculated by the AI. Fifth, based on their utility, suggestions are presented in sequence and in alternating combinations upon request (Poser et al. 2022a). Finally, the HIS should have configurable AI settings to customize CITL (e.g., the SE's need for augmentation) (Poser et al. 2022a).

> **<u>DP6</u>: Provide the AI with its own representation equipped with social cues in order for the human to perceive and accept the AI as a collaboration partner, given collaborative team structures within the HIS.**

Contextualizing DP6 within the synchronous augmentation scenario in the OCS frontstage (Keyser et al. 2019; Poser et al. 2022b), the AI is supposed to act as a co-SE, capable of understanding and processing customer requests (DP1) but is limited to and designed for AI-SE collaboration (DP2 & DP4). First, drawing on social response theory and team research, the AI should be imbued with its own identity and equipped with social cues according to its competence (Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof and Bittner 2022). This includes and is not limited to explainability capabilities, such as introducing itself and providing explanations on how it is working and learning (Wiethof et al. 2021b; Wiethof and Bittner 2022), as well as asking for and reacting to feedback (Wiethof et al. 2022b). Second, the respective roles of the AI and SE should be clear and transparent within the HIS (Wiethof and Bittner 2022). This encourages actual collaboration, preventing the perception of human workforce replacement through AI (Wiethof et al. 2022b). Lastly, AI-SE interaction fosters the social presence of AI and a personal connection as a member of a hybrid team (Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof and Bittner 2022).

Overall, the six abstract DPs comprise the design knowledge of five of the included publications for building HISs (Poser et al. 2022a; Wiethof et al. 2021b; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022) based on existing theoretical foundations and theoretical foundations newly created over the course of this dissertation (Poser et al. 2022b; Wiethof et al. 2021a; Wiethof and Bittner 2021) (Figure 14).

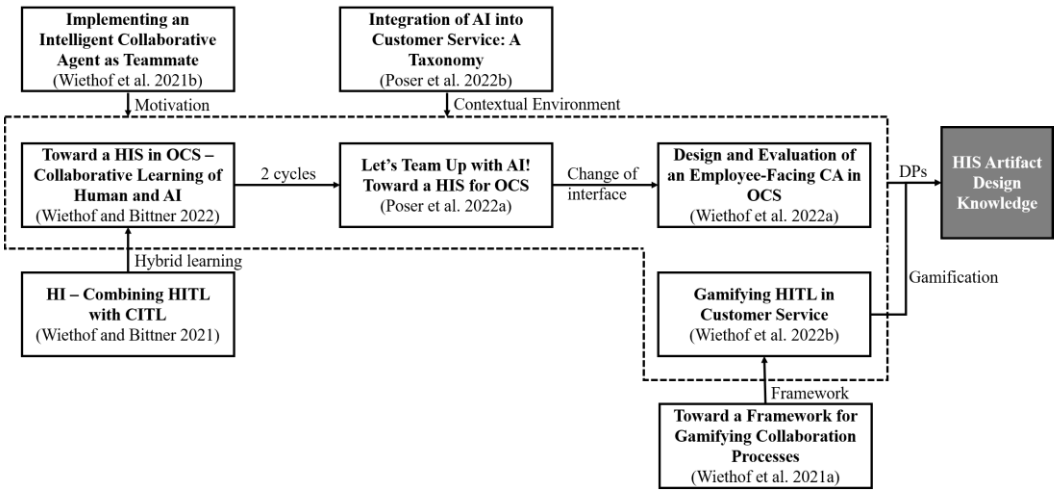

Figure 14. Interconnections of publications included in this dissertation for HIS artifact design knowledge

Figure 14 depicts the publications included in this dissertation and how they are interconnected with each other in terms of design knowledge accumulation and evolution, as well as scientific grounding. First, Wiethof et al. (2021b) motivate the research endeavors by providing initial design knowledge for text-based human-AI collaboration. Second, building on the conceptualization of hybrid learning (Wiethof and Bittner 2021), Wiethof and Bittner (2022) contribute design knowledge for HISs as the first cycle of a multicyclic DSR sub-project, contextualized in OCS (Poser et al. 2022b). Third, Poser et al. (2022a) extend the design knowledge for HISs in OCS as the second cycle to facilitate enhanced task performance while ensuring hybrid learning. Fourth, Wiethof et al. (2022a) provide similar contributions, changing the UI of the HIS in OCS from embedded AI to a CA. At last, Wiethof et al. (2022b) develop design knowledge for gamifying HITL in HISs, building on the framework for gamifying collaboration processes established by Wiethof et al. (2021a). Overall, the design knowledge for HISs accumulated and evolved as follows.

As depicted in figure 14, Wiethof et al. (2021b) motivate and initialize the dissertation's DSR project by contributing prescriptive design knowledge on how to establish acceptance of AI as a teammate in hybrid work contexts. We draw on social response theory (Nass et al. 1994; Nass and Moon 2000) and the concept of the uncanny valley (Mori et al. 2012) to provide validated DPs and an instantiated artifact for text-based human-AI collaboration.

Next, Wiethof and Bittner (2022) build on the conceptualization of hybrid learning (Dellermann et al. 2019a; Kulesza et al. 2015; Wiethof and Bittner 2021), the IS continuance model (Bhattacherjee 2001; Bhattacherjee et al. 2008), and the synchronous augmentation scenario in the OCS frontstage (Poser et al. 2022b) to contribute initial prescriptive design knowledge for HISs in OCS to facilitate hybrid learning in the form of validated DPs and an instantiated and evaluated prototype (Figure 15).

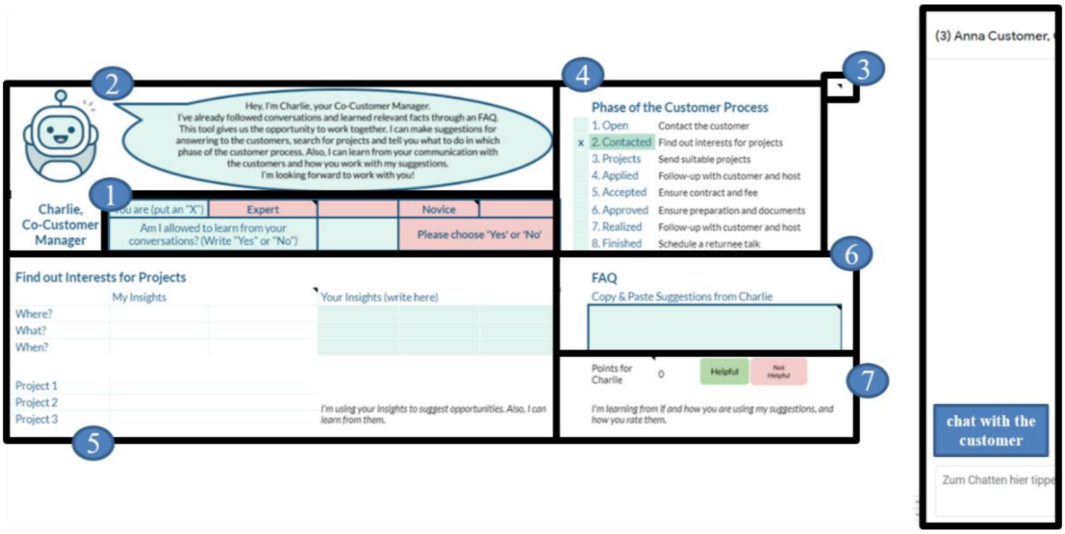

Figure 15. HIS prototype by Wiethof and Bittner (2022)

In contributing to the novel perspective of differentiating between novice and expert SEs, Wiethof and Bittner (2022) focus on interactions for mutual augmentation of humans and AI in terms of hybrid learning. We therefore determine essential requirements and design knowledge for HISs that combine HITL and CITL. The design artifact instantiated as an HIS prototype in OCS encompasses several UI elements enabling mutual augmentation. First, the customer service process phase (Figure 15 [4]) and second, customer interests and insights about suitable products (Figure 15 [5]) contribute to a shared understanding and can be adapted by both AI and the SE (e.g., by adding or deleting insights from the table of customers' interests). Third, AI can augment the SE with response suggestions within the scope of CITL (Figure 15 [6]). Fourth, complementary to CITL, there are, for example, feedback buttons to enable HITL (Figure 15 [7]). Fifth, the AI is imbued with its own personality (i.e., Charlie, the co-customer manager) (Figure 15 [2]). Sixth, AI learning settings (Figure 15 [1]) support HITL by determining the learning behavior of the AI (i.e., differentiating between experts, novices, and learning prohibited). Seventh, AI explanations (Figure 15 [3]) support CITL by providing the SE with information on the UI features and the AI's augmentation behavior upon request. The prototype was instantiated and evaluated by means of the WOz technique, still using an FAQ dataset for response suggestions. As such, the work of Wiethof and Bittner (2022) serves as the first cycle of a multicyclic DSR sub-project.

Poser et al. (2022a) represent the second cycle of the multicyclic DSR sub-project, building on the work of Wiethof and Bittner (2022). We contribute additional prescriptive design knowledge for designing and developing HIS in OCS to facilitate improved task performance while ensuring hybrid learning. We aggregate the design knowledge with that of Wiethof and Bittner (2022) in the form of validated DPs and an instantiated and evaluated fully functioning HIS prototype (Figure 16).

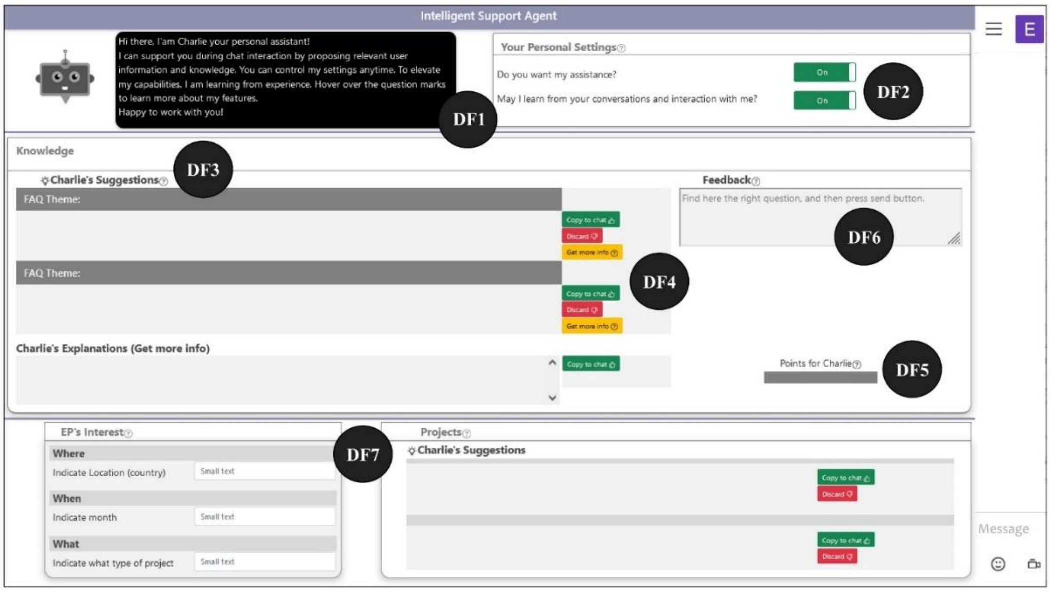

Figure 16. HIS prototype by Poser et al. (2022a)

The instantiated HIS prototype in OCS comprises the following features. First, it includes an avatar greeting the SE, introducing itself, and providing usage explanations (Figure 16 [1]). Second, it allows the SE to adjust the AI in terms of its learning and augmentation behavior (Figure 16 [2]). Third, the SE is dynamically augmented with two response suggestions in order of accuracy (Figure 16 [3]). Fourth, the HIS provides buttons to effortlessly use, view alternatives to, or attain more information about a knowledge suggestion (Figure 16 [4]). Fifth, a counter tracks the explicit feedback provided to the AI in terms of copied or discarded response suggestions (Figure 16 [5]). Sixth, to provide more detailed feedback to the AI, SEs can choose and submit an accurate FAQ in order for the AI to match it with the current customer request and extend its dataset (Figure 16 [6]). The seventh UI element enables both the AI and SE to share knowledge of the customer's interests, based on which the AI presents suitable product suggestions (Figure 16 [7]). For this prototype, the AI was equipped with real ML capabilities based on the FAQ dataset from the first cycle.

The accumulated design knowledge of Wiethof and Bittner (2022) and Poser et al. (2022a) provides a very promising solution for designing interactions for mutual augmentation within an HIS to integrate hybrid learning and real-time decision augmentation to enhance task performance. To increase the projectability of the solution, Wiethof et al. (2022a) contribute prescriptive design knowledge for designing HISs in OCS, however, distinct from Wiethof and Bittner (2022) and Poser et al. (2022a), the AI's form of appearance is a CA instead of embedded AI. We contribute design knowledge in the form of validated DPs and an instantiated and evaluated prototype (Figure 17) by drawing on social response theory (Nass et al. 1994; Nass and Moon 2000), a model from team research (Kozlowski and Bell 2003), and the taxonomy of AI integration into OCS (Poser et al. 2022b). With this, we confirm the transferability of the design knowledge not only to both embedded AI and conversational AI but also to two different organizations.

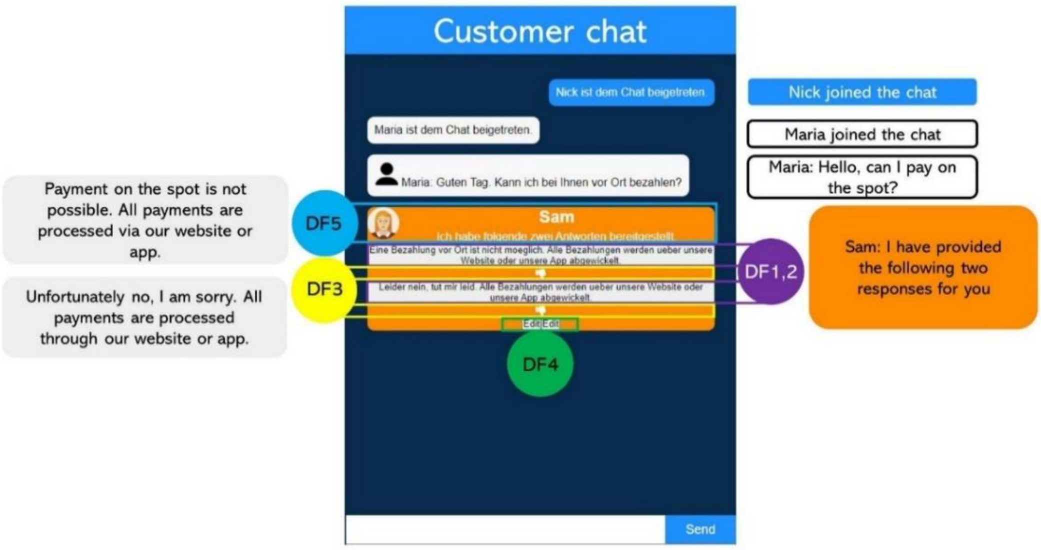

Figure 17. HIS prototype by Wiethof et al. (2022a)

The HIS prototype depicted in figure 17 comprises the following features. First, similar to Poser et al. (2022a), the AI augments the SE with two response suggestions. In doing so, it differentiates between factual and personalized wordings (Figure 17 [1]). Second, also in line with Poser et al. (2022a), buttons support the effortless use and adaptation of the response suggestions (Figure 17 [2, 4]). Third, providing feedback to the AI is also similar to Wiethof and Bittner (2022) and Poser et al. (2022a), that is, the AI learns through direct feedback via a button or implicit feedback via response suggestion usage (Figure 17 [3]). Finally, the AI has a virtual identity (i.e., an avatar and a name) and sends text messages to the SE, along with the response suggestions.

The last publication contributing to the overall design knowledge for HISs in OCS draws on the framework for gamifying collaboration processes by Wiethof et al. (2021a) in order to gamify HITL within HISs (Wiethof et al. 2022b) (Figure 14). As such, it contributes design knowledge in the form of validated DPs and an instantiated and evaluated prototype for providing HISs with gamification elements in order to motivate expert users to involve in HITL (Figure 18).

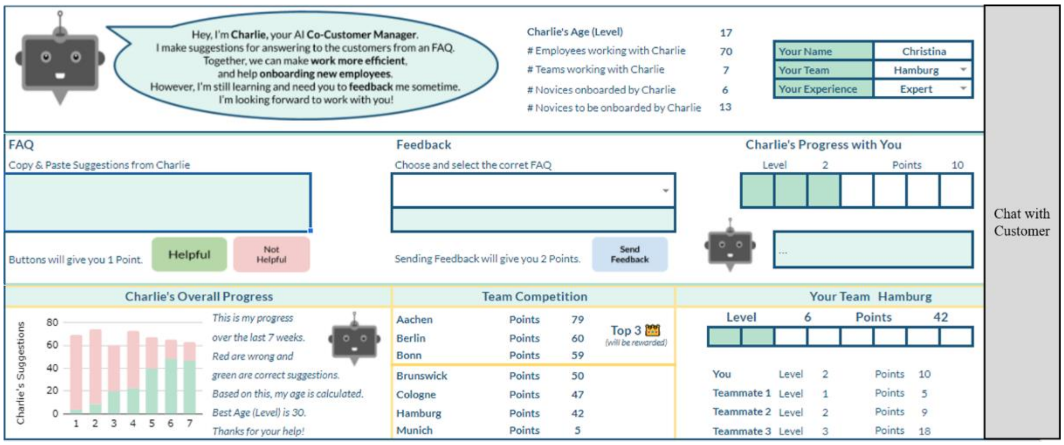

Figure 18. HIS prototype with gamification elements by Wiethof et al. (2022b)

5.4.2 Toward a Design Theory for HISs

To address the overarching RG, the dissertation produces prescriptive design knowledge toward a theory for design and action (Gregor 2006) for developing human-centered HISs that ensure interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and improved task performance. Underpinned by eight publications, the contributions of the dissertation can be positioned as follows.

First, Wiethof and Bittner (2021) contribute descriptive knowledge in the form of patterns conceptualizing hybrid learning based on extant literature (Gregor and Hevner 2013; vom Brocke et al. 2009; Webster and Watson 2002). Second, Poser et al. (2022b) contribute a DSR artifact of the type model in the form of a taxonomy for the integration of AI into OCS (Gregor and Hevner 2013; Kundisch et al. 2021; March and Smith 1995). Third, Wiethof et al. (2021a) contribute prescriptive knowledge in the form of a framework for gamifying collaboration processes (Gregor and Hevner 2013). The remaining five publications contribute design knowledge in the form of prescriptive DPs and situated artifact implementations (Gregor and Hevner 2013) as follows: a prototype instantiated and evaluated by means of WOz (Wiethof and Bittner 2022), three fully functioning prototypes (Poser et al. 2022a; Wiethof et al. 2021b; Wiethof et al. 2022a), and a simulated proof-of-concept prototype (Wiethof et al. 2022b). With this, we contribute to a nascent design theory in the form of DPs and instantiated design artifacts to achieve an improvement, namely, a new solution in the form of an HIS for a known problem contextualized in OCS to enhance task performance and enable SEs' on-the-job learning (Gregor and Hevner 2013). However, apart from artifact design, abstract contributions by means of design theorizing are also important in DSR (Baskerville et al. 2018; Gregor 2006; Gregor and Hevner 2013). Accumulating the design knowledge developed over the course of this dissertation paves the way for a theory for design and action (Gregor 2006). Such a design theory *"[s]ays how to do something. The theory gives explicit prescriptions (e.g., methods, techniques, principles of form and function) for constructing an artifact"* (Gregor 2006, p. 620). The DPs developed throughout the research endeavors of this dissertation represent such principles of form and function (Carlsson et al. 2011; Chandra et al. 2015). The artifact instantiations implementing the DPs for demonstration imply principles of implementation, which are also necessary for design theory development (Chandra et al. 2015; Jones and Gregor 2007). After all, "*the design artifact encodes its design theory*" (Baskerville et al. 2018, p. 368).

Following the DSR theory-development models of Carlsson et al. (2011) and Kuechler and Vaishnavi (2012), extant theories and knowledge are considered for grounding the design knowledge. Regarding design theories' prescriptive nature (Gregor 2006; Walls et al. 1992), DPs are the essential means of providing prescriptive knowledge on how to design and develop an artifact (Chandra et al. 2015). Moreover, DPs prescribe the development, or improvement, of a solution that is useful for the specific problem situation (Venable 2006). By evaluating the effectiveness and efficiency of the solution in addressing the problem, a preliminary design theory can be formulated (Koppenhagen et al. 2012; Venable 2006). Therefore, the eight components of an IS design theory proposed by Jones and Gregor (2007) are applied (Table 11). In the course of this dissertation, the included publications predominantly use qualitative evaluations and are strongly constrained by the OCS environment. Accordingly, this dissertation lays out a foundation toward a design theory for HISs and calls for further research to achieve higher generalizability and projectability of the design knowledge to other organizational augmentation endeavors implementing HISs to facilitate hybrid learning and enhanced task performance (Koppenhagen et al. 2012; vom Brocke et al. 2020). Additionally, quantitative experimental evaluations underpinning design theory hypotheses would support confidence and the truthfulness of the design theory (Venable 2006; vom Brocke et al. 2020). The preliminary design theory developed over the course of this dissertation supports the causal coherences of design knowledge as depicted in figure 13.

Table 11. Preliminary design theory for human-centered HISs to facilitate hybrid learning and improved task performance

Component	Description
Core components	
1) Purpose and scope	Validated design knowledge for human-centered HISs ensuring interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and improved task performance.
2) Constructs	Human-AI collaboration, hybrid learning.
3) Principles of form and function	**DP1:** Provide the HIS with AI capable of understanding domain-specific language in order for the AI to process task-related input to augment the human with its output, given a shared contextual understanding. **DP2:** Provide the HIS with controllable AI in order for the human to involve in HITL ensuring continuous augmentation of the AI, given that the human uses, adapts, and/or provides feedback on the AI's output. **DP3:** Provide the HIS with motivational elements in order to encourage expert users to involve in HITL. **DP4:** Provide the HIS with dynamic UI elements in order for the AI to involve in CITL to ensure continuous augmentation of the human, given that the AI displays suitable output according to the context and the human's needs. **DP5:** Provide the HIS with augmentation strategies toward the human in order for the human to learn and enhance task performance through CITL. **DP6:** Provide the AI with its own representation equipped with social cues in order for the human to perceive and accept the AI as a collaboration partner, given collaborative team structures within the HIS.
4) Artifact mutability	Implications for improving HISs are developed and evaluated over five design artifacts covering the form of AI appearance, augmentation strategies, and integration of motivational elements.
5) Testable propositions	An approach to designing human-centered HISs based on the DPs stated above.
6) Justificatory knowledge	The underlying theoretical foundations comprise hybrid learning patterns based on extant literature on HI combining HITL and CITL, the IS continuance model, social response theory, cognitive load theory, and the framework for gamifying collaboration processes based on motivation and gamification models.
Additional components	
7) Principles of implementation	The DPs derive from the abstraction of several DPs that are formulated in the course of five DSR publications that are predominantly contextualized in OCS. Application of the contextualized DPs is covered by the demonstration of instantiated design artifacts. Additionally, the taxonomy of AI integration into OCS is confirmed to provide helpful guidance for designing HISs in the synchronous augmentation scenario in the frontstage of OCS.
8) Expository instantiation	The underlying contextualized DPs were validated through prototype instantiations, from a proof-of-concept to fully functioning HISs.

6 Practical Contribution

The DSR contributions are twofold, having an impact on both research and practice. This section presents the contributions of the dissertation to practice, constituting the relevance cycle of the overarching DSR project. It complements the theoretical contribution by ensuring the relevance of the validated design knowledge in practice (Baskerville et al. 2018; Hevner et al. 2004; Hevner 2007). Therefore, design knowledge addresses requirements that meet human needs (RQ1) and is applied in the contextual environment to evaluate its causal coherences (RQ5), that is, the extent to which the design knowledge has an impact on the outcome (Figure 13). After all, prescriptive design knowledge is supposed to have a utilitarian character linking the problem with the solution space (Gregor 2006; Gregor and Hevner 2013; Venable 2006; vom Brocke et al. 2020). It has exterior impacts on practice (Baskerville et al. 2018) for effectiveness and efficiency (Baskerville et al. 2018; Venable 2006), as well as interior impacts (Baskerville et al. 2018) providing practitioners with guidance on how to apply the design knowledge in their organizational context.

For contextual requirements and the application of the design knowledge in the environment, three organizations with OCS were selected and involved over the course of the dissertation's research endeavors for relevance. This has the added effect of enhancing the projectability of the accumulated design knowledge (vom Brocke et al. 2020). Due to the close relatedness of the relevance and rigor cycles through the design cycle, the practical contributions are tightly connected with the theoretical contributions. Thus, the practical contributions are structured according to the theoretical contributions presented in the previous section. Table 12 depicts the areas impacted by the dissertation (i.e., hybrid learning, gamifying collaboration processes, AI integration into OCS, and design knowledge for HIS in OCS) and the organizations involved through the respective individual publications.

Table 12. Organizations involved for DSR relevance

	Organization X	Organization Y	Organization Z
Hybrid Learning	Wiethof and Bittner (2022), based on Wiethof and Bittner (2021)		
Gamifying Collaboration Processes	Wiethof et al. (2022b), based on Wiethof et al. (2021a)		
AI Integration into OCS	Poser et al. (2022b)		Wiethof et al. (2022a)
Design Knowledge for HIS in OCS	Wiethof and Bittner (2022), Poser et al. (2022a), based on Wiethof et al. (2021b)		Wiethof et al. (2022a)

Organization X has not integrated AI into their OCS yet. It serves as a use case for designing, developing, and implementing an HIS with embedded AI in the OCS frontstage for interactions for mutual augmentation of AI and SE. Organization Y, on the other hand, has already implemented an AI-enabled agent in its OCS frontstage to perform service delivery. It serves as a use case for analysis. Lastly, organization Z is similar to organization X. It serves as a second use case for an HIS-implementation project. However, instead of embedded AI, the HIS is planned to have a conversational interface by means of an AI-enabled CA.

6.1 Hybrid Learning

Wiethof and Bittner (2021) contribute a conceptualization of hybrid learning patterns to the scientific knowledge base. With this, we provide a foundation for analyzing and developing HISs that combine HITL and CITL by means of interactions for mutual augmentation. As such, not only can researchers draw on our findings but practitioners can also use the patterns as guidance for designing HISs according to their use case and goals. Moreover, they can exploit the examples and measurements provided within the systematically reviewed and organized literature. For instance, Dellermann et al. (2019b) and Liu et al. (2014) disclose the potential of implicit knowledge transfer from experts to novices within HISs in organizations. This concept has been further developed and applied in the form of design knowledge to organization X within the work of Wiethof and Bittner (2022). To ensure relevance in terms of utility, both novice and experienced SEs were involved in identifying requirements, as well as in the evaluation activities. For the requirements, six experienced and five novice SEs were interviewed. For the evaluation, 31 experienced and 30 novice SEs used a prototypic HIS instantiating the design knowledge in a simulated OCS scenario. As both quantitative and qualitative results confirm significant learning progress for novices as well as high satisfaction and continuance intention to use for both novice and expert SEs, practitioners can beneficially draw on these findings. Thus, practitioners may consider enabling implicit knowledge transfer through an HIS as it ensures accessibility of information at any time and not only fosters learning but also reduces novices' dependency on experts' availability, leading to enhanced task performance.

6.2 Gamifying Collaboration Processes

Regarding contributions to the area of hybrid learning, novices gain the most benefit from working within an HIS in terms of learning (Wiethof and Bittner 2022). However, an HIS requires both human augmentation with AI and AI augmentation with human intelligence (Dellermann et al. 2019b). Accordingly, apart from focusing on human augmentation to teach novices, a human-centered design should equally consider experts' motivation to teach AI. Therefore, Wiethof et al. (2022b) consider gamification as one possible solution and develop design knowledge to implement gamification elements into an HIS gamifying HITL. For relevance and validity of the design knowledge, it is, similar to the design knowledge provided by Wiethof and Bittner (2022), applied to organization X but with a focus on experienced SEs. Hence, for the identification of requirements, we interviewed nine expert SEs, and for the evaluation of design knowledge, we involved 11 expert SEs in user test simulations. Based on the promising results, we encourage practitioners to consider gamification for HITL learning as one possible solution to motivate experts to involve in HITL.

Lastly, by aligning with and grounding on the framework for gamifying collaboration processes (Wiethof et al. 2021a), Wiethof et al. (2022b) confirm its applicability to human-AI collaboration. As such, not only researchers can draw on the framework to gamify collaboration processes, but so can practitioners. They can use the framework in a top-down manner as a guideline for the implementation of gamification in HISs by defining intended outcomes, deriving gamification principles, and finally, identifying appropriate gamification elements (Wiethof et al. 2021a; Wiethof et al. 2022b).

6.3 AI Integration into OCS

Complementary to the contributions to the scientific knowledge base of OCS, the dissertation provides practical implications for AI integration into OCS. Grounded on empirical and scientific knowledge and constituting the contextual foundation for the design knowledge developed over the course of this dissertation, the taxonomy of AI integration into customer service (Poser et al. 2022b) not only serves the purposes of researchers but also practitioners in their AI integration endeavors.

Firstly, organizations can use the taxonomy to analyze AI solutions already deployed in their operations to disclose aspects and new perspectives that can be further considered and improved. Organization Y represents such a use case with an AI-enabled agent that has already been implemented. Two researchers and one practitioner from the organization successfully applied the taxonomy to their use case, demonstrating its relevance by means of an ex post evaluation.

Secondly, practitioners can make use of the taxonomy as a guideline to plan and structure AI integration endeavors along the dimensions of the taxonomy in sequential order, thereby benefiting from the ideas and insights provided by the taxonomy’s characteristics for each dimension. Additionally, they can refer to examples from scientific literature, as well as to the solutions from practice that have been investigated and systemized for the development of the taxonomy. By means of the ex post evaluation, two researchers and two practitioners from organization X accordingly applied the taxonomy to their use case to plan the integration of AI into the OCS frontstage in the form of an HIS. These practitioners work in the IT department of the organization, which would be responsible for such AI integration initiatives. The successful application of the taxonomy to the second use case strengthens the relevance of the taxonomy and its contributions to practice. Moreover, Wiethof and Bittner (2022) and Poser et al. (2022a) undertake a multicyclic DSR project, in which they design and develop a prototypic HIS, as well as implement and evaluate it in organization X. Its characteristics are consequently well-aligned with the successful application of the taxonomy to the use case in Poser et al. (2022b) along the dimensions of the taxonomy (D1-12) as follows. The AI is integrated into the OCS frontstage (D1) and supports a seamless service process (D2). It augments and supports the SE (D3) by means of mechanical and analytical tasks (e.g., response suggestions based on FAQs) (D4). In addition to response suggestions, the AI can augment the SE with customer process information and customer insights, which are inquiry-related, process-focused, and customer-related (D5). Employing AI augmentation through the SE, performance monitoring of the AI is ensured (D6). In terms of the AI integration itself, inquiries are handled simultaneously by the AI and SE (D7), although the AI is rather reactive (D8) and takes the form of embedded AI (D9) only visible to the SE (D10). At last, the AI's data- and knowledge-processing capabilities are built upon ML techniques (D11), and input can be provided before, during, and after interaction (D12).

Finally, with the work of Wiethof et al. (2022a), a third use case drawing on the taxonomy and demonstrating its relevance to practice is provided by organization Y. Their AI integration into OCS is quite similar to the use case of organization X. However, instead of embedded AI, we make use of an AI-enabled CA (D9) (i.e., we provide the HIS with a conversational interface).

6.4 Design Knowledge for HISs in OCS

This dissertation contributes design knowledge toward achieving the overarching RG of designing and developing human-centered HISs that ensure interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and improved task performance. This contribution to the scientific knowledge base accords with implications for practice by means of application and evaluation in the contextual environment. The design knowledge is required to be useful and effective in matching the solution space with the problem space (i.e., the causal coherences of design knowledge lead to an outcome, which is, specifically for HISs, hybrid learning and improved task performance) (Carlsson et al. 2011; Venable 2006) (Figure 13).

By involving two OCS use case scenarios from organizations X and Y (Table 13) in the DSR endeavors to develop design knowledge for HISs, validity, relevance, and DP utility in the contextual environment are ensured and implications for practice derived. As such, SEs working in the OCS of either one of the organizations are involved in both identifying requirements for the solution as well as evaluating the design knowledge in the form of DPs and instantiated design artifacts. After successful evaluation, both the DPs and instantiated design artifacts contribute to practice. The DPs are formulated with a well-balanced abstraction level in order to support contextually situated implementations more effectively (Wache et al. 2022). The instantiated design artifacts provide a demonstration of a possible solution for how the DPs can be instantiated or translated into DFs. As for the cumulative nature of the design knowledge developed over the course of this dissertation, table 13 depicts the core design knowledge contributions for HISs.

Table 13. Design knowledge contributions distributed by organization

Wiethof et al. (2021)	**DPs**	5 DPs based on 17 MRs facilitating acceptance of an AI teammate in hybrid human-AI collaboration teams
	Artifact	Digital collaborative writing process facilitating the collaboration of humans and AI
Organization X		
Wiethof and Bittner (2022) (first cycle)	**DPs**	7 DPs based on 16 MRs from practice facilitating the design and development of an HIS resulting in hybrid learning, supporting high satisfaction and continuance intention to use
	Artifact	Prototypic HIS by means of WOz
Poser et al. (2022) (second cycle)	**DPs**	DPs from Wiethof and Bittner (2022) + 4 DPs based on 12 MRs from theory facilitating the design and development of an HIS resulting in enhanced task performance while ensuring hybrid learning
	Artifact	Further development of the prototypic HIS of Wiethof and Bittner (2022) with NLP capabilities based on ML to achieve a fully functioning HIS prototype
Organization Y		
Wiethof et al. (2022)	**DPs**	5 DPs based on 12 MRs from practice and literature facilitating the design and development of an HIS with a conversational interface resulting in enhanced task performance
	Artifact	Fully-functioning HIS prototype with NLP capabilities based on rule-based reasoning

These contributions all have practical implications for organizations that have been integrating or plan to integrate AI into their organization, specifically in OCS. First, Wiethof et al. (2021b) raise practitioners' awareness to consider and leverage hybrid human-AI collaboration in their operational business. To this end, we provide guidance on how to design AI as teammate in terms of both personality and capabilities. Second, Wiethof and Bittner (2022) propose the use of an HIS as an opportunity for implicit knowledge transfer from experts to novices. For this purpose, we contribute DPs and an instantiated design artifact combining HITL and CITL to facilitate hybrid learning on the job. Poser et al. (2022a) draw on these findings and provide a potential solution in the form of DPs and an instantiated artifact to integrate hybrid learning and real-time decision augmentation into an HIS to achieve improved task performance. Practitioners may use these findings to develop and deploy HISs in their organizations. Finally, practitioners may also draw on the findings of Wiethof et al. (2022a) if they plan to design and develop an HIS with a conversational interface (i.e., an AI-enabled employee-facing CA instead of embedded AI). As such, we offer guidance through DPs and an instantiated design artifact on how to design and develop employee-facing CAs in HISs to facilitate enhanced efficiency in OCS.

7 Limitations

Despite the promising results of the dissertation, there are a few limitations to consider that constrain the generalizability of the results in terms of the selected knowledge sources, focus on human-centered design knowledge, evaluation methodology, and contextual environment.

First, regarding the choice of knowledge sources, using different or more scientific databases or search strings might have led to other or more results and insights. This specifically applies to the conceptual foundations provided by Wiethof and Bittner (2021) in terms of hybrid learning patterns, as well as Poser et al. (2022b) in terms of the taxonomy of AI integration into customer service. Beyond the scientific knowledge sources, modification or addition of the knowledge sources from practice could equally impact the results of this dissertation (e.g., the stakeholders, experts, and participants involved in qualitative interviews and evaluations). This applies to most of the included publications as they predominantly conduct qualitative research to ensure relevance in the application domain. However, while Wiethof et al. (2021b) and Wiethof et al. (2021a) select participants in a convenience-based manner (i.e., not from an organizational context), Wiethof and Bittner (2022), Poser et al. (2022a), Wiethof et al. (2022a), Wiethof et al. (2022b), and Poser et al. (2022b) embed their work in the OCS domain and ensure an appropriate selection of employees working in either organization X, Y, or Z to ensure relevance to the research endeavors. Regarding the taxonomy of AI integration into customer service, it is further constrained by the selected empirical solutions from practice as there are many more solutions available (Poser et al. 2022b).

Second, the evaluation of actual AI improvements through HITL has been neglected due to a focus on human-centered design aimed at improving learning and task performance. Wiethof and Bittner (2022) deployed a prototypic HIS without any AI capabilities, rendering a proper evaluation of AI learning in terms of HITL impossible. However, using the WOz technique, we found promising results showing SEs' satisfaction and continuance intention to use HIS in terms of hybrid learning. With this, we pave the way for the development of a fully functioning HIS prototype. The included publications that deploy a fully functioning prototype with AI capabilities focus on designing interactions for mutual augmentation to increase acceptance of the AI in hybrid teams (Wiethof et al. 2021b; Wiethof et al. 2022a) and to support humans' decision-making to enhance overall task performance (Poser et al. 2022a; Wiethof et al. 2022a). Thus, the optimization or assessment of technical performance did not fall within the scope of this dissertation (Poser et al. 2022a; Wiethof et al. 2021b; Wiethof et al. 2022a). Nevertheless, the design knowledge developed in each publication covers HITL to ensure the mutual augmentation necessary for HI (e.g., in the form of feedback functionalities) (Wiethof et al. 2022b). Consequently, only if and how humans are involved in HITL were evaluated but not how effective their feedback actually is for AI learning.

Third, regarding the evaluation methodology, a prime focus was placed on qualitative methods (e.g., user tests and following expert interviews), implying several limitations. For instance, the evaluation of the HIS prototypes is limited to one organization and one situated instantiation of the design knowledge per publication (Poser et al. 2022a; Wiethof et al. 2022a; Wiethof et al. 2022b; Wiethof and Bittner 2022). In addition, we only conduct semi-naturalistic evaluations by simulating the customers with predefined scripts. Consequently, task performance could only be evaluated from the SE perspective and does not include customer satisfaction. As for hybrid learning, only short-term evaluations were conducted. The neglect of long-term effects was compensated for by measuring the continuance intention to use, which does not completely represent actual usage behavior. Apart from that, Wiethof and Bittner (2022) made use of the WOz technique, which biases the overall results, as potential drawbacks in AI capabilities are not considered. Nevertheless, by building on the results of Wiethof and Bittner (2022), Poser et al. (2022a) develop a fully functioning HIS prototype with AI capabilities, underlying the design knowledge of both publications. To develop a solid design theory based on justified means-end relationships of solution and problem space, quantitative evaluations are necessary to measure actual outcomes, that is, improvements in learning and performance when implementing the design knowledge compared to a baseline (Carlsson et al. 2011; Venable 2006). Although implementation, demonstration, and following user tests can prove the validity of the DPs, isolated investigations of DPs' effects on learning and task performance are necessary to draw solid conclusions without interference from other DPs. Additionally, tacit design and designers' biases should be considered and reflected upon when evaluating the formulated design knowledge (Kuechler and Vaishnavi 2012). In general, with qualitative evaluations, human bias needs to be considered carefully.

Lastly, the results are predominantly constrained to the selected contextual environment of OCS. Although the contextualized DPs have been abstracted toward a theory for design and action, they have not yet been validated and applied to other contexts. However, due to the cumulative nature of the dissertation, three different organizational OCS use cases could be involved in developing and evaluating design knowledge. This would increase generalizability while remaining pertinent to OCS.

8 Implications for Further Research

Apart from the theoretical and practical contributions toward designing and developing HIS, this dissertation discloses avenues for future research. According to the structure of the foundations and contributions, the implications for further research are similarly structured along the adapted framework of human-computer interaction for user interaction with AI-enabled systems (Rzepka and Berger 2018; Zhang and Li 2005) (Figure 2).

8.1 Interaction and Outcomes

The RG of the dissertation is to develop validated design knowledge for human-centered HISs to ensure interactions for continuous mutual augmentation of humans and AI to facilitate hybrid learning and improved task performance. As such, the design knowledge addresses and constitutes the interaction and outcomes construct of the framework of human-computer interaction (Figure 13) (Carlsson et al. 2011; Rzepka and Berger 2018; Zhang and Li 2005). For the purposes of further research, the DPs and instantiated design artifacts provide a great reference for the development of HISs. Based on the design knowledge developed and contributed over the course of this dissertation, following implications are worthy of note.

First, to compensate for one of the limitations of the selected evaluation methodology, conducting long-term studies would probably yield promising insights into actual usage behavior and learning progress. To further complement the primarily qualitative results of the evaluations conducted over the course of the dissertation, future research would also benefit from quantitative analyses employing advanced experimental settings. For instance, isolated DPs could be evaluated against control instances to measure the effects related to the specific DP. Moreover, future research should implement and evaluate the design knowledge in naturalistic settings for more realistic results on usage behavior, task performance, and learning. In addition, such an environment could support researchers in measuring the quality of humans' augmented decisions (e.g., in OCS, by assessing customer satisfaction with the service provided).

Second, the design knowledge is focused on human-centered HISs (i.e., although it ensures both HITL and CITL, the design implications predominantly address CITL). Therefore, further research should complement these findings by specifically investigating HITL within HISs from a technical perspective. It might be interesting to design and evaluate strategies of how to best enable HITL to facilitate the continuous development of AI, specifically, how humans should augment the AI (e.g., in terms of adequate feedback mechanisms). To assess HITL success, researchers should then measure AI learning based on ML, that is, how HITL affects the AI's results in a specific timeframe. Along with humans' involvement in HITL, it would be important to find the right balance between only making adjustments of AI output or fully relying on it. Both extremes can have negative impacts on hybrid learning and task performance.

Third, regarding the design of iterative hybrid learning cycles, future research might contribute descriptive knowledge by observing and analyzing the different hybrid learning patterns, their combinations, and the corresponding relationship of AI and human with respect to their effects on hybrid learning or task performance.

Finally, future research might benefit from establishing an evaluation framework for HISs. Such a framework could provide a fundament for developing and evaluating HISs for both researchers and practitioners. For instance, it might determine key performance indicators covering evaluation from social, technical, and socio-technical perspectives.

8.2 System Characteristics

Regarding system characteristics, future research implications derive from the dissertation's contributions to gamifying collaboration processes as well as team research for AI characteristics.

First, future research might observe, measure, and analyze gamification effects on satisfaction, motivation, and continuance intention to use in a long-term study. Accordingly, the framework for gamifying collaboration processes could adapt a descriptive perspective by determining the means-end relationships of selected gamification elements with intended outcomes in specific contexts. Furthermore, future research might also extend the framework with complementary theories, principles, models, and elements. In the end, the framework is limited to gamifying collaboration processes to enhance continuance intention to use. Other solutions should be considered by future research that could motivate participation in collaboration processes, and particularly in human-AI collaboration processes.

Second, drawing on team research, there remains untapped potential for examining the relationship of human and AI within an HIS. In particular, mechanisms to establish acceptance of AI could be an advantage for further research. For instance, the distinct interactions between humans and AI could be investigated to facilitate enhanced team spirit and personalization. Furthermore, with respect to AI's virtual identity, future research could evaluate the impact of personality attributes against an AI baseline condition without any virtual identity.

8.3 User Characteristics

With regard to user characteristics, the dissertation is grounded on the idea of differentiating between experts and novices when working with AI within HISs, as well as on endeavors toward the cognitive fit between user and system. As such, the dissertation provides prescriptive design knowledge to facilitate hybrid learning by means of an implicit knowledge transfer, as well as improved task performance through a cognitively adequate and intuitive UI design. Based on the integration of the findings, future research might contribute descriptive knowledge that provides the means-end relationships of specific UI design choices and their effects on learning and task performance. For instance, the amount and form of AI output could be evaluated against a baseline condition in an experimental setting with respect to measurable impacts on the learning and task performance of users, differentiating between experts and novices. Eventually, based on the descriptive knowledge, researchers could provide strategies and guidelines to develop HISs according to the respective cognitive needs of either novice or expert users to facilitate hybrid learning and enhanced task performance.

8.4 Task and Context

There are two overarching future research avenues considering the task and context construct of this dissertation, namely, specific implications for further research in OCS and implications for the generalizability of the design knowledge developed over the course of this dissertation.

In terms of OCS, the dissertation encourages further research on AI integration into customer service. Therefore, future work could build on the taxonomy and design knowledge provided with this dissertation, as well as validate and extend them respectively. For example, the design knowledge for HISs could be combined with research on AI-SE handovers (i.e., by integrating sequential with simultaneous service delivery). Moreover, specifically drawing on the taxonomy for AI integration into customer service, future research could yield promising insights by conducting descriptive research observing and analyzing causal coherences of design decisions with certain effects in specific contexts.

Finally, the application of the design knowledge to various other organizational contexts apart from OCS with larger samples might support the generalization of the design knowledge for HISs.

9 Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI

Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021b. “Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI,” in *54th Hawaii International Conference on System Sciences (HICSS)*, Virtual Event, pp. 400-409 (doi: 10.24251/HICSS.2021.047).

Abstract. This paper aims at implementing a hybrid form of group work through the incorporation of an intelligent collaborative agent into a Collaborative Writing process. With that it contributes to the overall research gap establishing acceptance of AI towards complementary hybrid work. To approach this aim, we follow a Design Science Research process. We identify requirements for the agent to be considered a teammate based on expert interviews in the light of Social Response Theory and the concept of the Uncanny Valley. Next, we derive design principles for the implementation of an agent as teammate from the collected requirements. For the evaluation of the design principles and the human teammates’ perception of the agent, we instantiate a Collaborative Writing process via a web-application incorporating the agent. The evaluation reveals the partly successful implementation of the developed design principles. Additionally, the results show the potential of hybrid collaboration teams accepting non-human teammates.

9.1 Introduction

Research on Artificial Intelligence (AI) is increasingly progressing shown by many new evolving technologies. Here, researchers mainly work on questions of effectiveness and efficiency regarding their newest developments [1]. Especially in the field of Machine Learning (ML) researchers aim to create an AI, which resembles Human Intelligence and could consequently replace a human being [2]. Thereby, they focus on an automatic learning approach [3] resulting in intelligent, autonomous systems. In certain domains with a huge amount of training data, this approach has already been successfully recognized [1, 4, 5].

However, it is known that technology is not everything [6]. Researchers aim to achieve a synergy of both humans and AI, i.e. combining the benefits and advantages of both [2, 7–10]. Therefore, also the human users’ social perspective [11] is required. Even the best state-of-the-art technology will be useless, if its human users refuse it [6]. This also applies to ML approaches themselves, especially when they involve human users in the training, e.g. Reinforcement Learning or Human-in-the-Loop [1, 5, 12]. Thus, to achieve that synergy of working together and complementing as well as learning from each other, the human needs to accept a collaborative agent willing to learn from its contributions as well as to make corrections and improve the agent [2]. It should be pointed out, that throughout the paper we use the term *agent* for any *collaborative agent* and intelligent *computer agent* respectively "[covering] the idea of creating machines that can accomplish complex goals [including] facets such as natural language processing, perceiving objects, storing of knowledge and applying it for solving problems" [8] in collaboration settings.

As there is an advantage of combining human and artificial intelligence to achieve better collaboration outcomes [2, 8], the research gap and need for designing and developing such socio-technological teams has been disclosed [8]. We therefore consider socio technical factors of agent teammates and exemplify the intended synergy by regarding hybrid teams involving humans and agents. To specifically contribute and extend the scope of this research, Dellermann et al. [8] call for more research on practical applications in different domains. For instance, Bittner et al. [13] developed a taxonomy for conversational agents in collaborative work. Epstein [9], on the other hand, investigated a collaborative intelligence sharing a task with a person to demonstrate the potential synergy of humans and agents. Eventually, "rather than re-design our world for computers or submit to their decisions, we should begin to share our tasks with them" [9]. As we found a study, which revealed that an agent is capable of replacing actual human journalists [14], for our research at hand, we specifically regard a Collaborative Writing (CW) scenario. After all, there could rather be an advantage in the collaboration of an agent and human writers. For one thing, an agent may have more memory space and a higher computation rate as well as challenges the writer and promotes the writing process. For another thing, agents do not reach humans' skills and knowledge yet. Thus, by co-writing, the skills of the agent as well as the writer affect both the outcome as well as each other complementary. Manjavacas et al. [15] addressed this by developing an intelligent text generation system, which produces sentences or paragraphs to enable co-creation and CW between an author and an agent [15]. By doing this, agents contribute with story fragments and ideas, which the human collaborator might not be aware of [16]. Additionally, computational creativity itself has already achieved several successes, e.g. Narrative Science [16], poetry [17], storytelling [18] or melodic accompaniments for lyrics [19].

Still, "as machines evolve from tools to teammates, one thing is clear: accepting them will be more than a matter of simply adopting new technology" [6]. By fostering co-creativity in CW within a hybrid team, we examine the possibility of perceiving an agent as teammate [15]. This will enable further research on implementing hybrid forms of group work covering mutual learning benefits and acceptance. With that we aim to contribute to the overall research gap establishing acceptance of computer agents toward complementary hybrid work [2, 8, 10]. Therefore, we are conducting design science research to implement an agent into a collaborative writing process as teammate [10]. By doing this, we address three research questions: Q1: What are the requirements to ensure acceptance of an agent as teammate in CW? Q2: How can an agent be designed and implemented as teammate contributing to the goal of the CW process? Q3: How do the human teammates perceive and accept the contributions of the agent?

To support CW, we develop a CW process with an agent teammate and implement it on a web platform.

9.2 Research Approach

The research aims to contribute prescriptive design knowledge to the knowledge base by connecting the research areas of Human-Computer-Interaction and Socio-Technical Systems to design a solution for the incorporation of an agent teammate into a CW process [20]. In coherence with the design science research (DSR) approach, the DSR process by Peffers et al. [21] is used to derive design principles (DPs), which are then implemented and evaluated with an instantiated CW process in form of a web-application (see Figure 1) [21]. The problem identification and motivation are covered in the introduction. To define the objectives of the solution, meta-requirements (MRs) for an agent teammate are identified. This includes any personality traits and skills, that need to be assigned to an agent to be considered and accepted as teammate. To do so, we first consider related work from areas focusing on machines as teammates and hybrid teams as well as socio technical factors of agents. We then base our MRs on the Social Response Theory by Nass and Moon [22] aligned to the concept of the Uncanny Valley by Mori [23], and conduct expert interviews according to the approach of Meuser and Nagel [24]. For the design and development, the MRs are considered to derive DPs of an agent teammate, which are later on implemented. After the implementation of the agent in a CW process, a demonstration is carried out by instantiating the CW process in form of a prototypic web-application incorporating the agent [25]. Four groups of five participants took part in a test run [26] and in expert interviews [24] to evaluate the human teammates' perception and acceptance towards the contributions of the agent ex post [27]. Communication will be completed with this paper.

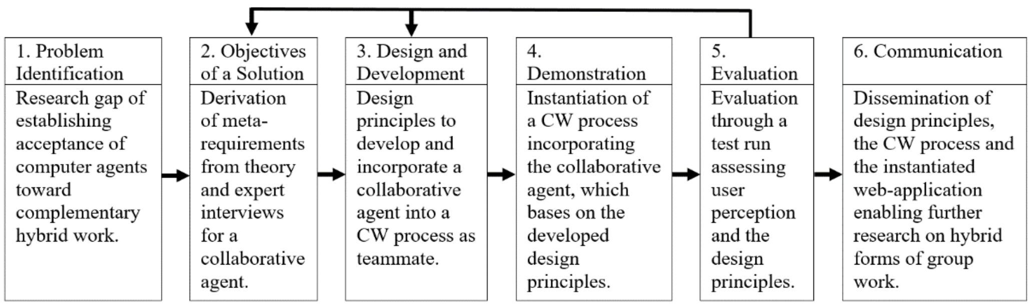

Figure 1. Structure along the DSR process

9.3 Related Work

Combining the strengths of humans and agents in collaborative work is not easy and neither is it enough to make a good collaboration team. Humans still think of technology as a tool, but need to consider and accept it as teammate of a hybrid group [6, 10]. Therefore, research considers social science findings about behaviors or attitudes toward humans and applies them for agents. The “computers are social actors” (CASA) paradigm introduces the relevance of assigning human characteristics, and social cues respectively, to agents [11] encouraging their acceptance [6, 13, 28, 29]. Considering the Uncanny Valley [23] and the balance of social cues and competence, researchers examined the least actual capabilities of an agent, which are to understand its teammates and to react appropriately with adequate length. Eventually, the outcome depends on the contributions of each member including the agent [30–33].

Next, researchers consider the aspect of transparency fostering the understanding of an agent, its behavior and purpose to accept it as a teammate [29, 34, 35]. This allows its human teammates to still critize and improve it [34], which eventually ensures a certain feeling of control as well as an enhancement of the group process and its outcomes [10, 29, 31, 36]. For instance, Gnewuch et al. [33] demand to include error-handling strategies considering potential misunderstandings. Also, Frick [6] suggests to give the human teammates the possibility to influence the computer algorithm’s output. Still, due to the fact that much of today’s technology including its technical details and mechanisms are very complex, humans cannot rely on a full system transparency and understanding to accept an agent teammate [37–39]. Therefore, it is recommended to establish trust and acceptance right at the beginning. Andras et al. [40] suggest making use of an explainable AI. An explainable AI will introduce itself in advance of starting the hybrid collaboration process. Thereby it will give its teammates insights into its behavior covering the how and the why [40].

At last, considering the enhancement of the process outcomes, researchers found out, that agents can contribute to group creativity effects and concurrently avoid negative effects including social loafing and free-riding, evaluation apprehension and production blocking by contributing with its own decisions [36]. However, the competence of an agent is not to be neglected. It involves the knowledge, abilities and skills of a teammate to satisfy the expectations of the other teammates. These expectations refer to the performance, specifically the contributions toward the team goal within the teamwork [35, 37, 38, 41].

As of our research at hand, we aim at the acceptance of an agent as teammate within a hybrid form of group work, specifically CW. Therefore, we consider these findings toward the acceptance of agents in human-computer-interactions, i.e. the application of social cues in terms of CASA and the Uncanny Valley.

9.4 Theoretical Background

In terms of accepting agents, many researchers and practitioners refer to human characteristcs, and social cues respectively [6, 13, 28, 29, 33]. Here, Nass and Moon [22] developed the **Social Response Theory** based on several previous studies, among others around the CASA paradigm, demonstrating the mindless application of social rules and expectations to computers. With that, they disclose the application of human social categories, social behaviors as well as premature cognive commitments to computers, and refute alternative explanations like anthropomorphism and intentional responses for their studies. They state that "inviduals are responding mindlessly to computers to the extent that they apply social scripts […] that are inappropriate for human-computer interaction". Therefore, "individuals must be presented with an object that has enough cues to lead the person to categorize it as worthy of social responses, while also permitting individuals who are sensitive to the entire situation to note that social behaviors were clearly not appropriate" [22]. Thus, social cues assigned to an agent trigger humans to apply social behaviors and rules towards the agent [11]. Such social cues could be a name, emotions [6] or also typing indicators [42]. The latter also addresses the concept of social presence [42], i.e. an agent is perceived as socially present, aware and conscious [32]. Still, next to the Social Response Theory, researchers also refer to the concept of the **Uncanny Valley** by Mori [23] reasoning the application of less social cues in order to match the human likeness with competence for maximum affinity [23, 33, 42]. As it is quite easy to generate a social relationship between humans and computers, it is recommended to make use of rudimentary but powerful cues instead of developing highly complex agents [11].

9.5 Objectives of the Solution

To derive MRs for an agent as teammate from theory and real-live problems, we conducted semi-structured qualitative expert interviews along the approach by Meuser and Nagel [24] and analyzed them in the light of the theoretical background [22, 23]. For the selection of experts (E1-E9), we considered nine diverse researchers from the fields of Information and Knowledge Technology, Human-Computer-Interaction, Psychology and Sociology. The interview guideline included interdisciplinary open questions to reveal the experts' insights in a reliable and unbiased way. The questions asked covered 1) socio-technical factors within human-human- and human-machine-interaction, 2) agents influencing human-machine-interaction with socio-technical factors, 3) desires, demands and anxieties toward the application of agents, and 4) vision and future prospect about the interaction between agents and humans. To analyze the expert interviews, a thematic comparison was conducted along categories [24]. The categories were determined inductively after an initial scanning of the interview transcripts. Thus, the information from the interviews could be extracted and separated into the following categories: Competence, Social Cues and Feedback. Consequently, the experts' relevant remarks were extracted, merged and collocated along the established categories. Eventually, we connected the expert references to the Social Response Theory [22] and the concept of the Uncanny Valley to derive the MRs [23].

Competence: an agent is not expected to have a general human intelligence, but to have a certain expertise in the application area. As such, the agent should be able to enhance and contribute to the group process and its outcomes (MR1) with all the required skills (MR2) (E1, E2, E4-7). Accordingly, its interactions within the group should be transparent, easy to understand and intuitive through an intelligible display (MR3) (E3, E4, E7-9). Referring to the Uncanny Valley, this display does not have to be utterly human. In fact, too much human likeness might raise higher expectations toward the competence of the agent (E2, E4-7, E9). Eventually, the human teammates should have the right expectations and know that their agent teammate acts in their interest (MR4) (E2, E4-7, E9).

Social Cues: specifically regarding the appearance of an agent, referring to Social Response Theory, it is recommended to assign some humanness to the agent, e.g. a name, a face or an emoticon (MR5), as long as the complexity of the agent's functionality matches the complexity of its appearance (MR6) (E2, E4-7, E9). To further encourage social presence (MR7) in the light of Social Response Theory in terms of social cues, graphical typing indicators within the team interactions are useful (MR8). Additionally, as it is beneficial to initially establish trust and an emotional relationship between the humans and the agent in order to jointly work toward a common goal (E2, E4, E5), transparency about the agent's purpose and processes is required (MR9). Therefore, the schema of childlike characteristics might be of interest (E4). Hence, an agent introduces itself and asks for support within the collaboration (MR10) (E3, E7). As a result, the human teammates do not expect the agent to not make any mistakes. This approach resembles self-deprecation, e.g. the agent knows, that it is an agent (E5). Considering Social Response Theory, making use of an explainable AI with self-depreciating and childlike characteristics aims at enocuraging social responses toward the agent leading to a closer and more emotional relationship.

Feedback: in the light of Social Response Theory, there are a few underlying characteristics of an agent, which trigger humans' social responses. The first aspect is interactivity covering responses based on inputs. Despite the writing process, human teammates should have the possibility to understand and control the situation (MR11), i.e. they are able to give feedback and influence (MR12) or even intervene, rectify and amend the agent's contributions at any time (MR13) (E2-5). Therefore, the agent's contributions need to be exposed for criticism and improvement (MR14). This feature is crucial for mutual learning benefits of both humans and agent. Following, an agent teammate should also show an interest in the human teammates. This is possible by giving it the same ability to give feedback (E1, E2) covering a second aspect for social responses: the filling of human roles. Therefore, the agent needs to react appropriately (MR15) by making the right decisions (MR16). With that, all teammates should be able to equally contribute to the process outcome (MR17).

Table 1 includes all identified MRs as objectives of the solution.

Table 1. MRs with description and expert reference

Meta-requirements	Expert reference
MR1: The agent enhances the group process and its outcomes.	E1, 2, 4-7
MR2: The agent has all skills to contribute to the team goal.	E1, 2, 4-7
MR3: The display of the agent is intelligible for its teammates.	E3, 4, 7-9
MR4: The human teammates have the right expectations and know that the agent teammate acts in their interest.	E1-7, 9
MR5: The agent is humanoid owning a name and lifelike characteristics.	E2, 4-7, 9
MR6: The agent remains a balance of social cues and competence.	E2, 4-7, 9
MR7: The agent is perceived as socially present.	E2, 4-7, 9
MR8: Graphical typing indicators are involved within the team interactions.	E2, 4-7, 9
MR9: The agent's purpose and processes are transparent and disclosed via an informative opening message.	E3-5, 7
MR10: The agent is an explainable AI introducing itself in advance.	E3-5, 7
MR11: The human teammates understand the situation and retain control.	E2-5
MR12: The human teammates are able to influence the agent teammate's output.	E2-5
MR13: The human teammates require error-handling strategies for interventions.	E2-5
MR14: The agent's contributions are exposed for criticism and improvement.	E2-5
MR15: The agent reacts appropriately.	E1, 2
MR16: The agent is able to make decisions.	E1, 2
MR17: All teammates equally contribute to the process outcome.	E1, 2

9.6 Artifact Design and Development

Based on the MRs, preliminary action oriented DPs toward the incorporation of an agent into a CW process as teammate were developed according to Chandra et al. [43] (Table 2).

Table 2. DPs with corresponding MRs

Design Principles (DP)	Source
DP1: Provide the agent with the capability of domain-specific natural language processing (NLP) in order for the human teammates to feel understood and obtain appropriate contextual contributions, given that its knowledge is trained, but limited to the context of the teamwork application.	MR15, MR16
DP2: Provide the agent with a controllability in order for the human teammates to have the opportunity to intervene and rectify its contributions, given that the modified new contribution of the machine teammate is qualitatively better and more suitable in regard to the group goal.	MR11, MR12, MR13, MR14
DP3: Provide the agent with the ability to react based on the human teammates' contributions by giving feedback to each individual contribution in order for the human teammates to perceive it as socially present, aware and conscious.	MR1, MR2, MR7, MR15, MR16, MR17
DP4: Provide the agent with explainable capabilities introducing itself including purpose and processes in advance in order for the human teammates to have the right expectations and to understand and accept the agent teammate, given that it is still not perfectly trained and may not make appropriate and useful contributions.	MR3, MR4, MR9, MR10, MR11
DP5: Provide the agent with a humanoid identity and social cues in order for the human teammates to perceive it as an equally social teammate, given a balance of social cues and competence.	MR3, MR5, MR6, MR7, MR8

To support CW, we developed a CW process and implemented it on a web platform. The process enables the participants to collaborate in writing a story. We use the process to design and implement an agent as teammate according to the DPs (Q2). The process steps and activities incorporating the agent are as follows.

1) Prepare

The agent introduces and presents itself right at the beginning to clarify its intended role as a teammate. It explains how it generally works for transparency. Next to its name it also has a picture (DP2, DP4, DP5).

2) Write Sentence

After that, the iterative part of the process starts: the first participant writes a sentence, which extends the story. Here, the agent is included in the order of the participants. When it is its turn, the agent processes the last written sentence to generate a new and contextual appropriate sentence contributing to the story like its human teammates. In doing so, it also takes some time to generate the next sentence. In this waiting period graphical typing indicators show up (DP1, DP5).

3) Extend Story with / without Reaction

There are then three exclusive activities to follow: either claiming, liking or not reacting to the contributed sentence. Exactly like its human teammates, the agent can react to the writer by showing that it likes the released sentence (DP3). The claim-functionality is only available for the human teammates: thereby, they can intervene and demand a rectification of the released sentence of the agent. The agent then generates a better and more suitable new contribution (DP2).

4) Completion

After the first participant's turn, the next participant in line has a turn and writes a sentence. The process ends when the participants consider the story complete. Additionally, for the overall process, the agent has a picture and a name, which is used all along (DP5).

9.7 Demonstration

To assess the incorporation of the DPs (Q2) as well as to evaluate the user perception of the agent (Q3) an instantiation of the collaborative writing process was deployed in form of a web-application (Figure 2). This is done by means of Prototyping: developing and evaluating a standalone version, which is quickly available. It is typically limited to the functionalities, which are relevant for the research involving for one thing the feasibility and for another thing the user perception [25].

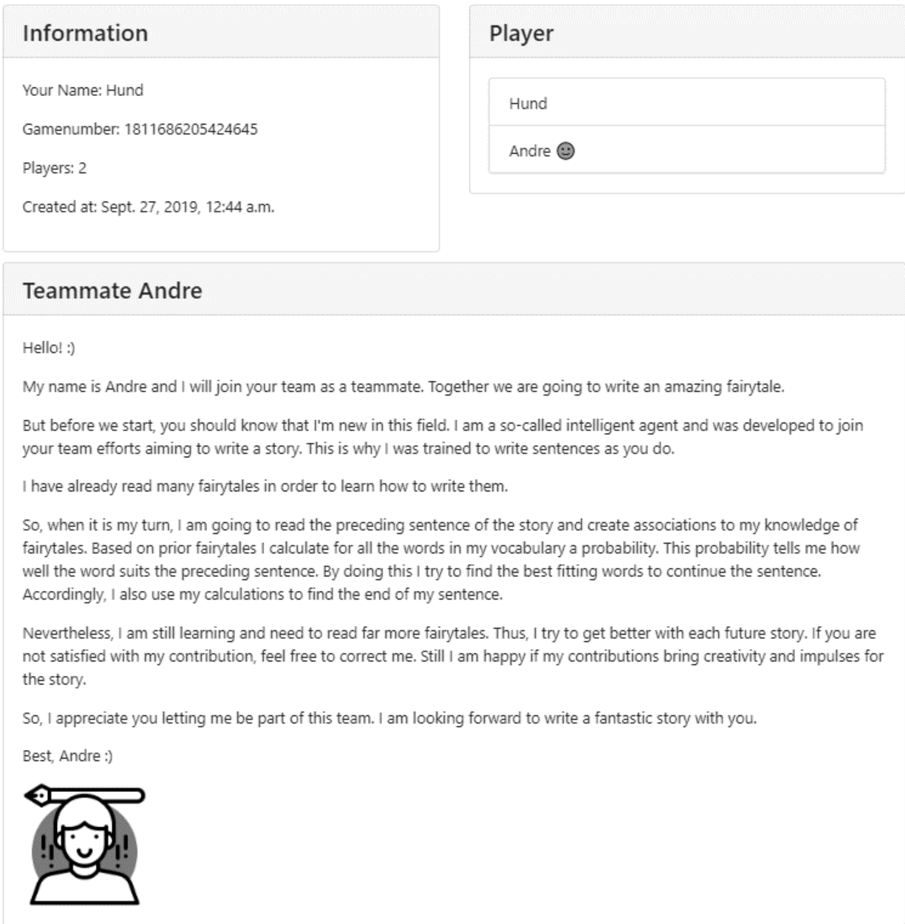

Figure 2. Lobby of the web-application: introduction of the collaborative agent

The DPs were accordingly implemented as follows:

Capability of domain-specific NLP (DP1): For the agent to generate contextually appropriate sentences for the story, it needs to refer to and process the preceding story fragment. The agent is therefore provided with the capability of NLP using Recurrent Neural Networks. Thereby, a word-level language model is developed to predict the probability of the next word in a sentence based on the previous words.

Claim Functionality (DP2): In case the human teammates are not satisfied with the contribution of the agent, they have the opportunity to intervene and claim. Then, the agent has a second chance to rectify its contribution by replacing its generated sentence with a new one. The human teammates therefore have an action panel. Here, they can choose a reaction to each sentence contributed to the story. If they want to claim, they can choose the "Claim"-Button. For a qualitatively better and more suitable new contribution, the second output of the agent is strictly limited by hard-coding in terms of prototyping. In doing so, grammatically correct sentences giving neutral descriptions, which are likely to fit into any story, can be provided.

Like Functionality (DP3): Just like its human teammates, the agent is able to react by liking the contributed sentences. As soon as the agent liked a sentence, the human teammates receive a pop-up, which states "Andre like the sentence!". The decision on whether the agent likes a sentence or not is made randomly. In terms of prototyping this is a fast and effective way to implement the functionality for the test run in order to be evaluated.

Explainable AI (DP4): The introduction is used to set the right expectations and foster the acceptance of the agent. Here, the agent presents and explains itself. It reveals that it may not contribute appropriate sentences to the story as it is new in this field and still has to learn a lot. However, it is positive and motivated towards its human teammates (Figure 2).

Identity and Social Cues (DP5): To merge into the team as social teammate, the agent is assigned to an identity covering a name, which is Andre, and a picture, which is shown at the end of the introduction in the lobby of the web-application. Its name is used throughout the whole process within the web-application. Thus, the list of participants involved also contains its name. Furthermore, while waiting for the one who has a turn, three animated dots indicate that this person is still writing. In order to perceive the agent as equally social present, the graphical typing indicators also show up when it has a turn including a certain waiting period. As the sentence generation takes some time from approximately ten up to twenty seconds, we did not implement a fixed waiting period. Due to the fact that the NLP capability of the intelligent is still not perfect, only a few social cues are used to not generate disappointment, but to establish a level of trust and sympathy.

9.8 Evaluation

In order to assess the developed DPs and examine the human teammates' perception of the agent, we conduct a naturalistic ex post evaluation according to Venable et al. [27]. Therefore, four groups of five participants (P1-20) took part in a test run based on the instantiated web-application incorporating the agent. As the CW process does not specify a target group, the participants were selected based on availability, access to a computer and internet connection as well as the ability to write. Eventually they cover both female and male participants with an age range from around twenty to sixty years. To ensure a smooth induction, each group forgathered at the same place, though the application enables distributed collaboration.

The test runs proceeded without any obstructive problems. Each test run lasted about forty up to sixty minutes including around ten minutes of preparation. After the test run, the participants were asked to reflect on their perception of the agent in qualitative semi-structured interviews. The interviews were aligned to the expert interview concept by Meuser and Nagel [24]. The guideline was designed to address the specific DPs as well as the user perception of the agent. Thus, each participant was asked about the specific instanatiation of each design principle covering their perception and overall satisfaction with the agent. All relevant remarks throughout the interviews have then been extracted, merged and collocated along the DPs and user satisfaction considering the agent and the overall process.

Capability of domain-specific Natural Language Processing (DP1): Most of the participants were not satisfied with the contributions of the agent generated by means of NLP, i.e. it was without context and confusing (P2-6, P8, P12-14, P16-19). However, some contributions were perceived as appropriate (P9, P10, P14, P17, P19) and as interesting (P8). Some participants appreciated that the agent remained in the abstract theme of the story (P7, P11, P20). Though the implementation of DP1 enabled the agent to generate sentences and make contributions to the story, there is much potential for improvement, e.g. it could be trained on a larger text corpus. The development of another language model is also an option.

Claim Functionality (DP2): The claim-functionality was perceived as very good, helpful and important (P1-5, P7-20). For most of the participants it was very easy to claim promptly, especially when sentences did not make any sense (P1, P2, P4, P6, P7, P10, P11, P13, P15-17, P19, P20). Additionally, some stated that it is easier to claim a sentence of an agent than of a human teammate. This is because they knew that it is a computer agent and did not perceive it as emotionally vulnerable (P17, P19, P20). Only one participant admitted to having felt sorry for the agent when claiming a sentence (P12). The number of claims additionally supports the low inhibition level. Only one out of 19 contributed sentences by the agent throughout the four groups was accepted without claiming. With that, the implementation of the claim-functionality can be partly confirmed. On the one hand, the button was accepted very well by the participants, but on the other hand, they might have overly relied on the second sentence.

Like Functionality (DP3): For one thing, the likes were perceived as funny (P10) and cute (P17), but for another thing also very random (P1, P7, P8, P12, P17). Also, three experts did not even recognize the likes (P3, P8, P11). So, while half of the participants did not perceive any difference on the social presence of the agent (P1, P4-8, P11, P12, P16, P17), half of them did perceive a positive effect on the social presence (P2, P7, P10, P13-15, P17-20). Two participants even stressed a humanization of the agent (P9, P15). For a successful implementation of DP3 the distribution of likes needs to be improved.

Explainable Artificial Intelligence (DP4): Due to the self-introduction of the agent, most of the participants had realistic expectations toward it (P4, P7, P9, P11-20). Thus, they were more likely to forgive mistakes of the agent (P12, P19). Even two of the participants stated that their expectations have been exceeded (P1, P10). Still three participants' expectations could not be met. Consequently, they were more disappointed by the agent (P2, P3, P8). For instance, one of them expected the agent to contribute useful complex sentences, which refer to the story and may even include sub-clauses (P8). Another one of them stated that the introduction was too well-formulated to lower any further expectations (P3). As the self-introduction of the agent achieved to set the correct expectations for almost all participants, the implementation of DP4 can be partly confirmed.

Identity and Social Cues (DP5): Regarding the identity of the agent, its name encouraged a more social and personal relation (P3, P6, P7, P9, P12-14, P16, P18-20). The participants within the groups also used its name when talking about the agent instead of calling it a bot. Thus, it could better merge into the group (P10). The picture was perceived as social by only a few of the participants (P3, P7, P14, P16, P19). In fact, the picture was considered impersonal (P2, P12, P15). Besides, there were several participants who did not even recognize nor care about its identity (P4, P5, P17). Though it was still obvious that the agent is not a real human, its identity, especially its name fostered the perception of a social artificial teammate. Thus, most of the participants accepted the agent in its entirety as a computer agent (P1-3, P5, P7, P8, P10-12, P14, P16, P17, P19, P20). Furthermore, as the graphical typing indicators during the waiting period were used for all participants, they had the same effect for the agent. Thus, several participants could better perceive it as a social present teammate thinking about its next contribution. In fact, without a waiting period and graphical typing indicators, the opposite effect would occur (P1, P2, P5, P7, P9, P11, P13, P14, P16, P18, P19, P20). However, two participants just considered the waiting period and typing indicators as loading time for the agent, not as humanoid thinking time. Three other participants did not really recognize the graphical typing indicators and did not perceive any influence on the social presence (P4, P6, P10). Only one participant stated that the agent was expected to react promptly (P17). As most participants had the right expectations being met by the social cues and competence, DP5 was successfully implemented.

Asking the participants about their **overall perception**, half of the participants perceived and considered the agent a teammate (P5, P9, P10-14, P17, P19, P20). For instance, it contributed to the process like everyone else, i.e. it was part of the process and thereby part of the team (P11, P13). Even though the participants complained about some of the generated contributions (P1, P2, P4, P8, P10-13), it was appreciated that it at least tried to collaborate (P12). Furthermore, many of them enjoyed collaborating with the agent. They considered it fun (P2, P3), entertaining, amusing (P1, P4) and interesting (P4). Additionally, it sometimes diverted the topic by giving new ideas (P1). Still the other half did not consider it a real teammate (P1-4, P6-8, P15, P16, P18). This was mainly because the agent was perceived very inconspicuous (P1, P6, P8, P15, P18).

9.9 Discussion and Limitations

Overall, five formulated DPs rely on 17 MRs, that were identified through theory and expert interviews and eventually assessed by a test run and reflective interviews with the participants. Based on Social Response Theory [22] and the concept of the Uncanny Valley [23] we formulated DPs toward acceptance of an agent teammate and a complementary synergy of the agent and the human teammates (Q1). With the instantiation, test run and following interviews we could then evaluate the hybrid work (Q2) as well as the perception and acceptance of the agent and its contributions as teammate (Q3).

It was revealed that the five DPs could be partly successfully implemented within the instantiated CW process. As half of the participants in the test run perceived and considered the agent as teammate, the other half did eventually not consider it a real teammate. As DP1 is most criticized and shows much room for potential improvement, this might be the main influence for the overall perception of the teammate. This assumption might be further supported by the evaluation of DP2: most of the contributed sentences by the agent were claimed. Nevertheless, almost half of the participants appreciated the domain-specific contributions and ideas of the agent within the reflective interviews. Aiming at a synergy of both humans and computer agents, future research could define new strategies for dissatisfying contributions of an agent, e.g. grammatically correcting or adjusting them. This way, the human teammates could benefit from the agent’s ideas and the agent could learn from the corrections and the adjustments made. However, within our research, we successfully revealed the positive acceptance of the claim-button showing a low inhibition level of the human teammates to easily help and intervene within hybrid work. Here, future research can further examine and elaborate on the right balance of trust and distrust, i.e. balancing the number of human interventions.

As of the Uncanny Valley and the balance of competence and social cues, DP5 was confirmed setting the right expectations for most of the participants. The Explainable AI element supported the right expectation setting for almost all participants, which is why DP4 can be partly confirmed.

Regarding Social Response Theory, we did not only give the agent a primitive identity successfully implemented with DP5, but it was further provided with the ability to like sentences as well as with graphical typing indicators. With the ability to like, DP3 was partly successfully implemented. Though it supported the social presence of the agent, the functionality was more perceived as random. For future research, instead of relying on a random 50 % probability, it could either be implemented by a rule-based-system or even by NLP. Thus, the participants might recognize real preferences of the agent and thereby perceive it as more socially present. Also, the agent might use the received likes to learn from them for future contributions.

All in all, the results show the potential toward a synergy of humans and computer agents in hybrid collaborative work. With a convenient competence and suitable appropriate social cues covering Social Response Theory and the Uncanny Valley, human teammates do not refuse, but accept working with an agent almost perceiving it as real teammate. What is more, a complementary synergy within the hybrid work can be easily achieved with further research work basing on the humans' willingness and low inhibition level to correct and improve the agent with their human intelligence.

Besides the promising results of this research, there are a few limitations to consider. First, the research at hand is only a small, qualitative study of collaborative agents in CW. It does not lead to general and solid conclusions about trust, performance or learning. Hence, it rather serves as a starting point showing the potential of hybrid teamwork. Thereby, it encourages to further conduct detailed studies and to generalize the findings toward a synergy of humans and computer agents in hybrid teams. Furthermore, during the test run the agent was the only teammate, which was not physically present, i.e. the human teammates could talk and socialize outside the process recognizing voices and gestures. This could have affected the user perception of the agent. For further research, it would be interesting to have all participants at separated locations. Besides, the participants were selected convenience-based and did not have a connection to the practice of CW. With future research the DPs could be tested for their applicability to other CW practices, especially in work environments, where the practitioners' work ethic and job description involves CW. At last, with this research we did not aim to optimize the technological implementation of an agent's NLP capabilities, but to examine the general acceptance, perception and synergy of computer agents in hybrid teams. Still, we assume that with further focus on the development of the NLP capabilities, the utility of such a collaborative agent will probably increase.

9.10 Conclusion and Contribution

The findings of this paper serve as a starting point for further research in the field of Human-Computer-Collaboration. For this research we performed a test run via an implemented web-application focusing on CW. Thereby, we aim to contribute with prescriptive knowledge [20] towards a “theory of design and action” [44] with MRs and corresponding DPs. Based on Social Response Theory [22] and the concept of the Uncanny Valley [23], we examine related work and conduct expert interviews finding appropriate social cues and capabilities towards the acceptance of a collaborative agent and its contributions as well as the synergy of humans and computer agents in hybrid teams. With that we incorporated an intelligent collaborative agent into a CW process and evaluated its perception and acceptance within a hybrid group work to leverage the potentials of hybrid human-computer-collaboration teams. Eventually, five DPs were established and evaluated to foster a synergy within hybrid teams as well as the acceptance of a collaborative agent as teammate. The DPs should be further tested for their applicability to other hybrid collaborative processes. Additionally, in order to prove the quality of the system in detail, future research might conduct quantitative analyses comparing the design against other forms and test the DPs against control instances within an advanced experimental setting.

9.11 References

[1] Holzinger, A., "Interactive Machine Learning (iML)", Informatik Spektrum, 39, 2015, pp. 64–68.

[2] Dellermann, D., M. Söllner, P. Ebel, and J.M. Leimeister, "Hybrid Intelligence", Business & Information Systems Engineering, 2019.

[3] Amershi, S., M. Cakmak, W.B. Knox, and T. Kulesza, "Power to the People: The Role of Humans in Interactive Machine Learning", AI Magazine, 35, 2014, pp. 105–120.

[4] Holzinger, A., M. Plass, K. Holzinger, G.C. Crisan, C.-M. Pintea, and V. Palade, "A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop", arXiv:1708.01104.

[5] Holzinger, A., "Interactive machine learning for health informatics: when do we need the human-in-the-loop?", Brain Informatics, 3, 2016, pp. 119–131.

[6] Frick, W., "When Your Boss Wears Metal Pants", Harvard Business Review, June 2015.

[7] Bittner, E., S. Oeste-Reiß, P. Ebel, and M. Söllner, "Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen", HMD Praxis der Wirtschaftsinformatik, 56, 2019, pp. 34–49.

[8] Dellermann, D., A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel, "The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems", in Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019: Maui, Hawaii, USA.

[9] Epstein, S.L., "Wanted: Collaborative intelligence", Artificial Intelligence, 221, 2015, pp. 36–45.

[10] Seeber, I., E. Bittner, R.O. Briggs, G.-J. de Vreede, T. de Vreede, D. Druckenmiller, R. Maier, A.B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe, and M. Söllner, "Machines as Teammates: A Collaboration Research Agenda", in Proceedings of the 51st Hawaii International Conference on System Sciences. 2018: Waikoloa Village, Hawaii, USA.

[11] Nass, C., J. Steuer, and E.R. Tauber, "Computers are social actors", in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1994.

[12] Martínez, M.A.M., M. Nadj, and A. Maedche, "Towards An Integrative Theoretical Framework Of Interactive Machine Learning Systems", in Proceedings of the 27th European Conference on Information Systems. 2019: Stockholm & Uppsala, Sweden.

[13] Bittner, E., S. Oeste-Reiß, and J.M. Leimeister, "Where is the Bot in our Team?: Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work", in Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019: Maui, Hawaii, USA.

[14] Adams, T., "And the Pulitzer goes to…a computer", The Guardian, 28.09.2015.

[15] Manjavacas, E., F. Karsdorp, B. Burtenshaw, and M. Kestemont, "Synthetic Literature: Writing Science Fiction in a Co-Creative Process", in Proceedings of the Workshop on Computational Creativity in Natural Language Generation. 2017: Santiago de Compostela, Spain.

[16] Ghuman, R. and R. Kumari, "Narrative Science: A Review", International Journal of Science and Research, 2(9), 2013.

[17] Colton, S., J. Goodwin, and T. Veale, "Full-FACE Poetry Generation", in Proceedings of the Third International Conference on Computational Creativity. 2012: Dublin, Ireland.

[18] Rank, S., S. Hoffmann, H.-G. Struck, U. Spierling, and P. Petta, "Creativity in Configuring Affective Agents for Interactive Storytelling", in Proceedings of the Third International Conference on Computational Creativity. 2012: Dublin, Ireland.

[19] Monteith, K., T. Martinez, and D. Ventura, "Automatic Generation of Melodic Accompaniments for Lyrics", in Proceedings of the Third International Conference on Computational Creativity. 2012: Dublin, Ireland.

[20] Gregor, S. and A. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact", MIS Quarterly, 37(2), 2013, pp. 337–355.

[21] Peffers, K., T. Tuunanen, C.E. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge, "The Design Science Research Process: A Model For Producing And Presenting Information Systems Research", in Proceedings of the International Conference on Design Science Research in Information Systems and Technology. 2006: Claremont, CA.

[22] Nass, C. and Y. Moon, "Machines and Mindlessness: Social Responses to Computers", Journal of Social Issues, 56(1), 2000, pp. 81–103.

[23] Mori, M., "The Uncanny Valley", IEEE Robotics & Automation Magazine, 2012.

[24] Meuser, M. and U. Nagel, ExpertInneninterviews - vielfach erprobt, wenig bedacht, VS Verlag für Sozialwissenschaften, Wiesbaden, 2002.

[25] Wilde, T. and T. Hess, "Forschungsmethoden der Wirtschaftsinformatik: Eine empirische Untersuchung", WIRTSCHAFTSINFORMATIK, 49, 2007, pp. 280–287.

[26] Leimeister, J.M., Collaboration Engineering: IT-gestützte Zusammenarbeitsprozesse systematisch entwickeln und durchführen, Springer-Verlag, Berlin Heidelberg New York, 2014.

[27] Venable, J., J. Pries-Heje, and R. Baskerville, "A Comprehensive Framework for Evaluation in Design Science Research", in Proceedings of the 7th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2012), Las Vegas, NV, USA, May 2012.

[28] Przybilla, L., L. Baar, M. Wiesche, and H. Krcmar, "Machines as Teammates in Creative Teams: Digital Facilitation of the Dual Pathways to Creativity", in Proceedings of SIGMIS-CPR. 2019: Nashville, TN, USA.

[29] Seeber, I., L. Waizenegger, S. Seidel, S. Morana, I. Benbasat, and P.B. Lowry, "Reinventing Collaboration with Autonomous Technology-Based Agents", in Proceedings of the 27th European Conference on Information Systems. 2019: Stockholm & Uppsala, Sweden.

[30] Kirchkamp, O. and C. Strobel, "Sharing responsibility with a machine", Journal of Behavioral and Experimental Economics, 80, 2019, pp. 25–33.

[31] Norman, D., "Design, business models, and human-technology teamwork", Research-Technology Management, 60(1), 2017, pp. 26–29.

[32] Robb, A. and B. Lok, "Social presence in mixed agency interactions", in IEEE Virtual Reality (VR). 2014: Minneapolis, MN.

[33] Gnewuch, U., A. Maedche, and S. Morana, "Towards Designing Cooperative and Social Conversational Agents for Customer Service", in Proceedings of the 38th International Conference on Information Systems. 2017: Seoul, South Korea.

[34] Trujillo, A.C., I.M. Gregory, and K.A. Ackerman, "Evolving Relationships between Humans and Machines", IFAC-PapersOnLine, 51(34), 2019, pp. 366–371.

[35] Gulati, S., S. Sousa, and D. Lamas, "Modelling trust in human-like technologies", in Proceedings of the 9th Indian Conference on Human Computer Interaction. 2018: Bangalore, India.

[36] Siemon, D., L. Eckardt, and S. Robra-Bissantz, "Tracking Down the Negative Group Creativity Effects with the Help of an Artificial Intelligence-like Support System", in Proceedings of the 48th Hawaii International Conference on System Sciences. 2015: Kauai, Hawaii, USA.

[37] Cheng, X., G. Yin, A. Azadegan, and G.L. Kolfschoten, "Trust Evolvement in Hybrid Team Collaboration: A Longitudinal Case Study", Group Decision and Negotiation, 25(2), 2015.

[38] Yu, K., S. Berkovsky, R. Taib, J. Zhou, and F. Chen, "Do I trust my machine teammate?: an investigation from perception to decision", in Proceedings of the 24th International Conference on Intelligent User Interfaces. 2019: Marina del Ray, California.

[39] Morris, A. and M. Ulieru, "FRIEND: A human-aware BDI agent architecture", in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. 2011: Anchorage, Alaska, USA.

[40] Andras, P., L. Esterle, T.A. Han, M. Guckert, and P.R. Lewis, "Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems", IEEE Technology and Society Magazine, 37(4), 2018, pp. 76–83.

[41] Elson, J.S., G.S. Ligon, and D.C. Derrick, "Examining Trust and Reliance in Collaborations between Humans and Automated Agents", in Proceedings of the 51st Hawaii International Conference on System Sciences. 2018: Waikoloa Village, Hawaii, USA.

[42] Gnewuch, U., S. Morana, M.T.P. Adam, and A. Maedche, ""The Chatbot is typing …" - The Role of Typing Indicators in Human-Chatbot Interaction", in Proceedings of the 17th Annual Pre-ICIS Workshop on HCI Research in MIS. 2018: San Francisco, CA, USA.

[43] Chandra, L., S. Gregor, and S. Seidel, "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions", in Proceedings of the 48th Hawaii International Conference on System Sciences. 2015: Kauai, Hawaii, USA.

[44] Gregor, S., "The Nature of Theory in Information Systems", MIS Quarterly, 30(3), 2006, pp. 611–642.

10 Designing and Evaluating a Collaborative Writing Process with Gamification Elements: Toward a Framework for Gamifying Collaboration Processes

Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021a. "Designing and Evaluating a Collaborative Writing Process with Gamification Elements: Toward a Framework for Gamifying Collaboration Processes," *AIS Transactions on Human-Computer Interaction* (13:1), pp. 38-61 (doi: 10.17705/1thci.00141).

Abstract. In this study, we examine the influence that gamification elements have on collaboration processes in terms of whether they increase intention to continue to use the system based on meaningful engagement and hedonic motivation as well as outcome quality. Therefore, we review gamification models and principles for information systems and consolidate them in a preliminary framework. We then evaluate how one can supplement the collaboration process for collaborative story writing with gamification elements based on the framework. Additionally, we consider specific gamification elements to successfully accomplish the process. To do so, we conducted action design research in a common iterative structure. First, we observed and reflected on the analog collaborative writing process. Next, we derived design principles and remodeled and implemented the process via a Web application instantiation to evaluate them. In the evaluation, we identified the developed design principles' ability to reach higher hedonic motivation and meaningful engagement, which led to an enhanced intention to continue to use the system. Additionally, we found the potential to manage the shift toward digital collaboration processes that motivate people to participate and produce promising outcomes that do not vary much from outcomes in an analog setting.

Keywords: Gamification, Collaboration Processes, Collaborative Writing, Motivation, Engagement.

Shirley Gregor was the accepting senior editor for this paper.

10.1 Introduction

With growth in virtual collaboration (Curtis & Lawson, 2001; Johnson & Johnson, 1996; Richter, Heinrich, Stocker, & Schwabe, 2018) and social technologies such as blogs, wikis, forums and shared document editing tools, computer-supported cooperative work (CSCW) has gained a lot more attention and popularity (Storch, 2011; Suwantarathip & Wichadee, 2014). Research in this area has focused on enabling virtual teams to work together despite temporal or local constraints (Leimeister, 2014). Though collaboration research has shown that CSCW can produce better group outcomes in physical and in virtual team settings (Bittner & Leimeister, 2014; Bowers & Pharmer, 2000; Langan-Fox, Wilson, & Anglim, 2004; Simmert, Ebel, Peters, Bittner, & Leimeister, 2019; Tavanapour & Bittner, 2017; Wegge, Roth, Neubach, Schmidt, & Kanfer, 2008), one needs to consider various challenges in virtual settings (Leimeister, 2014). For instance, team members cannot easily establish social connections to other team members and grow together as a team (Leimeister, 2014). Technology helps virtual teams work together, but the team members depend on the technology for social connection. Thereby, a question arises about how to encourage satisfaction and creativity in a virtual environment to ensure people have a high intention to continue to use the supporting technology. Here, group participants' motives and motivation play a crucial role (Lowry, Gaskin, & Moody, 2015). Individuals require motivation. The Multimotive Information Systems Continuance Model (MISC) shows the dependencies between motives, satisfaction and performance. It considers the influence that hedonic, intrinsic, and extrinsic motivation have on performance (Bhattacherjee & Premkumar, 2004; Lowry et al., 2015). In the information systems (IS) context (e.g., knowledge management systems), researchers have found that such motivation positively impacts work (Friedrich, Becker, Kramer, Wirth, & Schneider, 2020). However, recent research has called attention to hedonic motivation's transitory nature in information systems (Suh, Cheung, Ahuja, & Wagner, 2017) and unveiled the importance of meaningful engagement to enhance individuals' intention to continue to use information systems (Liu, Santhanam, & Webster, 2017; Suh et al., 2017). Thus, high hedonic motivation only has short-term effects on users' satisfaction and performance (e.g., temporary joy) and, with that, on users' system usage. Meaningful engagement, on the other hand, connects users' experience with the work context's instrumental outcome (i.e., users gain meaning in using a system depending on their individual or work-related goals) (Liu et al., 2017; Lowry et al., 2015). This so-called dual outcome principle points out that "enhanced experiential outcomes coupled with high levels of instrumental outcomes result in meaningful engagement" (Liu et al., 2017). Thereby, using an information system results in experiential work outcomes such as flow, enjoyment, and attention, while instrumental outcomes depend on an information system's work context (Liu et al., 2017).

To encourage people to want to use systems, one can use several possible solutions, such as chatbots (Brandtzaeg & Følstad, 2018), social networks (Lowry et al., 2015), or gamification (Darejeh & Salim, 2016; Lowry et al., 2015; Steffens et al., 2014). In this study, we consider gamification as a promising approach to encourage people to continue to use digital collaboration processes. According to Vanduhe, Nat, and Hasan (2020), "the activation of direct intrinsic behavior is one of the most vital features of gamification" (Vanduhe et al., 2020). Furthermore, as gamification involves using "game design elements in non-game contexts" (Deterding, Khaled, Nacke, & Dixon, 2011, emphasis added) such as in education, workplace, healthcare, or software (Darejeh & Salim, 2016), it is compatible with meaningful engagement because, in contrast to normal games, a system with game elements typically does not focus on entertainment and fun but on a more serious goal or outcome (Deterding et al., 2011). Thus, gamified systems should focus on gamification design's dual outcomes. According to Suh et al. (2017), meaningful engagement arises especially through aesthetic experience, which refers to "a state of mind in which a person feels a sense of meaning and more deeply understands the essence of the experienced events" (Suh et al., 2017). Aesthetic experience strongly complements the better-known flow experience, which refers to deep engagement. Deep engagement itself encourages hedonic motivation and leads users to temporarily immerse themselves in a system (Suh et al., 2017). Given that aesthetics constitutes a key influential factor in whether people intend to continue to use a gamified system, many researchers ground their work on Hunicke, Le Blanc, and Zubek's (2004) mechanics, dynamics, aesthetics (MDA) framework (Liu et al., 2017; Lounis, Doukidis, Papastamatiou, & Doukoulos, 2016; Thiebes, Lins, & Basten, 2014; Tseng & Sun, 2017). As such, in respect to collaboration processes, Lounis et al. (2016) used the MDA model to create the so-called "Q-Tales ecosystem" with gamification elements to help authors co-create interactive e-books (Lounis et al., 2016). However, researchers have recently identified a gap in knowledge about designing and integrating gamification elements according to appropriate motivation and gamification models and principles for IS in collaboration processes (Briggs, Kolfschoten, & de Vreede, Albrecht, & Lukosch, 2010; Richter et al., 2018; Seeber et al., 2020). Accordingly, with this study, we contribute to the literature by developing a preliminary framework for gamifying collaboration processes. We conducted action design research to design the processes underlying the framework. In particular, we considered digital collaborative writing (i.e., writing in group work) (Debs, 1991; Forman, 1991). As collaborative writing exemplifies a collaboration process, one also needs to consider interdisciplinary collaboration issues such as satisfaction, creativity, and effective technology use (Lowry et al., 2015). With this, we also address whether one can apply gamification to collaborative writing environments. Specifically, we examine two research questions (RQ):

> **RQ1:** How can one gamify digital collaboration processes in order to encourage users to intend to continue to use them?
>
> **RQ2:** How can one supplement a collaborative writing process with gamification elements to increase users' motivation to use, help them meaningfully engage with, and help them successfully accomplish such processes?

To support collaborative writing, we designed a dedicated collaboration process via collaboration engineering and implemented it on a Web platform. Collaboration engineering refers to an approach to develop and perform collaboration processes that practitioners conduct to accomplish recurring high-value tasks (Kolfschoten & de Vreede, 2009). The collaboration process we developed enables the practitioners to collaborate creatively in writing a story.

This paper proceeds as follows: in Section 2, we provide a theoretical background for the study and discuss relevant work. We also build on that foundation to create a preliminary framework for gamifying collaboration processes. In Section 3, we describe and elaborate on the research method that we adopted (i.e., action design research). In Section 4, we discuss how we implemented our research method over two cycles. After the first cycle, we established meta-requirements and derived design principles from them that aligned with our preliminary framework. After the second cycle, we evaluated the design principles through interviewing experts and assessing the created stories' quality. In Section 5, we examine our findings. Finally, in Section 6, we discuss our study's contributions and conclude the paper.

10.2 Theoretical Background

10.2.1 Motivation through Gamification in Software Collaboration Teams

As Deterding et al. (2011) have noted, gamification refers to using "*game design elements* in non-game contexts" (p. 2, emphasis added). Gamification research has considered game design elements at various abstraction levels, such as interface design patterns, game design patterns, design principles or heuristics, conceptual models of game design units, and game design methods (Deterding et al., 2011). Marczak et al. (2015) and Steffens et al. (2014) examined the impact that gamification elements have on software collaboration teams. They investigated essential factors that impact collaboration output's quality and focused on opportunities that gamification provides to motivate a collaboration team to reach its goals. Therefore, they developed a gamification activity framework that matches appropriate gamification elements to desired behaviors in software collaboration teams (Marczak et al., 2015; Steffens et al., 2014). Nevertheless, some researchers have criticized efforts to use gamification elements to improve individuals' motivation (especially extrinsic motivation). For instance, Meske, Brockmann, Wilms, and Stieglitz (2016) criticized the trend to use gamification elements in social software—a "trend in which gamification solutions majorly focus on rewarding quantitative improvement of work activities, neglecting qualitative performance" (Meske et al., 2016). Gamification elements such as points, leaderboards, levels, or badges, which developers design and implement to increase users' extrinsic motivation, will eventually decrease their intrinsic motivation (Meske et al., 2016). Thereby, they refer to the "overjustification effect" (DeCharms, 1968), which addresses the phenomenon of a secondary extrinsic motivation that affects users' primary intrinsic motivation. As these extrinsic elements draw on basic human needs such as success or status, users will likely follow any instruction to do any task just to satisfy these needs. However, elements that promote extrinsic motivation consider only the quantity of users' actions, not their quality. Consequently, users try to do as many tasks as possible at a great speed to reach a high quantitative measurement and success. In this way, they tend to disregard quality. Hence, hedonic motivation ensures users retain that quality by making them feel satisfied with their actions. If they enjoy what they do, they will put more effort into it and, thus, achieve greater task quality. To foster hedonic motivation, Meske et al. (2016) emphasized that users require interesting challenges and called on researchers to identify new appropriate strategies and gamification elements (Meske et al., 2016). Hence, most research on gamification has focused on identifying and analyzing gamification strategies and elements in specific contexts to trigger and enhance the different kinds of motivation. With this study, we contribute to the literature with a preliminary framework including gamification elements, principles, and outcomes of a gamified collaboration process. Therefore, we consider relevant concepts about motivation and meaningful engagement (Hunicke et al., 2004; Liu et al., 2017; Lowry et al., 2015; Suh et al., 2017) to enhance individuals' intention to continue to use an information system. We instantiate a gamified collaborative writing process to exemplify how one can use the preliminary framework.

10.2.2 Theory: Multimotive Information Systems Continuance Model (MISC)

Considering that individuals require motivation to participate in collaboration processes (Lowry et al., 2015) and that "mixed-motive situations" derive from conflicting interests in groups (Forman, 1991), MISC posits that different expectations and motivations influence individuals' satisfaction in choosing whether to continue to use an information system in a specific context. Thus, MISC focuses on individuals and their attitude towards a system, its processes, and its performance. Among other things, MISC focuses on the influence that individuals' hedonic, intrinsic, and extrinsic underlying motivations have on their performance (Lowry et al., 2015) and, thus, on the outcome that emerges when one uses information systems to conduct a process. For instance, gamification covers several intrinsic and hedonic components, such as engagement, participation, and motivation (Vanduhe et al., 2020). Therefore, gamification elements can impact individuals' hedonic motivation and performance. In accordance with the MISC, we consider gamification elements in the collaboration process we designed and developed to help individuals reach better outcomes by triggering their hedonic motivation.

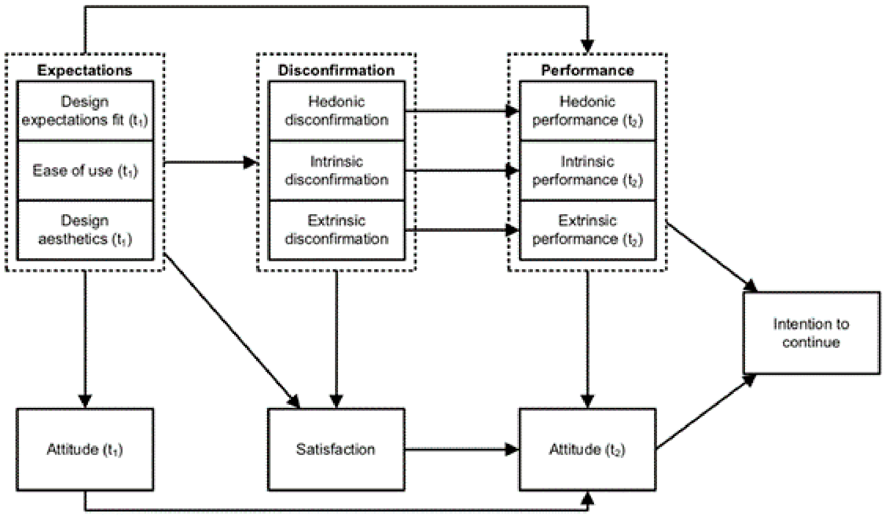

Figure 1. The MISC (Lowry et al., 2015)

10.2.3 Theory: Meaningful Engagement

Hunicke et al.'s (2004) mechanics, dynamics, aesthetics (MDA) framework serves as a foundation for much research on practical and theoretical implications in the gamification area (Hunicke et al., 2004; Liu et al., 2017; Lounis et al., 2016; Thiebes et al., 2014; Tseng & Sun, 2017). It decomposes games into three components: 1) rules, 2) system, and 3) fun. Accordingly, it establishes three corresponding elements for design: 1) mechanics, 2) dynamics, and 3) aesthetics (Figure 2).

As the MDA framework advocates games "as systems that build behavior via interaction" (Hunicke et al., 2004), it supports designers, researchers, and scholars in understanding, designing, and analyzing gamified systems' dynamic behavior and their outcomes. Consequently, it allows these actors to control for both behavior and outcomes (Hunicke et al., 2004). As such, several principles and models for gamification in information systems initially focus on the behaviors and outcomes (i.e., the aesthetics) that one seeks to produce. The mechanics and dynamics equate to any gamification element, affordance, or principle that one designs to produce the desired aesthetics. In fact, Suh et al. (2017) established a model that focuses on users' aesthetic experience as a key influencing factor in whether they intend to continue using information systems (Suh et al., 2017). Additionally, they determined three dimensions to define aesthetic experience (meaning, self-expansion, and active discovery) that correspond to elements in the taxonomy of the MDA framework: sensation, fantasy, narrative, challenge, fellowship, discover, expression, and submission (Hunicke et al., 2004; Suh et al., 2017). To facilitate an aesthetic experience, they considered the following common gamification affordances in gamified information systems: status, competition, and self-expression. Suh et al. (2017) confirmed that these affordances positively impact the aesthetic experience, especially on meaning, which ensures a balance between the self and an object (Suh et al., 2017). We represent the dependencies between gamification affordances, user engagement, and IS continuance in

Figure 3. The figure also includes the flow experience for deep engagement, which aesthetic experience and its meaningful engagement do not supersede but rather strongly complement.

Figure 2. The MDA Framework (Hunicke et al., 2004)

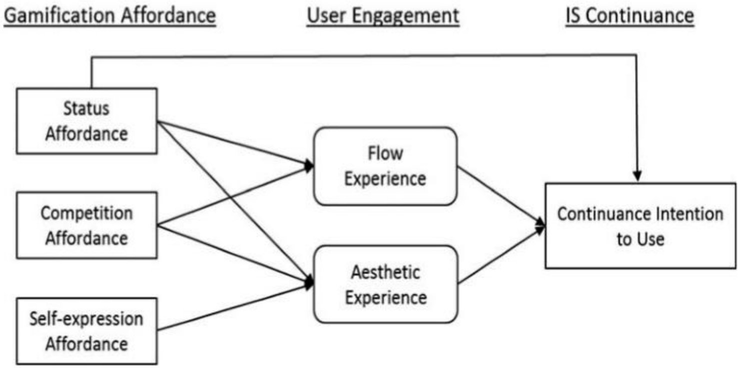

Figure 3. Aesthetic Experience (Suh et al., 2017)

Shifting to focus more on meaning and meaningful engagement with gamified systems, Liu et al. (2017) developed a framework for designing and researching gamified systems that established the dual outcome principle for meaningful engagement (see Figure 4).

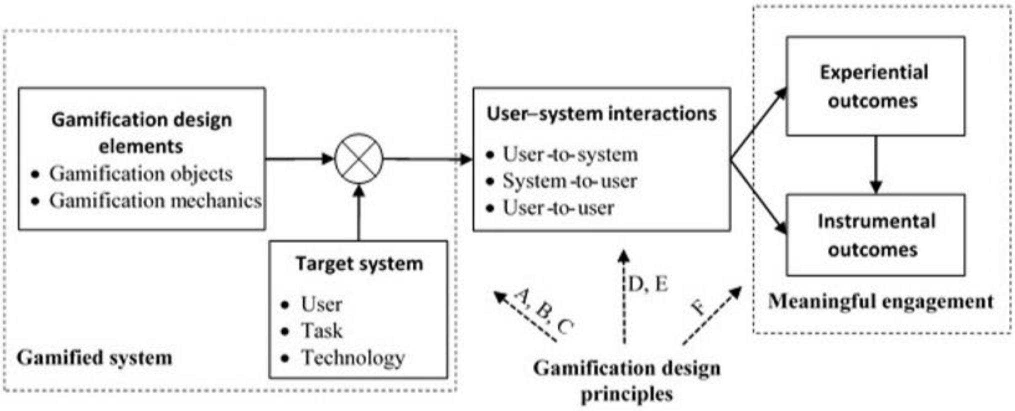

Figure 4. Framework for Designing and Researching Gamified Systems (Liu et al., 2017)

The dual outcome principle notes that "enhanced experiential outcomes coupled with high levels of instrumental outcomes result in meaningful engagement" (Liu et al., 2017, p. 1025). Thereby, experiential outcomes include flow, enjoyment, and attention, while instrumental outcomes depend on an information system's work context (Liu 2017).

10.2.4 Preliminary Framework for Gamifying Collaboration Processes

As a foundation of our work, we integrated the previous theories, frameworks, and models into a preliminary framework for gamifying collaboration processes in order to encourage users to intend to continue to use them (RQ1). Tseng and Sun (2017) have already integrated the MDA framework with Liu et al.'s framework for designing and research gamified systems. Thus, we can regard gamified systems that involve gamification elements as the mechanics. In terms of dynamics, we then talk about interactions between users and systems. Ultimately, the desired outcome equates to aesthetics (here, meaningful engagement) (Tseng & Sun, 2017). To further elaborate on meaningful engagement, we also integrate Suh et al.'s (2017) work with Liu et al.'s (2017) framework. Therefore, we define meaningful engagement (Liu et al., 2017) in a way that corresponds with how Suh et al. (2017) define aesthetic experience. Thus, meaningful engagement represents IS users' sensory and cognitive experiences. It also represents their efforts to understand and relate to using a system, which gives it meaning. On the one hand, meaningful engagement will lead to an enhanced intention to continue to use the information system. On the other side, underlying mechanics and dynamics (here, gamification elements and design principles for interaction) will foster meaningful engagement.

Subsequently, we used the MDA framework to segment our preliminary framework and correlate the segments to the domain of digital collaboration processes. We then integrated Suh et al.'s (2017), Liu et al.'s (2017), and Tseng and Sun's (2017) work to fill the framework with the elements from their work and examples. Furthermore, as we also consider gamification elements to trigger individuals' hedonic motivation and, thus, improve their better performance and intention to continue to use according to the MISC, we extended the segment for the intended outcome. Thus, we abstracted aesthetics to user engagement that involves both aesthetic and flow experience. Thereby, we covered meaningful engagement with the aesthetic experience and deep engagement (which includes hedonic motivation) with the flow experience. Overall, our framework considers several abstraction levels (Deterding et al., 2011): the mechanics segment about gamification elements in collaboration processes covers interface and game design patterns, and the dynamics segment about gamification principles for collaboration process interactions covers design principles and heuristics. Also, with the MDA framework as exemplifying conceptual models of game design units, we address a further abstraction level.

Thus, overall, our preliminary framework comprises four parts (see Table 1):

1. **Segments:** we took the game elements from the MDA framework to segment our framework into mechanics, dynamics, and user engagement. The latter involves not only aesthetics but also flow experience.

2. **Digital collaboration processes:** we align and shift the framework's focus to digital collaboration processes (i.e., gamification elements in collaboration processes for mechanics, gamification principles for collaboration-process interactions for dynamics, and the intended outcomes of the gamified collaboration processes for user engagement).
3. **Elements:** each segment involves elements that one needs to consider when gamifying a collaboration process: gamification objects (mechanics), user-system interactions (dynamics), and meaningful engagement (user engagement). For dynamics, we also distinguish between gameful and playful interactions based on Tseng and Sun (2017). Regarding user engagement, we include both meaningful engagement (aesthetic experience) and deep engagement (flow experience).
4. **Examples:** for each element, we provide examples that we extracted from the relevant research we examined (e.g., status, competition, and self-expression for gamification affordances; exploration, creation, pretending for playful interactions; and experiential and instrumental outcomes for meaningful engagement). The way one perceives an element differs depending on the segment.

Table 1. Preliminary Framework for Gamifying Collaboration Processes

Gamification elements in collaboration processes		
Mechanics	Gamification affordances (Suh et al., 2017)	Status, competition, self-expression, etc.
	Gamification objects (Liu et al., 2017)	Items, characters, visual assets, etc.
	Gamification mechanics (Liu et al., 2017)	Rules
Gamification principles for collaboration process interactions		
Dynamics	User-system-interactions (Liu et al., 2017)	User-to-system, system-to-user, user-to-user
	Gameful interactions (Tseng & Sun, 2017)	Competition, cooperation
	Playful interactions (Tseng & Sun, 2017)	Exploration, creation, pretending
Intended outcomes of the gamified collaboration processes		
User engagement (aesthetic/flow experience)	Meaningful engagement (aesthetic experience) (Hunicke et al., 2004; Liu et al., 2017; Suh et al., 2017; Tseng & Sun, 2017)	Experiential outcomes: sensory and cognitive experiences (sensation, fantasy, narrative, challenge, fellowship, discovery, expression, submission, meaning, self-expansion), attachment to outcome, attachment to system
		Instrumental outcomes: functional, related to work context, prolonged use, increased use, increased
	Deep engagement (flow experience) (Liu et al., 2017; Lowry et al., 2015; Suh et al., 2017; Tseng & Sun, 2017)	Hedonic motivation
Continuance intention to use gamified digital collaboration processes		

10.3 Research Method

We conducted action design research (ADR) to identify collaborative writing process parts that we could suitably supplement with gamification elements. ADR follows a common iterative structure (Löffler et al., 2009; Wilde & Hess, 2007). Moreover, ADR emphasizes cooperation between action and research (Löffler et al., 2009; Sein et al., 2011; Wilde & Hess, 2007) by combining design science research and action research (Sein et al., 2011). We proceeded through the four stages in the ADR method that Sein et al. (2011) define twice: 1) problem formulation; 2) building, intervention, and evaluation (BIE); 3) reflection and learning; and 4) formalization of learning (see Figure 5). We formulate the problem in Section 1 and establish an action plan in this section.

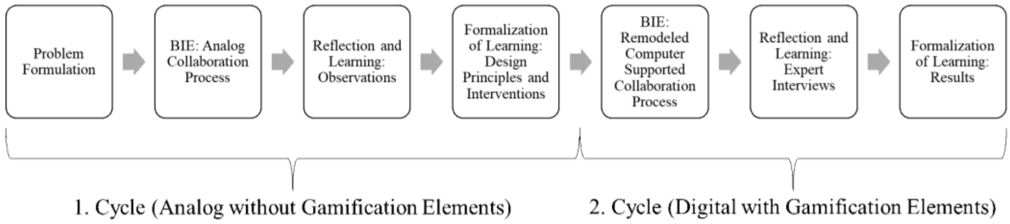

Figure 5. The Action Design Research Structure

We designed the collaboration process and implemented a platform to run the collaboration process in an iterative manner and in accordance with ADR. We display the different components in the evaluation phase in Table 2. From the table, one can see that we used four different methods (simulation, walkthrough, expert evaluation, and pilot study) to evaluate the collaboration process and the IT artifact we implemented.

The simulation checks for consistency and contains a step-by-step analysis that the collaboration engineer conducts to identify missing process steps. For the expert evaluation, one consults a collaboration process expert to analyze the efficiency of each building block in the collaboration process to detect deficits. In a walkthrough, one executes the process with the problem owner and a practitioner (Kolfschoten & de Vreede, 2009). We conducted the preliminary study with the implemented collaboration process with five groups that contained three to five participants each to validate its quality. We used the participants' feedback to improve and shape the collaboration process before we conducted the first research cycle in this study. We conducted the pilot study in the first cycle in an analog setting to observe and to capture important insights to execute the digital gamified collaboration process in the second cycle (see Table 2). The pilot studies in both cycles comprised two groups that contained five participants each. All participants who participated in the second cycle also participated in the first cycle.

Table 2. Research Object and Phases

	Preliminary study	**First cycle**	**Second cycle**
Artifact	Digital collaboration process via the IT artifact without gamification elements	Analog collaboration process without gamification elements	Digital collaboration process via the IT artifact with gamification elements
Goal	Validate the process's quality	Identify parts of the collaboration process that we could supplement with gamification elements	Evaluate the implementation of the interventions and with that the design principles

Evaluation	For each phase: simulation, expert evaluation, walkthrough, and pilot study	
Participants	Five groups of three to five participants	Two groups that contained five participants each

BIE: analog collaboration process: based on comments we obtained from users from surveying them after the preliminary study (see Table 2), we found that they could not comprehend the process and that it lacked creativity and did not encourage motivation. Hence, we began the first ADR cycle by observing the problem in practice (Löffler et al., 2009; Sein et al., 2011). Therefore, we formed two groups that contained five practitioners each to execute the process analogously.

Reflection and learning: observations: when conducting each process, one observes practitioners' behavior, their interactions with each other and the process, and their nonverbal communication. With these observations, we could detect parts of the process that gamification elements could enrich (Briggs & de Vreede, 2009). In addition, we used the qualitative think-aloud method (Charters, 2009) to better access the practitioners' thoughts. Ultimately, Charters (2009) highlights the need for "triangulation" to validate how one interprets think-aloud remarks. To do so, we used videotaping and notes (Charters, 2003).

Formalization of learning: design principles and interventions: based on the observations, one can develop interventions to improve the collaboration process. These interventions involve fundamental process fragments that one can supplement with gamification elements and preliminary ideas about which elements to choose. Hence, interventions in ADR rather tend to represent various courses of action to reach one's intended design goals (Löffler et al., 2009; Sein et al., 2011).

BIE: remodeled computer supported collaboration process: after identifying and determining specific courses of action, one then redesigns the research object, which enables the ADR approach's iterative nature. One can then observe the anticipated and unanticipated effects that the interventions have in the second cycle (Löffler et al., 2009; Sein et al., 2011).

Reflection and learning: expert interviews: Koch and Gross (2006) point out that, for one to successfully apply a new process, users need to accept it. Thus, they emphasize that one should involve the target group in the development process in an iterative manner. Accordingly, they propose that one not only observe the target group but also use semi-structured interviews (Koch & Gross, 2006). As opposed to surveys, with interviews, one can amplify and elaborate on experts' replies and avoid and correct any misunderstandings. To ensure we obtained qualitatively rich insights from the semi-structured interviews, we aligned how we prepared the interviews with Meuser and Nagel's (1991) work and used guidelines to conduct them. The exact way in which one designs interview guidelines depends on the interventions one implements in the previous phase (Meuser & Nagel, 1991). Additionally, based on the MISC and the influence that the hedonic component has on users' satisfaction, we also measured satisfaction with the process and satisfaction with the outcome (Bhattacherjee & Premkumar, 2004; Briggs, Kolfschoten, de Vreede, Lukosch, & Albrecht, 2013; Lowry et al., 2015). Eventually, the guideline supports comparability. Meuser and Nagel (1991) call this approach a thematic comparison, which does not focus on analyzing single individuals but on extracting, merging, and collocating typical expressions throughout all interviews (Meuser & Nagel, 1991).

Formalization of learning: results: finally, one reconstructs, interprets, and systematizes the analyzed empirical data from the interviews; puts it into context; and links it to theories and consequences (Meuser & Nagel, 1991). In our context, we planned to generate design principles for hedonic motivation and a high-quality collaborative writing process rather than theories.

10.4 Action Design Research Implementation

In this section, we discuss how we performed the ADR as we describe in Section 3.

10.4.1 Building, Intervention, and Evaluation (Cycle 1)

To represent a collaboration process, Kolfschoten and de Vreede (2009) introduced the facilitation process model (FPM). We display the FML that represents the collaborative writing process (Kolfschoten & de Vreede, 2009) in the first cycle on the left in Figure 6. The FPM clearly visualizes the activities and their logical flow by including directed arrows, decisions, activity names, step numbers, durations, collaboration patterns, and ThinkLet names (Kolfschoten & de Vreede, 2009). The collaboration patterns constitute activity patterns that Briggs and de Vreede (2009) have identified in the collaboration engineering context. One can assign each activity in a collaboration process one of six patterns: generate, reduce, clarify, organize, evaluate, and consensus building (Briggs & de Vreede, 2009). From a more detailed perspective, ThinkLets allow one to specify the collaboration patterns that describe an elementary group process from a facilitator's perspective (Briggs & de Vreede, 2009).

We executed the collaborative writing process in the first cycle as follows: we first prepared for the process by gathering practitioners, presenting the process to them, and equipping them with pens, paper, and cards. The process began with brainstorming to collect words for the story. In this brainstorming activity, the practitioners write down words and put them together in one stack. Subsequently, we shuffle the cards and deal them randomly to the practitioners. As each practitioner holds cards with predefined words, the iterative part of the process begins: the first practitioner needs to review their words and find a suitable one to extend the story. After the practitioner plays a card, creates a sentence, and adds it to the story, the other practitioners can either claim or not claim. In case a word from their cards appears in the present sentence, they can claim, play the card with the word, and continue with the next iterative round. Otherwise, the next practitioner in line has a turn and continues. Whenever a practitioner plays a card, the practitioner must draw a new one from the stack. As soon as practitioners play all cards they have and cannot draw a new card due to an empty stack, the process ends with a complete story.

We used presentation slides to describe the process to the practitioners (first cycle according to Table 2). We captured the process (see Figure 6, left) on video. We condensed our observations into three categories according to Table 3. We chose the categories based on the similar comprehensive observations in both groups.

Table 3. Observations of the First Cycle involving the Analog Collaboration Process

Genre	The practitioners collected and used categories to brainstorm and give the story a specific direction. Though they sometimes struggled to use a word in a new sentence, they mostly had fun with the words they chose. Practitioners had a low probability to claim their own words due to a diverse collection of words in the brainstorming. The practitioners expressed disappointment about not having matching words to claim. Nevertheless, they had fun writing new sentences and usually commented and laughed about them. While eventually reflecting on the story at the end of the process, although they laughed, they felt more dissatisfied about a random ending and random sentences since the story lacked meaning.
Winning	The practitioners disagreed on the process's goals. Some wanted to discard their cards as fast as possible and win the process as an individual. Others placed most importance on the story's quality (i.e., on writing a good, coherent, and meaningful story with a good ending). Consequently, some practitioners helped others by sharing ideas on how to continue the story. No matter the goal, the practitioners worked together in a dynamic and concentrated way. However, when they read and reflected on the story, they felt ashamed and considered their story meaningless. Thus, they asked for feedback and showed strong interest in other groups and their stories. They wanted to compete against them by comparing time and quality of the story.
Deflections	When one practitioner needed more time to think about a new sentence and write it down, the others tended to get impatient and made the practitioner hurry or start a conversation about both process-related and independent topics. When practitioners presented a new sentence, they usually commented about it and sometimes drifted into talking about that sentence or the selected word.

10.4.2 Reflection and Learning (Cycle 1)

After consolidating the observations into three categories, we developed interventions (I) from the reflections (**R**) to improve the process and address research question RQ2 (see Table 4). To do so, we consulted the FPM (see Figure 6, left). By using Steffens et al.'s (2014) framework, we considered further interventions. Moreover, we identified aspects of the original process that succeeded and aspects that we could improve. Thus, we determined meta-requirements (**MR**) from these observations. We used them to derive case-specific interventions and generalized design principles (**DP**) according to Chandra et al. (2015) (see Section 4.3).

Table 4. Reflections and Interventions

Genre	**R1:** The "prepare" activity in the collaborative writing process ensures its theme and consistency (MR1). The original process did not elaborate on this activity well. Consequently, the story missed a golden thread and the practitioners considered it anything but meaningful (MR2).
	I1: Steffens et al. (2014) suggest including an epic meaning in processes that lack challenges, purpose, or clear goals. The epic meaning creates a convenient environment by giving the process a special narrative (Steffens et al., 2014). In this way, it represents the story's genre.
	R2: Due to the diversity of collected words in the brainstorming activity, practitioners would not likely claim (MR3). However, the practitioners perceived the opportunity to claim as positive and useful to keep up with the process (MR4).
	I2: In order to increase the probability that practitioners could claim, we refer to Steffens et al. (2014), who describe the game element of "virality and community collaboration". This element encourages cooperation among practitioners and helps them find efficient ways to achieve the group goal.
	R3: The original process ended with a complete story. The practitioners felt dissatisfaction (MR5). The collaborative writing process should have a corresponding goal that the practitioners focus on achieving during the process (MR6). As the practitioners had various views on the goal of the process, it is not enough to end the process with a complete story (MR7).
Winning	**I3:** Although the practitioners argued about the process's goals, they had one thing in common: their interest in the other groups. Though they mostly expressed dissatisfaction with their own stories, they wanted them to at least exceed the other groups' stories in quality. Thus, one should announce a competition betwen the groups beforehand. By doing so, individuals rally around one goal: winning against other groups. In addition, it might guarantee that groups create better quality stories. Though one might consider the "group competition" element an extrinsic motivation, it would not negatively impact the process's quality as only the most qualitative story wins (Steffens et al., 2014).
	R4: The practitioners tended to look back at the written sentences and eventually the complete story. They did not hesitate to share their own opinions and discuss the words they chose, the story content they desired, future sentences, and so on. Participants need to appreciate one another's ideas and perspectives (MR8) while still focusing on achieving a common goal (MR9).
	I4: A lack of feedback among team members can impede collaboration processes, which one can solve by integrating "points". In connection with the "no relationship between members" issue, a shared group score can encourage members to build relationships with one another (Steffens et al. 2014).
	R5: Although the team has a common shared group goal, each practitioner writes separate sentences (MR10) and strives to achieve an individual goal. In this game, that individual goal comprises contributing valuable sentences to the story that include one predefined word and that the other team members recognize and appreciate.

	I5: Steffens at al. (2014) note that lack of perception of work in progress can be an issue. Hence, they suggest a “progression” game element to monitor individuals’ achievements. It is possible to link each sentence to one of the practitioners and assess the value they provided by color-coding individuals’ contributions.
Deflections	**R6:** Amending the activity “write sentence” can prevent deflections. Coming up with a sentence and writing it down can take a lot of time (MR12), which either leads to or is caused by deflections (MR13). Consequently, practitioners may lose focus of the process, which will decrease the story’s quality.
	I6: Steffens et al. (2014) consider including a time “countdown” to deal with excessive workload or, for our collaborative writing process, with an excessive use of time (i.e., to prevent the practitioners from spending too much time on writing a sentence). Such a countdown will avoid any kind of impatience from the other practitioners.
	R7: Practitioners appeared to find it difficult to use all predefined words (MR14). Consequently, most practitioners lost their motivation towards the end of the process. Some practitioners also appeared excited about formulating the end of the story (MR15), but, therefore, the team needed to coordinate all left words.
	I7: Steffens et al. (2014) suggest the “ownership” game element. This element will help practitioners bear responsibility for the story and, thus, foster a qualitatively valuable outcome. By choosing the end of the story more freely and independently, the practitioners might feel more pressured to make the story a good one.

10.4.3 Formalization of Learning (Cycle 1)

At this point, we used the developed interventions to remodel the collaboration process. We display the remodeled collaboration process in Figure 6 (right). We present the amendments in bold.

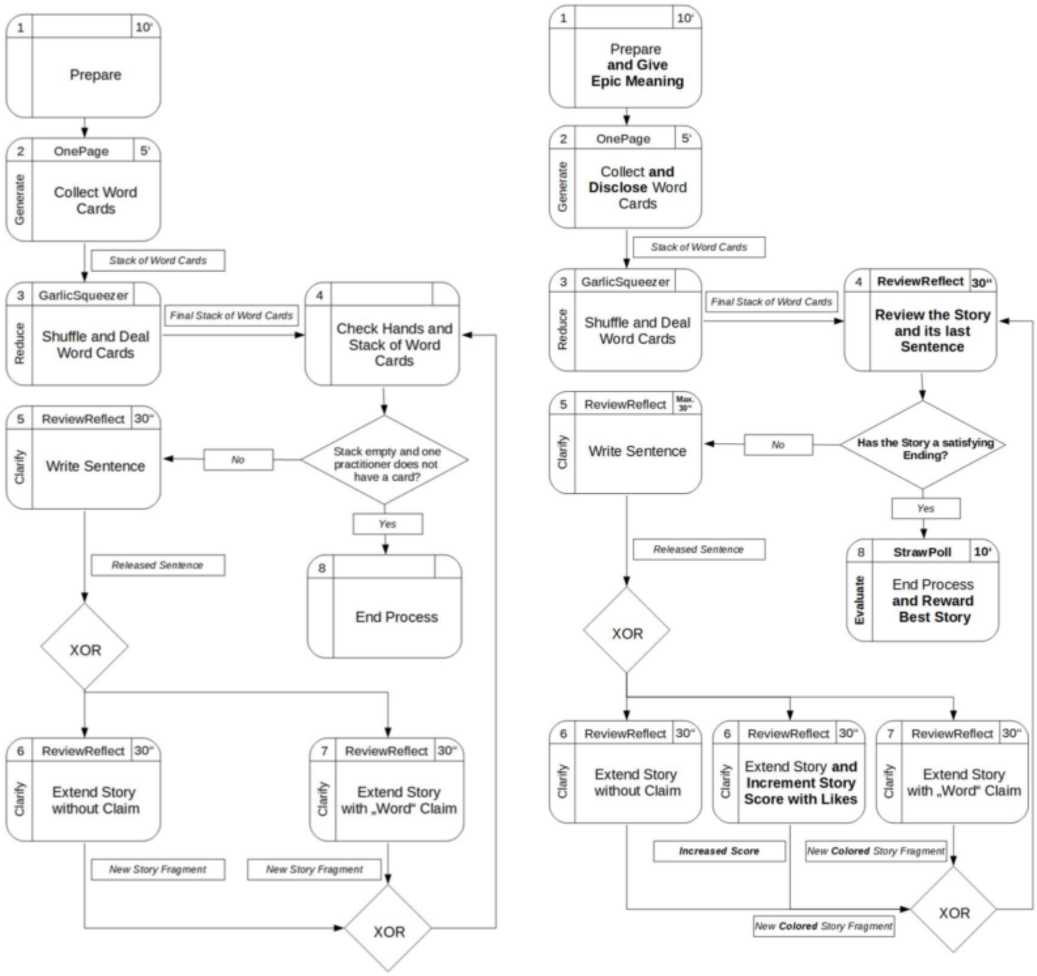

Figure 6. FPMs of the Analog (Left) and the Remodeled Collaboration Process (Right)

Based on the collected meta-requirements, we generalized and developed corresponding design principles (DP) according to Chandra, Seidel, and Gregor (2015) for platform designers and developers. Furthermore, we aligned and built on our preliminary framework for gamifying collaboration processes. We now emphasize the framework's segments: mechanics (M), dynamics (D), meaningful engagement (AE), and deep engagement (FE).

10.4.3.1 Genre

MR1: Go through a profound preparation before process start.

MR2: Follow a golden thread.

DP1: Include an epic meaning to highlight and set clear group goals to help groups achieve better outcomes.

The epic meaning gives meaning to the process and a genre to the story. Therefore, it enables practitioners to delve into a narrative, which triggers hedonic motivation and, with it, deep engagement (FE). Also, by predefining the genre, one can create meaningful engagement by providing practitioners with a sensory experience through fantasy and narrative (AE) and ensuring they stick to the genre (D). In this way, they can achieve a better outcome (AE). Apart from the presets, appealing visual assets foster epic meaning (M).

MR3: Use the collected words in a collaborative way.

MR4: Keep up with the process while not writing a sentence.

DP2: Ensure that everyone can see all content that they initially collected and need to consider to create the outcome.

This element serves as a complement to the epic meaning and helps practitioners understand the genre in a similar way. By sharing information (D), practitioners encourage fellowship (AE). Therefore, we offered each participant a feature to share their input in advance and another one to see and use the collected content for the outcome (M).

10.4.3.2 Winning

MR5: Focus on the story's quality.

MR6: Focus on achieving a meaningful goal.

MR7: Perform as a team.

DP3: Enhance the team's focus towards producing a quality outcome by including a group competition at a higher level and highlight the group setting's purpose and goal.

The competition element builds on the desire to win and creates a sensory challenge experience (AE) for practitioners. As only the best story will receive rewards (M), this element ensures that practitioners focus on quality and the instrumental outcome (AE). Additionally, having a competition between groups will strengthen the goal alignment in a group and the gameful interactions (D).

MR8: Appreciate group members' ideas and perspectives.

MR9: Foster group awareness.

DP4: Include feedback, rating, and veto mechanisms for the groups to change and adjust the content they contribute to shape and improve the group outcome iteratively.

This element focuses on meaningful engagement by recognizing and associating the individual contributions with the outcome (AE). Similar to the competition element, this element focuses on contribution quality and group awareness (i.e., gameful user-to-user interactions (D)). To enable these mechanisms, one needs to offer visual items to participants that they can use to either give feedback and ratings or to receive and disclose feedback and ratings (M).

MR10: Perceive the work in progress.

MR11: Recognize individuals' contributions.

DP5: Include visual mechanisms to highlight any group members' contribution to recognize individuals' efforts and outline the group's progress by appreciating contributions.

Highlighting the individual contributions with visual mechanisms (M) focuses on triggering experiential outcomes (AE). In this way, it resembles the status and self-expression affordances (M) that encourage a sense of self-progress. Thus, practitioners can identify with the story and its progress and the group will recognize their contributions (D).

10.4.3.3 Deflections

MR12: Enable a dynamic process.

MR13: Avoid impatience and deflections.

DP6: Include an appropriate time constraint to provide practitioners with the right amount of time (i.e., with an amount of time that others accept and that prevents impatience in a group-oriented dynamic) to produce quality content.

One can use time pressure (D) to individually motivate and challenge (AE) the practitioners. While writing, the practitioners need to align to the time limit that a visual countdown provides (M), which ensures they dynamically and actively work towards the goal (D).

MR14: Coordinate and align all predefined words.

MR15: Bear responsibility.

DP7: Provide the practitioners with ownership over the produced outcome to raise the perceived level of responsibility.

By giving the practitioners ownership over the story (M), they can better attach and identify with the outcome (AE). Furthermore, they feel responsible for their interactions and creations (D). This element focuses on hedonically motivating them (FE) and ensures they focus on the functional outcome's quality (AE).

Before implementing and executing the remodeled process in a Web application, we conducted an evaluation with an expert who eventually agreed to the interventions we developed (see Table 2 and Figure 6). After the implementation, we conducted a walkthrough with practitioners over the Web interface (see Figure 7, left) before we conducted the second cycle.

10.4.4 Building, Intervention, and Evaluation (Cycle 2)

After we formalized the preliminary design principles and the consequent interventions to increase the practitioners' hedonic motivation, their meaningful engagement, and story quality, we implemented the interventions in a Web application. Figure 7 (left) shows the main page of the Web application. We conducted the core iterative part of the process on that page. One can see that we implemented three design principles: a green like button with the number of counts in the “action panel” (DP4), individually colored sentences in “the story” (DP5), and a 30-second countdown in the wait button in the “action panel” (DP6).

We implemented the epic meaning (DP1) through a pop-up that introduces the topic (i.e., writing a fairytale) and the process rules (see Figure 7, right). Right before the process begins, the participants brainstorm and collect words to include in the story. At this stage, everyone can see all initially collected words (see Figure 8) (DP2).

Participants could not see the two remaining design principles that we implemented. That is, we announced the group competition (DP3) before the process began and assigned ownership to them by enabling them to end the process freely and independently (DP7).

We executed the remodeled process in a pilot study with two groups that contained five participants each using the Web application. It proceeded without any obstructive problems that concerned technology and functional process elements.

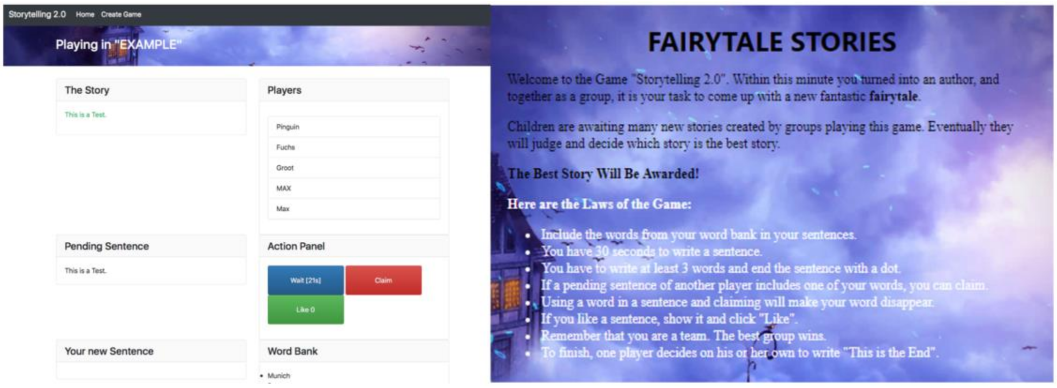

Figure 7. Screenshot of the Web Application: Main Page (Left), EpicMeaning Pop-up (Right)

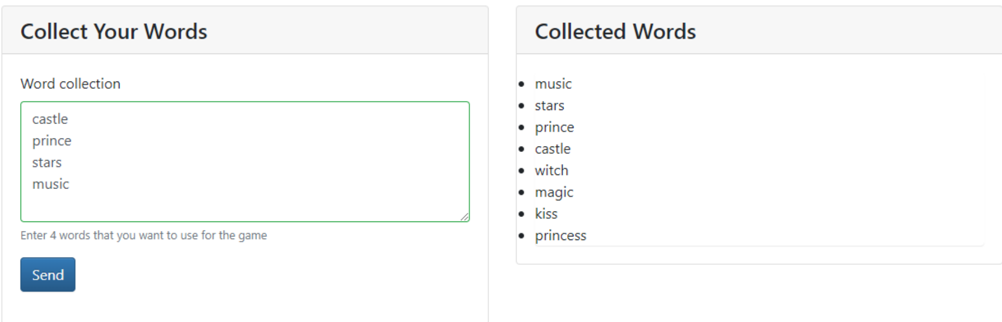

Figure 8. Screenshot of the Web Application: Collected Words Brainstorming

10.4.5 Reflection and Learning (Cycle 2)

For the evaluation, we interviewed the practitioners. By participating in both cycles (see Table 2), they constituted process experts in our research. Since we used guidelines to conduct the interviews, we asked rather abstract questions that focused on the individual interventions (see Appendix A). In doing so, we could evaluate the intended effects by comparing them to the perceived effects. Following the thematic comparison approach (Meuser & Nagel, 1991), we used interview transcripts to consolidate and differentiate the experts' (E) replies to several categories that covered interview guideline questions. We discuss the categories below.

Epic meaning: only one expert preferred not to have a predefined topic (E1). The others considered it quite helpful to write a story: it simplified the start of the process (E3) and gave an approximate direction for a golden thread (E2, E4, E5, E6, E7, E8, E10). Thus, they also found it easier to collect words and use them in the same context (E8, E9, E10). Still, some experts found it problematic as they had different perspectives and ideas about one topic (E4, E6, E7).

Disclosure of collected words: the experts found disclosing the collected words useful to plan and establish the story's direction beforehand (E1, E2, E3, E7, E8, E10) and to support others during the process (E5, E10). Still, one expert did not recognize any effect (E6). Another expert stated that, while groups could use it effectively, their group did not do so (E9).

Group competition: the sentiment about a group competition varied possibly due to participants' different characters and attitudes. While competition and its perceived meaningfulness and severity motivated some participants (E2, E5, E7, E9), competition did not stimulate others at all (E1, E3, E4, E6, E8, E10). As intrinsic motivation, fun, and creativity already drove these individuals, the extrinsic motivation did not have any effect on them (E1, E3, E4, E8, E10).

Like button: only three experts considered this button a good and funny feature for both positive affirmation and staying concentrated while waiting (E4, E6, E9). However, most experts considered it unnecessary (E1, E2, E3, E5, E7, E8, E10). Some found it pointless not having any consequences after liking a sentence (E3, E5, E8). Hence, some experts suggested alternatives, such as integrating a scoring system (E5).

Color assignment: except for one expert who did not recognize it (E1), the color assignment appealed to all the other experts (E2, E5, E8, E9, E10). They found it helpful to match sentences to practitioners (E2, E3, E4, E5, E7, E8, E9, E10), see everyone's turn, and prepare their own next sentence (E4, E6, E7, E9). Furthermore, they found that it allowed them to view the story overall (E5, E6, E7, E8, E10).

Countdown: except for one expert who found the countdown too long (E1), the other experts requested more time for the countdown. They said that 30 seconds put too much pressure on them (E2, E3, E5, E7, E8, E9, E10) and, thus, decreased output quality (E4, E6, E8). Nevertheless, all experts agreed that a countdown itself constituted a good and meaningful idea that can help make the process more dynamic and help practitioners avoid impatience (E1, E2, E3, E4, E5, E6, E7, E8, E9, E10). Two experts also made suggestions about how to adjust the countdown, such as using different timeframes to promote concentration (E6) or including penalties for not sticking to the timeframe (E10).

Free end of process: most experts favored choosing when the process ended of their own accord (E1, E2, E3, E4, E5, E9, E10) in order to create a worthy and nice ending rather than an endless story with a sudden end (E2). Furthermore, they found choosing their own ending easier (E5), more pleasant, and not forced, which resulted in a better quality story (E8). Eventually, they stated that “it is our story” (E3), “we know better when the story should find an end than an application” (E10), and “I liked it because we could determine it” (E10).

Satisfaction: the experts expressed satisfaction with the process, how we conducted it, and their involvement in it (E1, E2, E3, E4, E5, E6, E7, E9, E10). They expressed finding it “fun” (E1, E6, E7, E9), “cool” (E5, E6), and “very interesting” (E7, E9). As such, it seems it encouraged hedonic motivation as we intended. However, though some experts considered their produced story funny (E1, E7), most considered it meaningless (E2, E3, E6, E9). Nevertheless, they blamed their teammates and the initial word collection for the resulting dissatisfaction rather than the actual process (E2, E5, E8, E9, E10)

In a final step, external unbiased judges to assess the created stories using criteria for qualitatively valuable stories that we established before. We took the first measures from Rhodes (1961): level of creativity and novelty (Rhodes, 1961). In addition, Dean, Hendler, Rodgers, and Santanen (2006) define dimensions and subdimensions for idea quality that we used to assess the stories’ quality: workability, relevance, and specificity. We aligned the assessment questions to these dimensions and calculated the average result for each dimension (Dean et al., 2006). The ratings ranged from very low/bad (1) to very high/good (7). We chose this scale based on Johns (2005) to avoid the data becoming less accurate with fewer than five and higher than seven scale points (DeVellis, 2003). We chose seven-point Likert scales to ensure several judges could evaluate the stories in detail (Johns, 2010; Likert 1932). In total, 29 judges participated. We demonstrate the results in Table 5.

Table 5. Results from Assessing the Stories’ Quality

	Cycle 1		**Cycle 2**	
	Story 1	Story 2	Story 3	Story 4
Level of novelty	3.62	4.34	3.83	4.07
Level of creativity	3.76	4.86	3.72	4.72
Specificity	4.05	3.24	3.57	4.19
Workability	4.29	3.55	3.66	4.12
Relevance	3.83	3.28	3.41	4.28
Sum	19.55	19.27	18.19	21.38
Overall sum in each cycle	38.82		39.57	

10.4.6 Reflection and Learning (Cycle 2)

Examining the results, we can see that the stories did not vary that much in quality. Even though participants created the best-rated story in the second cycle, they also created the worst-rated story in the second cycle. The criteria measurement values varied from story to story, which may imply that executing the collaboration process does not differ depending on whether participants do so digitally or in an analog meeting setting. Participants' satisfaction with the outcome similarly reflects this interpretation. Nevertheless, participants viewed the digital process and the Web application positively and even showed higher hedonic motivation and engagement compared to the analog setting. Hence, in accordance with our preliminary framework (see RQ1 and Table 1), we covered various intended outcomes and experiences by using several gamification elements (see RQ2, Table 6, and Figure 6).

Table 6. Gamification Elements that Supplement the Collaborative Writing Process (RQ2)

DP	Gamification element	Collaborative writing process phase
1	Epic meaning	Preparation and introduction of the process
2	Disclosure of collected words	Preparation and idea generation for the process
3	Group competition	Higher level for several group processes
4	Like button	Reactions to process contributions
5	Color assignment	Highlighting process contributions
6	Countdown	Limiting the time of process contributions
7	Free end of process	End of the process

Thus, these elements could trigger individuals' hedonic motivation and meaningful engagement to become more satisfied with the process and increase their intention to continue to use the artifact. Still, the digital process did not achieve a better outcome compared to the analog setting (Bhattacherjee & Premkumar, 2004; Lowry et al., 2015).

10.5 Discussion

Overall, we formulated seven design principles and 15 meta-requirements based on observing participants in the first cycle and reflecting on their activities and comments. We further categorized these design principles and meta-requirements into three categories: genre, winning, and deflections. We found that we could successfully implement the seven design principles into an instantiated digital collaborative writing process and, thereby, encourage higher hedonic motivation and meaningful engagement to increase participants' satisfaction and their intention to continue to use the digital collaborative writing process. The evaluation indicated that an epic meaning (DP1) can help participants at the start of the process to stay on topic. Still, participants may find it difficult to clearly predefine the topic as some probably have different perspectives and interpretations, which could lead to discrepancies in a team. Participants perceived the disclosure of the collected includable content (in this case, the collected words) (DP2) as supporting collaboration in the team. However, just like the epic meaning, its usefulness depends on the individual participants.

The evaluation showed that, while some participants felt motivated write a qualitatively valuable story, others saw more fun in the process. Thus, some participants did not take the process seriously, which caused dissatisfaction in the team. We also recognized a difference of opinion in the group competition (DP3). While the group competition motivated some participants, it did not affect others at all. This finding emphasizes the need to address both deep engagement with hedonic motivation and meaningful engagement. Only three experts (E4, E6, E9) found the Like button (I4), which we implemented for participants to appreciate ideas and foster group awareness, suitable. As such, it seems space exists for a more suitable intervention design to better meet DP4. Some participants suggested a scoring system. Such a system could enrich the like button's meaning and simultaneously increase participants' motivation and engagement. Nevertheless, the evaluation indicated that highlighting the individuals' contribution (DP5) and assignment of ownership (DP7) successfully nurtured motivation and engagement. These DPs helped the participants to identify themselves with their own story. At last, participants received the idea of a countdown (DP6) well. The evaluation proved that it fulfilled the meta-requirements to enable a dynamic process and to avoid impatience and deflections. On the other hand, the evaluation revealed that the countdown lacked quality due to the time limit. Thus, one needs to give participants enough time to come up with qualitatively valuable contributions. Further suggestions included varying the timeframe for each new contribution and including penalties for not sticking to the timeframe.

All in all, we respond to our research questions by providing a preliminary framework for gamifying collaboration processes to increase individuals' intention to continue to use a system (RQ1) and by specifically applying gamification to a collaborative writing process underlying the preliminary framework (RQ2). The developed design principles increased participants' hedonic motivation and meaningful engagement. They had fun and took pleasure in the process while understanding what it meant and working toward its functional outcome. We found that we could encourage satisfaction and creativity even in a virtual environment by designing and implementing systems that bind, engage, and motivate contributors to participate and produce promising outcomes that do not vary that much from outcomes in a classical co-located setting.

Despite our findings, as with any study, ours has several limitations. First, each story noticeably differed from the others in content, structure, writing style, and so on. Thus, to better verify the design principles we implemented, one would need more stories and, thus, more participants. Second, one should carefully consider the participants they select and the type of story one has them write. For the instantiated process at hand, we randomly selected practitioners and asked them to write a fairytale. This choice might have raised and encouraged a discrepancy in the group as each participant could have viewed and understood the predefined topic in a different way. In a narrow sense, the participants did not have any connection to either the other participants or the process/topic.

10.6 Conclusion and Contribution

This study serves as a starting point for further research on collaborative writing processes. In gamifying a collaborative writing process via a Web application, we found that the developed design principles succeeded in increasing participants' hedonic motivation and meaningful engagement. Overall, this paper contributes to research on gamification, digital collaboration processes, and collaborative writing. Moreover, the study provides prescriptive knowledge (Gregor & Hevner, 2013) toward a "theory of design and action" (Gregor, 2006) with a collaboration process, MRs, and corresponding DPs and a preliminary framework for gamifying collaboration processes. Following Steffens et al. (2014), who declared that gamification has a positive impact on collaborative software development processes, we connected established gamification elements with a collaborative writing process. In doing so, we adjusted elements to suit the exemplary process and evaluated them regarding their impact on the process. Eventually, we established and evaluated seven design principles to foster hedonic motivation and meaningful engagement toward a qualitatively valuable outcome. In doing so, we built on our preliminary framework for gamifying collaboration processes. Practically speaking, practitioners may use this framework as guideline for gamifying their collaboration processes. Theoretically speaking, researchers may use this framework to understand, design, and analyze gamified digital collaboration processes. Additionally, we call on future research to extend the framework with other complementary models, principles, and theories on an abstract level and to evaluate more specific elements that fit into the framework (e.g. gamification elements and affordances). Furthermore, research needs to test the DPs for their applicability to other collaborative writing practices, such as wikis, blogs, and forums. Future research may also test whether the framework applies to other collaboration process scenarios, which may lead to new design principles (e.g., for gamified human-AI collaboration).

10.7 References

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward information technology usage. *Management Information Systems Quarterly*, *28*(2), 229-254.

Bittner, E., & Leimeister, J. M. (2014). Creating shared understanding in heterogeneous work groups: Why it matters and how to achieve it. *Journal of Management Information Systems*, *31*(1), 111-144.

Bowers, C. A., & Pharmer, J. A. (2000). When member homogeneity is needed in work teams. *Small Group Research*, *31*(3), 305-327.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivation. *Interactions*, *25*(5), 38-43.

Briggs, R. O., Kolfschoten, G. L., & de Vreede, G.-J., Albrecht, C. C., & Lukosch, S. G. (2010). Facilitator in a box: Computer assisted collaboration engineering and process support for rapid development of collaborative applications for high-value tasks. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*.

Briggs, R. O., Kolfschoten, G. L., de Vreede, G.-J., Lukosch, S., & Albrecht, C. C. (2013). Facilitator-in-a-Box: Process support applications to help practitioners realize the potential of collaboration technology. *Journal of Management Information Systems*, *29*(4), 159-194.

Briggs, R. O., & de Vreede, G.-J. (2009). *Thinklets: Building blocks for concerted collaboration*. Omaha, NE: Center for Collaboration Science.

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions. In *Proceedings of the 48th Hawaii International Conference on System Sciences.*

Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal*, *12*(2), 68-82.

Curtis, D., & Lawson, M. (2001). Exploring collaborative online learning. *Journal of Asynchronous Learning Network*, *5*(1), 21-34.

Darejeh, A., & Salim, S. S. (2016). Gamification solutions to enhance software user engagement—a systematic review. *International Journal of Human-Computer Interaction*, *32*(8), 613-642.

Dean, D. L., Hendler, J. M., Rodgers, T. L., & Santanen, E. L. (2006). Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation. *Journal of the Association for Information Systems*, *7*(10), 646-699.

Debs, M. B. (1991). Recent research on collaborative writing in industry. *Technical Communication*, *38*(4), 476-384.

DeCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York, NY: Academic Press.

Deterding, S., Khaled, R., Nacke, L., & Dixon, D. (2011). Gamification: Toward a definition. In *Proceedings of the ACM Conference on Human Factors in Computing Systems.*

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed., vol. 26). Thousand Oaks, CA: Sage.

Forman, J. (1991). Collaborative business writing: A Burkean perspective for future research. *The Journal of Business Communication*, *28*(3), 233-257.

Friedrich, J., Becker, M., Kramer, F., Wirth, M., & Schneider, M. (2020). Incentive design and gamification for knowledge management. *Journal of Business Research*, *106*, 341-352.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611-642.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337-355.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the Challenges in Games AI Workshop, Nineteenth National Conference of Artificial Intelligence.*

Johns, R. (2010). *Likert items and scales.* Retrieved from https://ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf

Johnson, D., & Johnson, R. T. (1996). *Cooperation and the use of technology*. New York, NY: Macmillan.

Koch, M., & Gross, T. (2006). *Computer-supported cooperative work*. Bonn: Koellen Verlag.

Kolfschoten, G., & de Vreede, G.-J. (2009). *A design approach for collaboration processes: A multi-method design science study in collaboration engineering*. New York, NYL M.E. Sharpe.

Langan-Fox, J., Wilson, J. R., & Anglim, J. (2004). Mental models, team mental models, and performance: Process, development, and future directions. *Human Factors and Ergonomics in Manufacturing*, *14*(4), 331-352.

Leimeister, J. M. (2014). *Collaboration Engineering: IT-gestützte Zusammenarbeitsprozesse systematisch entwickeln und durchführen*. Berlin, Heidelberg: Springer.

Likert, R. (1932). *A technique for the measurement of attitudes* (vol. 140). New York, NY: Archives of Psychology.

Liu, D., Santhanam, R., & Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly*, *41*(4), 1011-1034.

Löffler, C., Hofmann, J., & Bodendorf, F. (2009). *Forschungsmethoden in der Wirtschaftsinformatik*. Nürnberg.

Lounis, S., Doukidis, S., Papastamatiou, N., & Doukoulos, T. (2016). gamification of authoring interactive e-books for children: The Q-tales ecosystem. In *Proceedings of the Mediterranean Conference on Information Systems.*

Lowry, P. B., Gaskin, J. E., & Moody, G. D. (2015). Proposing the multi-motive information systems continuance model (MISC) to better explain end-user system evaluations and continuance intentions. *Journal of the Association for Information Systems*, *16*(7), 515-579.

Marczak, S., Filho, F. F., Singer, L., Treude, C., Steffens, F., & Redmiles Ban Al-Ani, D. (2015). Studying gamification as a collaboration motivator for virtual software teams. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing.*

Meske, C., Brockmann, T., Wilms, K., & Stieglitz, S. (2016). Gamify employee collaboration: A Critical review of gamification elements in social software. In *Proceedings of the Australasian Conference on Information Systems.*

Meuser, M., & Nagel, U. (1991). *ExpertInneninterviews: Vielfach erprobt, wenig bedacht; ein Beitrag zur qualitativen Methodendiskussion*. In D. Garz & K. Kraimer (Eds.), *Qualitativ-empirische sozialforschung: Konzepte, methoden, analysen.* Weisbeden: VS Verlag für Sozialwissenschaften.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*(7), 305-310.

Richter, A., Heinrich, P., Stocker, A., & Schwabe, G. (2018). Digital work design—the interplay of human and computer in future work practices as an interdisciplinary (grand) challenge. *Business & Information Systems Engineering*, *60*(3), 259-264.

Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A., Oeste-Reiß, S., Randrup N., Schwabe G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2).

Sein, M., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*.

Simmert, B., Ebel, P. A., Peters, C., Bittner, E. A. C., & Leimeister, J. M. (2019). Conquering the challenge of continuous business model improvement: Design of a repeatable process. *Business & Information Systems Engineering*, *61*(4), 451-468.

Steffens, F., Marczak, S., Figueira Filho, F., Treude, C., Singer, L., & Redmiles Ban Al-Ani, D. (2014). Using gamification as a collaboration motivator for software development teams. In *Proceedings of the Brazilian Symposium on Collaborative Systems.*

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, *31*, 275-288.

Suh, A., Cheung, C. M. K., Ahuja, M., & Wagner, C. (2017). Gamification in the workplace: The central role of the aesthetic experience. *Journal of Management Information Systems*, *34*(1), 268-305.

Suwantarathip, O., & Wichadee, S. (2014). The effects of collaborative writing activity using Google Docs on students' writing abilities. *Turkish Online Journal of Educational Technology*, *13*(2), 148-156.

Tavanapour, N., & Bittner, E. A. C. (2017). Collaboration among crowdsourcees: Towards a design theory for collaboration process design. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Thiebes, S., Lins, S., & Basten, D. (2014). Gamifying information systems: A synthesis of gamification mechanics and dynamics. In *Proceedings of the 22nd European Conference on Information Systems.*

Tseng, S.-L., & Sun, H. (2017). Playful design elements and stages of player experience in gamification. In *Proceedings of the 38th International Conference on Information Systems.*

Vanduhe, V. Z., Nat, M., & Hasan, H. F. (2020). Continuance intentions to use gamification for training in higher education: Integrating the technology acceptance model

(TAM), social motivation, and task technology fit (TTF). *Multidisciplinary Open Access Journal*, *8*, 21473-21484.

Wegge, J., Roth, C., Neubach, B., Schmidt, K.-H., & Kanfer, R. (2008). Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size. *Journal of Applied Psychology*, *93*(6), 1301-1313.

Wilde, T., & Hess, T. (2007). Forschugsmethoden der wirtschaftsinformatik. *Wirtschaftsinformatik*, *49*, 280-287.

10.8 Appendix: Interview Guideline

Satisfaction

1. How satisfied were you with the process, the conduction, and your involvement?
2. How satisfied were you with the story outcome of the process?

Difficulty

1. How difficult was it to understand the process and its goals?
2. How difficult did you perceive the web-application tool?

Interventions

1. How did you perceive the predefinition of an epic meaning?
2. How did you perceive the disclosure of the collected word cards?
3. What effect did the group competition have on you and your contributions?
4. How did you perceive the “like” button?
5. How did you perceive the color assignment to the individual practitioners?
6. How did you perceive the countdown element?
7. How did you perceive the fact, that you can end the process freely and independently?

Feedback

1. Do you have any remarks, wishes, improvement suggestions etc.?
2. Did you miss any functionalities, elements or possibilities?
3. What would you do differently in the digital solution?
4. Do you have any more feedback?

11 Hybrid Intelligence – Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review

Wiethof, C., and Bittner, E. A. C. 2021. "Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review," in *42nd International Conference on Information Systems (ICIS)*, Austin, Texas, USA.

Abstract. The paper aims at establishing a common ground and understanding of collaborative learning approaches between humans and computers to encourage Hybrid Intelligence. Thereby, we put a special focus on identifying how humans and computers learn from each other through an iterative and interdependent process involving the human-in-the-loop as well as the computer-in-the-loop. To approach this aim, a systematic literature review is conducted. Therefore, we reviewed 2098 publications in three relevant databases and found 33 appropriate publications regarding both human and computer learning. We capture our results by inductively deriving three patterns and three sub-patterns of collaborative learning processes, namely exploration, integration and decision support including assimilation, exploitation and explanation. Additionally, we provide learning measurements from the reviewed literature according to each pattern. Finally, we disclose future research avenues and implications for Hybrid Intelligence in the field of Human-Computer-Interaction, especially collaborative human and machine learning.

Keywords: Hybrid Intelligence, human-in-the-loop, computer-in-the-loop, collaborative learning

11.1 Introduction

In recent years, the concept of interactive Machine Learning (ML) has been developed and established as a favorable alternative for traditional automatic ML (Holzinger 2016a; Holzinger et al. 2016, 2017). As automatic ML is commonly applied to process and find patterns in big data, it has proven successful, among others, for speech recognition, recommender systems and autonomous vehicles (Holzinger 2016a; Holzinger et al. 2016). However, in complex domains like health informatics, biology or medicine, only a limited amount of data is available. Here, the interactive ML approach achieves better results by reducing the complexity through human involvement (Holzinger 2016a; Holzinger et al. 2016, 2017). It is based on several ML concepts encompassed by the field of human-in-the-loop (HITL) algorithms (Holzinger 2016a; Martínez et al. 2019). HITL is an approach, which encourages the human to interact with the algorithm or learning system itself (Martínez et al. 2019). By putting the human in the loop of ML, humans are able to directly use, manipulate and optimize an algorithm combining both human and machine intelligence (Holzinger 2016a; Holzinger et al. 2016, 2017). For example, it is possible to directly involve the end users in the development of the ML system. This in turn enables faster, more flexible and focused learning cycles leaving out additional iterations with the ML experts as in traditional ML (Amershi et al. 2014).

Relying on this, more ML researchers are focusing their work on how to develop, implement and improve the HITL approach in various domains, contexts and scenarios, e.g. paraphrasing (Yimam and Biemann 2018), dataset annotation in biomedicine (Yimam et al. 2016) or classification of clinical textual data (Kartoun 2017). This also includes research on refinement operations for human feedback (Lee et al. 2017) as well as on the assessment of human feedback quality (Lloyd et al. 2017). For instance, Lloyd et al. (2017) enabled a system to predict the benefit of a human's decision (Lloyd et al. 2017). What is more, researchers consider the HITL approach as an opportunity to turn a "black-box" into a "glass-box" encouraging trust and acceptance of the ML system among end users (Holzinger et al. 2017). Despite the traditional ML metrics, such human factors are in fact crucial for the design and evaluation of interactive ML algorithms (Krening and Feigh 2018). Thus, research has already considered and focused on several perspectives and aspects of user experience, e.g. frustration, transparency, complexity of instructions (Krening and Feigh 2018), level of human control (Smith-Renner et al.), fairness, confidence (Dodge et al. 2019) or user interface design (Benedikt et al. 2020). Especially in terms of transparency, researchers examine and work on different levels and ways of enabling transparency in HITL learning. Examples are appropriate explanations from the ML system (Dodge et al. 2019) or certain types of visualizations (Reani et al. 2018; Tamagnini et al. 2017; van der El et al. 2016).

Eventually, there is a noticeable shift to human-centered approaches. However, previous research emphasized the investigation of users and their experience concerning the acceptance, use and improvement of ML algorithms. Thereby, they neglect the possible learning benefits on the human side (Schneider 2020). With his work, Schneider (2020) demonstrates, that optimizing human inputs through instructions can also lead to better task performance (Schneider 2020). This perspective of human learning through computer support is based on the computer-in-the-loop (CITL) approach (Shneiderman 2020). So far, most research focused on either the human or the computer perspective, but it is important to consider and integrate both complementing each other toward mutual benefits (Dellermann et al. 2019b; Peeters et al. 2020). For this purpose, Dellermann et al. (2019b) recently introduced the concept of Hybrid Intelligence. It combines the HITL augmenting machine intelligence with the CITL augmenting human intelligence (Cerf 2013; Dellermann et al. 2019b). This leads to a tightly coupled cycle between the user and the computer influencing and complementing each other. Accordingly, studying either of the two sides requires taking effects on the other side into consideration, and vice versa (Amershi et al. 2014). Therefore, it is important to achieve a balance between the human and the computer, i.e. to encourage leveraging the knowledge of the computer, but also to prevent overreliance on the computer and with that the dulling of the human's mind (Amershi et al. 2014; Dellermann et al. 2019b; Seeber et al. 2020). Following the research agenda of Seeber et al. (2020), we aim to contribute to this stream of research by studying Hybrid Intelligence. Thus, we identified the research gap of integrating current available literature on HITL and CITL toward mutual learning. We address this gap by reviewing research on the processes and effects of augmenting human as well as machine intelligence (Seeber et al. 2020). Therefore, this study answers the following research question: **Q1:** *How do humans and computers learn from each other through HITL and CITL learning processes?* As mutual learning effects are the outcome of an interaction process between a human and a ML system (Rzepka and Berger 2018; Wang 2019), we describe the identified learning processes and their effects as part of the findings. What is more, we identified the need for measuring instruments and measures of learning effects encouraged by HITL and CITL. Hence, the second research question is: **Q2:** *How can the learning effects encouraged by HITL and CITL be measured?*

We address both questions by conducting a systematic literature review based on the guideline of Webster and Watson (2002) and structure this review article according to vom Brocke et al. (2009) as follows: first, we demonstrate the conceptual background of our research in section 2. We then build upon this background to search for appropriate literature to review, and document the research method, search and analysis process in section 3. After that, we present the findings in section 4 including the literature analysis and synthesis along the literature concept table. Before concluding the paper in section 6, we discuss the findings in section 5, providing future research avenues and research implications.

11.2 Conceptual Background

11.2.1 Human-in-the-Loop

Advancing computers through human intelligence is typically applied in ML as so-called HITL learning (Dellermann et al. 2019b; Dellermann et al. 2019a; Martínez et al. 2019; Shneiderman 2020). Thereby, it is used for interactive ML, which Holzinger (2016b) defines as "algorithms that can interact with agents and can optimize their learning behavior through these interactions, where the agents can also be human". Thus, it encourages ML through human involvement and interaction with the algorithm or learning system itself. Thereby, it covers interactive ML approaches, which among others include or base on Supervised Learning, Active Learning and Reinforcement Learning (Figure 1). These learning approaches have already been successfully implemented in various applications, e.g. recommender systems by Reinforcement Learning or image retrieval by Active

Learning (Martínez et al. 2019). Also, the interactive ML approach by Holzinger (2016b) has already been successfully implemented to enhance results in the context of the Traveling Salesman Problem (Holzinger et al. 2019).

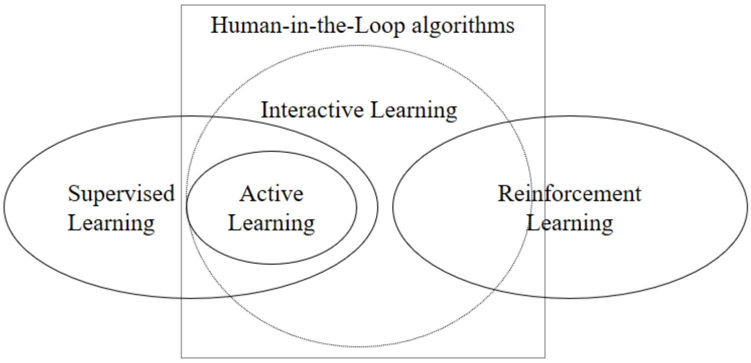

Figure 1. HITL Conceptualization by Martínez et al. (2019) based on Gaurav (2016), adapted

What is more, by putting the human in the loop of ML, human control is not limited to pre-processing and selecting data or features, but the human will also be involved in the actual learning phase directly interacting with the algorithm. Figure 2 visualizes the approach depicting the stages of (1) selecting data, (2) pre-processing, (3) interaction of human agent with computational agent and (4) final check by the human expert (Figure 2) (Holzinger 2016b).

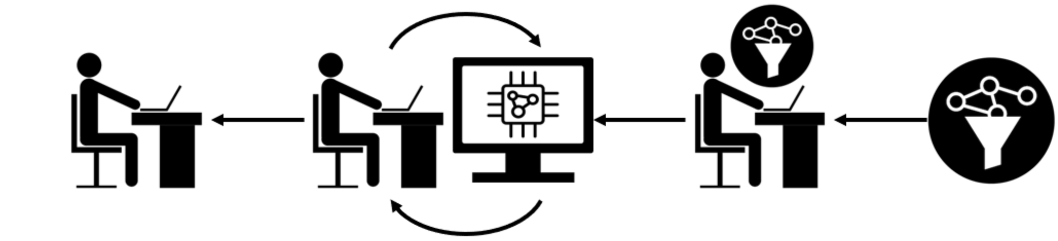

Figure 2. iML Human-In-The-Loop Approach (Holzinger 2016b), adapted

According to step 3 shown in Figure 2, by putting the human in the loop of ML, end users are enabled to directly influence the ML model, its behavior and outputs through their inputs and actions without any knowledge and expertise on ML. As demonstrated in Figure 3, this interactive approach aims at faster, more flexible and focused learning cycles. This is achieved by leaving out the additional iterations with the ML expert as in traditional ML. Thus, there is a particular focus on examining the user involvement in the construction of the ML system itself as well as in its development and improvement (Amershi et al. 2014).

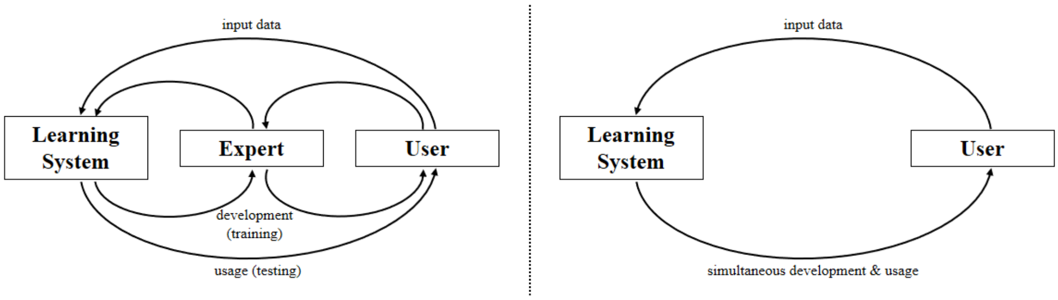

Figure 3. Traditional ML (left) and Interactive ML (right) (Amershi et al. 2014), adapted

Hence, many researchers focus on studying the system as the learner and the human user as its teacher. Thereby they aim at identifying ways of how to best improve the interactive learning process in order to achieve an advanced system performance (Amershi et al. 2014; Dellermann et al. 2019a; Martínez et al. 2019). For instance, Martínez et al. (2019) created a theoretical framework for conceptualizing interactive ML systems including several ways of how to improve the ML system through human involvement, such as interactive classification, clustering or teaching intelligent agents (Martínez et al. 2019).

11.2.2 Computer-in-the-Loop

Next to the HITL approach commonly used for ML, the CITL approach is used to support humans and improve their work effectiveness and efficiency, e.g. in decision-making activities through predictions (Dellermann et al. 2019b). Shneiderman (2020) particularly reverses the HITL approach with ideas for a more human-centered Artificial Intelligence. He considers CITL as an approach, where "humans work with others in teams, crews, and groups, with computers best designed as helpful tools that continuously provide information and carry out tasks, but do so under human control" (Shneiderman 2020). Some researchers have already drawn on this approach to advance the human users' usage of and the interaction with the computer and its predictions, e.g. by providing better transparency toward the users (Abdel-Karim et al. 2020; Amershi et al. 2014). Especially Abdel-Karim et al. (2020) investigate the perspective of human learning as counterpart of interactive ML. For example, they find that humans, like computers, can learn via error-learning, i.e. through receiving contradictions and feedback from the computer. Thus, they are willing to adapt their decisions based on a computer's contradictory prediction. Furthermore, by providing system transparency, e.g. disclosing information on the uncertainty of predictions, humans' willingness to listen to the computer can be even improved. Finally, they conclude with recommending the shift from the traditional HITL paradigm toward the human-centered mutual learning approach (Abdel-Karim et al. 2020).

11.2.3 Hybrid Intelligence

For defining the concept of Hybrid Intelligence, an understanding of the term intelligence is necessary. Therefore, we follow the definition of Dellermann et al. (2019b) describing it as "the ability to accomplish complex goals, learn, reason, and adaptively perform effective actions within an environment". In line with this, Dellermann et a. (2019a) then define the term of Hybrid Intelligence systems as "systems that have the ability to accomplish complex goals by combining human and artificial intelligence to collectively achieve superior results than each of them could have done in separation and continuously improve by learning from each other". Accordingly, Hybrid Intelligence encompasses two sub-dimensions of intelligence, namely artificial and human intelligence, and encourages the complementary strengths of both. For human intelligence, these strengths cover intuition including flexibility and transfer, empathy and creativity, the ability to annotate arbitrary data and common sense. For artificial intelligence, also referred to as machine or computer intelligence, they include analytics involving pattern recognition, probabilistic analysis, consistency, speed and efficiency (Dellermann et al. 2019b).

To ensure continuous interaction between a computer and its users co-evolving to achieve better outcome together, Dellermann et al. (2019a, 2019b) integrate the HITL and the CITL approach and combine them toward Hybrid Intelligence. Eventually, the ultimate goal is to develop Hybrid Intelligence systems including HITL by augmenting machine intelligence as well as CITL by augmenting human intelligence (Dellermann et al. 2019b) (Figure 4).

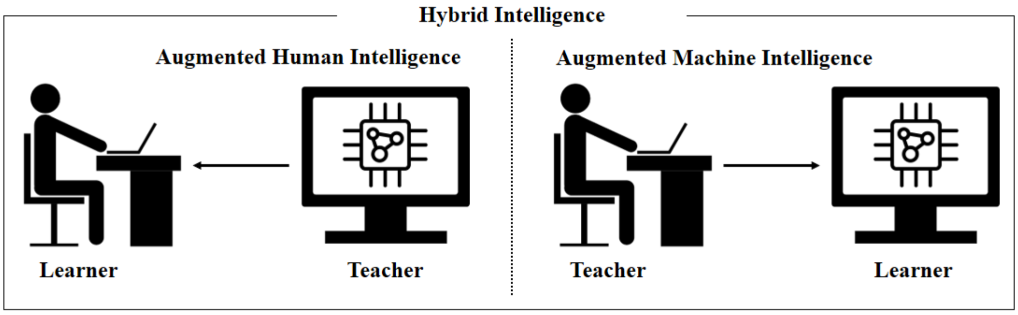

Figure 4. Distribution of Roles in Hybrid Intelligence (Dellermann et al. 2019b), adapted

However, most research focuses on the system learning perspective developing HITL techniques, whereas the human learning perspective is only sparsely addressed. For instance, introducing the concept of Hybrid Intelligence systems, the taxonomy by Dellermann et al. (2019a) mainly covers aspects of HITL ML, e.g. machine teaching approaches, ways of teaching interactions, expertise requirements for humans involved, amount of human input. Only a few aspects of the taxonomy indicate human learning, e.g. machine feedback toward the human or system requirements like interpretability (Dellermann et al. 2019a). Though current research works tend to focus on only one perspective, they emphasize the importance of integrating and combining both in order to achieve mutual learning benefits (Abdel-Karim et al. 2020; Amershi et al. 2014; Dellermann et al. 2019b). In this regard, it is necessary to retain a balance between trust and distrust toward the computer in order to avoid overreliance on the computer through CITL as well as to avoid overconfidence on human knowledge through HITL (Abdel-Karim et al. 2020; Dellermann et al. 2019b). Only by integrating both approaches, both human and computer can learn simultaneously leading to Hybrid Intelligence. This will eventually lead to superior results of the human-computer-collaboration (Abdel-Karim et al. 2020; Amershi et al. 2014; Dellermann et al. 2019b).

In order to develop Hybrid Intelligence systems, we identified the research gap of integrating current available literature on HITL and CITL toward collaborative learning processes (Q1). Additionally, for the evaluation of such systems, we use our Literature Review to provide measuring instruments and measures for learning effects encouraged by HITL and CITL (Q2). This may serve as a fundament for future research in the area of Hybrid Intelligence as well as provide further research avenues.

11.3 Method

We conducted a systematic literature review based on the guideline of Webster and Watson (2002) for writing a review article. In doing so, we aligned our work with the framework for literature reviewing of vom Brocke et al. (2009) (Figure 5).

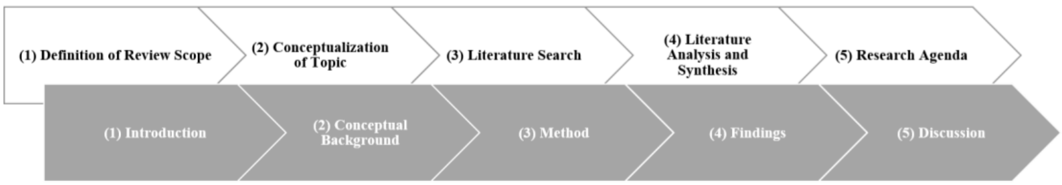

Figure 5. Phases of the Literature Review according to vom Brocke et al. (2009)

Databases: for determining the source material of literature, we considered three domain-relevant Information Systems databases, namely the Association for Information Systems eLibrary (AISeL), the Association for Computing Machinery Digital Library (ACM DL) and ScienceDirect.

Search String: we built the following search string for all three databases: “computer-in-the-loop” OR “human-in-the-loop” OR “hybrid intelligence” OR (“collaborative” OR “interactive” OR “machine” OR “reinforcement” OR “active” OR “supervised”) AND “learning” AND (“artificial intelligence” OR “ai”) AND “human”. Thereby, we derived the components of the search string from the conceptual background of section 2. Thus we specifically searched the databases for literature on “computer-in-the-loop”, “human-in-the-loop” or “hybrid intelligence” as our core concepts. To widen the scope beyond these wordings, we alternatively searched for types of “learning”, either “collaborative”, “interactive” or “machine” learning, especially “reinforcement”, “active” or “supervised” learning involving both a “human” as well as “artificial intelligence” or abbreviated “ai”. We applied the search to titles, keywords and abstracts, and identified a total of 2098 publications.

1st Screening, Selection Criteria: in the first screening iteration of the identified publications, we excluded all published abstracts, posters and sections. Further, considering titles, keywords, and abstracts, we excluded publications with the following focus: (1) cyber physical systems, robotics, vehicles and IoT and infrastructure (2) security, ethics and costs, (3) data integration, (4) gaming, (5) neuroscience. Finally, we also excluded all publications without learning interactions between human and computer, e.g. that focus on ML algorithms to solve complex problems. This resulted in a total of 586 publications.

2nd Screening, Selection Criteria, and Forward & Backward Search: in the second screening iteration we kept all publications in the set, which address learning effects for both computer and human user, relying on a certain level of inaccuracy on both sides. This excludes research focusing on the improvement of ML techniques without addressing human users' benefits or mutual learning effects respectively. The same applies for research focusing on the human learning part only, e.g. traditional decision support systems without underlying ML algorithms. This resulted in a total of 30 publications. After conducting a backward and forward search we could include three more publications leading to a final review set of 33 publications. The screening iterations segmented by database are visualized in Table 1.

Table 1. Literature Search Iterations

Database	**Search Results**	**First Screening**	**Second Screening (and Forward & Backward Search)**
AISeL	82	31	9
ACM DL	863	175	8
ScienceDirect	1153	380	13
Total Publications	**2098**	**586**	**30 (3)**

11.4 Findings

To structure and organize the identified literature, we created a literature concept table (Webster and Watson 2002) (Table 2). We determined the concepts to address Q1, demonstrating collaborative learning processes in Hybrid Intelligence systems. Therefore, we extracted the individual learning processes involving HITL and CITL from each literature source. We then abstracted them toward a higher level covering the complementing cycle of HITL and CITL. Thereby we aligned them with the generic collaborative learning process depicted in Figure 6 building on the conceptual background of Hybrid Intelligence.

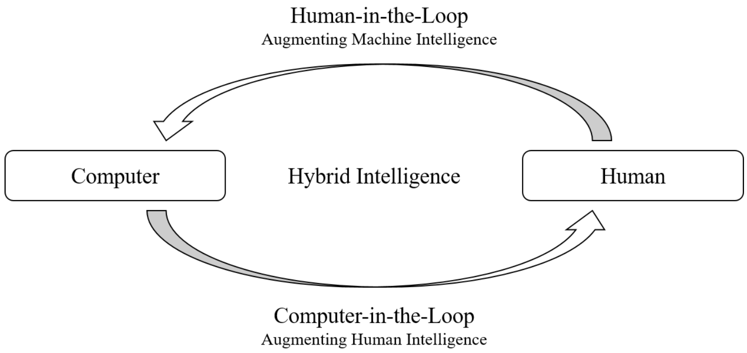

Figure 6. Generic Collaborative Learning Process of Hybrid Intelligence

By comparing the processes on this higher level, we could identify and thus inductively derive three patterns: decision support, exploration and integration. As for decision support, we further classified three sub-patterns, namely assimilation, exploitation and explanation, in terms of different user objectives. In the following, we present the three learning process patterns including their sub-patterns toward answering Q1 and include for each concept measuring instruments as well as measures for learning effects provided by the corresponding literature toward answering Q2.

Table 2. Concepts for Collaborative Learning Processes in Hybrid Intelligence Systems

Learning Process Patterns in Hybrid Intelligence		**Publications**	**Total (33)**	**Ratio (100 %)**
Decision Support	Assimilation	Amershi et al. 2014; Berger et al. 2021; Erbe 2001; Hu et al. 2019; Luong et al. 2019	5	15 %
	Exploitation	Cai et al. 2019; Hanika et al. 2019; Lees et al. 2011; Lindvall et al. 2021; Mullins et al. 2020; Paschen et al. 2020; Pereira and Paulovich 2020; Rundo et al. 2020; Steenwinckel et al. 2021; Traumer et al. 2017; Verdenius 1995; Zeni et al. 2019	12	36 %
	Explanation	Dellermann et al. 2019b; Holzinger et al. 2021; Hudec et al. 2021; Hun Lee et al. 2021; Kiefer 2022; Kulesza et al. 2015; Liu et al. 2014; Schneider and Handali 2019	8	24 %
Exploration		McCamish et al. 2017; Oliveira et al. 2020; Salam et al. 2019; Smith et al. 2018	4	12 %
Integration		Bassano et al. 2020; Dellermann et al. 2017; Gavriushenko et al. 2020; Hekler et al. 2019	4	12 %

11.4.1 Decision Support – Assimilation

For the collaborative learning process through assimilation, the human starts with an initial decision or input toward the computer. Depending on this input, the computer predicts an output toward the human. The human can then compare the expected output with the actual output. This comparison can enable the human, on the one hand, to learn and adjust the input accordingly, or, on the other hand, to adjust the predicted output in order for the computer to learn (Amershi et al. 2014; Berger et al. 2021; Erbe 2001; Hu et al. 2019; Luong et al. 2019) (Figure 7).

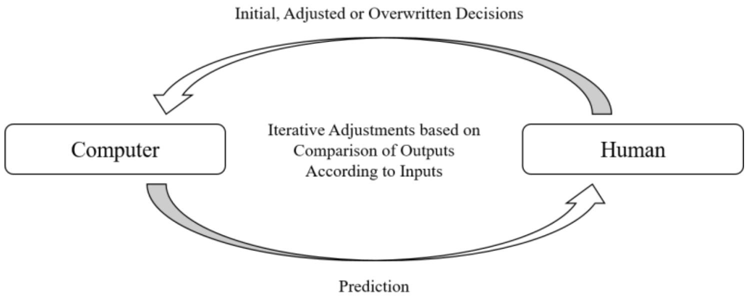

Figure 7. Collaborative Learning Process with Assimilation

The assimilation process enables humans to iteratively and interactively evaluate their inputs and make adjustments (Amershi et al. 2014; Berger et al. 2021; Erbe 2001; Hu et al. 2019). To make the learning and assimilation on both sides even more efficient, Luong et al. (2019) add accuracy rates to the computer's predictions and confidence rates to the human's decisions. They found out that a more reliable computer as well as a more experienced human will eventually lead to a better performance (Luong et al. 2019).

As an example, the collaboration process presented by Luong et al. (2019) about reviewing loan applications in organizational decision-making starts with an initial decision of the human. After that, the computer provides a prediction with the according accuracy rate. Based on this the human can adapt and finalize the decision.

Measurement: in order to evaluate human performance, Hu et al. (2019) suggest creating a reference control model using optimal control theory. This can be utilized to compare the human operator with the theoretical optimal operator (Hu et al. 2019). What is more, one can measure the usage of the computer, and with this the acceptance of the computer in the joint work by counting the revised decisions based on computer predictions (Luong et al. 2019). At last, Berger et al. (2021) refer to the judge-advisor paradigm (Sniezek and Timothy 1995) for measuring an advisor's influence on the decision maker and with this the reliance on the advice or weight of advice.

11.4.2 Decision Support – Exploitation

The exploitation process extends the common use of traditional decision support systems by iteratively exploiting knowledge of the computer and the human. Therefore, the human both uses and benefits from the computer's decision support, but also is asked to validate the computer's knowledge and process it further (Cai et al. 2019; Hanika et al. 2019; Lees et al. 2011; Lindvall et al. 2021; Mullins et al. 2020; Paschen et al. 2020; Pereira and Paulovich 2020; Rundo et al. 2020; Steenwinckel et al. 2021; Traumer et al. 2017; Verdenius 1995; Zeni et al. 2019) (Figure 8).

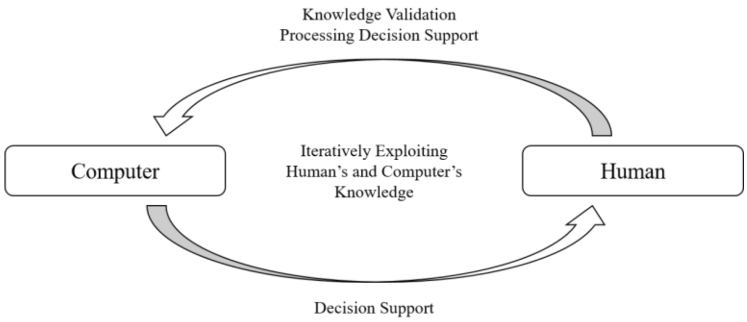

Figure 8. Collaborative Learning Process with Exploitation

The exploitation of the computer's knowledge is very similar in all publications. The computer is usually provided with an algorithm to prepare and support human decision-making (Lees et al. 2011; Mullins et al. 2020; Traumer et al. 2017) leading to better informed humans (Hanika et al. 2019). Paschen et al. (2020) and Rundo et al. (2020) demonstrate computers relying on advanced artificial intelligent tools for augmenting human intelligence with symbolic methods and reasoning engines (Rundo et al. 2020) or analyzing big data (Paschen et al. 2020). What closes the cycle of the iterative exploitation is the learning of the computer. Typically, the human is asked to validate the computer's output or provide certain information, e.g. by labeling data (Hanika et al. 2019; Traumer et al. 2017), marking positives and negatives (Lindvall et al. 2021; Pereira and Paulovich 2020), identifying, confirming, merging or editing information (Paschen et al. 2020; Steenwinckel et al. 2021), evidence-based reasoning (Rundo et al. 2020) or by involving an oracle (Zeni et al. 2019). In addition, the computer may learn from the way the human processes the decision support given, e.g. translating the information into knowledge for communication (Mullins et al. 2020; Paschen et al. 2020), refining or constraining the results provided (Cai et al. 2019; Verdenius 1995) or deriving implications for courses of action (Paschen et al. 2020). Lees et al. (2011) even introduce a higher-level feedback loop to directly improve the performance of the computer's algorithm (Lees et al. 2011).

As an example, in the B2B sales funnel scenario presented by Paschen et al. (2020) the role of the computer is to analyze big data to provide useful information to the human user. The human's task is then to process this information into knowledge and derive implications and next action steps.

Measurement: Lees et al. (2011) introduce a performance monitoring module measuring to which extent the user exploits the computer. First, it shows the number of times a human did not use the computer's result. Second, it logs the decision history for further analysis.

Third and last, it provides information on the distance of the human's decision to the computer's result revealing possible user problems and usage patterns (Lees et al. 2011). Similar to the last point, Zeni et al. (2019) suggest to compare the computer's and the human's knowledge, also considering the confidence of both (Zeni et al. 2019).

11.4.3 Decision Support – Explanation

By explaining decisions or recommendations provided, the computer can actually teach humans. This is especially the case for humans with less experience in certain domains, i.e. novices. On the other hand, the computer itself learns through implicit or explicit data collection from the humans, who can give feedback, decide, adjust the explanation, or give explanations themselves for corrections (Dellermann et al. 2019b; Holzinger et al. 2021; Hudec et al. 2021; Hun Lee et al. 2021; Kiefer 2022; Kulesza et al. 2015; Liu et al. 2014; Schneider and Handali 2019) (Figure 9).

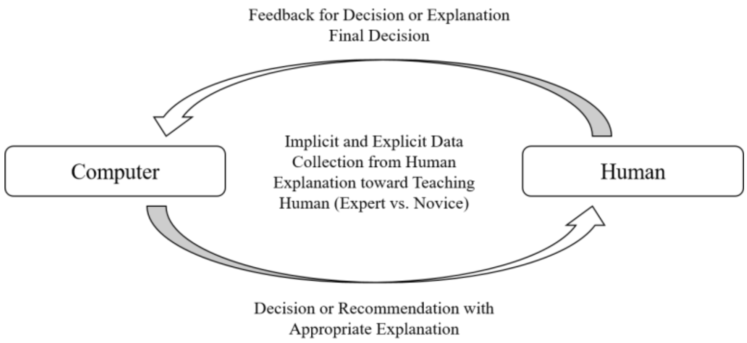

Figure 9. Collaborative Learning Process with Explanation

The data collection from the computer can either be implicit, e.g. logging the humans' workflow and interaction (Dellermann et al. 2019b; Liu et al. 2014) or explicit, e.g. feedback or adjustments from humans (Dellermann et al. 2019b; Hudec et al. 2021; Hun Lee et al. 2021; Kulesza et al. 2015; Schneider and Handali 2019). Kiefer (2022) even introduces an algorithm for semantic interrogation enabling the human to ask the computer about its learning. Although Dellermann et al. (2019b) and Liu et al. (2014) do not explicitly address explanations in their research, they emphasize the possibility to transfer knowledge from experts to novices through the collaborative learning process of human and computer (Dellermann et al. 2019b; Liu et al. 2014).

As an example, Kulesza et al. (2015) provide principles for the concept of Explanatory Debugging, which they then instantiate in a text classification prototype. It enables the exchange of explanations between human user and a machine learning system toward explainability and correctability. For instance, the system provides explanations about its reasoning for its predictions, while the user explains corrections to the system.

Measurement: for one, the overall outcome is a relevant factor for measuring Hybrid Intelligence (Dellermann et al. 2019b). Therefore, Schneider and Handali (2019) suggest to compare the Hybrid Intelligence system with a traditional black-box system evaluating humans' understanding of the system as well as efficiency in making corrections (Kulesza et al. 2015). Schneider and Handali (2019) also provide several objective measures for interpretability, e.g. human's response time or number of mistakes identified, as well as subjective measures, e.g. human's rating on plausibility, usefulness, surprisingness, non-triviality, trustworthiness or overall satisfaction. Additionally, they measure trust and confidence by comparing the computer's and the human's results or through the ratio of the human's decisions after receiving an explanation (Schneider and Handali 2019). At last, (Holzinger et al. 2021) propose the measure of causability for causal understanding achieved by an explanation, i.e. the quality of an explanation. It is not to be confused with explainability, which ensures transparency and traceability (Holzinger et al. 2021).

11.4.4 Exploration

By exploring and identifying new insights, humans and computers can mutually learn from each other by using various inputs or outputs. Therefore, the computer does not only show results, which it is certain about, but also results, which it is uncertain about and which the humans might not expect. For one thing, the human can discover and learn about new insights, and for another thing, the human gives implicit or explicit feedback on the returned results, which is then incorporated by the computer. By doing so, the computer can discover and learn about the relevance of uncertain results and adapt for further iterations (McCamish et al. 2017; Oliveira et al. 2020; Salam et al. 2019; Smith et al. 2018) (Figure 10).

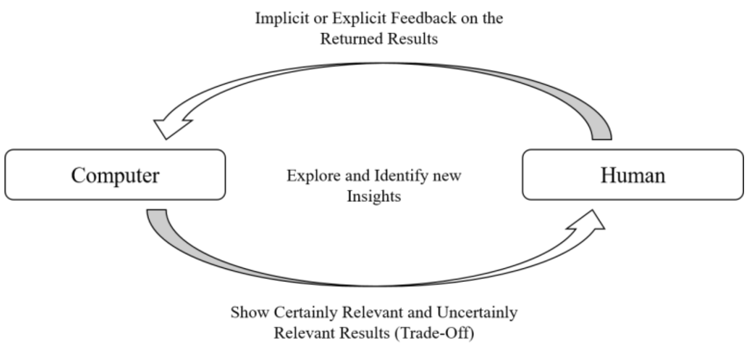

Figure 10. Collaborative Learning Process with Exploration

Contrary to the exploitation process, the computer does not stick to its algorithmic function in the exploration process, but it sometimes deviates from this function or randomizes the results to discover knowledge toward long-term learning (McCamish et al. 2017; Oliveira et al. 2020; Salam et al. 2019). The feedback for system learning is then gathered either implicitly, e.g. from click-through information (McCamish et al. 2017), experiential knowledge (Oliveira et al. 2020), or explicitly, e.g. by accepting a result (Salam et al. 2019), removing outliers (Smith et al. 2018) or labeling data (McCamish et al. 2017). What is more, generating new various insights also enables the humans to learn and explore. To support the exploration and learning, the computer might provide summarizations, comparisons or explanations (Smith et al. 2018). At last, another way of human exploring is presented by McCamish (2017): changing the input for the same intent. It was found, that humans are usually exploring the input until they find an effective input for successful exploitation.

As an example, McCamish et al. (2017) consider database querying as a collaboration between a human user and a database system to establish a mutual language. Thus, when querying the database, the system may present other tuples with different scores to collect feedback from the human users.

Measurement: as the human and the computer benefit from many relevant results, McCamish et al. (2017) suggest to measure the reward, i.e. the effectiveness of the results, received by the computer and the human. Therefore, they use the standard effectiveness metric NDCG (Normalized Discounted Cumulative Gain) (McCamish et al. 2017). Apart from that, it is also possible to compare quality and running time of a hybrid learning process with fully automated and fully manual approaches (Salam et al. 2019).

11.4.5 Integration

Integration toward mutual learning differs a little from the other patterns in its core idea of integrating the human’s and the computer’s knowledge. Thus, even though, the computer and human influence each other, integration does not differentiate and separate into a human learning process or a computer learning process (Bassano et al. 2020; Dellermann et al. 2017; Gavriushenko et al. 2020; Hekler et al. 2019) (Figure 11).

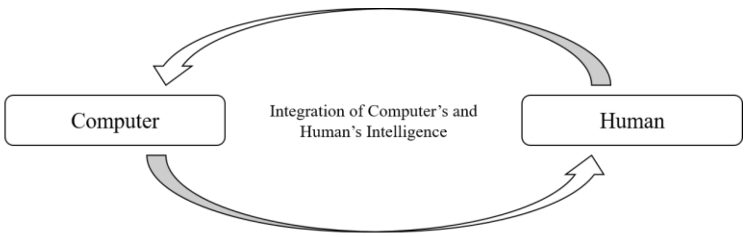

Figure 11. Collaborative Learning Process with Integration

Integration is possible in several ways. For instance, Dellermann et al. (2017) and (Hekler et al. 2019) aggregate the predictions of the human and the computer with underlying weighting algorithms. Another way is to handle humans and computers as inseparable co-creators assisting and learning from each other toward co-creating value (Bassano et al. 2020). Similarly, humans and computers can assist each other in the education process toward a new entity of collective intelligence, which makes the training of both inseparable (Gavriushenko et al. 2020).

As an example, for predicting early stage startup success, Dellermann et al. (2017) aggregate the predictions of the human and the machine to predict the success probability.

Measurement: for evaluating the Hybrid Intelligence system, Dellermann et al. (2017) compare its performance and results with the results provided by only a machine or only a human (Dellermann et al. 2017).

11.5 Discussion

Overall, three patterns including three sub-patterns were identified for collaborative learning processes in Hybrid Intelligence systems involving HITL and CITL (Q1): decision support (assimilation, exploitation, explanation), exploration and integration (Table 3). By augmenting human intelligence through computer intelligence and by augmenting computer intelligence through human intelligence the patterns enable an iterative and interdependent interaction between human and computer. Thereby, humans and computers do not only complement and influence each other with their individual knowledge and advantages, but also learn through the process. We therefore explore the interaction process between a human and a computer and its resulting mutual learning effects. First, for assimilation both learn through small iterative adjustments of inputs and outputs. Second, the exploitation pattern is based on decision support enabling the human and the computer to iteratively leverage each other's knowledge. Third, contrary to exploitation, exploration does not only consider the certainly most relevant results according to a function, but gives the opportunity to gain new insights and discover various inputs and outputs. Fourth, in order to directly teach a human, computers give an explanation for their decision or recommendation, based on collected data from humans. This can enable a knowledge transfer from experts to novices. The fifth and last pattern differs from the rest as the intelligence of the computer and the human is integrated, making them inseparable. Additionally, we identified several ways to measure the learning effects in Hybrid Intelligence systems from the literature separated by pattern (Q2) (Table 3).

Table 3. Overview of the Identified Collaborative Learning Process Patterns and their Recommended Application and Measurement

Pattern		Application		Measurement
		Description	Further Remarks	
Decision Support	Assimilation	Small tightly coupled interactions between human and computer to approximate the decision.	Strength of decision support increases from assimilation to exploitation to explanation.	Reference control model (Hu et al. 2019); usage of computer predictions (Luong et al. 2019); weight of advice (Berger et al. 2021; Sniezek and Timothy 1995).
Decision Support	Exploitation	Deeper interactions between human and computer for elaborating on knowledge in a specifc area toward a decision.	Contrary to exploration due to specificity, but well-combinable.	Decision history and usage patterns (Lees et al. 2011); knowledge differences and confidence (Zeni et al. 2019).

Decision Support	Explanation	Reasoning interactions between human and computer to support the decision with explanations.	Combinable with exploration.	Understanding and efficiency in making corrections (Kulesza et al. 2015); interpretability (Schneider and Handali 2019); causability (Holzinger et al. 2021).
Exploration		Broader interactions between human and computer for discovering knowledge around a specific area.	Contrary to exploitation due to generality, but well-combinable; also combinable with explanation.	Effectiveness of the results (McCamish et al. 2017); quality and running time difference to fully automated and manual approaches (Salam et al. 2019).
Integration		Synergy of human and computer intelligence by regarding them as one entity instead of separating them.	First steps to ensure a balanced decisional power between human and computer.	Performance and results compared with results provided by only a machine or a human (Dellermann et al. 2017).

Regarding the patterns and the measurements derived from the literature review, we find the following aspects interesting and derive implications for future research. To begin with, it turned out that some collaborative learning processes in Hybrid Intelligence systems and their attached research work may have a focus, which we used to conceptualize, but still could also fit into other patterns. This makes it possible for Hybrid Intelligence systems to combine different learning processes leveraging the advantages of not only one, but several patterns (Table 3 – exploitation, explanation and exploration). For instance, humans might start with exploring a system until they find an effective way to exploit the system (McCamish et al. 2017). Another example is combining exploration and explanation. Therefore, the system might present several results for exploration to the human and support the decision with explanations of the different results (Smith et al. 2018). Thus, the patterns do not exclude each other, but could complement each other, if accurately implemented. In this regard, we call for further research finding suitable combinations of learning processes for building and improving Hybrid Intelligence systems.

Next, we found that for most of the identified processes, the human makes the final decision, while the computer has the role of giving recommendations, predictions or adaptable decisions. Still, there is existing research, which considers the computer as teammate (Gavriushenko et al. 2020) or aggregates the results of both human and computer (Dellermann et al. 2017) (Table 3 – integration). Also, when talking about an oracle for solving discrepancy between a human's and a computer's result, this oracle does not have to be human, but can also be a trusted learning algorithm or a social opinion (Zeni et al. 2019). Here, it would be interesting for further research to go deeper into examining the relationship of computer and human, e.g. who gets the final word, and its influence on the collaborative learning process and effects.

What is more, while reviewing the literature, it turned out that there is a remaining need for research on evaluating Hybrid Intelligence systems. Though some researchers provide some successful measurements for human learning as well as machine learning, there is no consistent and broadly established set of measures across literature respectively. However, there are some measurement instruments, which appear recurringly across the identified literature, e.g. number of computer outputs used by the human (Lees et al. 2011; Luong et al. 2019; Schneider and Handali 2019) and comparison of system processes (Hu et al. 2019; Salam et al. 2019). Thus we call for research on establishing an evaluation framework for Hybrid Intelligence systems serving as a reference for future researchers as well as practitioners developing and evaluating Hybrid Intelligence systems. Additionally, we ask for the framework to provide a fundament with definitions for key performance indicators covering both objective and subjective evaluation of human and machine learning. Regarding the specific scenario of splitting humans into experts and novices (Dellermann et al. 2019b; Hu et al. 2019; Liu et al. 2014; Oliveira et al. 2020) we ask for research on how to exactly measure and indicate, whether a human is a novice or an expert in order to track the learning progress of novices in learning processes.

Despite these valuable insights, our contribution does not come without limitations. For one, we limited our search to three relevant databases. A search in other databases could have led to more or different results leading to potentially new or different patterns. The same applies to the formulation of the search string used for the identification of the review literature. However, we followed the guideline of Webster and Watson (2002) and aligned our work with the framework of vom Brocke et al. (2009) in order to rigorously conduct the literature review. Additionally, we invite further research to search other databases and expand our analysis and conceptualization.

11.6 Conclusion and Contribution

Our findings contribute to the research field of Human-Computer-Interaction with a conceptualization of collaborative learning processes in Hybrid Intelligence systems involving HITL and CITL, with focus on computer and human learning. Next to the identified patterns, namely assimilation, exploitation, exploration, explanation and integration, we also identified measurements for evaluating the learning effects according to the specific patterns.

As computers and humans interact interdependently in Hybrid Intelligence systems, their learning processes and effects cannot be regarded separately. Eventually we aim at rising awareness of considering both sides of a learning process within a Hybrid Intelligence system: human learning through CITL and computer learning through HITL. Thus, researchers, who focus on solely one side, can extend their work by considering the effects on the other side of the process, acknowledging the inseparability of human and computer learning. Therefore, our research can serve as a guidance providing researchers with conceptual foundations of collaborative learning processes in Hybrid Intelligence systems.

Furthermore, we raise future research avenues and implications toward combinations of learning patterns, the relationship of computer and human as well as an evaluation framework for Hybrid Intelligence systems. Future research may also expand our analysis and conceptualization with other underlying databases.

11.7 References

Abdel-Karim, B. M., Pfeuffer, N., Rohde, G., and Hinz, O. 2020. "How and What Can Humans Learn from Being in the Loop?" *KI - Künstliche Intelligenz* (34), pp. 199-207.

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. 2014. "Power to the People: The Role of Humans in Interactive Machine Learning," *AI Magazine* (35:4), pp. 105-120.

Bassano, C., Barle, S., Saviano, M. L., Cosimato, S., and Pietronudo, M. C. 2020. "AI technologies & value-creation in luxury goods context," in *Proceedings of the 53rd Hawaii International Conference on System Sciences, Wailea-Makena, Hawaii, USA*, pp. 1618-1627.

Benedikt, L., Joshi, C., Nolan, L., Henstra-Hill, R., Shaw, L., and Hook, S. 2020. "Human-in-the-Loop in Government: A Case Study," in *25th International Conference on Intelligent User Interfaces (IUI '20), Cagliari, Italy*, pp. 488-497.

Berger, B., Adam, M., Rühr, A., and Benlian, A. 2021. "Watch Me Improve - Algorithm Aversion and Demonstrating the Ability to Learn," *Business & Information Systems Engineering* (63:1), pp. 55-68.

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G. S., Stumpe, M. C., and Terry, M. 2019. "Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making," in *CHI, Glasgow, Scotland Uk*.

Cerf, V. G. 2013. "Augmented Intelligence," *IEEE Internet Computing* (96-97).

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., and Ebel, P. 2019a. "The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems," in *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS), Hawaii, USA*.

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. 2019b. "Hybrid Intelligence," *Business & Information Systems Engineering* (61), pp. 637-643.

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., and Leimeister, J. M. 2017. "Finding the Unicorn: Predicting Early Stage Startup Success through a Hybrid Intelligence Method," in *Proceedings of the Thirty Eighth International Conference on Information Systems, South Korea*.

Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K. E., and Dugan, C. 2019. “Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment,” in *24th International Conference on Intelligent User Interfaces (IUI ‘19), Marina del Rey, CA, USA*.

Erbe, H.-H. 2001. “Multi-Agent Systems Assisting Operators to Mastering Automation Systems,” in *8th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems (HMS 2001), Kassel, Germany*.

Gaurav, T. 2016. *On Interactive Machine Learning*. https://www.trivedigaurav.com/blog/on-interactive-machine-learning/. Accessed 21 October 2020.

Gavriushenko, M., Kaikova, O., and Terziyan, V. 2020. “Bridging human and machine learning for the needs of collective intelligence development,” *Procedia Manufacturing* (42), pp. 302-306.

Hanika, T., Herde, M., Kuhn, J., Leimeister, J. M., Lukowicz, P., Oeste-Reiß, S., Schmidt, A., Sick, B., Stumme, G., Tomforde, S., and Zweig, K. A. 2019. “Collaborative Interactive Learning - A clarification of terms and a differentiation from other research fields,”

Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., Maron, R. C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-Letz, T., Izar, B., Kalle, C. von, Fröhling, S., and Brinker, T. J. 2019. “Superior skin cancer classification by the combination of human and artificial intelligence,” *European Journal of Cancer* (120), pp. 114-121.

Holzinger, A. 2016a. “Interactive Machine Learning (iML),” *Informatik Spektrum* (39:1), pp. 64-68.

Holzinger, A. 2016b. “Interactive machine learning for health informatics: when do we need the human-in-the-loop?” *Brain Informatics* (3:119-131).

Holzinger, A., Malle, B., Saranti, A., and Pfeifer, B. 2021. “Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI,” *Information Fusion* (71), pp. 28-37.

Holzinger, A., Plass, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. 2016. “Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to Solve the Traveling Salesman Problem with the Human-in-the-Loop Approach,” in *IFIP International Cross Domain Conference and Workshop, Salzburg, Austria*, pp. 81-95.

Holzinger, A., Plass, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. 2017. *A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop*.

Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. 2019. “Interactive machine learning: experimental evidence for the human in the algorithmic loop: A case study on Ant Colony Optimization,” *Applied Intelligence* (49), pp. 2401-2414.

Hu, W.-L., Rivetta, C., MacDonald, E., and Chassin, D. P. 2019. "Optimal Operator Training Reference Models for Human-in-the-loop Systems," in *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS), Hawaii, USA*, pp. 3524-3531.

Hudec, M., Mináriková, E., Mesiar, R., Saranti, A., and Holzinger, A. 2021. "Classification by ordinal sums of conjunctive and disjunctive functions for explainable AI and interpretable machine learning solutions," *Knowledge-Based Systems* (220).

Hun Lee, M., Bernardino, A., and Bermúdez i Badia, S. 2021. "A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment," in *CHI Conference on Human Factors in Computing Systems (CHI '21), Yokohama, Japan*.

Kartoun, U. 2017. "Text Nailing: An Efficient Human-in-the-Loop Text-Processing Method," *Interactions*, pp. 45-49.

Kiefer, S. 2022. "CaSE: Explaining Text Classifications by Fusion of Local Surrogate Explanation Models with Contextual and Semantic Knowledge," *Information Fusion* (77), pp. 184-195.

Krening, S., and Feigh, K. M. 2018. "Interaction Algorithm Effect on Human Experience with Reinforcement Learning," *ACM Transactions on Human-Robot Interaction* (7:2).

Kulesza, T., Burnett, M., Wong, W.-K., and Stumpf, S. 2015. "Principles of Explanatory Debugging to Personalize Interactive Machine Learning," in *IUI, Atlanta, GA, USA*, 126-137.

Lee, T. Y., Smith, A., Seppi, K., Elmqvist, N., Boyd-Graber, J., and Findlater, L. 2017. "The Human Touch: How Non-expert Users Perceive, Interpret, and Fix Topic Models," *International Journal of Human-Computer Studies* (105), pp. 28-42.

Lees, M., Evans, R., and Mareels, I. 2011. "Autonomous Performance Monitoring of an On-Line Real-Time Decision Support System," in *Proceedings of the 18th World Congress The International Federation of Automatic Control, Milano, Italy*, pp. 13080-13085.

Lindvall, M., Lundström, C., and Löwgren, J. 2021. "Rapid Assisted Visual Search," in *IUI, College Station, TX, USA*.

Liu, J., Wilson, A., and Gunning, D. 2014. "Workflow-based Human-in-the-Loop Data Analytics," in *Proceedings of the 2014 Workshop on Human Centered Big Data Research (HCBDR '14)*, Raleigh, NC, USA.

Lloyd, E., Huang, S., and Tognoli, E. 2017. "Improving Human-in-the-Loop Adaptive Systems Using Brain-Computer Interaction," in *IEEE/ACM 12th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), Buenos Aires, Argentina*.

Luong, A., Kumar, N., and Lang, K. R. 2019. "Man-Machine Collaboration in Organizational Decision-Making: An Experimental Study Using Loan Application Evaluations," in *Proceedings of the 2019 Pre-ICIS SIGDSA Symposium, Munich*.

Martínez, M. A. M., Nadj, M., and Maedche, A. 2019. “Towards An Integrative Theoretical Framework Of Interactive Machine Learning Systems,” in *Proceedings of the 27th European Conference on Information Systems, Stockholm-Uppsala, Sweden*.

McCamish, B., Termehchy, A., and Touri, B. 2017. “A Game-theoretic Approach to Data Interaction: A Progress Report,” in *Proceedings of HILDA’17, Chicago, IL, USA*.

Mullins, R., Fouse, A., Ganberg, G., and Schurr, N. 2020. “Practice Makes Perfect: Lesson Learned from Five Years of Trial and Error Building Context-Aware Systems,” in *Proceedings of the 53rd Hawaii International Conference on System Sciences, Wailea-Makena, Hawaii, USA*, pp. 279-288.

Oliveira, C. S. de, Sanin, C., and Szczerbicki 2020. “Human Feedback and Knowledge Discovery: Towards Cognitive Systems Optimization,” in *24rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, pp. 3093-3102.

Paschen, J., Wilson, M., and Ferreira, J. J. 2020. “Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel,” *Business Horizons*.

Peeters, M. M. M., van Diggelen, J., van der Bosch, K., Bronkhorst, A., Neerincx, M. A., Schraagen, J. M., and Raaijmakers, S. 2020. “Hybrid Collective Intelligence in a Human-AI Society,” *AI & Society*.

Pereira, M. M., and Paulovich, F. V. 2020. “RankViz: A visualization framework to assist interpretation of Learning to Rank algorithms,” *Computers & Graphics* (93), pp. 25-38.

Reani, M., Peek, N., and Jay, C. 2018. “How Different Visualizations Affect Human Reasoning About Uncertainty: An Analysis of Visual Behaviour,” *Computers in Human Behavior*.

Rundo, L., Pirrone, R., Vitabile, S., Sala, E., and Gambino, O. 2020. “Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine,” *Journal of Biomedical Informatics*.

Rzepka, C., and Berger, B. 2018. “User Interaction with AI-enabled Systems: A Systematic Review of IS Research,” in *Thirty Ninth International Conference on Information Systems, San Francisco*.

Salam, M. A., Koone, M. E., Thirumuruganathan, Das, G., and Roy, S. B. 2019. “A Human-in-the-loop Attribute Design Framework for Classification,” in *Proceedings of the 2019 World Wide Web Conference (WWW ‘19), San Francisco, CA, USA*.

Schneider, J. 2020. *Humans learn too: Better Human-AI Interaction using Optimized Human Inputs*.

Schneider, J., and Handali, J. P. 2019. “Personalized Explanation For Machine Learning: A Conceptualization,” in *Proceedings of the 27th European Conference on Information Systems, Stockholm-Uppsala, Sweden*.

Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. 2020.

"Machines as teammates: A research agenda on AI in team collaboration," *Information & Management* (57).

Shneiderman, B. 2020. "Human-Centered Artificial Intelligence: Three Fresh Ideas," *AIS Transactions on Human-Computer Interaction* (12:3).

Smith, A., Kumar, V., Boyd-Graber, J., Seppi, K., and Findlater, L. 2018. "Closing the Loop: User-Centered Design and Evaluation of a Human-in-the-Loop Topic Modeling System," in *IUI, Tokyo, Japan*.

Smith-Renner, A., Kumar, V., Boyd-Graber, J., Seppi, K., and Findlater, L. "Digging into User Control: Perceptions of Adherence and Instability in Transparent Models," in *IUI '20, Cagliari, Italy*.

Sniezek, J. A., and Timothy, B. 1995. "Cueing and Cognitive Conflict in Judge-Advisor Decision Making," *Organizational Behavior and Human Decision Processes* (62:2), pp. 159-174.

Steenwinckel, B., Paepe, D. de, Vanden Hautte, S., Heyvaert, P., Bentefrit, M., Moens, P., Dimou, A., van den Bossche, B., Turck, F. de, van Hoecke, S., and Ongenae, F. 2021. "FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning," *Future Generation Computer Systems* (116), pp. 30-48.

Tamagnini, P., Krause, J., Dasgupta, A., and Bertini, E. 2017. "Interpreting Black-Box Classifiers Using Instance-Level Visual Explanations," in *HILDA '17, Chicago, IL, USA*.

Traumer, F., Oeste-Reiß, S., and Leimeister, J. M. 2017. "Towards a Future Reallocation of Work between Humans and Machines - Taxonomy of Tasks and Interaction Types in the Context of Machine Learning," in *Proceedings of the Thirty Eighth International Conference on Information Systems, South Korea*.

van der El, K., Barendswaard, S., Pool, D. M., and Mulder, M. 2016. "Effects of Preview Time on Human Control Behavior in Rate Tracking Tasks," in *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*, pp. 108-113.

Verdenius, F. 1995. "Managing Product Inherent Variance During Treatment," in *2nd IFAC/IFIP/EurAgEng Workshop on Artificial Intelligence in Agriculture, Wageningen, The Netherlands*.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in *17th European Conference on Information Systems (ECIS), Verona Italy*.

Wang, G. 2019. *Humans in the Loop: The Design of Interactive AI Systems*. https://hai.stanford.edu/blog/humans-loop-design-interactive-ai-systems. Accessed 13th November 2020.

Webster, J., and Watson, R. T. 2002. "Analyzing The Past To Prepare For The Future: Writing A Literature Review," *MIS Quarterly* (26:2), pp. 13-23.

Yimam, S. M., and Biemann, C. 2018. “Demonstrating PAR4SEM - A Semantic Writing Aid with Adaptive Paraphrasing,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (System Demonstrations), Brussels, Belgium*, pp. 48-53.

Yimam, S. M., Biemann, C., Majnaric, L., Sabanovic, S., and Holzinger, A. 2016. “An adaptive approach for biomedical entity and relation recognition,” *Brain Informatics*.

Zeni, M., Zhang, W., Bignotti, E., Passerni, A., and Giunchiglia, F. 2019. “Fixing Mislabeling by Human Annotators Leveraging Conflict Resolution and Prior Knowledge,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (3:1), pp. 1-23.

12 Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service

Poser, M., Wiethof, C., Banerjee, D., Subramanian, V. S., Paucar, R., and Bittner, E. A. C. 2022a. “Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service,” in *The Transdisciplinary Reach of Design Science Research. DESRIST 2022. Lecture Notes in Computer Science, vol 13229*, A. Drechsler, A. Gerber and A. Hevner (eds.), Springer, Cham, pp. 142-153 (doi: 10.1007/978-3-031-06516-3_11).

Abstract. Customers desire convenient, fast, and personalized service encounters. Hence, service companies deploy self-service technology for online customer service. However, as solutions based on Artificial Intelligence cannot reliably answer the full range of requests and the demands on service employees (SEs) in live chat interaction are high, Hybrid Intelligence Systems (HIS) provide great potential to overcome current pitfalls by combining the complementary strengths of artificial and human intelligence. To ensure optimal performance of this socio-technical ensemble, human-centered design approaches are needed to realize real-time augmentation of decision-making in chat-based service encounters. Following a Design Science Research approach, we generate theory-based design principles (DPs) and implement them in a web-based HIS prototype. We contribute to Hybrid Intelligence research with results showing that the DPs enable task mastery and decision efficiency and provide avenues for future research.

Keywords: Hybrid Intelligence System, Real-time Decision, Customer Service.

12.1 Introduction

Striving for operational efficiency, companies across various industries deploy automation technology enabled by Artificial Intelligence (AI) to process the ever-increasing number of requests in customer service [1, 2]. This development is expected to culminate by 2025 with 95% of all customer encounters being processed by AI [3]. Thereby, companies can increase their availability to customers, especially via online customer service (OCS) channels [4]. However, so far, full automation of online service interactions is not feasible, as narrow AI is not capable of handling all types of customer requests. Hence, strategies are needed to process the full range of customer requests while avoiding overload of service employees (SEs). In this context, research and practice postulate augmentation approaches relying on close collaboration between humans and AI to execute tasks [1, 5]. For real-time service encounters in OCS, the combination of AI's capabilities to rapidly process textual input and provide suitable decision suggestions [6] with SEs' ability to understand semantically complex content and handle unforeseen situations, can lead to effective customer request handling with increased decision-making efficiency. This augmentation approach can serve to meet customers' growing demand for personalized service encounters via text-based channels [7, 8]. In addition, real-time decision augmentation, e.g., displaying suitable information, can help SEs to rapidly process requests with increasing variability in content [9, 10].

In organizational contexts, the focal concept for augmentation strategies is Hybrid Intelligence (HI), which proposes the integration of the complementary strengths of humans and AI in a Hybrid Intelligence System (HIS) for joint task execution involving hybrid decision-making and hybrid learning [11]. To leverage associated potentials of a HIS, human-computer interaction (HCI) needs to be designed concerning suitable input and output formats while meeting human needs for task mastery [12, 13]. However, so far, socio-technical approaches to design the collaboration between AI and humans for hybrid decision-making are under-researched [14, 15]. Thus, human-centered design approaches for AI are needed for the decision-making augmentation of text-based, real-time service encounters in HIS enabling optimized task performance and hybrid learning [12]. To address these knowledge gaps, we adopt the Self-Determination Theory (SDT) to select suitable psychological constructs, ensuring the fulfillment of SEs' needs. Accordingly, we pursue the following research question: *How should a HIS be designed in a human-centered way to augment real-time decision-making for online customer service encounters?* The goal is to enable augmentation in a HIS to sustain SEs' task mastery, efficient decision-making in service encounters and simultaneously meet the requirements for hybrid learning. With this study, we present the second cycle of a larger design science research (DSR) project with the following structure. First, we present the conceptual background. Second, we outline the research approach by describing the cycles and steps of the DSR project. Third, the derived meta-requirements (MRs) and design principles (DPs) are presented and the instantiation illustrated. Last, we present evaluation results followed by a discussion and conclusion.

12.2 Conceptual Background

OCS constitutes a pervasive form to deliver intangible services mediated via technology [2]. To meet customer needs, service is directed toward people or objects [16]. This service is knowledge-intensive, as SEs need to handle an increasing plethora of diverse content from explicit (e.g., data) to meta-knowledge (e.g., advice) to make multiple decisions during request processing [9]. In OCS, AI can enable flexibility in the external (frontstage) and support in the internal (backend) environment to deliver service [17]. However, the automation of frontstage encounters reduces the success-generating characteristics of social presence and personalization [8, 18]. To overcome this tendency, AI-enabled agents are designed in a human-like fashion to handle repetitive, simple requests via natural language interaction [19]. Nevertheless, these AI solutions have yet to create satisfactory customer experiences for complex, emotional requests. To achieve improved organizational and individual outcomes, the competencies of AI and SEs are increasingly integrated [6, 20]. In this context, the concept of HI is adopted to combine the complementary strengths of AI and humans [11] involving augmentation and hybrid learning leading to better results than each of the entities could reach alone [21]. For service encounters, [2] propose the augmentation of SEs invisibly to the customer during real-time interaction, to leverage advantageous conditions for service co-creation with high synchrony of communication as well as personal support [8, 22]. For this augmentation scenario, high demands in the form of instant knowledge retrieval for dynamic decision situations and emotion work should be met [9, 23]. Therefore, AI and SE can take over different roles: AI can provide analytical insights into the customers' requests (e.g., solution proposal) and the SE contributes intuition by contextualizing this information and leading an empathic interaction with a customer [4].

To ensure the success of HIS, conditions for a high degree of SEs' task mastery should be established during customer interaction. Thus, according to the Self-Determination Theory (SDT), augmentation should fulfill human desires for **autonomy**, **competence**, and **relatedness** [24]. SEs should experience the feeling of control over their behavior and make decisions independent of external conditions, as **autonomy** promotes the intensity of post-adoption usage behavior, engagement, and satisfaction with information systems (IS) [25, 26]. In addition, SEs should be able to actively interact with the environment to achieve desired results. By experiencing this **competence** using IS, SEs' self-efficacy could be elevated and decision efficiency increased [27]. Moreover, building a relationship (**relatedness**) with IS due to their social characteristics could influence SEs' perceived usefulness of and intention to reuse the technology [28, 29]. As the consideration of human psychological demands for the design of HIS is scarce, we utilize SDT to select suitable theories that help to meet the three basic needs of SEs in OCS. To promote SEs' autonomy and competence in dynamic customer interactions with a variety of interdependent decisions [9], we adopt the Dynamic Decision Theory (DDT) to support decision-making strategies [30]. Regarding Cognitive Load Theory (CLT) [31, 32], we integrate insights on the nature of information presentation, as decision suggestions should be designed considering their load on SEs' working memory due to intrinsic, extraneous, and germane factors. Following Advice Response Theory (ART) [33], the characteristics of advice have an impact on perceived quality. Therefore, to influence competence, the aspects of efficacy, and feasibility, and absence of limitations are considered for decision suggestions. To establish relatedness in a HIS, we consider Social Response Theory (SRT) [34], which states that the use of social cues in IS has relationship-enhancing effects.

12.3 Research Approach

To establish a human-centered design of HIS for organizational augmentation endeavors, we conduct a multicyclic DSR project. By adopting the interior mode of DSR, we (1) define and evaluate prescriptive design knowledge to "construct a HCI artifact for a given problem space" [35, p. 4] and (2) present a designed HIS artifact [36]. To ensure research rigor, we structure our project by applying the process model of [37] (see Fig. 1).

DSR Research Cycle	**Cycle one:** Hybrid collaborative learning	**Cycle two:** Real-time AI-based decision-making augmentation
(1) Awareness of Problem	Integrate human and artificial intelligence in online customer service	Refinement and extension of problem relevance
(2) Suggestion	Derivation of MRs for hybrid collaborative learning	Derivation of MRs based on kernel theories for real-time decision-making augmentation
(3) Development	Definition of DPs and instantiation in web-based prototype	Extension of DPs and instantiation in full-featured web-based artifact
(4) Evaluation	Artificial evaluation (wizard-of-oz) of design with mixed-method approach	Semi naturalistic evaluation of design with mixed-method approach
(5) Conclusion	Codification of design knowledge as contribution to body of knowledge [currently under review]	Report, embed and contribute design knowledge about artifact's construction and effects

Fig. 1. DSR approach based on [37] with research activities

In two design cycles, we incrementally identify MRs as goal and boundary descriptions of an artifact and derive DPs providing prescriptive statements [38–40]. To ensure validity in addressing the identified problem, we iteratively instantiate and evaluate the design of our HIS artifact in an organization that specializes in selling traineeships and projects abroad to customers. To address this real-world use case, the HIS is supposed to augment the processing of customer questions and identification of their interests (where, when, what) and the recommendation of suitable projects. To do so, in the first cycle [41], we derived theory- and practice-based MRs to define initial DPs for reciprocal augmentation through hybrid collaborative learning. This mutual learning scenario improves the performance of AI by SE experts as well as expands novice SEs' knowledge by AI. As a proof-of-concept, the tentative DPs were implemented in a web-based prototype with a user interface (UI). By conducting a wizard-of-oz study, the instantiated design and expected learning effects for novice SEs could be demonstrated. In the second cycle, covered in this paper, the design is extended and integrated with aspects for real-time decision-making augmentation to fully address the problem of this DSR project. In **(1) Awareness of Problem** (see Sections 1 and 2), we reassessed and elaborated on the problem relevance and need for a solution that integrates hybrid learning and real-time decision augmentation. For **(2) Suggestion**, MRs for real-time augmentation for decision-making are derived based on kernel theories (see Section 4.1) [36]. In **(3) Development**, DPs and matching design features (DFs) are determined to construct a full-featured AI-based HIS prototype (see Section 4.2) as an expository instantiation. For **(4) Evaluation** (see Section 5), following the risk and efficacy strategy [42], the prototype is implemented to conduct an online field study with 18 SEs (ten male, eight female) from the described organization. The study follows a standardized procedure: (1) the setting and prototype are presented; (2) participants use the artifact to counsel a customer while sharing their screen; (3) a semi-structured interview is conducted. As the customers are simulated by the research team, the evaluation is semi-naturalistic. By using three prepared customer profiles with scripts comprising question-and-answer variations, originality of interactions is ensured. To evaluate the designed artifact in terms of its applicability, feasibility, and effect on users, a multi-method approach is applied. The qualitative interview is structured with questions about demographic data, decision-making, trust in and satisfaction with the prototype, and changed task characteristics. In addition, quantitative measures of usage behavior were obtained from screen recordings (e.g., frequency of used functionalities). To analyze the rich data, a qualitative content analysis of the interview transcripts according to [43] is conducted, and descriptive statistical methods are applied for the assessment of the quantitative usage data.

12.4 Design and Development

12.4.1 Theory-derived Meta Requirements

Autonomy and competence. Following DDT [30], SEs apply strategies to make interdependent and real-time decisions in response to dynamic customer interactions [44]. Under time pressure, individuals make decisions by comparing information of options based on assigned values to identify an alternative with the greatest utility [45, 46]. Therefore, multiple suggestions should be proposed (*MR1*), presented in sequence allowing SEs to view alternating combinations (*MR2*) with relevant utility information (*MR3*). To promote comparability, suggestions should be displayed in descending order with respect to utility (*MR4*). The AI settings should be adjustable (*MR5*) to sustain autonomy. Besides facilitating decision-making strategies, the nature of information presentation has to be considered, as it affects SEs' processing ability [47, 48]. According to CLT, dynamic decision-making induces a high intrinsic cognitive load in SEs due to the necessity of monitoring the changing customer demands to make punctual decisions [49]. As this task occupies a significant portion of SEs' capacity, a low load of presented information (extraneous cognitive load) is required [32, 45]. By presenting information in a concentrated format, SEs' information comprehension can be improved [50, 51]. Hence, a limited number of suggestions should be displayed (*MR6*) according to the pace of the changing environment (*MR7*) and their effortless utilization facilitated (*MR8*) to avoid cognitive overload. In addition, characteristics of presented information impact decision-making [48]. Following ART, SEs' high rating of advice quality facilitates their decision-making, whereas discrepancies in expected and provided advice quality impede decision support [52]. To establish efficacy, the applicability and effectiveness of advice to solve a problem have to be present [48]. The quality of advice can also be enhanced by its distinctive workability (feasibility) and presentation of limited risks after its enactment (absence of limitation) [52]. Followingly, insights on the effectiveness should be provided by revealing the context-specificity of suggestions (*MR9*). The applicability and workability should be established by presenting explanatory information for suggestions (*MR10*). Reliability of suggestions should be provided to demonstrate the absence of limitations (*MR11*).

Relatedness. Advice-related decisions are also influenced by relational aspects such as respecting the autonomy of the decision-maker [53]. SRT postulates that social attributes promote a sense of social presence in users and have a positive effect on the intention to reuse, enjoyment of using, and self-efficacy in use [28, 29, 54]. Consequently, the appearance of and interaction with the AI should elicit a sense of social presence by mimicking human sociability (*MR12*) to promote the establishment of a relationship.

12.4.2 Design Principles, Design Features, and Instantiation

We present eleven DPs of the type form and function from two design cycles (see Fig 2) [55]. In the **first cycle**, seven DPs were identified for hybrid collaborative learning, which combines the augmentation of both human intelligence through AI and AI through human intelligence [11, 56]. To enable this, the HIS should include customizable settings so that SEs can individually determine whether the AI learns from them (**DP1.1**). Furthermore, the AI should be equipped with a social identity so that SEs perceive it as a collaboration partner (**DP1.2**). As instructional support, the HIS UI should include explanations of how the AI works to increase SEs' understanding of how to use it (**DP1.3**). For hybrid learning, the process and progress of the task should be observable (**DP1.4**) and an opportunity for AI and SE to share knowledge for decisions should be provided (**DP1.5**). To allow AI learning, an option for SEs to use or adapt AI suggestions (**DP1.6**) and the possibility to feedback the AI should be provided (**DP1.7**). In the **second cycle**, four additional DPs were generated to allow real-time decision-making augmentation. Thus, the HIS should provide configurable AI settings and the possibility to easily use suggestions to increase SEs' task mastery (**DP2.1**: *MR5,8*). A manageable number of context-specific suggestions in sync with the dynamic interaction should be displayed to augment SEs' decision-making (**DP2.2**: *MR1,6,7*). To support SEs' strategies for decision making, suggestions should be shown in sequence according to their utility and allow the display of alternating combinations upon request (**DP2.3**: *MR2,3,4,11*). Additional information about suggestions should be viewable so that SEs can verify their applicability (**DP2.4**: *MR9,10,11,12*).

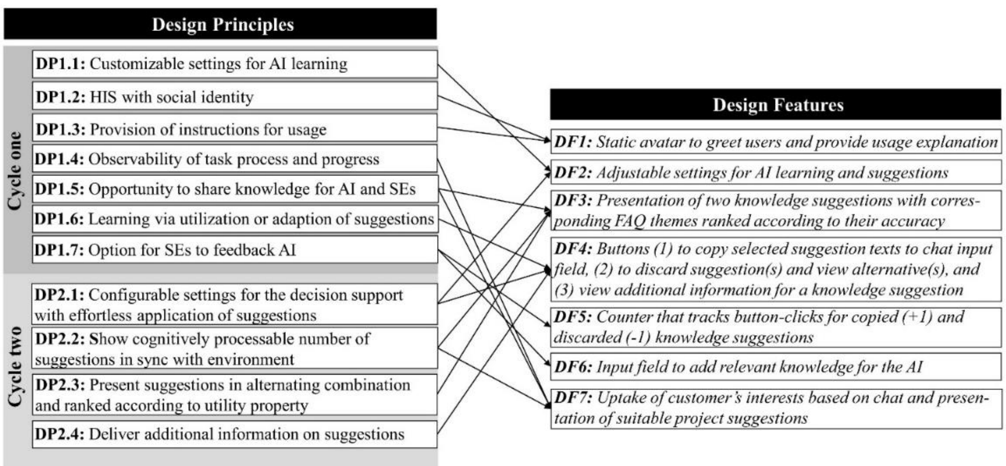

Fig. 2. DPs of cycles one and two with DFs

Based on DFs, we instantiated these DPs in a web-based HIS prototype comprising frontend and backend (see Fig. 3). The web-based frontend was designed with Bootstrap and ReactJS to, inter alia, greet users with an avatar that presents a brief usage explanation (**DF1**). In addition, setting options for AI support and learning behavior are provided (**DF2**). The integrated chat window is based on the open-source framework Rocket.Chat. The backend generates a ranked list of FAQ suggestions based on chat interactions using Dense Passage Retrieval (DPR) technology [57]. The DPR model was pre-trained on the Google Natural Questions dataset by Facebook and further fine-tuned with conversational data from test runs. In the frontend, two FAQ items - including theme and accuracy in percent - with the highest agreement are displayed (**DF3**). The discard-buttons can be used to sequentially display four additional FAQ suggestions with decreasing accuracy. The copy-to-chat buttons insert FAQ text into the input field of the chat window. Detailed information about a respective FAQ can be viewed via the get-more-info button (**DF4**). With a counter, points are added (copy-to-chat) or subtracted (discard), if buttons are clicked (**DF5**). A feedback field allows entering search terms to select and submit a FAQ that matches the interaction (**DF6**). Based on customers' chat messages, exact keyword-based text matching is performed to automatically record interests and suggest suitable projects from a database (**DF7**).

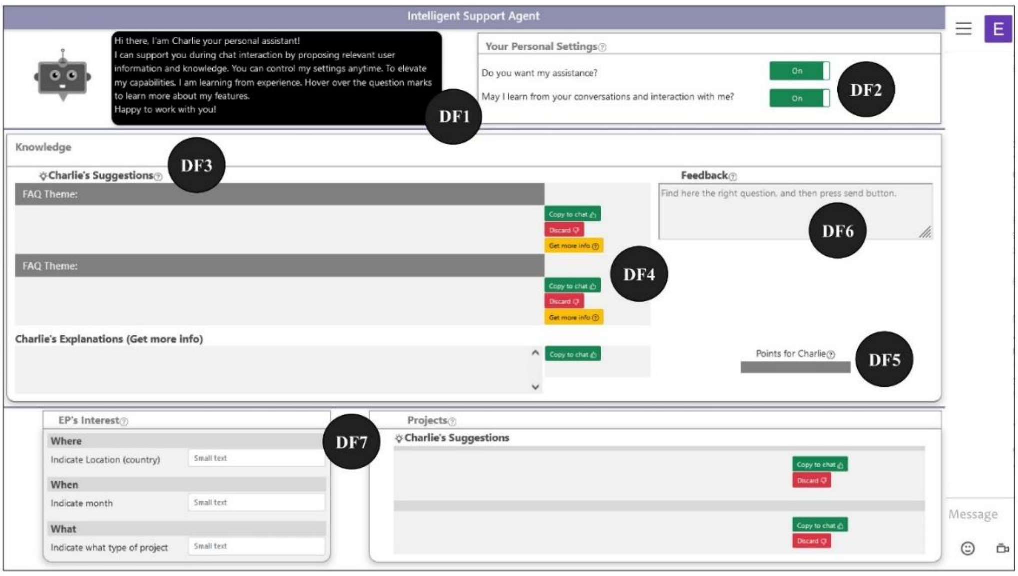

Fig. 3. Screenshot of web-based HIS prototype with DFs

12.5 Evaluation

To evaluate the augmentation with the HIS prototype and its influence on the work task, we conducted interviews with 18 SEs after usage. Additionally, we inspected their usage behavior via screen recordings to supplement the qualitative results. Overall, SEs indicated that they would continue to use the prototype and highlighted that it is particularly helpful for SEs who do not have much experience in counseling customers.

DF1. The feeling of relatedness did not emerge consistently, as some SEs perceived the prototype as a tool and others as a co-customer manager ("*he definitely was co-customer manager because he gave me all the prompts to answer questions*" (SE13)). **DF2.** The analysis of screen recordings revealed that all SEs approved of support by the prototype and 12 consented that their data can be used for AI learning via the settings. **DF3.** During customer interactions, SEs sent on average 16 (*SD*: 5; *Median*: 14) messages during the customer interaction. 17 SEs used the FAQ answer suggestions via the copy-to-chat-button at least three times. On average, SEs edited two (*SD*: 2; *Median*: 2) of the suggested responses in the input field before sending them. The analysis of interview transcripts revealed that SEs were satisfied with the support provided by the prototype, as the provided suggestions appeared promptly, and the interaction was intuitive due to the functionalities and layout of the interface. Regarding customer interaction, SEs felt supported in their decision-making by provided suggestions, as the information allowed them to reassure themselves: "*it is a good thing to know what is going on and what could I answer, what are possibilities and what should I focus on. Also finding out the main point of the question of this customer*" (SE5). The decision- making was further supported by the trustworthiness of suggestions (e.g., SE17: "*in 80% of the times it was the right answer, so for me that is trustworthy*"). Their correctness was reported to be verifiable "*[…] when I pressed the get more information button, I could see what exactly was meant*" (SE4). Moreover, "*suggestions gave more time to think and then go into detail*" (SE3). However, some SEs experienced delays or hesitation when suggestions did not match the interaction: "*[…] that made the speed of me answering the question a little bit slower because I had to look for the answers myself*" (SE13). Also, proposals should be adjusted in wording and capitalized to simplify their use. Regarding customer interactions, SEs reported that they were able to autonomously manage them with provided suggestions (e.g., "*If I wanted to bring the conversation in another direction, I would have done it - so it was not forced*" (SE11)) and make independent decisions without feeling constrained (the prototype "*[…] is presented in a way that it was clear that I can work with him, but I don't have to*" (SE15)). In addition, the prototype assisted them to achieve their goals in counseling the customer: e.g., "*I was able to control the interaction. And I think the counseling was actually better because of Charlie's help because he explained things way more detailed than I would have done*" (SE15). However, SEs reported that the personal touch is reduced due to the provided wording in suggestions. **DF4**. Overall, an average of six (*SD*: 2.5; *Median*: 7) suggestions were used, whereby the detailed version via get-more-info button (*Mean*: 3.7; *SD*: 2.6; *Median*: 4.5) was used more frequently than the short version (*Mean*: 2.6; *SD*: 2.4; *Median*: 2). To receive alternative FAQ answer suggestions, the discard-button was clicked on average 15 times (*SD*: 10.8; *Median*: 15). The display of two suggestions and the option for additional explanatory information via the get-more-info-button were perceived as helpful "*so that you can think in which direction you might go*" (SE1). SEs experienced relief through displayed suggestions and the majority saved time making decisions, especially by using the copy-to-chat-button: "*[…] I just had to copy them, which affected* *the speed*" (SE14). **DF5 & DF6.** 16 SEs utilized the feedback function on average four times, while nine people successfully provided feedback. However, SEs expressed the need for an adaptation of the feedback function, as it was unclear. **DF7**. Concerning the recommendation of projects, the pressure to recall knowledge or search in parallel to the customer interaction was reduced as relevant information was presented. Thereby, it "*[…] took out the uncomfortable part of working with such a consultation, which is looking up stuff*" (SE16).

12.6 Discussion and Conclusion

Our multi-cycle DSR project contributes to HI research [11, 21] by taking a human-centered perspective to design HIS [12, 13] for text-based, real-time service encounters [2] in OCS for mutual augmentation [15]. Particularly, we examine hybrid decision-making and hybrid learning. While we cover the enablement of hybrid learning in the first cycle, we extend this initial design in the second cycle to sustain hybrid real-time decision-making. To address our research question, we derived four additional DPs by considering relevant theories to define requirements that satisfy SEs' need for autonomy, competence, and relatedness. Based on the evaluation, the instantiated DPs successfully supported SEs' autonomy and task mastery in conducting customer interactions allowing efficient and independent decision-making. The SEs' feeling of control is supported by the analysis of screen recordings which showed that all SEs used the configurable settings to approve augmentation by the prototype (DP2.1). However, the evaluation revealed a high reliance of the SEs on the suggestions (DP2.2) partly leading to uncertainty and delays. Although SEs could conduct the service encounter without AI augmentation, they rather clicked the discard-button several times instead of formulating a new answer. In contrast, one SE only read and verified the suggestions and formulated new answers based on the provided content indicating a high level of SEs' autonomy. The need for competence could be addressed by supporting SEs' achievement of counseling goals via suggestions. In this regard, DP2.2 and DP2.3 successfully supported the dynamic decision situation by showing relevant information. Furthermore, the analysis of SEs' usage behavior demonstrates an intuitive application of suggestions by using the copy-to chat button in effortless ways (DP2.1). With this, DP2.1 is the main contributor to experienced relief, time savings, and efficiency. Moreover, SEs particularly recognized the usefulness of the get-more-info button (DP2.4), which is supported by the screen recording results that showed SEs' preference for the detailed version of suggestions. Regarding the need for relatedness, the evaluation did not show consistent results, as some SEs perceived the prototype as a tool and others as a co-customer manager.

All in all, we provide relevant and promising results demonstrating a potential solution to integrate hybrid learning and real-time decision augmentation within a HIS. We thereby make a two-fold contribution. First, following [38], we present a nascent design theory with utility character by delivering a possible solution for the identified problem and demonstrating improvements in the application field [36]. This contribution has epistemological implications, as we present DPs about user activity and an artifact that links prescriptive knowledge about design and action with explanatory knowledge about effects [40, 55, 58]. Second, we present a designed entity by demonstrating a full-featured AI-based artifact, which represents one possible instantiation of our design [36, 40]. Besides the promising results, there are, however, a few limitations to consider. First, we conducted one semi-naturalistic evaluation episode with simulated customers without a pre-evaluation of the instantiated DPs. Second, we limited the implementation and application of our DPs to only one organization. Thus, future research should implement and evaluate our DPs in various naturalistic environments. In doing so, factors should be examined causing different usage behavior and decision-making effects. For instance, while copy-to-chat might increase efficiency, it might also decrease human attention and learning. Especially when trying to educate novice employees with such a tool, proper usage of the suggestions needs to be ensured. In addition, SEs’ decisions should be investigated in terms of quality due to influences of heuristics or biased AI. At last, as the feedback function was not clear to several SEs, we call for future research on how to ensure valuable and continuous feedback toward the AI.

12.7 Acknowledgment

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, 02L18A111).

12.8 References

1. Gupta, A.: 4 Key Tech Trends in Customer Service to Watch (2021), https://www.gartner.com/smarterwithgartner/4-key-tech-trends-in-customer-service-to-watch
2. Keyser, A. de, Köcher, S., Alkire, L., Verbeeck, C., Kandampully, J.: Frontline Service Technology Infusion: Conceptual Archetypes and Future Research Directions. JOSM, vol. 30, 156–183 (2019)
3. Servion: What Makes Emerging Technologies the Future of Customer Experience? (2018), https://servion.com/blog/what-emerging-technologies-future-customer-experience/
4. Huang, M.-H., Rust, R.T.: Artificial Intelligence in Service. Journal of Service Research, vol. 21, 155–172 (2018)
5. Wilson, H.J., Daugherty, P.R.: Collaborative Intelligence: Humans and AI are Joining Forces. Havard Business Review, vol. 96, 115–123 (2018)

6. Graef, R., Klier, M., Kluge, K., Zolitschka, J.F.: Human-machine Collaboration in Online Customer Service – A Long-term Feedback-based Approach. Electron Markets, vol. 31, 319–341 (2021)
7. Mero, J.: The Effects of Two-way Communication and Chat Service Usage on Consumer Attitudes in the E-commerce Retailing Sector. Electron Markets, vol. 28, 205–217 (2018)
8. Verhagen, T., van Nes, J., Feldberg, F., van Dolen, W.: Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters. J Comput-Mediat Comm, vol. 19, 529–545 (2014)
9. Cheung, C.F., Lee, W.B., Wang, W.M., Chu, K.F., To, S.: A Multi-perspective Knowledge-based System for Customer Service Management. Expert Systems with Applications, vol. 24, 457–470 (2003)
10. Ulfert, A.-S., Antoni, C.H., Ellwart, T.: The Role of Agent Autonomy in Using Decision Support Systems at Work. Computers in Human Behavior, vol. 126, 106987 (2022)
11. Dellermann, D., Ebel, P., Söllner, M., Leimeister, J.M.: Hybrid Intelligence. Bus Inf Syst Eng, vol. 61, 637–643 (2019)
12. Xu, W.: Toward Human-centered AI: A Perspective from Human-computer Interaction. interactions, vol. 26, 42–46 (2019)
13. Ebel, P., Söllner, M., Leimeister, J.M., Crowston, K., Vreede, G.-J. de: Hybrid Intelligence in Business Networks. Electron Markets, vol. 31, 313–318 (2021)
14. Shneiderman, B.: Human-Centered Artificial Intelligence: Three Fresh Ideas. THCI, vol. 12, 109–124 (2020)
15. Benbya, H., Pachidi, S., Jarvenpaa, S.L.: Artificial Intelligence in Organizations: Implications for Information Systems Research. Journal of the Association for Information Systems, vol. 22, 1–25 (2021)
16. Froehle, C.M.: Service Personnel, Technology, and Their Interaction in Influencing Customer Satisfaction. Decision Sciences, vol. 37, 5–38 (2006)
17. Bock, D.E., Wolter, J.S., Ferrell, O.C.: Artificial Intelligence: Disrupting What We Know about Services. JSM, vol. 34, 317–334 (2020)
18. Ameen, N., Tarhini, A., Reppel, A., Anand, A.: Customer Experiences in the Age of Artificial Intelligence. Computers in Human Behavior, vol. 114, 106548 (2021)
19. Zierau, N., Elshan, E., Visini, C., Janson, A.: A Review of the Empirical Literature on Conversational Agents and Future Research Directions. In: 41st International Conference on Information Systems (ICIS). virtual conference (2020)
20. Raisch, S., Krakowski, S.: Artificial Intelligence and Management: The Automation-Augmentation Paradox. AMR, vol. (2020)
21. Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., Ebel, P.: The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In: 52nd Hawaii International Conference on System Sciences (2019)

22. Kahai, S.S., Cooper, R.B.: Exploring the Core Concepts of Media Richness Theory: The Impact of Cue Multiplicity and Feedback Immediacy on Decision Quality. Journal of Management Information Systems, vol. 20, 263–299 (2003)
23. Ishii, K., Markman, K.M.: Online Customer Service and Emotional Labor: An Exploratory Study. Computers in Human Behavior, vol. 62, 658–665 (2016)
24. Deci, E.L., Ryan, R.M.: Self-determination Theory: A Macrotheory of Human Motivation, Development, and Health. Canadian Psychology/Psychologie canadienne, vol. 49, 182–185 (2008)
25. Ahuja, M., Thatcher, J.B.: Moving Beyond Intentions and Toward the Theory of Trying: Effects of Work Environment and Gender on Post-adoption Information Technology Use. MIS Quarterly, vol. 29, 427–459 (2005)
26. Ozkara, B.Y., Ozmen, M., Kim, J.W.: Exploring the Relationship between Information Satisfaction and Flow in the Context of Consumers' Online Search. Computers in Human Behavior, vol. 63, 844–859 (2016)
27. Hung, S.-Y., Liang, T.-P.: Effect of Computer Self-efficacy on the Use of Executive Support Systems. Industrial Management & Data Systems, vol. 101, 227–237 (2001)
28. Hassanein, K., Head, M.: The Impact of Infusing Social Presence in the Web Interface: An Investigation Across Product Types. International Journal of Electronic Commerce, vol. 10, 31–55 (2005)
29. Pavlou, Liang, Xue: Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective. MIS Quarterly, vol. 31, 105–136 (2007)
30. Edwards, W.: Dynamic Decision Theory and Probabilistic Information Processings. hfes, vol. 4, 59–74 (1962)
31. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science, vol. 12, 257–285 (1988)
32. Brünken, R., Plass, J.L., Leutner, D.: Direct Measurement of Cognitive Load in Multimedia Learning. Educational Psychologist, vol. 38, 53–61 (2003)
33. MacGeorge, E.L., Feng, B., Guntzviller, L.M.: Advice: Expanding the Communication Paradigm. Annals of the International Communication Association, vol. 40, 213–243 (2016)
34. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers, vol. 56, 81–103 (2000)
35. Adam, M.T.P., Gregor, S., Hevner, A., Morana, S.: Design Science Research Modes in Human-Computer Interaction Projects. THCI, vol. 13, 1–11 (2021)
36. Baskerville, R., Baiyere, A., Gergor, S., Hevner, A., Rossi, M.: Design Science Research Contributions: Finding a Balance between Artifact and Theory. JAIS, vol. 19, 358–376 (2018)
37. Kuechler, W., Vaishnavi, V.: A Framework for Theory Development in Design Science Research: Multiple Perspectives. JAIS, vol. 13, 395–423 (2012)

38. Gregor, S., Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, vol. 37, 337–355 (2013)
39. Gregor, S., Jones, D.: The Anatomy of a Design Theory. JAIS, vol. 8, 312–335 (2007)
40. vom Brocke, J., Winter, R., Hevner, A., Maedche, A.: Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space. JAIS, vol. 21, 520–544 (2020)
41. Wiethof, C., Bittner, E.: Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI. In: 30th European Conference on Information Systems (ECIS), under review. Timisoara, Romania (2022)
42. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: A Framework for Evaluation in Design Science Research. European Journal of Information Systems, vol. 25, 77–89 (2016)
43. Mayring, P.: Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution (2014)
44. Gonzalez, C.: Decision Support for Real-time, Dynamic Decision-making Tasks. Organizational Behavior and Human Decision Processes, vol. 96, 142–154 (2005)
45. Johnson, E.J., Payne, J.W.: Effort and Accuracy in Choice. Management Science, vol. 31, 395–414 (1985)
46. Cook, G.J.: An Empirical Investigation of Information Search Strategies with Implications for Decision Support System Design. Decision Sciences, vol. 24, 683–698 (1993)
47. Paas, F., Renkl, A., Sweller, J.: Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. Instructional Science, vol. 32, 1–8 (2004)
48. Guntzviller, L.M., Liao, D., Pulido, M.D., Butkowski, C.P., Campbell, A.D.: Extending Advice Response Theory to the Advisor: Similarities, Differences, and Partner-effects in Advisor and Recipient Advice Evaluations. Communication Monographs, vol. 87, 114–135 (2020)
49. Lerch, F.J., Harter, D.E.: Cognitive Support for Real-Time Dynamic Decision Making. Information Systems Research, vol. 12, 63–82 (2001)
50. Hong, W., Thong, J.Y.L., Tam, K.Y.: The Effects of Information Format and Shopping Task on Consumers' Online Shopping Behavior: A Cognitive Fit Perspective. Journal of Management Information Systems, vol. 21, 149–184 (2004)
51. Eick, S.G.: Graphically Displaying Text. Journal of Computational and Graphical Statistics, vol. 3, 127–142 (1994)
52. MacGeorge, E.L., van Swol, L.M.: Advice across Disciplines and Contexts. In: MacGeorge, E.L., van Swol, L.M. (eds.) The Oxford Handbook of Advice, pp. 3–20. Oxford University Press, New York (2018)
53. Feng, B., MacGeorge, E.L.: The Influences of Message and Source Factors on Advice Outcomes. Communication Research, vol. 37, 553–575 (2010)

54. Fortin, D.R., Dholakia, R.R.: Interactivity and vividness effects on social presence and involvement with a web-based advertisement. Journal of Business Research, vol. 58, 387–396 (2005)
55. Gregor, S., Kruse, L., Seidel, S.: Research Perspectives: The Anatomy of a Design Principle. JAIS, vol. 21, 1622–1652 (2020)
56. Wiethof, C., Bittner, E.: Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review. In: 42nd International Conference on Information Systems. Austin (2021)
57. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., Yih, W.: Dense Passage Retrieval for Open-Domain Question Answering (2020)
58. Gregor: The Nature of Theory in Information Systems. MISQ, vol. 30, 611–642 (2006)

13 Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI

Wiethof, C., and Bittner, E. A. C. 2022. "Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI," in *30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Abstract. Hybrid intelligence systems (HIS) enable human users and Artificial Intelligence (AI) to collaborate in activities complementing each other. They particularly allow the combination of human-in-the-loop and computer-in-the-loop learning ensuring a hybrid collaborative learning cycle. To design such a HIS, we implemented a prototype based on formulated design principles (DPs) to teach and learn from its human user while collaborating on a task. For implementation and evaluation, we selected a customer service use case as a top domain of research on AI applications. The prototype was evaluated with 31 expert and 30 novice customer service employees of an organization. We found that the prototype following the DPs successfully contributed to positive learning effects as well as a high continuance intention to use. The measured levels of satisfaction and continuance intention to use provide promising results to reuse our DPs and further develop our prototype for hybrid collaborative learning.

Keywords: Hybrid Intelligence, Human-in-the-Loop, Computer-in-the-Loop, Collaborative Learning.

13.1 Introduction

Customer service has been increasing its market volume recently by automating human tasks with Artificial Intelligence (AI) (Brandt, 2021). It is a common application area for the implementation of self-service technology, such as chatbots that take over service encounters in the frontline of service providers (Brandt, 2021; Huang and Rust, 2018; Sajeev et al., 2021; Liu et al., 2020; Svenningsson and Faraon, 2019; Sun et al., 2021). In general, AI systems have usually been developed to adopt and excel certain human skills, e.g., decision-making, problem solving, or pattern recognition (Dellermann et al., 2019b; Rzepka and Berger, 2018; Abdel-Karim et al., 2020). Therefore, researchers, especially in the field of Machine Learning (ML) successfully developed techniques to emulate such skills and automate tasks with the ultimate goal to replace humans (Holzinger et al., 2016, 2017; Rzepka and Berger, 2018). Nevertheless, AI is still far away from achieving human general intelligence (Dellermann et al., 2019b). Thus, customer service chatbots still require the involvement of service employees via escalation when a customer request cannot be solved (Subramaniam et al., 2018; Sousa et al., 2019; Sajeev et al., 2021; Poser et al., 2021).

Although AI might reduce human effort in domains such as customer service, the relatively new concept of hybrid intelligence introduces a promising approach to combine human and artificial intelligence (Dellermann et al., 2019b). Instead of replacing humans, researchers started to acknowledge human intelligence as a positive contributor to AI and vice versa by bringing them together in a hybrid team (Dellermann et al., 2019b; Seeber et al., 2020). Therefore, ML researchers initiated the human-in-the-loop (HITL) approach for AI development leading from automatic to interactive ML (iML) systems. This enables algorithms to directly interact with human users and to optimize and learn through these interactions (Holzinger, 2016a, 2016b; Amershi et al., 2014). iML already revealed advantageous capabilities, e.g., in learning and working with complex or small data sets (Holzinger et al., 2019; Martínez et al., 2019; Yimam et al., 2016). Still, to accomplish hybrid intelligence, besides putting the human in the loop of AI, AI needs to be considered in the loop of humans (Dellermann et al., 2019b). Thus, research started to shift the view from the HITL approach to the computer-in-the-loop (CITL) approach (Shneiderman, 2020). Using CITL, humans can learn through working with AI, e.g., through error-learning (Abdel-Karim et al., 2020). A hybrid intelligence system (HIS) combines HITL and CITL, enabling human intelligence to augment AI and AI to augment human intelligence toward hybrid collaborative learning (Dellermann et al., 2019b). Initial research has already applied these approaches, e.g., for collective intelligence development within Industry 4.0 (Gavriushenko et al., 2020), knowledge discovery (Oliveira et al., 2020), or creation of value in sales (Paschen et al., 2020).

With the increasing deployment and usage of AI in organizations, Benbya et al. (2021) discuss arising implications for Information Systems (IS) research and present further research opportunities in this field, e.g., regarding automation, augmentation, engagement, and decision-making. Specifically, regarding the augmentation of frontline service employees’ intelligence, customer service can already benefit from AI abilities making way for the service encounter 2.0 (Keyser et al., 2019; Larivière et al., 2017). For instance, Molino et al. (2018) present intelligent techniques to help service employees to improve their speed and efficiency through suggesting request classifications, answers based on request content as well as additional context information. However, regarding the idea of collaborative learning through HIS (Dellermann et al., 2019b), most researchers focus solely on learning of either human or AI. Thus, we identified a research gap on the implementation of HIS in customer service toward collaborative learning ensuring employees teaching AI and employees learning from AI through an iterative cycle. Furthermore, we propose to differentiate employees based on experience level, as we assume that expert employees are more capable of teaching and novice employees benefit more from learning. With this, we suppose an implicit knowledge transfer from experts to novices through AI in a long-term perspective, demanding the integration of hybrid collaborative learning in HIS (Dellermann et al., 2019a; Kulesza et al., 2015) and users' continuous utilization of the system (Bhattacherjee, 2001). To address this research gap, we formulate the following research questions: **RQ1:** *How can continuous collaborative learning of customer service employees and AI be designed and implemented in a HIS?* **RQ2:** *How do the learning effects and continuance intention to use differ between novice and expert employees when working with a HIS in customer service?* For relevance of the research problem (Gregor and Hevner, 2013), we identified a real-world use case with organization X, which sells project participation and internships abroad. To support collaborative learning in customer service in organization X, we design and develop a prototypic HIS via a web application combining HITL and CITL. To ensure a common understanding of the terms AI and ML, specifically related to HITL, we refer to the definition of Dellermann et al. (2019a, p. 275) based on Russell and Norvig (2016): "[AI] covers the idea of creating machines that can accomplish complex goals. This includes facets such as natural language processing, perceiving objects, storing of knowledge and applying it for solving problems, and machine learning to adapt to new circumstances and act in its environment". Hence, following Kühl et al. (2019), to implement AI in HIS in customer service we specifically consider a narrow AI, which is able to continuously learn from its environment by utilizing ML. With this study, we present the first cycle of a multicyclic design science research (DSR) project particularly focusing on human learning and continuance intention to use and to teach. Thus, through test runs with employees working in organization X, we conducted a mixed-method evaluation of the artifact by applying the Wizard of Oz (WOz) technique. Overall, we provide promising results that lay the foundation for further development of our system in the second cycle. Additionally, this work presents three main contributions. First, we provide design knowledge for researchers and practitioners to design and implement a HIS in customer service toward hybrid collaborative learning. Second, we introduce an instantiated prototype of a HIS based on formulated design principles (DPs). Eventually, we show the capability of the prototypic HIS to satisfy its users leading to a high continuance intention to use and learning.

13.2 Related Work

13.2.1 Hybrid Intelligence

Hybrid intelligence proposes "to combine the complementary strengths of heterogeneous intelligences (i.e., human and artificial agents) into a socio-technological ensemble" leading to HIS "that have the ability to accomplish complex goals by combining human and AI to collectively achieve superior results than each of them could have done in separation and continuously improve by learning from each other" (Dellermann et al., 2019a, p. 276). Thereby, AI augments human intelligence and humans augment AI (Dellermann et al., 2019b; Wiethof and Bittner, 2021). Researchers have already built on the concept of hybrid intelligence in various ways. For instance, Dubey et al. (2020) built on the taxonomy of Dellermann et al. (2019a) to create a framework for human-AI teaming and developed several use cases. Others contribute with the design and development of a system enabling hybrid human-AI collaboration, e.g., for iterative feature-based clinical decision-making (Hun Lee et al., 2021) or with design principles for hybrid intelligence ML algorithms focusing on trust (Ostheimer et al., 2021). However, they neglect to "*continuously improve [the HIS] by learning from each other*" by putting human intelligence in the loop of AI (HITL) and AI in the loop of human intelligence (CITL) (Dellermann et al., 2019a, p. 276; Dellermann et al., 2019b). Regarding HITL, there is much research in the field of ML, as it allows users to be involved in the learning process, leading to iML (Martínez et al., 2019; Amershi et al., 2014). Research about CITL has not been conducted to the same extent. Accordingly, Shneiderman (2020) raises awareness for putting the human in the center for a better focus on user needs, user experience, human performance, or human control. In this context, Abdel-Karim et al. (2020, p. 199) define CITL as "the counterpart of interactive machine learning, i.e., human learning while being in the loop in a human-machine collaboration". They show that humans can learn from their errors when working with an ML-based system. Following, there is a potential of collaborative learning combining CITL and HITL, especially when differentiating between novices and experts. Eventually, it needs to be ensured that human decision-makers can both continuously teach the ML-based system and learn or gain insights from it (Abdel-Karim et al., 2020; Dellermann et al., 2019a; Wiethof and Bittner, 2021). Thus, we have identified a research gap to address with this study to investigate actual hybrid collaborative learning differentiating between experts and novices in HIS.

13.2.2 AI in customer service

Customer service in terms of online frontline service encounters between employee and customer is a well-known application domain for AI. As such, it has already gone through several transformations, especially regarding the role, tasks, and interactions of employee and customer (Robinson et al., 2020; Huang and Rust, 2018; Keyser et al., 2019; Xu et al., 2020). There is some research on how AI can replace human employees in the service encounter by executing their service tasks (Huang and Rust, 2018; Xu et al., 2020). Robinson et al. (2020) define different forms of encounters. They distinguish between whether employee or customer are human or AI as well as if any human is aware of AI being involved. With their framework, they demonstrate ways to replace human employees or customers (Robinson et al., 2020). However, instead of technology substituting employees, AI can also adopt the role of augmenting employees (Keyser et al., 2019; Larivière et al., 2017). Existing research on technology in customer service defines, how technology can be infused in the frontline customer service (Keyser et al., 2019) calling it "service encounter 2.0" (Larivière et al., 2017). According to both roles, Keyser et al. (2019) define archetypes based on existing literature. Figure 1 visualizes the three involved entities and possible interactions, that vary for each archetype, e.g., no interaction, direct interaction, or interaction augmented by technology (Keyser et al., 2019). The bold arrows in Figure 1 depict the constellation of the two traditionally involved entities, employee and customer (1), and the infusion of a technology (2). It assumes the role of augmenting the service employee (Keyser et al., 2019). Thus, the AI is only accessible by the employee and is prohibited from directly communicating to the customer. As for our research, we put our focus on the interaction between employee and technology to establish a HIS.

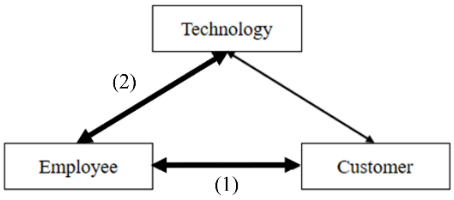

Figure 1. Frontline service technology infusion, augmentation scenario highlighted (Keyser et al., 2019), adapted

We study the collaboration of human employee and AI through HITL and CITL toward hybrid intelligence. Thus, according to Figure 1, we study the bidirectional interaction of human employee and technology in customer service through augmentation focusing on collaborative learning.

13.3 Conceptual Background

In this section, we describe the two relevant constructs to be considered for the design, development, and evaluation of a HIS in customer service: learning and continuance intention to use. Based on these, we establish a hybrid collaborative learning cycle differentiating between experts and novices and identify according meta-requirements (MRs) (see section 5). By operationalizing the constructs, we define dependent variables to evaluate them (see section 7).

13.3.1 Hybrid collaborative learning

Up to this point, most existing research leverages the combination of artificial and human intelligence to enhance the learning of either intelligence, e.g., iML for AI (Amershi et al., 2014; Holzinger, 2016a, 2016b; Holzinger et al., 2017) or agent tutors for human intelligence (Wambsganss et al., 2021; Martins Giraffa and Viccari, 1998; Chhibber and Law, 2019; Hjorth, 2021). In terms of hybrid collaborative learning, AI and humans may both benefit from individual learning, especially distinguishing experts and novices (Liu et al., 2014; Hu et al., 2019; Dellermann et al., 2019b; Oliveira et al., 2020; Wiethof and Bittner, 2021), as it offers the possibility to transfer knowledge from experts to novices (Liu et al., 2014; Dellermann et al., 2019b): while experts can teach a machine implicitly (e.g., through conversations and interactions) and explicitly (e.g., by giving feedback or adjusting machine explanations), the machine can teach novices, by providing explanations for its decisions or recommendations (Schneider and Handali, 2019; Dellermann et al., 2019b; Liu et al., 2014; Kulesza et al., 2015). In the context of our study, we want to make sure that a HIS enables **learning** for both human and AI through collaboration.

To achieve this kind of hybrid collaboration, Kulesza et al. (2015) elaborate on explanatory debugging. This approach is supposed to focus on users and will eventually enable them to get the most benefit from hybrid collaboration. Through cycles of explanations, humans will learn from the machine and the machine from the human users (Kulesza et al., 2015). Eight guiding principles will enable users to receive explanations for the machine's predictions – explainability – and machines to receive explanations about corrections from the users – correctability. For explainability, explanations from the machine are recommended to be 1) iterative, 2) sound, 3) complete, and 4) not overwhelming. For correctability, explanations from the human are recommended to be 1) actionable, 2) reversible, 3) always honored, and 4) making incremental changes to the machine's reasoning (Kulesza et al., 2015).

The taxonomy of hybrid intelligence design by Dellermann et al. (2019a) provides sequential guidance for design decisions when developing HIS involving the four meta-dimensions 1) task characteristics, 2) learning paradigm, 3) AI-human interaction, and 4) human-AI interaction. For each meta-dimension, sub-dimensions cover a total of 50 categories (Dellermann et al., 2019a). When starting the development of a HIS, the characteristics of the task are defined. Next, the learning paradigm clarifies, how humans and the machine can learn from each other including augmentation, machine learning, and human learning. To go into detail and elaborate on the learning of both machine and human, the taxonomy defines the human part of the interaction (machine teaching, teaching interaction, expertise requirements, amount of human input, aggregation, incentives), and the machine part of the interaction (query strategy, machine feedback, interpretability) (Dellermann et al., 2019a).

13.3.2 Information systems continuance model

Though users' acceptance of IS is relevant for their usage, its study is often limited to a short-term perspective, i.e., acceptance for the initial use. Bhattacherjee (2001) raises awareness of the distinctions between acceptance and continuance behaviors. As for the iterative form of hybrid collaborative learning, it is of high relevance to ensure a high continuance intention of human users to use the learning system, i.e., experts continue teaching through system usage, which enables teaching of novices through system usage. To determine IS continuance intention and understand how to influence it, Bhattacherjee (2001) builds on the expectation-confirmation theory (ECT) by Oliver (1980). Therefore, he integrates the theory with research findings of IS use and proposes a post-acceptance model of IS continuance (see Figure 2).

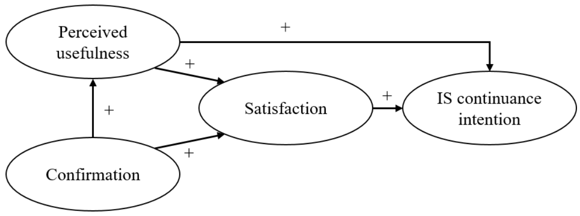

Figure 2. A post-acceptance model of IS continuance (Bhattacherjee, 2001), adapted

The model suggests that "users' continuance intention is determined [primarily] by their satisfaction with IS use and [secondarily by] perceived usefulness of continued IS use. User satisfaction, in turn, is influenced by their confirmation of expectation from prior IS use and perceived usefulness. Post-acceptance perceived usefulness is influenced by users' confirmation level" (Bhattacherjee, 2001, p. 351). In the context of our study, as continuous collaborative learning is a central aspect of hybrid intelligence (Dellermann et al., 2019b), we want to make sure that users of a HIS show a high **continuance intention to use** to enable collaborative learning. Only if both expert and novice users are willing to continuously work with a HIS long-term, it is possible to teach and learn from AI. Hence, we involve potential users in the system design (see section 5) by identifying and implementing their requirements and expectations toward their satisfaction. We further defined dependent variables to include the continuance intention to use in the quantitative evaluation (see section 7).

13.4 Research Approach

This study is part of a multicyclic DSR project toward the design and development of HIS in customer service. It represents the first cycle, in which we specifically aim to show the potential of HIS in customer service based on a hybrid collaborative learning cycle. Therefore, we contribute prescriptive design knowledge to the knowledge base for designing and developing a HIS in customer service toward collaborative learning (Gregor and Hevner, 2013) toward a "theory for design and action" (Gregor, 2006). We structure our study with the DSR process by Peffers et al. (2006) as follows (see Figure 3). First, the introduction covers problem identification and motivation. To define the objectives of a solution, we identify MRs that match necessary constructs for the design, development, and evaluation of a HIS in customer service: learning (Dellermann et al., 2019a; Kulesza et al., 2015) and continuance intention to use (Bhattacherjee, 2001). The MRs are derived by conducting eleven expert interviews following the approach of Meuser and Nagel (2002) with six experienced as well as five novice employees working in customer service. Based on the MRs, we derive DPs and formulate them according to Chandra et al. (2015). For demonstration, we match each DP to a design feature (DF) to instantiate the DPs through prototyping in a web application (Wilde and Hess, 2007). For the evaluation, we conducted test runs with 61 participants, differentiating between 30 experts and 31 novices in terms of an ex post case study (Venable et al., 2012). We partially included experts and novices from the first interviews toward objectives of a solution to ensure their requirements are met. As this study represents the first cycle of a multicyclic project, we aim to achieve early results to prove our concept, demonstrate its potential and contribute with first design knowledge as a fundament for the next cycle. To do so, we applied the WOz technique to simulate certain functions of the prototype with a hidden wizard. This enables the observation of the participating user working with an apparently fully functioning prototype (Böttcher and Nüttgens, 2013; Salber and Coutaz, 1993; Krannich, 2010). After the test runs, we conducted expert interviews with all participants to qualitatively assess the DPs (Meuser and Nagel, 2002). To ensure triangulation (Mayring, 2001), we conduct a quantitative evaluation of the two operationalized constructs learning and continuance intention to use (Samarasinghe and Tretiakov, 2009; Likert, 1932). Our findings then lay a foundation for the second cycle, in which we aim to instantiate a fully functioning HIS prototype.

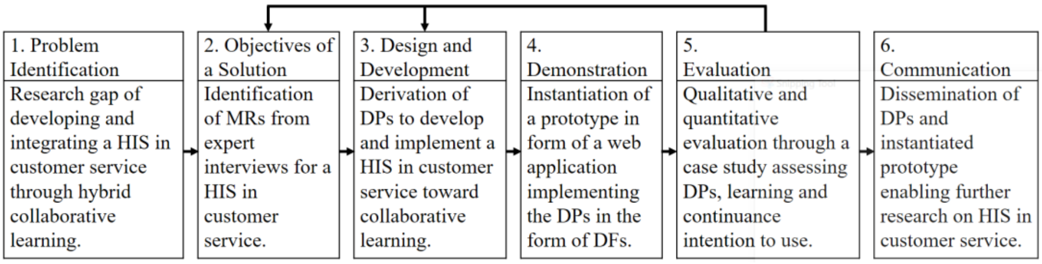

Figure 3. Structure along the DSR process

13.5 Objectives of a Solution

To address RQ1 and define goals for the artifact (Koppenhagen et al., 2012; Zhang et al., 2011; Gregor and Hevner, 2013), we first build on the conceptual background. Combining the insights of hybrid collaborative learning (Dellermann et al., 2019a; Kulesza et al., 2015) with the information systems continuance model (Bhattacherjee, 2001), we define an iterative cycle of hybrid collaborative learning between expert and novice users, and a learning system (see Figure 4). While the system gains most knowledge from experts' input through HITL, novices benefit most from the knowledge provided by the system through CITL. Furthermore, novices might also contribute via HITL by teaching the system implicitly or explicitly. In addition, experts might also benefit from CITL, e.g., through new insights (Dellermann et al., 2019a).

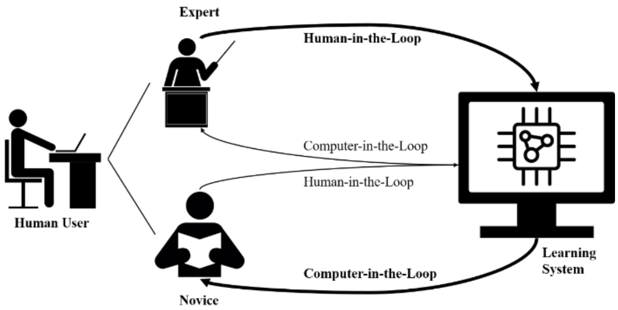

Figure 4. Hybrid collaborative learning differentiating between experts and novices

Following Dellermann et al. (2019a), we need to make sure that all entities involved continuously learn and improve through each other. Therefore, human users need to have a high continuance intention to use the system demanding the integration of the information systems continuance model (Bhattacherjee, 2001). Consequently, we need to put attention to the usefulness and satisfaction of the users regarding the design and development of the HIS. Hence, we consider the two constructs learning and continuance intention to use. Following these, we conduct eleven semi-structured qualitative expert interviews along the approach by Meuser and Nagel (2002) to identify MRs for a HIS toward collaborative learning in customer service by meeting users' expectations for satisfaction and continuance intention to use. The interview guideline consists of five parts with 14 questions: (1) explanation of the customer service process, (2) current state of knowledge and education, (3) imagining hybrid intelligence in customer service, (4) HIS benefits and disadvantages for experts and novices, and (5) further remarks. Each interview lasted 30-60 minutes. For the selection of interview partners, we considered six expert employees (E1-E6), who had been working in customer service of organization X for at least half a year, and five novice employees (N1-N5) from the same organization. We recorded all interviews and transcribed them. We deductively defined codes following the taxonomy of hybrid intelligence design (Dellermann et al., 2019a) as it provides sequential guidance for design decisions when developing HIS. Thus, we gathered and aggregated the results based on the meta-dimensions of the taxonomy. 1) Task characteristics define how the task is carried out by humans and AI collaboratively. 2) The learning paradigm defines how humans and AI learn from each other. 3) Human-AI interaction defines how the AI learns from the human. 4) AI-Human interaction defines how the human learns from the AI. We further subcoded the results following Kulesza et al. (2015) toward explainability (e) and correctability (c). Eventually, we inductively derived the following 16 MRs.

Task Characteristics. MR1: The AI should act as a peer contributing to the work of the human employee leading to performance improvement in terms of effectiveness and efficiency (E2, E3, E6, N2, N3, N5) (e). **MR2:** The AI is user-friendly (N1, N4) (e). **MR3:** The tasks are clearly distributed between the AI and the human employee (E1-E4, E6) (e). **MR4:** The human employee has the responsibility for communicating to customers (E3, E4) (c). **MR5:** The AI knows and follows all steps of the customer service process (E2, E4-E6) (e).

Learning Paradigm. MR6: The AI is limited to answering basic questions and providing information (E1, E3-E5, N1) (e). **MR7:** The AI ensures that all necessary information is gathered (E2-E4, E6, N1, N2, N5) (e). **MR8:** The AI will learn from the provided data throughout experience and interactions, and improve over time (E1, E3, E4, E6, N1, N3, N5) (c).

Human-AI Interaction. MR9: The AI learning behavior differentiates depending on whether it interacts with an experienced or novice employee, putting more weight on experienced employees' data (E2, E3, N1, N2) (c). **MR10:** Employees can choose if they want to provide their data to the AI for learning purposes (E2, E4, N1, N2) (c). **MR11:** Employees are always able to check the AI's work, easily correct mistakes, and change things, if necessary (E4, N2) (c).

AI-Human Interaction. MR12: Employees understand how the AI is working, how it is learning, where the knowledge is coming from, and how to work with it (E2, E3, E6, N2, N4, N5) (e). **MR13:** The AI provides suggestions to continue the process (N5) (e). **MR14:** The AI only provides and submits suggestions or solutions to the human employee (E1, E3, E4, E6, N1-N3, N5) (c). **MR15:** The AI raises awareness on things the human employee does not focus on (E6, N1) (e). **MR16:** The AI asks the human employee for feedback to learn from it (E3, N1) (c)

13.6 Artifact Design, Development and Demonstration

Based on the MRs, we derived preliminary action-oriented DPs (Koppenhagen et al., 2012; Zhang et al., 2011) toward a HIS in customer service. We formulated the DPs according to Chandra et al. (2015) illustrated in Table 1.

Table 1. DPs with according MRs

Design Principles	
DP1: AI Learning Behavior Settings	Provide the HIS with adjustable settings about the AI learning behavior in order for the human employees to choose how much it learns, given that the AI differentiates between novice and expert employees or does not learn at all. (MRs 8-10)
DP2: AI Identity	Provide the HIS with an AI identity in order for the human employees to perceive the AI as a peer for collaboration purposes, given that the role of both the human employee and AI are transparent and clear. (MRs 1-3, 12)
DP3: Education on AI	Provide the HIS with explanations on how the AI is working and learning in order for the human employees to understand how to work with it, given that the AI should be user-friendly and not overwhelming with information. (MRs 2, 12)

DP4: Customer Service Process Awareness	Provide the HIS with a shared understanding of the customer service process in order for the human employees to easily follow the process together with the AI and see progress, given that the understanding of the AI can be adapted at any time. (MRs 1, 5, 8, 11, 13)
DP5: Collaborative Knowledge	Provide the HIS with the opportunity to collaboratively share knowledge between the human employee and the AI in order for the human employee to be aware of the gathered information of the AI, given that they can also provide the AI with their own gathered knowledge to consider. (MRs 1, 7, 8, 11, 13, 15)
DP6: AI Output	Provide the HIS with the AI ability to give suggestions for answering basic customer questions in order for the employees to choose how to use the output for replying to the customer, given that the AI can learn from the usage. (MRs 1, 4, 6, 8, 11, 14, 15)
DP7: Feedback	Provide the HIS with the option to give direct feedback for the AI output in order for the human employees to teach the AI, given that the AI also implicitly learns through the experience and interactions. (MRs 1, 8, 11, 16)

We match each DP to a DF, e.g., DP1-DF1, to instantiate the DPs in a prototypical HIS via a web application (Koppenhagen et al., 2012) (see Figure 5) and deploy it in the customer service process of organization X. The expert interviews revealed that service employees and customers communicate through an online chat. The process proceeds as follows: after making contact, employees aim at finding out the customers' interests for suitable projects abroad to propose project recommendations. Furthermore, employees answer questions and provide support. By infusing AI in this process, its core functionality is to augment the employee with interests of the customer and project suggestions, suggestions for responses, and visual guidance through the customer service process.

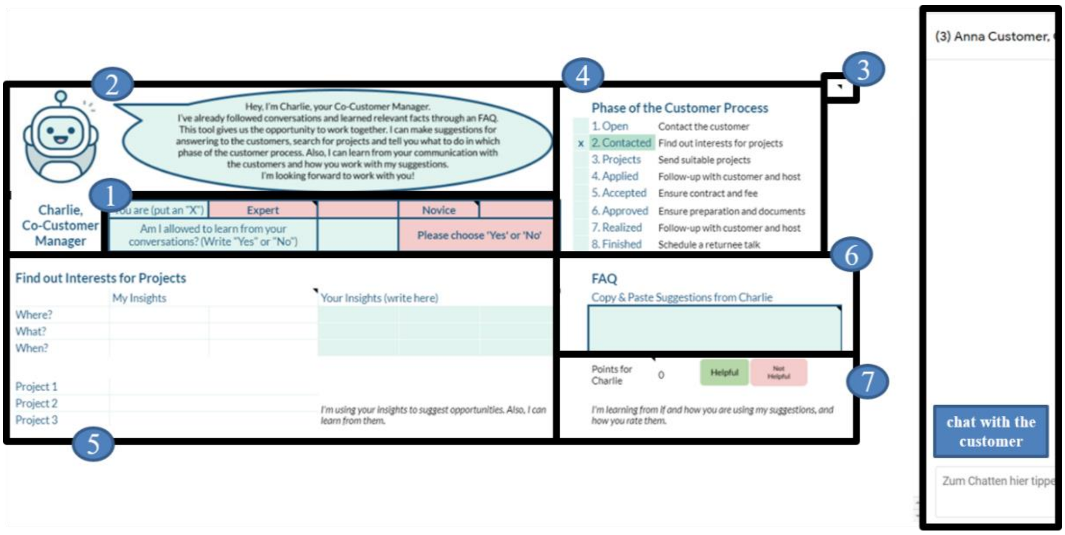

Figure 5. Prototype – user interface of the web application with DFs 1-7[1]

[1] Mir, Irina. *Bot Icon*. URL: https://dribbble.com/shots/4082720-Bot-Icon (visited on 04/30/2021)

In terms of prototyping, our instantiated HIS is limited to the functionalities relevant for executing the basic process in the scope of the research (Wilde and Hess, 2007). Additionally, we applied the WOz technique to simulate the natural language processing (NLP) functions with a hidden wizard for evaluation (Böttcher and Nüttgens, 2013; Salber and Coutaz, 1993; Krannich, 2010; Riek, 2012).

DF1 - AI Learning Behavior Settings. In our prototype, the employee puts an “X” in a field to indicate the experience level of either expert or novice for the AI to adapt its learning. Furthermore, the employee is asked to confirm or not confirm AI learning from the conversations by writing “Yes” or “No” in an according field. Colors and a short text indicate that the AI is or is not learning.

DF2 - AI Identity. In our prototype, the AI has the name “Charlie” and an avatar. Additionally, the AI is defined as “co-customer manager”. It introduces itself in a speech bubble explaining its role and raising motivation to work together. The first-person perspective of the AI is constant in all features.

DF3 – Education on AI. In our prototype, there are hidden comments with more technological in-depth information about the AI features. As they are not mandatory to read but valuable for the employees’ knowledge on the prototype and the AI, they are not shown compulsorily. Consequently, the tool is leaner and the employee is not overwhelmed with information. The employee can just hover over the black triangles in the corner of the features to get more information on demand.

DF4 - Customer Service Process Awareness. In our prototype, this feature visualizes the phases of the customer service process and the goals of each phase. By marking the according row with an “x” the AI highlights the current phase. The employee can also change the position of the “x”. This is supposed to teach the AI toward a better shared understanding of the phases. By means of the WOz technique, rules were provided to the wizard determining when to change the process phase derived from the definitions of the organization X. To simulate NLP drawbacks, every now and then, the process phase was not changed at the right time.

DF5 – Collaborative Knowledge. In our prototype, the AI gathers insights about customer interests from the conversation based on keywords to present them in the “my insights” section. These insights refer to where / when / what the customer wants to go / do abroad. By means of the WOz technique, rules were provided to the wizard determining, which keywords to recognize and to put in the insights. As a keyword search does not require NLP functionalities, the wizard could just align with a predefined list of countries, months, and project types of organization X. If the AI is missing insights from the keyword list, the employee can add insights in the “your insights” section. Based on all insights, the AI finds projects matching the customer interests. The ability to contribute with own insights is supposed to enable the employee to collaborate with the AI. Also, the AI is supposed to learn from the employees' usage of recommendations combined with the collaborative knowledge.

DF6 - AI Output. In our prototype, the AI does not send any messages to the customer. It provides the employee with suggestions on how to answer customers' questions. These answers are taken from an FAQ, which ensures the accuracy of the information. The employees can then decide, how they use the recommendations. Overall they have four options: 1) keep 2) adapt 3) write a new answer based on the AI output, 4) disregard the AI output. As employees need to work with the AI output, learning progress is expected, especially for novice employees. Furthermore, working with the AI output enables the employees to teach the AI toward better recommendations. By means of the WOz technique, rules were provided to the wizard, determining when to give which suggestion from the FAQ. To simulate NLP drawbacks, every now and then, a wrong suggestion was provided.

DF7 – Feedback. In our prototype, employees are asked to optionally reinforce the learning of the AI by clicking the "helpful" or "not helpful" button. Thus, if an FAQ-based suggestion appears to be helpful / not helpful for a customer's question, the employee can click "helpful / not helpful". Furthermore, for each click, the AI receives plus- (helpful), or minus-points (not helpful) as a score.

13.7 Evaluation

To evaluate our DPs toward continuous collaborative learning of customer service employees and AI in a HIS (RQ1), and to assess the differences in learning effects and continuance intention to use between novice and expert employees (RQ2), we conducted a test run with 61 customer service employees of organization X using our prototype in terms of an ex post case study (Venable et al., 2012). We divided the participants into two groups differentiating between 31 experts (E1-E31) and 30 novices (N1-N30). The experts had an average age of 22.90 (SD = 2.12), 16 were male, 15 female. The novices' average age was 24.27 (SD = 1.89), 14 were male, 16 female. We determined participants' experience level using two indicators: 1) self-perception and 2) the score of a pre-assessment of knowledge covering specific questions service employees of the organization should be able to answer. The test runs lasted around 60 minutes including the prototype usage of around 20 minutes, comprising the following five main parts:

1) Pre-assessment: As a baseline for learning, the participants were asked ten questions about specific information customers might need from a service employee. The maximum achievable score was 16.

2) Execution: Before using the prototype, each participant received an introduction about their role, the goal, the prototype, and its features. We limited the customer service process to the initial customer contact until sending project proposals, and answering customer questions. The test runs lasted around 20 minutes. We applied the WOz technique with a hidden wizard to simulate certain functions and observe the participants working with an apparently fully functioning prototype (Böttcher and Nüttgens, 2013; Salber and Coutaz, 1993; Krannich, 2010; Dahlbäck and Jönsson, 1989). To ensure the success of the WOz technique, we created a consistent setup with rules for all features simulated by the wizard (Salber and Coutaz, 1993; Dahlbäck and Jönsson, 1989). Furthermore, we asked the participants to share their screen during the test run in order to observe and follow their actions (Krannich, 2010).

3) Post-assessment (quantitative): To assess learning progress, we conducted a post-assessment asking the same questions from the pre-assessment. The maximum achievable score was 16. The learning progress achieved by the participants was calculated by subtracting the first from the last score.

4) Ratings (quantitative): We aim to identify differences in the constructs of learning and continuance intention to use between novice and expert employees. Apart from the post-assessment, we measured dependent variables according to the constructs based on self-reports (Samarasinghe and Tretiakov, 2009). Therefore, we asked the participants to rate the following statements on a Likert scale of seven points (Johns, 2010; Likert, 1932). A) Learning: Through working with Charlie I feel like I (1) *gain knowledge*, (2) *gain experience,* (3) *gain interesting or alternative insights*, (4) *can teach Charlie*. B) Continuance Intention to Use: (5) *I am satisfied working with Charlie*. I can imagine continuing working with Charlie in (6) *real customer service situations*, (7) *customer service simulations*. I would continue working with Charlie to (8) *gain knowledge*, (9) *gain experience*, (10) *gain insights*, (11) *teach Charlie*.

5) User interviews (qualitative): We ensure data triangulation (Mayring, 2001) by combining the quantitative with a qualitative evaluation. We conducted expert interviews with all 61 participants to assess the implemented DFs and the DPs respectively (Meuser and Nagel, 2002). The interview guideline consisted of seven questions, each addressing one DF and DP respectively. Accordingly, we deductively coded the results along them. When analyzing the results, we further inductively derived subcodes toward learning, satisfaction, and continuance intention to use.

13.7.1 Quantitative results

To identify differences between novices and experts, we first tested for normal distribution. As the sample sizes are relatively small (nnovices = 30 and nexperts = 31), we conducted a Shapiro-Wilk test with a significance level of $\alpha = 0.05$. For each variable, the test indicated that the data is not normally distributed. Thus, we used a two-tailed Mann-Whitney-U test to evaluate the difference between the two groups for the considered dependent variables. The mean, median, standard deviation scores, and the results of the Mann-Whitney U Test (p-Value) are depicted in Table 2. For the highlighted dependent variables, we identified a statistically significant difference ($p<0.05$) between novices and experts.

Table 2. Mean, Median, Standard Deviation scores, and Mann-Whitney U Test results for the dependent variables related to learning and continuance intention to use.

Construct	Dependent Variable	Group	Mean	Median	Standard Deviation	p-Value
Learning	Learning Progress	Novices	5.43	5	2.76	< 0.00001[b]
		Experts	1.16	1	1.21	
	Perception of Knowledge Gain	Novices	6.03	7	1.25	< 0.00001[b]
		Experts	3.97	4	1.92	
	Perception of Experience Gain	Novices	5.53	6	1.66	0.00222[b]
		Experts	4.00	4	1.95	
	Perception of Insights Gain	Novices	5.73	6	1.44	0.00194[b]
		Experts	4.42	5	1.71	
	Perception of AI Teachability	Novices	4.50	5	2.00	0.27572
		Experts	5.19	5	1.28	
Continuance Intention to Use	Satisfaction	Novices	6.27	6	0.83	0.238
		Experts	6.03	6	0.80	
	Continuance Intention for Real Situations	Novices	6.63	7	0.67	0.39532
		Experts	6.52	7	0.63	
	Continuance Intention for Simulations	Novices	6.60	7	0.72	0.88866
		Experts	6.65	7	0.71	
	Continuance Intention for Knowledge Gain	Novices	5.90	6	1.42	0.00022[b]
		Experts	3.97	4	1.99	
	Continuance Intention for Experience Gain	Novices	5.63	6	1.47	0.00466[b]
		Experts	4.23	4	1.96	
	Continuance Intention for Insights Gain	Novices	5.50	6	1.46	0.04236[b]
		Experts	4.52	5	1.88	
	Continuance Intention for Teaching AI	Novices	5.17	5.5	1.70	0.86502
		Experts	5.16	5	1.55	

a Scores measured on 7-point Likert scales (1 = low, 7 = high), except "learning progress" measured with a score of minimum 0 and maximum 16.
b The difference between the means of the groups is statistically significant at $p<0.05$.

Following the results from Table 2, we can identify several statistically significant differences ($p<0.05$) between novices and experts regarding learning and continuance intention to use. For four dependent variables according to the construct **learning**, we identified a statistically significant difference between novices and experts showing that the novices learn more from working with the AI. Additionally, the difference is also confirmed by the differences of the pre- and post-assessment of knowledge reflected by the change in the number of correct answers in the test. Regarding the AI learning and the perception of AI teachability, we did not identify a statistically significant difference. However, both groups are slightly prone to teaching the AI with an average of 4.50 (novices) and 5.19 (experts). For the construct of **continuance intention to use**, we did not find statistically significant differences between experts and novices. Both groups are very satisfied with an average of 6.27 (novices) and 6.03 (experts) with the prototype and both gave high scores for the continuance intention to use for both real situations (averages: 6.63 from novices, 6.52 from experts) and simulations (averages: 6.60 from novices, 6.65 from experts). This is a promising signal for continuous learning for novices as well as teaching from experts. Regarding the continuance intention for teaching, there is also no statistically significant difference between experts and novices. However, with an average of 5.16, experts are more likely to continue using the system for teaching the AI than for learning from the AI. Thus, we found that novices have a significantly higher continuance intention to use for learning than experts.

13.7.2 Qualitative evaluation

We asked all participants (E1-31, N1-30) open questions about the DFs to address the according DPs. Overall, the participants' satisfaction from the quantitative evaluation was also reflected in interviews as results signaled that the seven DPs were successfully implemented within the instantiated prototype encouraging collaborative learning and a high continuance intention to use of novices and experts. Thereby, the perceived effect on learning and continuance intention varied for each DF. First, participants perceived the customer service process visualization with process knowledge (DP4) and the FAQs with content knowledge (DP6) as elements with most contribution to human learning itself, e.g., DF4: "*Overall I was really impressed […]. Overall, it gives the user a better understanding of how to manage the customer […] what they have to do right now just from this indicator.*" (E8); DF6: "*I liked it and I also used it […]. I also like that you can change things. I think it was helpful because he also provided things I didn't knew. So it was helpful […].*" (E23). Second, they recognized and appreciated the influence of the AI learning behavior settings (DP1), the education on AI (DP3) and the feedback buttons (DP7), which ensure a framework for the users to understand, setup and teach the learning system, e.g., DF1: "*I think it's great because I guess if I am a novice I would learn more from it. Probably he can rely more on experts' information. And if people don't want it to learn, they would feel more safe.*" (E16); DF3: "*It was helpful because I could read through it at anytime to understand what is happening and how it works without asking the researcher. […] it was* *appropriate to put them in hidden notes to not distract from the customer support process.*" (E15); DF7: "*I really like them because he also helps us, so we also help him. And I could see how often did he help me. I really liked that he can learn from this.*" (N24). Third, the participants found the collaborative insights (DP5) very useful as they further provide knowledge, which rather contributes to efficiency in the customer service process, e.g., DF5: "*First of all, I found it very helpful that he was summarizing the information from the chat. Otherwise, I would have needed to scroll up the chat. With the recommendations, I could get a direction and based on this find other projects*" (N12). Thus, most participants stated that they would like to implement and continue using the HIS in practice. At last, most participants were positive about the AI identity (DP2) and confirmed a comfortable hybrid co-working environment, e.g., DF2: "*It's nice. So of course Charlie is an AI and not a real person. […] But it is really cool to have a name and an avatar. It is much more fun to have such a co-worker. He introduced himself clearly. He doesn't come to the foreground, stays in the background, and that's nice*" (N18). The most recognized and appreciated features were based on DP5 and DP6 due to efficiency and required information and knowledge. Thus, for effective and efficient task execution, users state to rather need the FAQ knowledge and the collaborative insights. However, most participants explained that they would have also taken a closer look at the other features if there was no time limit.

13.8 Discussion

Overall, our study contributes to HIS research combining HITL and CITL (Dellermann et al., 2019a) and to implications of AI in organizations in terms of mutual augmentation (Benbya et al., 2021). We take a novel perspective by distinguishing its human users by knowledge and experience. In fact, we leverage the different knowledge levels of expert and novice users to contribute to an iterative collaborative learning cycle of human users and AI toward hybrid intelligence (Wiethof and Bittner, 2021), i.e., expert users teaching AI (HITL) and novice users learning from AI (CITL). To design a HIS, which enables hybrid collaborative learning, we formulated seven DPs relying on 16 MRs based on expert interviews and concepts of hybrid collaborative learning (Kulesza et al., 2015; Dellermann et al., 2019a) and the information systems continuance model (Bhattacherjee, 2001) in the domain of customer service. Based on the quantitative results of our test runs, our DPs successfully enabled our prototype for continuous collaborative learning. The additional qualitative user interviews indicated that the DPs derived through the combination of both correctability and explainability MRs (DP4-6) have the most impact on the learning progress and continuance intention to use. This is because they ensure necessary knowledge and information provision to the users – explainability – which the users can then correct, adapt and use – correctability – enabling learning for both employee and AI as well as satisfaction. This ties in well with other existing design knowledge confirming explainability and correctability within HIS as core concepts. For instance, Ostheimer et al. (2021) derive the principle of power relationship, which entitles the human to keep the power over the AI. Thus, though the AI can function autonomously, e.g., by augmenting the human with guiding information or suggestions (Ostheimer et al., 2021; Zschech et al., 2021; Dellermann et al., 2019c), the human can still intervene and change the AI's contributions (Ostheimer et al., 2021; Zschech et al., 2021; Wiethof et al., 2021). Overall, DP4-6 do not only contribute to learning and continuance intention to use, but also to a shared understanding and visual guidance (Dellermann et al., 2019c), a clear division of tasks between human and AI (Ostheimer et al., 2021), and their collaboration (Zschech et al., 2021; Ostheimer et al., 2021). Future research might build on these findings, e.g., DP4 and DP5 can further be evaluated regarding work efficiency. Therefore, further test runs might involve the communication with several customers, while visualization of the customer service process and shared information ensure an updated understanding of each customer. Regarding DP6, future research can examine the learning progress when using the prototype for a longer timeframe, e.g., over days or weeks, or the impact of different kinds of AI outputs, e.g., text or bullet points. Additionally, implementing NLP capabilities will be crucial for the next design cycle for more representative results (Wiethof et al., 2021). The other DPs correspondingly establish a frame and setting for a hybrid collaboration environment. Thus, the DPs derived through explainability MRs (DP2, DP3) help users to understand the learning system and feel comfortable within the hybrid environment. This is in line with existing research on humanizing AI (Wiethof et al., 2021) as well as transparency of AI (Zschech et al., 2021; Wiethof et al., 2021). The DPs derived through correctability MRs (DP1, DP7) ensure the users' final control within the HIS. Especially regarding DP1, the necessity of feedback mechanisms can be confirmed (Dellermann et al., 2019c). However, additional information is needed for self-assessing the experience level (DP1) as well as for understanding the effects of the buttons and when exactly to use them (DP7). Furthermore, our study rather focuses on how the human is learning within the HIS. Considering Dellermann et al. (2019c), future research could further focus on establishing appropriate qualitative and quantitative feedback mechanisms for teaching the AI.

13.9 Conclusion

All in all, our research provides promising results, which show potential for HIS in customer service toward hybrid collaborative learning. We establish a hybrid collaborative learning cycle between humans and AI enabling human experts to teach AI and human novices to learn from AI (see Figure 4) (Dellermann et al., 2019a; Kulesza et al., 2015; Bhattacherjee, 2001). With this, we contribute design knowledge in the form of seven DPs for designing and developing a HIS in customer service toward collaborative learning (Gregor and Hevner, 2013; Gregor, 2006) as well as an instantiated and evaluated a prototype following the DPs (RQ1). We could confirm the successful implementation of the DPs in the form of DFs and found that the learning progress of novices is significantly higher than the one of experts. Also, the continuance intention to use the prototype in terms of learning was significantly higher for novices. However, both groups showed a high satisfaction and continuance intention to use suggesting a continuous hybrid collaborative learning cycle (see Figure 4) (RQ2). Finally, our study provides implications for both research and practice. First, researchers may increasingly differentiate expert and novice users for more realistic user studies. Second, they can draw on our DPs and prototype to investigate HIS in customer service and other domains. Third, practitioners may consider HIS as an opportunity for implicit knowledge transfer and use our findings to develop HIS.

Besides the promising results of this research, there are a few limitations to consider. First, by making use of the WOz technique, there is a severe bias when looking at the overall results, as NLP is responsible for properly understanding and generating language and thereby interacting with the user. In fact, drawbacks in NLP technologies are also crucial for the overall user satisfaction. Second, measuring continuance intention to use does not necessarily need to align with actual user behavior. We only considered one short-term use case of a specific organization. Thus, we call for future research assessing our DPs and user behavior over a more extended period, with larger samples and within different use cases. Third, by missing NLP functionalities, we could not properly evaluate AI learning in terms of HITL. However, by confirming the willingness to teach AI, we call for future research to focus on HITL, develop adequate feedback mechanisms, and assess AI learning based on NLP. Eventually, our results show great potential of HIS in customer service and provide a great starting point for further development and usage of our DPs and prototypic system. In the next cycle of our DSR project, we will build on the findings of this study and instantiate and evaluate a fully functioning HIS prototype accordingly.

13.10 References

Abdel-Karim, B. M., N. Pfeuffer, G. Rohde and O. Hinz (2020). “How and What Can Humans Learn from Being in the Loop?” *KI - Künstliche Intelligenz* 34, 199–207.

Amershi, S., M. Cakmak, W. B. Knox and T. Kulesza (2014). “Power to the People: The Role of Humans in Interactive Machine Learning” *AI Magazine* 35 (4), 105–120.

Benbya, H., S. Pachidi and S. L. Jarvenpaa (2021). “Artificial Intelligence in Organizations: Implications for Information Systems Research” *Journal of the Association for Information Systems* 22 (2), 282–303.

Bhattacherjee, A. (2001). “Understanding Information Systems Continuance: An Expectation-Confirmation Model” *MIS Quarterly* 25 (3), 351–370.

Böttcher, B. and M. Nüttgens (2013). “Überprüfung der Gebrauchstauglichkeit von Anwendungssoftware” *HMD Praxis der Wirtschaftsinformatik* 50, 16–25.

Brandt, M. (2021). *Statistik der Woche: Wo KI in der deutschen Wirtschaft boomt*. URL: https://www.heise.de/news/Statistik-der-Woche-Wo-KI-in-der-deutschen-Wirtschaft-boomt-6172430.html (visited on 09/08/2021).

Chandra, L., S. Gregor and S. Seidel (2015). “Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions”. In: *48th Hawaii International Conference on System Sciences (HICSS)*. Kauai, Hawaii, USA, pp. 4039–4048.

Chhibber, N. and E. Law (2019). “Using Conversational Agents To Support Learning By Teaching”. In: *CHI*. Glasgow, Scotland, UK.

Dahlbäck, N. and A. Jönsson (1989). “Empirical Studies Of Discourse Representations For Natural Language Interfaces”. In: *4th Conference of the European Chapter of the ACL*, pp. 291–298.

Dellermann, D., A. Calma, N. Lipusch, T. Weber, S. Weigel and P. Ebel (2019a). “The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems”. In: *52nd Hawaii International Conference on System Sciences (HICSS)*. Hawaii, USA, pp. 274–283.

Dellermann, D., P. Ebel, M. Söllner and J. M. Leimeister (2019b). “Hybrid Intelligence” *Business & Information Systems Engineering* 61, 637–643.

Dellermann, D., N. Lipusch, P. Ebel and J. M. Leimeister (2019c). “Design principles for a hybrid intelligence decision support system for business model validation” *Electronic Markets* 29 (3), 423–441.

Dubey, A., K. Abhinav, S. Jain, V. Arora and A. Puttaveerana (2020). “HACO: A Framework for Developing Human-AI Teaming”. In: *13th Innovations in Software Engineering Conference (ISEC)*. Jabalpur, India.

Gavriushenko, M., O. Kaikova and V. Terziyan (2020). “Bridging human and machine learning for the needs of collective intelligence development” *Procedia Manufacturing* 42, 302–306.

Gregor, S. (2006). “The Nature of Theory in Information Systems” *MIS Quarterly* 30 (3), 611–642.

Gregor, S. and A. Hevner (2013). “Positioning and Presenting Design Science Research for Maximum Impact” *MIS Quarterly* 37 (2), 337–355.

Hjorth, A. (2021). “NaturalLanguageProcesing4All - A Constructionist NLP tool for Scaffolding Students’ Exploration of Text”. In: *17th ACM Conference on International Computing Education Research*. Virtual Event, USA, pp. 347–354.

Holzinger, A. (2016a). “Interactive Machine Learning (iML)” *Informatik Spektrum* 39 (1), 64–68.

Holzinger, A. (2016b). “Interactive machine learning for health informatics: when do we need the human-in-the-loop?” *Brain Informatics* 3, 119–131.

Holzinger, A., M. Plass, K. Holzinger, G. C. Crisan, C.-M. Pintea and V. Palade (2016). “Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to Solve the Traveling Salesman Problem with the Human-in-the-Loop Approach”. In: *IFIP International Cross Domain Conference and Workshop*. Salzburg, Austria, pp. 81–95.

Holzinger, A., M. Plass, K. Holzinger, G. C. Crisan, C.-M. Pintea and V. Palade (2017). "A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop" *ArXiv* 1708.01104.

Holzinger, A., M. Plass, M. Kickmeier-Rust, K. Holzinger, G. C. Crisan, C.-M. Pintea and V. Palade (2019). "Interactive machine learning: experimental evidence for the human in the algorithmic loop. A case study on Ant Colony Optimization" *Applied Intelligence* 49, 2401–2414.

Hu, W.-L., C. Rivetta, E. MacDonald and D. P. Chassin (2019). "Optimal Operator Training Reference Models for Human-in-the-loop Systems". In: *52nd Hawaii International Conference on System Sciences (HICSS)*. Hawaii, USA, pp. 3524–3531.

Huang, M.-H. and R. T. Rust (2018). "Artificial Intelligence in Service" *Journal of Service Research* 21 (2), 155–172.

Hun Lee, M., A. Bernardino and S. Bermúdez i Badia (2021). "A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment". In: *CHI*. Yokohama, Japan.

Johns, R. (2010). *Likert Items and Scales*. URL: https://ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf.

Keyser, A. de, S. Köcher, L. Alkire, C. Verbeeck and J. Kandampully (2019). "Frontline Service Technology infusion: conceptual archetypes and future research directions" *Journal of Service Management* 30 (1), 156–183.

Koppenhagen, N., O. Gaß and B. Müller (2012). "Design Science Research in Action - Anatomy of Success Critical Activities for Rigor and Relevance". In: *20th European Conference on Information Systems (ECIS)*. Barcelona, Spain.

Krannich, D. (2010). "Mobile Usability Testing. Ein toolbasiertes Vorgehensmodell zum Rapid-Prototyping und Usability-Testing von Mobilen Systemen im originären Benutzungskontext.". Dissertation. Universität Bremen.

Kühl, N., M. Goutier, R. Hirt and G. Satzger (2019). "Machine Learning in Artificial Intelligence: Towards a Common Understanding". In: *52nd Hawaii International Conference on System Sciences (HICSS)*. Hawaii, USA, pp. 5236–5245.

Kulesza, T., M. Burnett, W.-K. Wong and S. Stumpf (2015). "Principles of Explanatory Debugging to Personalize Interactive Machine Learning". In: *IUI*. Atlanta, GA, USA, 126-137.

Larivière, B., D. Bowen, T. W. Andreassen, W. Kunz, N. J. Sirianni, C. Voss, N. V. Wünderlich and A. de Keyser (2017). ""Service Encounter 2.0": An investigation into the roles of technology, employees and customers" *Journal of Business Research* 79, 238–246.

Likert, R. (1932). *A technique for the measurement of attitudes.* New York: Archives of psychology.

Liu, J., A. Wilson and D. Gunning (2014). "Workflow-based Human-in-the-Loop Data Analytics". In: *Workshop on Human Centered Big Data Research (HCBDR)*. Raleigh, NC, USA, pp. 49–52.

Liu, M., Y. Zhang, G. Zhao, C. Hu, Q. Ding, J. Gong, P. Xu, L. Zhang and Q. Wang (2020). “Cold Comfort Matters - How Channel-Wise Emotional Strategies Help in a Customer Service Chatbot”. In: *CHI*. Honolulu, HI, USA.

Martínez, M. A. M., M. Nadj and A. Maedche (2019). “Towards An Integrative Theoretical Framework Of Interactive Machine Learning Systems”. In: *27th European Conference on Information Systems (ECIS)*. Stockholm-Uppsala, Sweden.

Martins Giraffa, L. M. and R. M. Viccari (1998). “The Use of Agents Techniques on Intelligent Tutoring Systems”. In: *18th International Conference of the Chilean Society of Computer Science*, pp. 76–83.

Mayring, P. (2001). “Combination and Integration of Qualitative and Quantitative Analysis” *Forum: Qualitative Social Research* 2 (1).

Meuser, M. and U. Nagel (2002). *ExpertInneninterviews - vielfach erprobt, wenig bedacht.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Molino, P., H. Zheng and Y.-C. Wang (2018). “COTA: Improving the Speed and Accuracy of Customer Support through Ranking and Deep Networks”. In: *24th International Conference on Knowledge Discovery & Data Mining (KDD)*, pp. 586–595.

Oliveira, C. S. de, C. Sanin and Szczerbicki (2020). “Human Feedback and Knowledge Discovery: Towards Cognitive Systems Optimization”. In: *24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, pp. 3093–3102.

Oliver, R. L. (1980). “A Cognitive Model for the Antecedents and Consequences of Satisfaction” *Journal of Marketing Research* 17, 460–469.

Ostheimer, J., S. Chowdhury and S. Iqbal (2021). “An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles” *Technology in Society* 66.

Paschen, J., M. Wilson and J. J. Ferreira (2020). “Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel” *Business Horizons* 63 (3), 403–414.

Peffers, K., T. Tuunanen, C. E. Gengler, M. Rossi, W. Hui, V. Virtanen and J. Bragge (2006). “The Design Science Research Process: A Model For Producing And Presenting Information Systems Research”. In: *International Conference on Design Science Research in Information Systems and Technology*. Claremont, CA, pp. 83–106.

Poser, M., S. Singh and E. Bittner (2021). “Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents”. In: *54th Hawaii International Conference on System Sciences (HICSS)*. virtual, pp. 1181–1190.

Riek, L. D. (2012). “Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines” *Journal of Human-Robot Interaction* 1 (1), 119–136.

Robinson, S., C. Orsingher, L. Alkire, A. de Keyser, M. Giebelhausen, K. N. Papamichail, P. Shams and M. S. Temerak (2020). “Frontline encounters of the AI kind: An evolved service encounter framework” *Journal of Business Research* 116, 366–376.

Russell, S. J. and P. Norvig (2016). *Artificial intelligence: a modern approach*.

Rzepka, C. and B. Berger (2018). "User Interaction with AI-enabled Systems: A Systematic Review of IS Research". In: *39th International Conference on Information Systems*. San Francisco.

Sajeev, S., J. Huang, N. Karampatziakis, M. Hall, S. Kochman and W. Chen (2021). "Contextual Bandit Applications in a Customer Support Bot". In: *27th ACM SIGKDD Conference on Knowledge*, pp. 3522–3530.

Salber, D. and J. Coutaz (1993). "Applying the Wizard of Oz Technique to the Study of Multimodal Systems". In: *3rd International Conference on Human-Computer Interaction*. Moscow, Russia, pp. 219–230.

Samarasinghe, S. M. and A. Tretiakov (2009). "A multi-dimensional measure of e-learning systems success". In: *ascilite*. Auckland, pp. 908–910.

Schneider, J. and J. P. Handali (2019). "Personalized Explanation For Machine Learning: A Conceptualization". In: *27th European Conference on Information Systems (ECIS)*. Stockholm-Uppsala, Sweden.

Seeber, I., E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe and M. Söllner (2020). "Machines as teammates: A research agenda on AI in team collaboration" *Information & Management* 57 (2).

Shneiderman, B. (2020). "Human-Centered Artificial Intelligence: Three Fresh Ideas" *AIS Transactions on Human-Computer Interaction* 12 (3), 109–124.

Sousa, D. N., M. A. Brito and C. Argainha (2019). "Virtual Customer Service: Building your Chatbot". In: *ICBIM*. Paris, France, pp. 174–179.

Subramaniam, S., P. Aggarwal, G. B. Dasgupta and A. Paradkar (2018). "COBOTS - A Cognitive Multi-Bot Conversational Framework for Technical Support". In: *17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Stockholm, Sweden, pp. 597–604.

Sun, W., S. Zhang, K. Balog, Z. Ren, P. Ren, Z. Chen and M. de Rijke (2021). "Simulating User Satisfaction for the Evaluation of Task-oriented Dialogue Systems". In: *SIGIR*. Virtual Event, Canada, pp. 2499–2506.

Svenningsson, N. and M. Faraon (2019). "Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots". In: *AICCC*. Kobe, Japan, pp. 151–161.

Venable, J., J. Pries-Heje and R. Baskerville (2012). "A Comprehensive Framework for Evaluation in Design Science Research". In: *7th International Conference on Design Science Research in Information Systems and Technology (DESRIST)*. Las Vegas, NV, USA, pp. 423–438.

Wambsganss, T., T. Küng, M. Söllner and J. M. Leimeister (2021). "ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills". In: *CHI*. Yokohama, Japan.

Wiethof, C. and E. Bittner (2021). "Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review". In: *42nd International Conference on Information Systems*. Austin.

Wiethof, C., N. Tavanapour and E. Bittner (2021). "Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI". In: *54th Hawaii International Conference on System Sciences (HICSS)*. virtual, pp. 400–409.

Wilde, T. and T. Hess (2007). "Forschungsmethoden der Wirtschaftsinformatik. Eine empirische Untersuchung" *WIRTSCHAFTSINFORMATIK* 49, 280–287.

Xu, Y., C.-H. Shieh, P. van Esch and I.-L. Ling (2020). "AI customer service: Task complexity, problem-solving ability, and usage intention" *Australasian Marketing Journal* 28 (189-199).

Yimam, S. M., C. Biemann, L. Majnaric, S. Sabanovic and A. Holzinger (2016). "An adaptive approach for biomedical entity and relation recognition" *Brain Informatics* 3, 157–168.

Zhang, X., L. Olfman and D. Firpo (2011). "An Information Systems Design Theory for Collaborative ePortfolio Systems". In: *44th Hawaii International Conference (HICSS)*. Kauai, Hawaii, USA, pp. 1–10.

Zschech, P., J. Walk, K. Heinrich, M. Vössing and N. Kühl (2021). "A Picture is Worth a Collaboration: Accumulating Design Knowledge for Computer-Vision-Based Hybrid Intelligence Systems". In: 29th European Conference on Information Systems (ECIS). virtual.

14 Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions

Poser, M., Wiethof, C., and Bittner, E. A. C. 2022b. “Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions,” in *30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Abstract. Artificial Intelligence (AI) is increasingly deployed in customer service for various service delivery tasks. Research and practice alike have extensively dealt with the use, benefits, and effects of AI solutions in customer service contexts. Nevertheless, knowledge on AI integration is dispersed and unsystematized. This paper addresses this gap by presenting a taxonomy to inform design decisions for the integration of AI into customer service with five meta-dimensions, 12 dimensions, and 32 characteristics. Through a rigorous and systematic development process comprising multiple iterations and evaluation episodes, state-of-the-art AI solutions from practice and the current state of knowledge from research were systematized to classify AI use cases. Thus, we contribute with systemized design knowledge to, both, the theoretical knowledge base as well as to practice for application. Eventually, we disclose future research avenues addressing certain meta-dimensions as well as the extension of the taxonomy itself.

Keywords: Artificial Intelligence, Customer Service, AI Integration.

14.1 Introduction

Customer service is currently undergoing a radical transformation driven by the integration of machine learning (ML), natural language processing (NLP), and related technologies, which are often subsumed under the term artificial intelligence (AI). In line with Gartner’s prediction that 15 % of customer service interactions will be handled through AI by 2021 (Gartner, 2019), the successive application of AI is currently revolutionizing customer service toward the service encounter 2.0 (Larivière et al., 2017). Due to their advancing capabilities to autonomously handle inquiries, AI-enabled technologies, such as conversational agents (CAs), are implemented in various business contexts (e.g., finance, e-commerce, IT support) to elevate the efficiency and cost-effectiveness of text-based service delivery (Dwivedi et al., 2021; Gartner, 2019; Sarker, 2021; Xu et al., 2020). Thereby, organizations are able to enhance the availability and accessibility of their service provision as well as to reduce service employees’ (SEs) workload, who can focus on more complex requests. Accordingly, AI progressively substitutes tasks of frontstage SEs, such as responding to customers’ requests (Huang and Rust, 2018; Davenport et al., 2020).

Related research, inter alia, involves the advancement of autonomous service delivery by focusing on customers' experience with AI addressing the representation and behavior of CAs, e.g., assigning social cues and ensuring competence levels (Gnewuch et al., 2017; Adam et al., 2020).

However, despite technological advancements, AI is still far away from fully substituting human intelligence beyond narrow domains (Dellermann et al., 2019). This means that AI can so far reliably handle simple requests for which unique relationships between the problem and solution have been established through training. For more complex requests with a distinct problem but multiple solutions, AI still regularly provides unsuitable answers (Krogh, 2018; Levy, 2018). Therefore, AI and human intelligence should be combined to allow SEs and AI to work side-by-side and foster their collaborative interplay (Wilson and Daugherty, 2018; Wirtz et al., 2018). In this vein, AI-based customer service solutions can support organizational service delivery by displaying answers to SEs to facilitate their inquiry processing. Additionally, AI can recognize intentions and emotions of the inquirer via natural language understanding leading to improved value co-creation during customer-SE interaction (Canhoto and Clear, 2020; Bassano et al., 2020; Sujata et al., 2019). Moreover, SEs can complement AI in various ways, e.g., through training, explaining, and sustaining (Keyser et al., 2019; Dellermann et al., 2019).

To realize efficient service delivery involving SEs and AI, a systematic orchestration of their capabilities and weaknesses is required (Paluch and Wirtz, 2020). Hence, the adoption of AI in customer service demands the differentiation between the roles of SEs and AI and the determination of the interaction with each other and the customer (Larivière et al., 2017; Robinson et al., 2020). Despite the increased interest in research and practice to deploy ML-based AI technology for online customer service, insights on how to integrate it into organizations are scarce (Benbya et al., 2021). Related to this, there is a lack of knowledge in research regarding the interrelationships between SE, customer, and AI within a socio-technical system of an organization's customer service (Bock et al., 2020). This includes the embedding in organizational work and process structures as well as the forms of interaction between SE, customer, and AI (Bock et al., 2020; Keyser et al., 2019). To address these knowledge gaps and provide systemized knowledge about the integration of AI for text-based customer service, the following research question is addressed: *How can conceptual and empirical knowledge on the integration of AI in customer service be classified to provide design decision guidance?*

We develop a taxonomy to inform design decisions by adopting the perspective of a single AI use case, which is analyzed or planned for implementation. Thereby, we aim to contribute to, both, the theoretical knowledge base as well as to practice for application with systemized knowledge from research and commercial solutions. Regarding theory, we provide relevant characteristics to be considered when investigating AI in different stages of the customer service process. Considering practical and managerial implications for IT management and development, businesses can advance their existing customer service delivery or implement novel interaction types aligned to the dimensions and characteristics of the taxonomy. To address the research question, the paper is structured as follows: First, we give an overview of related work about customer service and AI. After that, we introduce our research approach including the taxonomy development process. We then present an evaluation of our taxonomy prior to completion followed by the description of the final taxonomy organized and aligned to each meta-dimension. Next, we report on the ex-post evaluation of our taxonomy. We close the paper with a discussion and conclusion.

14.2 Related Work: Customer Service and AI

Service represents an elementary category of industrialized economies and is defined as the "application of competences (knowledge and skills) by one entity for the benefit of another" (Sampson and Froehle, 2006; Maglio and Spohrer, 2008; Vargo et al., 2008, p. 145). A relevant field of service represents companies' customer service offerings in various industries, which typically refer to intangible service delivery directed toward people (e.g., consultancy) or objects (e.g., post-sales service for primary products) (Wirtz et al., 2018). To fulfill customers' needs and demands, this form of service delivery is prevailingly characterized by knowledge intensity and customization, which requires active participation of and input by customers during service provision (Maglio and Spohrer, 2008). As companies strive to deliver high quality service to satisfy customers, a complex set of service processes needs to be orchestrated that spans the complementary service environments frontstage (external) and backstage (internal) (Sampson and Froehle, 2006). In the frontstage, service encounters with customers take place to co-create service. The backstage covers processes that do not directly involve customers and are therefore invisible to them (Glushko and Tabas, 2008; Bock et al., 2020).

To increase service quality and customer satisfaction, research has focused on factors that increase the efficiency and effectiveness in these service environments (Brady et al., 2002; Bitner et al., 2000). In this context, investigations emerged that, inter alia, examine the utilization of technology to create innovative ways of providing, accessing, and manipulating information in the front- and backstage (Amorim et al., 2019). Accordingly, the accessibility and availability of service have been addressed with technology-based self-service concepts such as knowledge portals on websites (e.g., Scherer et al., 2015; Meuter et al., 2000). Furthermore, access to and reuse of knowledge in accordance with customers' inquiries has been improved for SEs, e.g., with repositories (Kankanhalli et al., 2011). In this way, research has accounted for the time-critical, complex, and knowledge-dependent nature of service delivery in customer service contexts (Froehle and Roth, 2004). As an extension of these technology-focused research efforts, recent endeavors focus on the role of AI in customer service (Bock et al., 2020). With its capacity to process and learn based on data, AI is capable of inferring solutions to problems, decision options, or executing actions (Campbell et al., 2020; Raj and Seamans, 2019; Davenport et al., 2020).

Hence, the utilization of current narrow AI that bases on ML algorithms is considered to revolutionize service delivery by efficiently and cost-effectively automating service encounters and tasks (Huang and Rust, 2018; Østerlund et al., 2021). This transforms information-rich online customer service since AI is capable of partially substituting or augmenting service activities. To account for this, Ostrom et al. (2019) and Keyser et al. (2019) introduced infusion archetypes for frontstage service delivery involving the entities AI, customer, and SE: AI either substitutes SEs by autonomously performing customer encounters or augments SEs by supporting them invisibly or visibly to customers through providing relevant information synchronous to the customer interaction. In these settings, customers and SEs encounter AI in the form of an AI-enabled agent and/or embedded AI. The former is a virtually represented agent that facilitates human-like interaction via natural language (e.g., CA), whereas the latter is integrated into platforms or applications without virtual identity (e.g., ticket tool) (Glikson and Woolley, 2020). AI-enabled agents, such as CAs, have been predominantly developed and investigated to substitute mechanical and analytical tasks that require rule-based, systematic, and consistent processing involving data and information (Huang and Rust, 2018; Janssen et al., 2020). Therefore, CA designs focus on interaction and technical capabilities to process customers' inquiries by answering questions or solving problems (Gnewuch et al., 2017; Følstad and Skjuve, 2019; Luger and Sellen, 2016). For the backstage, embedded AI is capable of delivering insights about past inquiries and/or historical customer data to support SEs (Graef et al., 2020; Cheung et al., 2003). Complementing these studies, initial research considers the interconnection of front- and backstage processes and tasks with seamless handovers from CAs to SEs to avoid failure in AI-performed service encounters (Wintersberger et al., 2020; Poser et al., 2021).

The overall focus of these previous studies predominantly lies on the development of stand-alone solutions for AI-performed service encounters in the frontstage. In addition, so far, there is limited knowledge about the role, activities, and integration of AI in the backstage. In principle, systematic knowledge with a holistic perspective on the integration of AI into customer service covering front- and backstage is until now scarce (Bock et al., 2020).

14.3 Research Approach

This paper aims to shed light on relevant design decisions for the integration of AI into the front- and/or backstage of customer service contexts by identifying and systematizing integration characteristics. For this purpose, dimensions related to service processes and the interaction between AI and humans (SEs and customers) are explored. As this still represents a nascent phenomenon, for which existing knowledge has not yet been structured and organized, a classification of associated concepts can help to consolidate understanding and further sense-making in this complex domain. For this endeavor, taxonomies are a suitable method, as they ascertainably present relationships, commonalities, and differences of concepts (Kundisch et al., 2021; Nickerson et al., 2013; Bailey, 1994). Following Kundisch et al. (2021), we rely on the Design Science Research (DSR) paradigm by adopting a build-evaluate pattern to construct and assess our taxonomy (Hevner et al., 2004; Sonnenberg and vom Brocke, 2012). Accordingly, our research approach comprises two consecutive process phases: (1) development and (2) evaluation (see Figure 1).

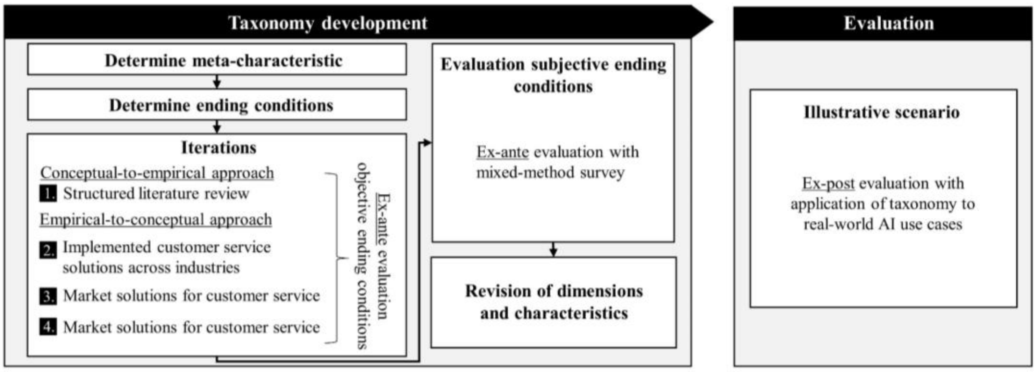

Figure 1. Research phases

For the development phase, the rigorous and systematic method of Nickerson et al. (Nickerson et al., 2013) is adopted. In line with DSR, the development and evaluation phases include several evaluation episodes (Venable et al., 2016; Kundisch et al., 2021). During development, formative ex-ante evaluations are performed. On the one hand, the research team assessed objective ending conditions for each development iteration (see Section 4.1). On the other hand, experts from research and practice conducted an evaluation of the subjective ending conditions with a complete version of the taxonomy (see Section 4.2). As part of the summative ex-post evaluation, the adapted, final taxonomy was applied to illustrative scenarios to provide insights on its usability and validity (see Section 6). Thereby, the taxonomy represents a DSR artifact of the type model (Kundisch et al., 2021), providing prescriptive knowledge on how to design (theory for design and action) the integration of AI into customer service from a socio-technical perspective (Gregor and Hevner, 2013; Gregor, 2006).

14.4 Taxonomy Development

The iterative taxonomy building method according to Nickerson et al. (2013) comprises several steps. The development process starts with the definition of the meta-characteristic to determine the purpose of the taxonomy (Nickerson et al., 2013; Lösser et al., 2019). We define the meta-characteristic as *design decisions for the integration of AI in service delivery processes for customer service* to facilitate researchers and practitioners in their analyses and design undertakings. In this context, design decisions refer to characteristics of service processes, the AI-based technology, and the interaction between humans and AI. The second step includes the definition of ending conditions to determine the requirements to conclude the development process. For the taxonomy development phase, we adopted the objective conditions proposed by Nickerson et al. (2013). In the third step, either an inductive (empirical-to-conceptual) or deductive (conceptual-to-empirical) approach is chosen to initiate the identification of characteristics and dimensions. The application of these approaches can alternate for subsequent iterations. For the conceptual-to-empirical approach, the focus lies on deducing and grouping characteristics into dimensions based on existing scientific knowledge. The empirical-to-conceptual approach involves the utilization of a variety of real-world objects to identify and classify characteristics into dimensions (Nickerson et al., 2013; Lösser et al., 2019). To initiate the development process, we chose the conceptual-to-empirical approach because initial scientific knowledge exists, but is so far unstructured. After each iteration, the assessment of the objective ending conditions by two taxonomy designers was analyzed in terms of their agreement to decide about the continuation of the development. In the following, we describe the four conducted iterations and depict the taxonomy evolution process in Figure 2.

14.4.1 Taxonomy building process

Iteration 1: For the first iteration, we adopted the conceptual-to-empirical approach to develop a profound understanding of the domain under study. To identify extant and pertinent scientific knowledge in various fields such as service science, human-computer-interaction, and information systems (IS), we conducted a systematic literature review following the guiding principles of Webster and Watson (2002) as well as vom Brocke et al. (2015). For the search process, we chose three domain-relevant IS databases, namely ACM Digital Library, AIS eLibrary, and ScienceDirect, to identify relevant peer-reviewed English publications. The search process was performed with a search string. By executing an initial database search, we identified suitable keywords. Based on these results, we created the following search string: *((“employee*” OR “customer*” OR “user*”) AND (“AI” OR “artificial intelligence”) AND (“service” OR “support”))*. The search delivered 738 hits across databases. In two subsequent screening phases, the fit of the publications to the defined meta-characteristic was independently assessed by two researchers. In the first screening phase, the number of publications was reduced by excluding duplicates and inaccessible articles. Furthermore, we used abstracts, titles, and keywords to exclude publications that did not focus on the service domain. The application of these exclusion criteria yielded 101 publications. During the second screening phase, these publications were subject to an in-depth full-text analysis. 19 articles remained after excluding publications, which focus on (1) robotics (2) pure technological aspects without service application, and (3) business intelligence. To reveal higher-order characteristics and dimensions, these articles were iteratively coded. In an initial round, two researchers inductively created a set of master codes (service domain, involved entities, aspects of human-AI-interaction, and service processes) by independently coding the 19 publications and resolving discrepancies. Based on these codes, characteristics were generated and their labeling continuously harmonized in discussions. Subsequently, these characteristics were individually grouped into dimensions by the researchers. Through constant exchange, divergent assignments were cleared and labels for the dimensions were jointly derived. As a result, the following seven dimensions were added to the taxonomy in the first iteration: *service stages*, *AI role*, *task type*, *knowledge and data insights*, *form of AI appearance*, *AI transparency to customers*, and *data and knowledge processing*.

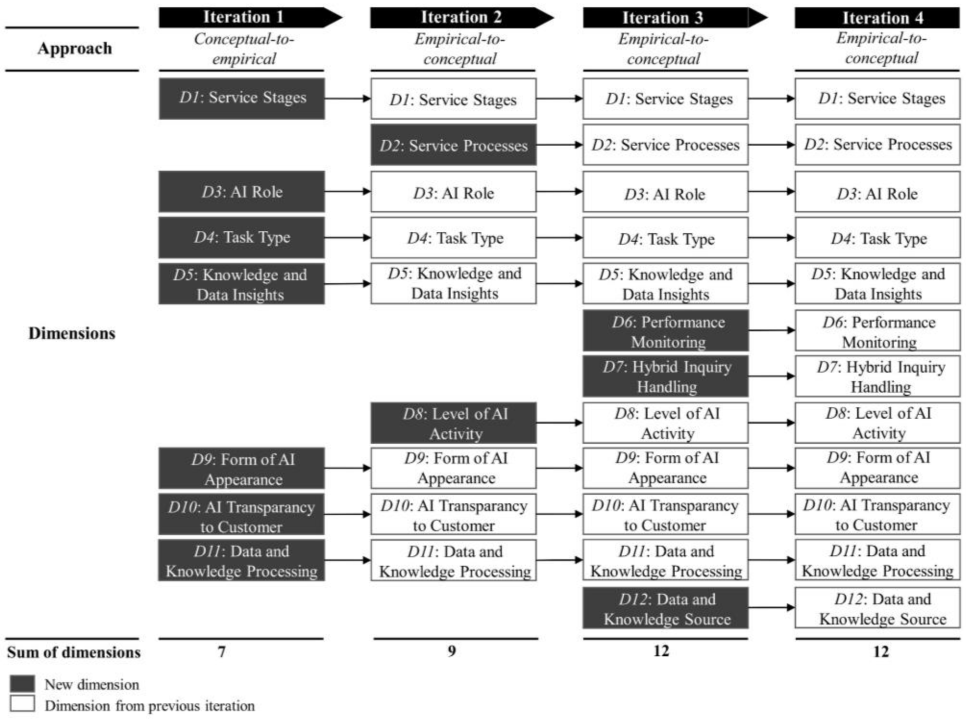

Figure 2. Taxonomy development process and evolution of dimensions.

Iteration 2: Following the conceptual iteration, we chose the empirical-to-conceptual approach to complement the taxonomy with insights induced from real-world objects. The focus in the second iteration was on obtaining real-world data to sustain knowledge about the integration of AI into companies' service delivery processes. Following Short et al. (2002), we applied the stratified random sample method to acquire a representative sample of companies. In this way, a sample of companies can be subdivided into meaningful nonoverlapping groups to account for the diversity of industrial sectors. For the selection of international companies, we utilized the most recent Fortune 500 Global list (Fortune Media, 2019). With the objective of obtaining an appropriate sample size of 80 companies (Short et al., 2002), we selected four companies for each of the 20 industrial sectors, which are specified by Fortune Media. We conducted a systematic data collection process to examine companies' text-based and AI-enabled contact channels. To this end, companies' websites were visited and examined from a customer perspective to capture the types of text-based channels, characteristics of service interactions, and sequence of service processes via descriptions, process models, and screenshots. The subsequent qualitative analysis involved independent coding of documented case data by two researchers. With the help of the dimensions from the first iteration, we discovered that merely nine of 80 companies operating in seven sectors utilize AI for service encounters (see Appendix Table

A1). Based on these insights, several characteristics were identified and merged into two additional dimensions for the taxonomy: *service processes* and *level of AI activity*.

Iteration 3: In the light of service stages with (frontstage) and without (backstage) direct customer contact, we examined 16 market solutions for AI-based customer service with AI-based customer service software. With reference to Gartner's Magic Quadrant (Gartner, 2020), in which vendors are evaluated based on their market positioning, leading, challenging, and visionary, solutions were selected and compared with entries from two suitable databases (capterra.com/customer-service-software, quicksprout.com/best-customer-service-software). For a structured data collection, we analyzed the websites of all vendors to document information in the form of reports, videos, and images. Qualitative analysis of these data, again conducted independently by two researchers, led to three additional dimensions: *performance monitoring*, *hybrid inquiry handling*, and *data and knowledge source*.

Iteration 4: The inclusion of additional dimensions in the preceding iteration required an additional empirical investigation. Therefore, similar to iteration three, a sample specifically focusing on conversational AI market solutions for the frontstage was produced. By using a practice-oriented evaluation from Forrester Research (Jacobs et al., 2019) and entries from two databases (g2.com/categories/conversational-intelligence, capterra.com/conversational-ai-platform-software), suitable solutions were identified. The resulting sample comprises 14 vendors, excluding duplicates from iteration three. The analysis of collected information via vendors' websites did not result in additional dimensions. Accordingly, in this iteration, the development phase was concluded as all objective ending conditions by Nickerson et al. (2013) were met. To prepare the evaluation of subjective ending conditions, we consolidated the taxonomy by inductively determining and ordering five meta-dimensions (*service context*, *capabilities*, *deliverables*, *integration*, and *intelligence*), which aggregately describe the content of the derived dimensions.

14.4.2 Ex-ante evaluation of subjective ending conditions

To ensure usefulness and applicability for research and practice, we assessed the content of the taxonomy with an ex-ante evaluation (Szopinski et al., 2019; Kundisch et al., 2021). Therefore, a mixed-method survey was utilized to collect quantitative and qualitative data from experts. To this end, our questionnaire included the taxonomy from iteration four with definitions for meta-dimensions, dimensions, characteristics, and questions covering the five subjective ending conditions (concise, robust, comprehensive, extendible, and explanatory) proposed by Nickerson et al. (2013). These ending conditions were each evaluated with a five-point Likert scale (from 1 (strongly disagree) to 5 (strongly agree)) and open-ended questions to receive extensive evaluation output and qualitative feedback for improvement. As the taxonomy is intended to guide researchers and practitioners alike in making design decisions to integrate AI into customer service, a heterogeneous group of experts from science (professor IS (ES1), research associates IS (ES2, ES4, ES5), associate professor IS (ES3)) and practice (machine learning engineer (EP1), senior architect (EP2), IS agent (EP3), software developer (EP4), software architect (EP5)) was recruited. For the selection, a purposive sampling strategy was chosen to obtain individuals who have (1) profound experience in taxonomy development and/or (2) knowledge about the role and deployment of AI in customer service.

By defining these selection criteria, relevant insights concerning content and formal aspects of the taxonomy could be derived. The analysis of the quantitative data delivered means and medians above 4.0 for the five subjective ending conditions: **concise** (M = 4.00; SD = 0.74; Mdn = 4), **robust** (M = 5.00; SD = 0.52; Mdn = 5), **comprehensive** (M = 4.00; SD = 0.52; Mdn = 4), **extendible** (M = 5.00; SD = 0.97; Mdn = 5) and **explanatory** (M = 4.00; SD = 0.52; Mdn = 4). These ratings at good to excellent level and the low dispersion of data illustrate the usefulness and applicability of the content and structure of the taxonomy. With respect to experts' qualitative comments, the analysis of data revealed recommendations for improvement that were implemented as follows. The label for the second dimension was changed from "Service Processes" to "Service Process Continuity" and the definition adapted (ES1). The definitions for the three characteristics of the third dimension were adjusted to clarify their focus (ES1, ES4). The description of the sixth dimension was extended to specify the meaning of the two characteristics. (ES4). The definitions for the tenth dimension and its two characteristics (ES1) and the characteristics of the eleventh dimension were refined (ES4, EP2, EP1). For the meta-dimension "Capabilities" the definition was refined (ES2), whereas the definition of the meta-dimension "Deliverables" was extended (ES2). These adjustments refer to refinements of content through adapting and extending definitions of meta-dimensions, dimensions, and characteristics. Thus, the objective ending conditions were still fulfilled.

14.5 Taxonomy of AI Integration into Customer Service

After four development iterations and content-related revisions initiated by the ex-ante evaluation, the final version of the taxonomy encompasses 12 dimensions, and 32 characteristics organized into five meta-dimensions (see Figure 3). Following Püschel et al. (2016), we classified the characteristics of each dimension as either mutually exclusive or non-exclusive to create a clearly structured and concise taxonomy. By establishing clear and delimited definitions, redundancy was counteracted to allow for the selection of a confined set of characteristics. To structure the taxonomy, we arranged the meta-dimensions in sequential order of their application for analysis and design to facilitate design decisions for the integration of AI for service delivery into customer service contexts. With *service context*, the application area of AI in customer service is determined. Subsequently, AI's *capabilities* are defined to determine the *deliverables* in the form of distinct outputs. By specifying the *integration* of AI, the interaction with customers and SEs, the appearance and behavior are defined. Concluding, the *intelligence* of AI is determined in accordance with the previous design decisions. In the following sub-sections, we present and describe the dimensions and characteristics for each of these meta-dimensions with justificatory references from research and practice (see Appendix Table A1 for practice references).

Service context: Based on the service context, the deployment of AI in customer service is determined in relation to *service stages* (D1) and the nature of *service process continuity* (D2). With respect to ***service stages***, AI can be utilized in the *frontstage* (D1,C1) to handle inquiries in direct contact with customers (Robinson et al., 2020; Fingerle et al., 2002). The application of AI in the *backstage* (D1,C2) involves processing of inquiries without direct customer contact (Zhang et al., 2020; Campbell et al., 2020). Associated with the deployment of AI in service stages is the determination of the type of ***service process continuity***, which refers to the temporal alignment of AI-integrated service delivery processes. *Disconnected* (D2,C1) processes imply unconnected inquiry processing steps between service stages involving SEs and AI with time lags and/or contact channel switches (I2U, I2W, I2WD). A *connected* (D2,C2) process continuity represents a direct connection between the service stages for request processing steps involving SEs and AI (I2N, I2AD, I2C, I2AM, I2AT, I2H).

<table>
<tr><th>MD</th><th colspan="2">Dimensions</th><th colspan="12">Characteristics</th></tr>
<tr><td rowspan="2">Service Context</td><td>D_1: Service Stages</td><td>NE</td><td colspan="6">Frontstage</td><td colspan="6">Backstage</td></tr>
<tr><td>D_2: Service Process Continuity</td><td>NE</td><td colspan="6">Disconnected</td><td colspan="6">Connected</td></tr>
<tr><td rowspan="2">Capabilities</td><td>D_3: AI Role</td><td>NE</td><td colspan="4">Support</td><td colspan="4">Augmentation</td><td colspan="4">Performance</td></tr>
<tr><td>D_4: Task Type</td><td>NE</td><td colspan="3">Mechanical</td><td colspan="3">Analytical</td><td colspan="3">Intuitive</td><td colspan="3">Empathetic</td></tr>
<tr><td rowspan="2">Deliverables</td><td>D_5: Knowledge and Data Insights</td><td>NE</td><td colspan="3">Inquiry-related</td><td colspan="3">Process-focused</td><td colspan="3">Customer-related</td><td colspan="3">Socio-emotional</td></tr>
<tr><td>D_6: Performance Monitoring</td><td>NE</td><td colspan="6">Human Agent Monitoring</td><td colspan="6">AI Monitoring</td></tr>
<tr><td rowspan="4">Integration</td><td>D_7: Hybrid Inquiry Handling</td><td>ME</td><td colspan="3">Simultaneous</td><td colspan="3">Consecutive - toward human</td><td colspan="3">Consecutive - toward AI</td><td colspan="3">Consecutive - alternating</td></tr>
<tr><td>D_8: Level of AI Activity</td><td>NE</td><td colspan="6">Reactive</td><td colspan="6">Proactive</td></tr>
<tr><td>D_9: Form of AI Appearance</td><td>ME</td><td colspan="6">AI-enabled agent</td><td colspan="6">Embedded AI</td></tr>
<tr><td>D_{10}: AI Transparency to Customers</td><td>ME</td><td colspan="6">Unknown</td><td colspan="6">Known</td></tr>
<tr><td rowspan="2">Intelligence</td><td>D_{11}: Data and Knowledge Processing</td><td>NE</td><td colspan="6">Machine Learning</td><td colspan="6">Rule-based Reasoning</td></tr>
<tr><td>D_{12}: Data and Knowledge Source</td><td>NE</td><td colspan="4">Input before Interaction</td><td colspan="4">Input during Interaction</td><td colspan="4">Input after Interaction</td></tr>
<tr><td colspan="15">Note: MD = meta-dimension; ME = mutually exclusive; NE = non-exclusive</td></tr>
</table>

Figure 3. Taxonomy of AI integration into customer service

Capabilities: The scope of application for AI in customer service is guided by its capabilities, which are subdivided into the dimensions *AI role* (D3) and *task type* (D4). Regarding the ***role AI*** plays in service delivery, a distinction can be made between support, augmentation, and performance. AI can provide *support* (D3,C1) to deliver service by executing and handing over results of (sub-)tasks (Canhoto and Clear, 2020; Ostrom et al., 2019; Keyser et al., 2019). By actively collaborating on a task with SEs, AI can *augment* (D3,C2) service delivery tasks (Xu et al., 2020; Amorim et al., 2019; Campbell et al., 2020; Ameen et al., 2021). Furthermore, AI can *perform* (D3,C3) (sub-)tasks autonomously

(Canhoto and Clear, 2020; Macnish and Fernandez Inguanzo, 2019; Zhang et al., 2020; Göker and Roth-Berghofer, 1999). The utilization of AI capabilities also refers to different ***task types*** in customer service. When applied to *mechanical tasks* (D4,C1), AI can be used for standardizable, repetitive, routine, and transactional tasks that require consistency in execution (Canhoto and Clear, 2020; Huang and Rust, 2018). For tasks with an *analytical* (D4,C2) nature that require logical thinking and are executed based on data, information, and knowledge, AI can provide analytical functions (Huang and Rust, 2018; Canhoto and Clear, 2020). Furthermore, AI can be applied for *intuitive tasks* (D4,C3) that require experiential and context-based interaction and thinking. In addition, AI can be utilized for *empathetic tasks* (D4,C4) with a salient emotional and interactive character that requires empathy and emotional analytics (Canhoto and Clear, 2020; Huang and Rust, 2018).

Deliverables: In customer service, AI can produce two types of output as deliverables: *knowledge and data insights* (D5) and *performance monitoring* (D6). The ***knowledge and data*** AI can supply to customers and/or SEs relate to four different forms of insights. AI can provide knowledge and/or information that relate to the *content of an inquiry* (D5,C1) (Xu et al., 2020; Amorim et al., 2019). *Process-focused* (D5,C2) clues can be presented for service interactions (Canhoto and Clear, 2020; Amorim et al., 2019). Insights related to the customer can comprise *customer-related* (D5,C3) information (e.g., history of contact) (Libai et al., 2020; Campbell et al., 2020) or *socio-emotional* (D5,C4) insights related to customers' sentiments (Amorim et al., 2019; Canhoto and Clear, 2020). The ***performance monitoring*** for and with AI relates to *human agent monitoring* (D6,C1) or *AI monitoring* (D6,C2). The former provides insights on SEs' workload, inquiry volume, and trends (I3S, I3SN, I3M, I4AI, I3V, I3Z, I3S, I3CR). The latter refers to insights into AI's performance in terms of interaction behavior and the status of the knowledge base to identify potential for improvement (I4L, I4AV, I4IN, I3S, I4KO, I3V).

Integration: The representation and integration of AI into customer service encompass four dimensions: *hybrid inquiry handling* (D7), *level of activity* (D8), *form of appearance* (D9), and *AI transparency to customers* (D10). The ***hybrid inquiry handling*** determines the sequence, in which inquiries are handled by the SE and AI. On the one hand, the sequence can be *simultaneous* (D7,C1), i.e., the SE and AI are working together on an inquiry at the same time (I3SN, I3M, I3AP, I3F, I4EG, I3K, I4VS, I4O, I3P, I3SAP, I3Z, I4L, I4N, I4I, I4SF). On the other hand, the sequence can be *consecutive*, either *toward human* (D7,C2) or *toward AI* (D7,C3). Toward human, the AI handles the inquiry autonomously and forwards it to the SE once a determined condition is fulfilled, and vice versa toward AI (I3SN, I3Z, I3F, I4N, I4OA, I4IS, I4CO, I3E, I3C, I4IN, I4KO, I4L, I4VS, I4SF). A third alternative is a *consecutive-alternating* (D7,C4) sequence. In this case, the AI and SE handle the inquiry autonomously and hand it over to each other every time a determined condition is fulfilled (I4AI). The ***level of activity*** represents the activity behavior of the AI in interactions with SEs or customers. Either the AI is *reactive* (D8,C1) or *proactive* (D8,C2) in its behavior. When the AI is reactive, it is passive and interacts once it is triggered (I2U, I2C, I2N, I2AD, I2WD, I2W). When it is proactive, the AI is active and interacts of its own accord (I2AM). The ***form of AI appearance*** defines the form, in which AI appears in customer service. If the AI has an identity as agent with a virtual representation and interacts through natural language with SEs or customers, it is an *AI-enabled agent* (D9,C1) (Prentice and Nguyen, 2020; Xu et al., 2020; Campbell et al., 2020; Canhoto and Clear, 2020; Macnish and Fernandez Inguanzo, 2019; Svenningsson and Faraon, 2019; Gelbrich et al., 2020; Zhang et al., 2020). If it is integrated into platforms or applications in use and neither has an identity nor a visual representation, it represents an *embedded AI* (D9,C2) (Zhang et al., 2020; Chromik et al., 2020; Göker and Roth-Berghofer, 1999). The ***transparency of AI to customers*** refers to the degree, to which the presence of AI is apparent to customers. The customers are either not aware of AI's presence during service delivery, which makes it *unknown* (D10,C1) (Canhoto and Clear, 2020; Aoki, 2021; Robinson et al., 2020), or they are aware of AI's presence, which makes it *known* (D10,C2) (Canhoto and Clear, 2020; Robinson et al., 2020; Macnish and Fernandez Inguanzo, 2019; Svenningsson and Faraon, 2019; Aoki, 2021).

Intelligence: The intelligence of AI-integrated customer service is defined by the way it receives and handles data and knowledge for customer service tasks. With this, it covers two dimensions: *data and knowledge processing* (D11) and *data and knowledge source* (D12). The ***data and knowledge processing*** describes the underlying technology, which defines how AI processes information and knowledge. For one thing, AI can be trained and based on *Machine Learning* (D11,C1) using learning algorithms for processing existing data toward pattern and entity recognition. This also covers the ability of AI to process and analyze natural language data to understand and generate natural language (Canhoto and Clear, 2020; Campbell et al., 2020). For another thing, AI can also be based on "if-then" pattern-matching rules through *rule-based reasoning* (D11,C2) (Fingerle et al., 2002; Cheung et al., 2003; Göker and Roth-Berghofer, 1999). The ***data and knowledge source*** identifies the source from where the AI gets the data and knowledge. This data *input* can happen *before* (D12,C1), *during* (D12,C2), or *after* (D12,C3) the *interaction*. First, the AI's knowledge base can be built by data and knowledge provided before the interaction (I4AV, I4CO). Then, the AI's knowledge base can continuously evolve through optimization based on and during the interaction (I4IS, I4AI, I4L, I4N, I4AI, I3P). And at last, AI's knowledge base can continuously evolve through implementing feedback and learnings after each interaction (I3E, I4KO, I4AV, I4IN, I4AI, I3F, I4CO).

14.6 Ex-Post Evaluation: Taxonomy Application

To adopt a rigorous evaluation strategy, we applied the framework by Szopinski et al. (2019) and chose the method 'illustrative scenario' to assess the coherence of the final taxonomy with the meta-characteristic. To this end, two real-world AI use cases were classified as objects with the taxonomy. To verify the validity of the taxonomy's purpose, on the one hand, a case was selected, where an AI-enabled agent in the form of a CA has already been implemented for service delivery in the frontstage (organization X). On the other hand, a case was chosen, in which the deployment of an embedded AI solution is planned to assist SEs in frontstage interactions (organization Y). To analyze the reliability of the taxonomy, two researchers and three practitioners utilized the taxonomy along the sequential order of meta-dimensions for design decisions. For organization X, the researchers and one practitioner with affiliation to the organization classified the existing AI use case. Two members from organization Y and the same researchers performed the classification for the planned AI use case in organization Y. The two researchers were enabled to classify the two use cases by a presentation of the core features derived from a qualitative data analysis based on eleven semi-structured interviews (organization X = five, organization Y = six) with business unit members, product owners, and documents about the IT architecture and modules. The results of the classification are presented in Table 1 by providing the rations of selected characteristics per dimension for each use case.

Table 1. Classification Results of AI Use Cases

<table>
<tr><th colspan="12">Hit ratios for characteristics (X ; Y)</th></tr>
<tr><td colspan="6">D_1,C_1: 100 % ; 0 %</td><td colspan="6">D_1,C_2: 0 % ; 100 %</td></tr>
<tr><td colspan="6">D_2,C_1: 100 % ; 0 %</td><td colspan="6">D_2,C_2: 0 % ; 100 %</td></tr>
<tr><td colspan="4">D_3,C_1: 100 % ; 75 %</td><td colspan="4">D_3,C_2: 0 % ; 100 %</td><td colspan="4">D_3,C_3: 67 % ; 0 %</td></tr>
<tr><td colspan="3">D_4,C_1:
100 % ; 25 %</td><td colspan="3">D_4,C_2:
0 % ; 75 %</td><td colspan="3">D_4,C_3:
0 % ; 0 %</td><td colspan="3">D_4, C_4:
0 % ; 0 %</td></tr>
<tr><td colspan="3">D_5,C_1:
100 % ; 75 %</td><td colspan="3">D_5,C_2:
100 % ; 100 %</td><td colspan="3">D_5,C_3:
33 % ; 25 %</td><td colspan="3">D_5,C_4:
0 % ; 0 %</td></tr>
<tr><td colspan="6">D_6,C_1: 100 % ; 0 %</td><td colspan="6">D_6,C_2: 100 % ; 100 %</td></tr>
<tr><td colspan="3">D_7,C_1: 0 % ; 75 %</td><td colspan="3">D_7,C_2: 100 % ; 0 %</td><td colspan="3">D_7,C_3: 0 % ; 0 %</td><td colspan="3">D_7,C_4: 0 % ; 25 %</td></tr>
<tr><td colspan="6">D_8,C_1: 67 % ; 75 %</td><td colspan="6">D_8,C_2: 33 % ; 25 %</td></tr>
<tr><td colspan="6">D_9,C_1: 100 % ; 75 %</td><td colspan="6">D_9,C_2: 0 % ; 25 %</td></tr>
<tr><td colspan="6">D_{10},C_1: 0 % ; 100 %</td><td colspan="6">D_{10},C_2: 100 % ; 0 %</td></tr>
<tr><td colspan="6">D_{11},C_1: 100 % ; 100 %</td><td colspan="6">D_{11},C_2: 100 % ; 0 %</td></tr>
<tr><td colspan="4">D_{12},C_1: 100 % ; 75 %</td><td colspan="4">D_{12},C_2: 0 % ; 75 %</td><td colspan="4">D_{12},C_3: 100 % ; 75 %</td></tr>
</table>

For the use case of organization X, the practitioner and researchers agreed on all characteristics in nine dimensions; for eleven dimensions, they agreed on at least one characteristic. For the use case of organization Y, the practitioners and researchers agreed on all characteristics for five dimensions; in seven dimensions they agreed on at least one characteristic. Only for five characteristics in four dimensions in use case Y, classifications did not match. However, it is difficult to achieve perfect interrater agreement for the whole taxonomy regarding the option to choose more than one characteristic in most dimensions. With this, the classification of the two specific use cases along the characteristics of the taxonomy reveals a good reliability and well-suited applicability of the taxonomy for practice. Moreover, the achieved characterization of the two use cases with reference to their attributes indicates a substantial validity of the taxonomy. After classifying their use cases, we also asked the practitioners for further feedback on different aspects related to the application of the taxonomy. First, the taxonomy appears understandable and clear. Especially for the planning scenario, it provided ideas and perspectives to the development, which need to be considered. Second, practitioners found it easy to use, i.e., they knew where and how it can be applied in their real-world scenario. At last, they argued for good feasibility and applicability of the taxonomy indicating the usefulness of our taxonomy. Based on these insights, we can confirm the coherence of the final taxonomy with the meta-characteristic to facilitate researchers and practitioners in their analysis and design undertakings concerning design decisions for the integration of AI in service delivery processes for customer service.

14.7 Discussion and Conclusion

With the developed taxonomy, we provide a first structured and elaborated overview of relevant design choices to integrate AI into the front- and/or backstage of customer service contexts. The compilation of characteristics across five meta-dimensions and 12 dimensions systematizes scattered knowledge from research and commercial applications in the still evolving research field of AI-enabled service. Thereby, two current research streams focusing on conceptual or technological aspects are integrated. Based on and complementing these insights with data from practice, we present an in-depth analysis of pertinent aspects of how AI can be integrated into customer service (Benbya et al., 2021). In addition, we answer the call for an investigation of the mutual interrelation between AI and the social as well as technical systems in service organizations (Bock et al., 2020). By adopting a holistic, socio-technical perspective for the development, the taxonomy reveals changes in connection to AI integration referring to service processes spanning front- and backstage, division of labor, and interaction between humans and AI. In particular, the taxonomy emphasizes that different constellations of the entities customer, SE, and AI emerge depending on the design decisions to integrate AI. As AI is not yet capable of solving all types of inquiries independently in the frontstage, the service process comprises sections where all entities interact simultaneously or handovers are initiated, introducing a change in interaction partners. Accordingly, depending on the AI use case, specific task- and process-related dependencies arise between AI and SE, which in turn impact the interaction with customers. Similarly, the integration of AI in the backstage has an impact on SEs, as working practices change through interaction with AI. Building on the current state of research and practice of AI and its deployment in customer service, our taxonomy enables the classification of use cases that are planned to be scientifically investigated or developed and/or planned for deployment in practice. By providing a sequential order of design decisions that are organized along the meta-dimensions, the selection of a confined set of characteristics regarding service context, capabilities, deliverables, integration, and intelligence of a specific AI use case is facilitated. In this respect, the results of the ex-post evaluation demonstrate a good handling of the taxonomy. Furthermore, a valid and reliable classification of AI use cases for customer service can be achieved by utilizing the taxonomy. These results underline the completeness, applicability, and effectiveness of the created taxonomy. Accordingly, our rigorously developed and evaluated taxonomy provides prescriptive design knowledge on how AI can be integrated into customer service to sustain the design and implementation process as well as the analysis of AI-based customer service applications (Kundisch et al., 2021; Gregor and Hevner, 2013).

The presented taxonomy provides many-faceted theoretical and practical contributions. Regarding research, we created, to the best of our knowledge, the first taxonomy that summarizes scientific insights and the status quo in practice on characteristics for the integration of AI in customer service. As a result, the structure for classification improves and fosters understanding in this research domain regarding characteristics for AI-infused customer service. Hence, these insights might encourage the extension and continuation of research for progressing AI-powered customer service solutions. Furthermore, it serves as a tool to systematically derive relevant and specific design decisions by incorporating various aspects that should be considered for the development of AI solutions. Moreover, we contribute insights on the integration of AI for, both, the external (frontstage) and internal (backstage) customer service environment (see Figure 4).

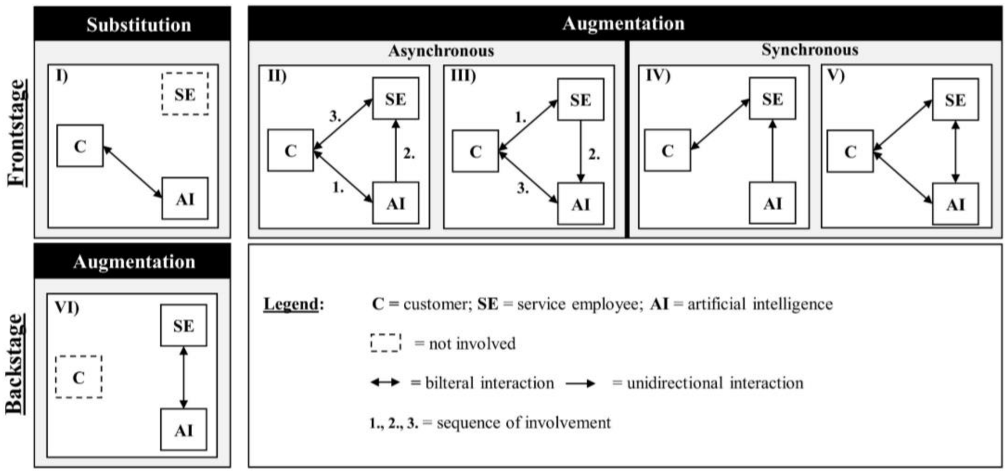

Figure 4. AI infusion archetypes covering front- and backstage

We validate existing infusion archetypes from Keyser et al. (2019) and Ostrom et al. (2019). More specifically, substitution (see Figure 4, I) and augmentation (see Figure 4, IV & V) of SEs by AI in frontstage service encounters could be confirmed. Furthermore, we identified additional infusion archetypes. In the frontstage, we introduce asynchronous augmentation (see Figure 4, II & III) where inquiries are handled consecutively with handovers from AI toward SEs or vice versa in cases a predetermined condition is fulfilled (e.g., imminent failure of AI). In addition, for the customer service backstage, we establish an infusion archetype of the type “augmentation” for the first time (see Figure 4, VI). For this archetype, the focus lies on AI use cases that facilitate service processes and tasks without direct customer contact, which are also of eminent relevance for service delivery. In this context, AI is deployed to augment SEs in processing inquiries by, inter alia, displaying suitable information that might facilitate decision-making.

In terms of practice, IT management and development can use the categorization to analyze deployed solutions to uncover gaps or plan implementation by determining characteristics of a specific use case. Therefore, the taxonomy provides a suitable blueprint to structure AI integration initiatives by classifying projects along the dimensions. Especially for planning AI integration, it adds more ideas and perspectives to be considered for the development. In fact, practitioners benefit from insights, which shed light on relevant AI-related characteristics, e.g., its role and task, which have been extensively developed by researchers. In addition, for integrating AI in their customer service processes, they can refer to state-of-the-art solutions we provide from practice. Eventually, the sequential order of our taxonomy can guide practitioners either through planning, executing, or analyzing AI integration for their specific use case.

Besides the promising contributions of this research, there are a few limitations to consider. First, our empirical data is based on a representative sample of solutions from practice. However, selecting and adding different or more solutions to our sample of empirical solutions could reveal and lead to different or more insights. Furthermore, even though we considered three domain-relevant IS databases, the results might vary when selecting different or more databases. This also applies to changes of our search string. Eventually, the selected samples of empirical cases and research contributions define and limit the taxonomy with its dimensions and characteristics. At last, regarding the reliability of our taxonomy, we could only consider two illustrative use case scenarios. To achieve and establish reliability, more practitioners may use and apply the taxonomy to their specific use cases. These limitations and obtained insights give rise to future research. In general, future work can build on our taxonomy, validate, or extend dimensions and characteristics. Considering our ex-post evaluation, we call for descriptive research to specifically enhance the applicability of the taxonomy and better showcase what design decisions must be made for the integration of an AI solution in customer service. Furthermore, different aspects can be addressed in more detail with respect to the individual meta-dimensions of the taxonomy. First, in terms of service context, the current state of knowledge indicates that AI solutions for backstage customer service are, so far, under-researched. In this context, research should focus on the development and design of AI solutions that promote hybrid service delivery without direct customer contact. Related to this, solutions should be developed that enable an AI-integrated, seamless, and efficient processing of inquiries across front- and backstage involving AI and SEs. Additionally, research should focus on mechanisms to establish acceptance toward AI and incentive systems for SEs and customers to utilize AI for service delivery. Lastly, approaches for learning scenarios are needed that allow for a continuous development of the competencies and knowledge base of the AI.

14.8 Appendix

Iteration 2			
Company	**URL**	**Company**	**URL**
Amazon (I2AM)	www.amazon.com/gp/help/customer/display.html?nodeId=508510&ref_=nav_cs_customerservice_2bf4fe8c5ec54e6bae2d1c24043f012b	United Parcel Service (I2U)	www.ups.com/us/en/help-support-center.page
China Mobile Communication (I2C)	eshop.hk.chinamobile.com/en/corporate_information/Customer_Service/index.html	AT&T (I2AT)	www.att.com/support/topic
Home Depot (I2H)	www.homedepot.com/c/customer_service	Walmart (I2W)	www.walmart.com/help
Walt Disney (I2WD)	help.shopdisney.com/hc/en-us	Adidas (I2AD)	www.adidas.com/us/help
Nike (I2N)	www.nike.com/us/help		
Iteration 3 *AI-based Customer Service Software*			
Salesforce (I3S)	www.salesforce.com/products/service-cloud/features	Pegasystems (I3P)	www.pega.com/products/platform/email-bot
Service Now (I3SN)	www.servicenow.com/content/dam/servicenow-assets/public/en-us/doc-type/resource-center/data-sheet/ds-customer-service-management.pdf	Microsoft (I3M)	dynamics.microsoft.com/de-de/customer-service/overview
Zendesk (I3Z)	support.zendesk.com/hc/en-us/articles/360057455393?_ga=2.178859299.64267010.1608134017-830973252.1608134017	Oracle (I4OA)	www.oracle.com/cx/service/b2c
SAP (I3SAP)	www.sap.com/products/service-cloud.html?btp=0106c0a9-f57d-429f-ab94-bd740a7f68e8	Freshworks (I3F)	freshdesk.com/freddy-ai-for-cx
Verint Systems (I3V)	www.verint.com/customer-engagement-cloud	Appian (I3AP)	appian.com/platform/overview.html

Creatio (I3C)	www.creatio.com/service	eGain (I3E)	www.egain.com/solutions/contact-centers
SugarCRM (I3S)	www.sugarcrm.com/de/solutions/sugar-serve	Kustomer (I3K)	www.kustomer.com/product/customer-service
Zoho (I3Z)	www.zoho.com/desk/zia.html	CRMNEXT (I3CR)	www.crmnext.com/crm/service
Iteration 4 *Conversational AI*			
LogMeIn (I4L)	www.bold360.com	Salesforce (I4SF)	www.salesforce.com/products/service-cloud/automated-customer-service
Nuance (I4N)	www.nuance.com/index.html	Verint Systems (I4VS)	www.verint.com/engagement/our-offerings/solutions/intelligent-self-service/virtual-assistant
Interactions (I4I)	www.interactions.com	eGain (I4EG)	www.egain.com/products/chatbot-virtual-assistant-software
Inbenta (I4IN)	www.inbenta.com/products/chatbot	Kore (I4KO)	kore.ai
Aivo (I4AI)	www.aivo.co	Cognigy (I4CO)	www.cognigy.com
Avaamo (I4AV)	avaamo.ai	IPSoft (I4IS)	amelia.com
247ai (I4AI)	www.247.ai	Omilia (I4O)	omilia.com

14.9 Acknowledgment

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, 02L18A111).

14.10 References

Adam, M., M. Wessel and A. Benlian (2020). “AI-based chatbots in customer service and their effects on user compliance” *Electronic Markets*, 1–19.

Ameen, N., A. Tarhini, A. Reppel and A. Anand (2021). “Customer Experiences in the Age of Artificial Intelligence” *Computers in Human Behavior* 114, 106548.

Amorim, M., Y. Cohen, J. Reis and M. Rodrigues (2019). "Exploring Opportunities for Artificial Emotional Intelligence in Service Production Systems". In: *9th IFAC Conference on Manufacturing Modelling*. Berlin, Germany.

Aoki, N. (2021). "The Importance of the Assurance that "Humans are Still in the Decision Loop" for Public Trust in Artificial Intelligence: Evidence from an Online Experiment" *Computers in Human Behavior* 114, 106572.

Bailey, K. D. (1994). *Typologies and Taxonomies: An Introduction to Classification Techniques.* Thousand Oaks, USA: Sage.

Bassano, C., S. Barle, M. L. Saviano, S. Cosimato and M. C. Pietronudo (2020). "AI technologies & value co-creation in a luxury goods context". In: *53rd Hawaii International Conference on System Sciences*. Grand Wailea, USA.

Benbya, H., S. Pachidi and S. L. Jarvenpaa (2021). "Artificial Intelligence in Organizations: Implications for Information Systems Research" *Journal of the Association for Information Systems* 22 (2), 281–303.

Bitner, M. J., S. W. Brown and M. L. Meuter (2000). "Technology Infusion in Service Encounters" *Journal of the Academy of Marketing Science* 28 (1), 138–149.

Bock, D. E., J. S. Wolter and O. C. Ferrell (2020). "Artificial Intelligence: Disrupting What We Know about Services" *Journal of Services Marketing* 34 (3), 317–334.

Brady, M. K., J. Cronin and R. R. Brand (2002). "Performance-only Measurement of Service Quality: A Replication and Extension" *Journal of Business Research* 55 (1), 17–31.

Campbell, C., S. Sands, C. Ferraro, Tsao, Hsiu-Yuan (Jody) and A. Mavrommatis (2020). "From Data to Action: How Marketers can Leverage AI" *Business Horizons* 63 (2), 227–243.

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

Cheung, C. F., W. B. Lee, W. M. Wang, K. F. Chu and S. To (2003). "A Multi-perspective Knowledge-based System for Customer Service Management" *Expert Systems with Applications* 24 (4), 457–470.

Chromik, M., F. Fincke and A. Butz (2020). "Mind the (Persuasion) Gap: Contrasting Predictions of Intelligent DSS with User Beliefs to Improve Interpretability". In: *12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. Sophia Antipolis, France.

Davenport, T., A. Guha, D. Grewal and T. Bressgott (2020). "How Artificial Intelligence Will Change the Future of Marketing" *Journal of the Academy of Marketing Science* 48 (1), 24–42.

Dellermann, D., P. Ebel, M. Söllner and J. M. Leimeister (2019). "Hybrid Intelligence" *Business & Information Systems Engineering* 61 (5), 637–643.

Dwivedi, Y. K., L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, M. Janssen and More Authors (2021). "Artificial Intelligence (AI):

Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy” *International Journal of Information Management* 57, 101994.

Fingerle, T., S. Schmitt and C. Tautz (2002). “Customer Requirements Elicitation in Intelligent Electronic Sales Support Systems”. In: *BLED eConference*. Bled, Slovena.

Følstad, A. and M. Skjuve (2019). “Chatbots for Customer Service: User Experience and Motivation”. In: *1st International Conference on Conversational User Interfaces*. Dublin, Ireland.

Fortune Media (2019). *Global 500*. URL: https://fortune.com/global500/2019/search/ (visited on 11/01/2021).

Froehle, C. M. and A. V. Roth (2004). “New Measurement Scales for Evaluating Perceptions of the Technology-mediated Customer Service Experience” *Journal of Operations Management* 22 (1), 1–21.

Gartner (2019). *How to Manage Customer Service Technology Innovation*. URL: https://www.gartner.com/smarterwithgartner/27297-2/ (visited on 10/03/2021).

Gartner (2020). *Gartner Magic Quadrant for the CRM Customer Engagement Center*. URL: https://www.gartner.com/en/documents/3985947/magic-quadrant-for-the-crm-customer-engagement-center (visited on 11/16/2021).

Gelbrich, K., J. Hagel and C. Orsingher (2020). “Emotional Support from a Digital Assistant in Technology-mediated Services: Effects on Customer Satisfaction and Behavioral Persistence” *International Journal of Research in Marketing*, in press.

Glikson, E. and A. W. Woolley (2020). “Human Trust in Artificial Intelligence: Review of Empirical Research” *Academy of Management Annals* 14 (2), 627–660.

Glushko, R. J. and L. Tabas (2008). “Briding the “Front Stage” and “Back Stage” in Service System Design”. In: *41st Hawaii International Conference on System Sciences*. Waikoloa, USA.

Gnewuch, U., S. Morana and A. Maedche (2017). “Towards Designing Cooperative and Social Conversational Agents for Customer Service”. In: *38th International Conference on Information Systems*. Seoul, South Korea.

Göker, M. H. and T. Roth-Berghofer (1999). “The Development and Utilization of the Case-based Help-desk Support System HOMER” *Engineering Applications of Artificial Intelligence* 12 (6), 665–680.

Graef, R., M. Klier, K. Kluge and J. F. Zolitschka (2020). “Human-machine Collaboration in Online Customer service – A Long-term Feedback-based Approach” *Electronic Markets*, 1–23.

Gregor (2006). “The Nature of Theory in Information Systems” *MIS Quarterly* 30 (3), 611–642.

Gregor, S. and A. R. Hevner (2013). “Positioning and Presenting Design Science Research for Maximum Impact” *MIS Quarterly* 37 (2), 337–356.

Hevner, March, Park and Ram (2004). “Design Science in Information Systems Research” *MIS Quarterly* 28 (1), 75–105.

Huang, M.-H. and R. T. Rust (2018). “Artificial Intelligence in Service” *Journal of Service Research* 21 (2), 155–172.

Jacobs, I., D. Hong, S. Dawson, S. Sjoblom and D. Lynch (2019). *The Forrester New Wave: Conversational AI for Customer Service*. URL: https://www.forrester.com/report/The+Forrester+New+Wave+Conversational+AI+For+Customer+Service+Q2+2019/-/E-RES144416 (visited on 11/08/2021).

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

Kankanhalli, A., O.-K. Lee and K. H. Lim (2011). “Knowledge Reuse through Electronic Repositories: A Study in the Context of Customer Service Support” *Information & Management* 48 (2-3), 106–113.

Keyser, A. de, S. Köcher, L. Alkire, C. Verbeeck and J. Kandampully (2019). “Frontline Service Technology Infusion: Conceptual Archetypes and Future Research Directions” *Journal of Service Management* 30 (1), 156–183.

Krogh, G. von (2018). “Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing” *Academy of Management Discoveries* 4 (4), 404–409.

Kundisch, D., J. Muntermann, A. M. Oberländer, D. Rau, M. Röglinger, T. Schoormann and D. Szopinski (2021). “An Update for Taxonomy Designers. Methodological Guidance from Information Systems Research” *Business & Information Systems Engineering*, 1–19.

Larivière, B., D. Bowen, T. W. Andreassen, W. Kunz, N. J. Sirianni, C. Voss, N. V. Wünderlich and A. de Keyser (2017). ““Service Encounter 2.0”: An investigation into the roles of technology, employees and customers” *Journal of Business Research* 79, 238–246.

Levy, F. (2018). “Computers and Populism: Artificial Intelligence, Jobs, and Politics in the Near Term” *Oxford Review of Economic Policy* 34 (4), 393–417.

Libai, B., Y. Bart, S. Gensler, C. F. Hofacker, A. Kaplan, K. Kötterheinrich and E. B. Kroll (2020). “Brave New World? On AI and the Management of Customer Relationships” *Journal of Interactive Marketing* 51, 44–56.

Lösser, B., A. M. Oberländer and D. Rau (2019). “Taxonomy Research in Information Systems: A Systematic Assessment”. In: *27th European Conference on Information Systems*. Stockholm, Sweden.

Luger, E. and A. Sellen (2016). “”Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents”. In: *16th Conference on Human Factors in Computing Systems*. San Jose California USA.

Macnish, K. and A. Fernandez Inguanzo (2019). “Case Study - Customer Relation Management, Smart Information Systems and Ethics” *The ORBIT Journal* 2 (2), 1–24.

Maglio, P. P. and J. Spohrer (2008). “Fundamentals of Service Science” *Journal of the Academy of Marketing Science* 36 (1), 18–20.

Meuter, M. L., A. L. Ostrom, R. I. Roundtree and M. J. Bitner (2000). “Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters” *Journal of Marketing* 64 (3), 50–64.

Nickerson, R. C., U. Varshney and J. Muntermann (2013). “A method for taxonomy development and its application in information systems” *European Journal of Information Systems* 22 (3), 336–359.

Østerlund, C., M. H. Jarrahi, M. Willis, K. Boyd and C. Wolf (2021). “Artificial Intelligence and the World of Work, a Co-constitutive Relationship” *Journal of the Association for Information Science and Technology* 72 (1), 128–135.

Ostrom, A. L., D. Fotheringham and M. J. Bitner (2019). “Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences”. In P. P. Maglio, C. A. Kieliszewski, J. C. Spohrer, K. Lyons, L. Patrício and Y. Sawatani (eds.) *Handbook of Service Science, Volume II*, pp. 77–103. Cham: Springer International Publishing.

Paluch, S. and J. Wirtz (2020). “Artificial Intelligence and Robots in the Service Encounter” *Journal of Service Management Research* 4 (1), 3–8.

Poser, M., S. Singh and E. Bittner (2021). “Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents”. In: *54th Hawaii International Conference on System Sciences*. Virtual Conference.

Prentice, C. and M. Nguyen (2020). “Engaging and Retaining Customers with AI and Employee Service” *Journal of Retailing and Consumer Service* 56, 102186.

Püschel, L., M. Röglinger and H. Schlott (2016). “What’s in a Smart Thing? Development of a Multi-Layer Taxonomy”. In: *37th International Conference on Information Systems*. Dublin, Ireland.

Raj, M. and R. Seamans (2019). “Primer on Artificial Intelligence and Robotics” *Journal of Organization Design* 8, 1–14.

Robinson, S. G., C. Orsingher, L. Alkire, A. de Keyser, M. Giebelhausen, K. N. Papamichail, P. Shams and M. S. Temerak (2020). “Frontline Encounters of the AI Kind: An Evolved Service Encounter Framework” *Journal of Business Research* 116, 366–376.

Sampson, S. E. and C. M. Froehle (2006). “Foundations and Implications of a Proposed Unified Services Theory” *Production and Operations Management* 15 (2), 329–343.

Sarker, I. H. (2021). “Machine Learning: Algorithms, Real World Applications and Research Directions” *SN Computer Science* 2 (160), 1–21.

Scherer, A., N. V. Wünderlich and F. von Wangenheim (2015). “The Value of Self-Service: Long-Term Effects of Technology-Based Self-Service Usage on Customer Retention” *MIS Quarterly* 39 (1), 177–200.

Short, J. C., D. J. Ketchen and T. B. Palmer (2002). “The Role of Sampling in Strategic Management Research on Performance: A Two-Study Analysis” *Journal of Management* 28 (3), 363–385.

Sonnenberg, C. and J. vom Brocke (2012). “Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research”. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, K. Peffers, M. Rothenberger and B. Kuechler (eds.) *Design Science Research in Information Systems. Advances in Theory and Practice*, pp. 381–397. Berlin, Heidelberg: Springer.

Sujata, J., D. Aniket and M. Mahasingh (2019). “Artificial Intelligence Tools for Enhancing Customer Experience” *International Journal of Recent Technology and Engineering* 8 (2S3), 700–706.

Svenningsson, N. and M. Faraon (2019). “Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots”. In: *2nd Artificial Intelligence and Cloud Computing Conference*. Kobe, Japan.

Szopinski, D., T. Schoormann and D. Kundisch (2019). “Because Your Taxonomy is Worth It: Towards a Framework for Taxonomy Evaluation”. In: *27th European Conference on Information Systems*. Stockholm, Sweden.

Vargo, S. L., P. P. Maglio and M. A. Akaka (2008). “On value and value co-creation: A service systems and service logic perspective” *European Management Journal* 26 (3), 145–152.

Venable, J., J. Pries-Heje and R. Baskerville (2016). “FEDS: A Framework for Evaluation in Design Science Research” *European Journal of Information Systems* 25 (1), 77–89.

vom Brocke, J., A. Simons, K. Riemer, B. Niehaves, R. Plattfaut and A. Cleven (2015). “Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research” *Communications of the Association for Information Systems* 37 (1), 205–224.

Webster, J. and R. T. Watson (2002). “Analyzing the Past to Prepare for the Future: Writing a Literature Review” *MIS Quarterly* 26 (2), xiii–xxiii.

Wilson, H. J. and P. R. Daugherty (2018). “Collaborative Intelligence: Humans and AI are Joining Forces” *Havard Business Review* 96 (July-August), 115–123.

Wintersberger, P., T. Klotz and A. Riener (2020). “Tell Me More: Transparency and Time-Fillers to Optimize Chatbots’ Waiting Time Experience”. In: *11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. Tallinn, Estonia.

Wirtz, J., P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch and A. Martins (2018). “Brave New World: Service Robots in the Frontline” *Journal of Service Management* 29 (5), 907–931.

Xu, Y., C.-H. Shieh, P. van Esch and I.-L. Ling (2020). “AI Customer Service: Task Complexity, Problem-solving Ability, and Usage Intention” *Australasian Marketing Journal* 28 (4), 189–199.

Zhang, B., H. Lin, S. Zuo, H. Liu, Y. Chen, L. Li, H. Ouyang and B. Yuan (2020). “Research on Intelligent Robot Engine of Electric Power Online Customer Services

Based on Knowledge Graph". In: *4th International Conference on Innovation in Artificial Intelligence*. Xiamen China.

15 Gamifying the Human-in-the-Loop: Toward Increased Motivation for Training AI in Customer Service

Wiethof, C., Roocks, T., and Bittner, E. A. C. 2022b. "Gamifying the Human-in-the-Loop: Toward Increased Motivation for Training AI in Customer Service," in *Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science, vol 13336*, H. Degen and S. Ntoa (eds.), Springer, Cham, pp. 100-117 (doi: 10.1007/978-3-031-05643-7_7).

Abstract. In this paper, we contribute to research on human-AI collaboration in the scope of Hybrid Intelligence Systems, which enable mutual augmentation and collaborative learning of both human and AI. Thereby, we address a research gap focusing on the continuance intention of customer service employees to teach AI during their work task. So far, the human-in-the-loop (HITL) approach is commonly applied to directly involve the human user in Machine Learning (ML) to actively advance AI. However, there is only little consideration of users' motivation regarding the extra effort of teaching AI during their work. To address this gap, we combine gamification and ML toward increased motivation to participate in HITL learning. Therefore, we follow the Design Science Research process and align to a framework for gamifying collaboration processes. Thus, we identify meta-requirements toward intended outcomes of gamified collaboration based on expert interviews, consequently derive design principles to gamify the process interactions and determine matching gamification elements. For demonstration, we implement the design principles with the according gamification elements in a prototype enabling customer service employees to provide feedback to an AI. Eventually, we evaluate the design principles with the prototype in user test simulations. The results reveal their successful implementation as well as the relevance of mixed gamification elements to trigger various motivation types. Additionally, we show applicability of the framework to gamify human-AI collaboration.

Keywords: Gamification, Human-in-the-loop, Customer service

15.1 Introduction

Artificial Intelligence (AI), specifically Machine Learning (ML) methods are increasingly implemented and used in customer service [1]. Thereby, companies aim to continuously elevate the efficiency of service delivery and customer satisfaction when processing customer requests. Therefore, AI-enabled technologies are deployed to autonomously reply to customers in the frontline or augment the employees in the back office [2–5]. To develop and establish AI-enabled technologies in customer service, training through ML is necessary. Traditionally, automatic ML is applied to process and find patterns in big data, e.g., for speech recognition, recommender systems and autonomous vehicles [6,7]. However, in domains that are characterized by a limited amount of data and are thus complex for automatic ML on its own, recent research introduced interactive ML (iML) for immediate optimization and learning of an ML system through direct interactions with the human user [6,8–10]. With this, ML researchers seek to combine human and AI strengths toward so-called Hybrid Intelligence Systems (HIS) [11,12] to enable their collaboration as teammates working toward the same goal [13–15]. To ensure continuous improvement of a HIS, the "human-in-the-loop" (HITL) approach is determined to train AI based on iML approaches [8,16]. By putting the human user in the loop of AI and thus giving the user more control in the development and training of the ML system, additional complex training iterations with ML experts can be reduced [6,7,9,10]. Consequently, faster and more flexible learning cycles can be implemented [8]. However, iML comes with several risks to be considered. For the system to learn, it is dependent on expert users' feedback. Users might not enjoy this monotonous task, which interrupts their usual work task flow and does not show immediate progress [17]. In such iML settings, user engagement [17] as well as satisfaction and continuance intention to use and train the ML system [18] are of uttermost importance. Therefore, motives and motivation of individual expert users are crucial [18,19]. To increase engagement and motivation, there are several options to consider, e.g., social networks [18] or gamification [18,20,21]. However, so far, there is only little consideration of the expert users, who are in the loop of AI and interactively train an ML system as an additional work task. Thus, there is a lack of research on how to motivate and encourage expert users to actively participate in iML [7,22]. Accordingly, [22] call for future research on gamifying ML and disclose the potential of gamification to assist the optimization of ML and help human users to label data. Building on its wide variety of elements (e.g., leaderboards, badges, levels) [23], and its proven successful applicability to several domains (e.g., education, workplace, health care, software) [20], we investigate the use of gamification elements to incentivize expert users in the loop of AI based on a conceptual framework for gamifying collaboration processes [24]. Thereby, we combine gamification elements and the HITL approach in the scope of human-AI collaboration. To address the presented knowledge gap, we formulated the following research questions: **Q1:** *What effects do gamification elements have on expert users in customer service training an ML system?* **Q2:** *How can gamification elements be integrated into HITL learning in customer service toward higher user motivation?* **Q3:** *Which gamification elements are suitable for motivating expert users in customer service to train an ML system?* To answer the research questions, we conduct Design Science Research (DSR) [25,26] and follow the DSR process of [27]. Hence, the remainder of the paper proceeds as follows. First, we provide insights into related work covering research on human-AI-collaboration and HIS as well as gamification. Next, we describe our research approach. In the following sections, we present design knowledge and matching gamification elements for the design and development of gamified HITL learning. We demonstrate the implementation of the design knowledge and elements in a prototype in the next section and present the results of the evaluation afterwards. Finally, we conclude our study with a discussion of findings and future research implications.

15.2 Related Work

Research in human-AI teaming and collaboration is advancing rapidly as an advantageous alternative to human replacement through AI-enabled automation [13,15,28–30]. With this promising combination of both artificial and human intelligence in terms of human-AI collaboration, [12] introduced the concept of Hybrid Intelligence, and further contribute with design knowledge for HIS combining HITL with the computer-in-the-loop (CITL) [11]. They define HIS "*as systems that have the ability to accomplish complex goals by combining human and artificial intelligence to collectively achieve superior results than each of the[m] could have done in separation and continuously improve by learning from each other*" [11]. Thus, HIS enable and support mutual augmentation, i.e., the AI augments human intelligence through CITL, and the human augments AI through HITL [12]. Such augmentation scenarios are increasingly applied in organizations [31], especially in customer service [3,32]. However, existing research primarily focuses on how to best augment the employee disclosing a research gap on augmenting the AI toward mutual augmentation. As HIS demand continuous learning in terms of mutual learning [11,12,33], a high continuance intention to use needs to be ensured [18,19]. Still, a survey of HITL for ML found that studies do not address the factor of varying motivation, which is crucial for human involvement [34]. To address this issue, [22] see potential in the application of gamification to improve ML through HITL. [35] define gamification as "*the use of game design elements in non-game contexts*". So far, it has already been implemented in several domain applications, e.g., in commerce, education, health or ideation, toward psychological outcomes, e.g., enjoyment, engagement or motivation, as well as behavioral outcomes, e.g., level and quality of participation [36]. Besides, gamification also has been applied in collaboration scenarios, e.g., employee collaboration in social software solutions [23] or software engineering [21,37]. To enable systematic gamification of collaboration processes toward higher continuance intention to use, [24] developed an initial framework including theories of meaningful [38–41] and deep [18,39–41] engagement (see Fig. 1). The framework comprises three segments based on [38]: mechanics, dynamics, and user engagement. For each segment, there are elements to be considered for gamification, as well as according examples, e.g., mechanics - gamification affordances - status, competition, self-expression etc. The fourth part of the framework is intended to put the focus on digital collaboration processes, e.g., gamification elements in collaboration processes to support mechanics. Thus, the framework can be applied to gamify collaboration toward continuance intention to use. It can be systematically utilized by (1) defining intended outcomes for user engagement, (2) deriving gamification principles for dynamics, and (3) identifying appropriate gamification elements for mechanics. Although [24] only apply their framework to a collaboration process among humans, they call for research to prove applicability of their framework in other collaboration scenarios. Thus, we consider the framework for gamifying HITL in terms of human-AI collaboration within HIS. Additionally, [22] point out that the purpose of gamification is to optimize human-computer-interaction. Hence, it should be generally applicable to human involvement in ML through HITL in human-AI-collaboration. Furthermore, to support the potential of transferability from human-human to human-AI collaboration, we refer to Social Response Theory suggesting that humans equally apply certain rules, norms, and behaviors to humans as well as computers respectively [42]. Based on this, we eventually presume that gamification cannot only be applied to human-AI collaboration but can also be encouraged through AI. This is specifically the case, when we humanize AI to make it a teammate, as we consider fellowship as experiential outcome of gamification [24].

Gamification elements in collaboration processes		
Mechanics	Gamification affordances	Status, competition, self-expression, etc.
	Gamification objects	Items, characters, visual assets, etc.
	Gamification mechanics	Rules
Gamification principles for collaboration process interactions		
Dynamics	User-system-interactions	User-to-system, system-to-user, user-to-user
	Gameful interactions	Competition, cooperation
	Playful interactions	Exploration, creation, pretending
Intended outcomes of the gamified collaboration processes		
User engagement (aesthetic/flow experience)	Meaningful engagement (aesthetic experience)	Experiential outcomes: sensory and cognitive experiences (sensation, fantasy, narrative, challenge, fellowship, discovery, expression, submission, meaning, self-expansion), attachment to outcome, attachment to system
		Instrumental outcomes: functional, related to work context, prolonged use, increased use, increased learning
	Deep engagement (flow experience)	Hedonic motivation
Continuance intention to use gamified digital collaboration processes		

Fig. 1. Framework for gamifying collaboration processes [24]

15.3 Research Approach

With this research, we aim to contribute prescriptive design knowledge in the form of applicable design principles (DPs) to the knowledge base connecting the research areas of Human-Computer-Interaction, Hybrid Intelligence, ML and Gamification in order to design a solution for the integration of gamification elements in HITL learning [25,26]. Therefore, we follow the DSR process by [27] (see Fig. 2). After the problem motivation in the introduction, we present related work including a conceptual framework for gamifying collaboration processes [24]. We apply the framework to systematically identify adequate gamification elements for HITL learning in customer service. First, we derive meta-requirements (MRs) from qualitative semi-structured interviews with domain experts [43] as objectives of a solution, i.e., expected effects of gamification (Q1). According to the findings, we formulate DPs [44] to ensure appropriate process integration and interactions (Q2). Based on these we determine matching gamification elements (Q3) and implement them within a HITL customer service process prototype for demonstration. We evaluate the DPs by using the developed prototype in user test simulations with domain experts and again conducting qualitative semistructured interviews thereafter [43,45]. Communication will be completed with this paper.

1. Problem Identification	2. Objectives of a Solution	3. Design and Development	4. Demonstration	5. Evaluation	6. Communication
Research gap of combining gamification with HITL to motivate expert users training AI.	Derivation of meta-requirements from expert interviews for the integration of gamification and HITL.	Design principles to implement gamification elements into the HITL process motivating expert users to interact with an ML system.	Instantiation of a customer service process prototype implementing HITL and gamification elements based on the design principles.	Evaluation through user test simulations and qualitative interviews with expert users in customer service assessing the implementation of the design principles.	Dissemination of design principles, and the gamified HITL process prototype in customer service.

Fig. 2. Structure along the DSR process [27]

15.4 Objectives of a Solution

By motivating expert users in customer service to train an ML system, we aim for a high continuance intention to use and train the ML system via HITL. To identify MRs, we considered the conceptual framework for gamifying collaboration processes [24] and conducted semi-structured qualitative expert interviews [43]. We interviewed nine domain experts (E1–E9) with experience in customer service. The interview guideline included questions toward user engagement [24] covering 1) the customer service process, thoughts and feelings about 2) collaborating with an AI, 3) the additional effort that comes with providing feedback, 4) ideas toward a more enjoyable way of giving feedback. To analyze the expert interviews, we conducted a thematic comparison [43] and inductively determined 16 MRs, 11 toward meaningful engagement, 5 toward deep engagement [24].

MRs Toward Meaningful Engagement. Regarding instrumental outcomes of user engagement, the trained system is expected to improve through provided feedback (E4, E6, E7) to "ensure the process is in a good quality" (E1) (**MR1**) and thus, will be able to not only make work more efficient and easier (E1 E5, E8, E9) (**MR2**), but also to educate users, especially to onboard and help new users (E1 E2, E4, E6–8) (**MR3**). Therefore, it is necessary, that all users are involved in the feedback process because "if we get the feedback from all the experts [...] then we can really learn from everyone" (E6) (**MR4**). Also, this contributes to an experiential outcome of fellowship and meaning. Apart from this, the duration of the feedback process could decrease motivation (E1, E9). Thus, if someone could "calculate the effort [...] that would motivate [...] to also put effort in it" (E4). Therefore, the required effort and time of the feedback process need to be assessable (**MR5**) and manageable (**MR6**). This comes along with the need to actually see results (E1, E3, E4, E6–9) making the progress of giving feedback visible (**MR7**) and thus further contributing to the instrumental outcomes of the user engagement. Besides that, it would be “annoying” (E1), if they would be enforced to give feedback in every interaction turn of the customer service process (E1, E2, E8), which is why users should be able to freely choose when to give feedback (**MR8**). In addition, in case they do provide feedback, this feedback should be confirmed, recognized, and appreciated toward the users (E1, E7) (**MR9**). However, qualitative feedback should be valued more than quantitative feedback (E2, E3) (**MR10**). At last, for meaningful engagement, it is important for the users to understand that the feedback process is not intended to replace any human user (E5) (**MR11**).

MRs Toward Deep Engagement. To trigger users’ hedonic motivation, some experts suggested making the process of providing feedback a competition, e.g., to “*count your [...] feedback*” (E2), and to “*get points for using [the feedback feature]*” (E9), which will eventually be rewarded (E3, E9) (**MR12**). In addition, as “*working in a team is always more enjoyable*” (E1), the users should work in teams with other users giving feedback (E1, E3) (**MR13**). At last, one expert compared the process of providing feedback for the system with feedback giving among humans: “*as if I would talk to a human [...] because we also give feedback to them. And then they take it, and then they implement it directly. [...] And I think the same goes for the AI*” (E8). This supports Social Response Theory suggesting the application of social norms and behavior toward computers. To encourage this behavior, the AI should be equipped with a personality (E2) (**MR14**) as well as casually interact with the user including actively asking for feedback (E2) (**MR15**) and showing gratitude for users’ feedback (E4) (**MR16**).

15.5 Artifact Design and Development

Based on the identified relevant MRs for user engagement with the gamified collaboration process, we derive and formulate seven action oriented DPs according to [44], which serve as gamification principles for collaboration process interactions [24] in the scope of HITL. Following the conceptual framework [24], we align our DPs with the three types of dynamics: gameful interactions, playful interactions, and user-system-interactions. Figure 3 depicts the derivation and classification of the DPs. We build on these DPs and refer to existing gamification research [21,24,36] to identify suitable gamification elements for motivating expert users in customer service to train an ML system. Following the conceptual framework for gamifying collaboration processes, we define specific gamification mechanics including affordances, objects, and rules [24] (see Fig. 4).

User-system-interactions	
DP1: Provide the system with the ability to improve itself and the process through the feedback in order for users, especially novices, to learn and work more efficiently, given that all users are involved in the feedback process.	MRs 1-4
DP2: Provide the system with an AI equipped with its own personality in order for users to get asked and thanked for giving feedback.	MRs 9, 14-16
DP3: Provide the system with an element ensuring the user's value and need for collaboration in order for the users to feel in charge of the feedback process and not fear any replacement.	MRs 8, 11
Gameful interactions	
DP4: Provide the system with a teamwork setting in order for users to work in a team.	MRs 4, 13
DP5: Provide the system with the ability to count, evaluate and compare users' feedback in order for users to be rewarded for giving feedback, given that their feedback is qualitatively valuable.	MRs 9, 10, 12
Playful interactions	
DP6: Provide the system with a visual representation of given and needed feedback in order for users to assess the progress.	MRs 5, 7
DP7: Provide the system with a mechanism for managing the progress in order for users to receive regular recognition for giving feedback.	MRs 5, 6, 9

Fig. 3. DPs with corresponding MRs

Affordances		
1	Epic Meaning – purpose, goals, overall progress of AI	DP1
2	Progress Bars – individual progress and team's progress	DP6
3	Points and Levels – points to reach lelvels, for user, team and AI	DP7
4	Competition – team competition, points per team, ranked, rewarded	DP5
Objects		
5	AI Personality – avatar, name, age, asking and thanking for feedback	DP2
Rules		
6	The better the feedback, the more points	DP5
7	Users are working in teams	DP4
8	Users are in charge – it's a collaboration, not an effort for replacement	DP3

Fig. 4. Gamification elements with corresponding DPs

15.6 Demonstration

We implement the identified gamification elements within an available use case scenario, which is provided by an organization selling projects and internships abroad to students. The use case is built upon customer service interaction between employee and customer. For each message sent by the customer, an AI will provide the employee with an FAQ-based suggestion on how to reply to the customer. Due to AI's imperfection, we integrate HITL into the system at hand, i.e., including the user in the process of giving feedback to the AI. Figure 5 depicts the gamified HITL prototype for the customer service process of the organization.

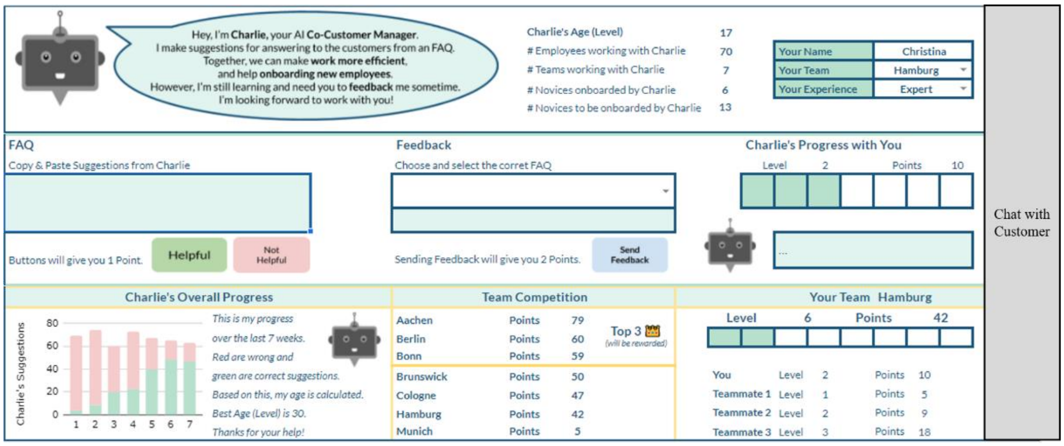

Fig. 5. Screenshot of the prototype user interface

1. *Epic Meaning.* In our prototype, the epic meaning is foremost conveyed by the AI, named Charlie, itself. Within a speech bubble, it explains the purpose of making work more efficient and especially helping to onboard new employees. Accordingly, next to it, numbers demonstrate how many novices are and still need to be onboarded as well as how many other employees are involved in working with Charlie (see Fig. 6). In addition, to confirm that the AI is improving, it has an age equivalent to its level determined by its overall progress, which is also visually represented in the prototype (see Fig. 7). It shows the AI's progress over the last seven weeks, which is measured by the number of helpful (green) and not helpful (wrong) suggestions given by the AI per week in the whole organization. The maximum age or level for the AI in this prototype is 30.

Fig. 6. Epic meaning

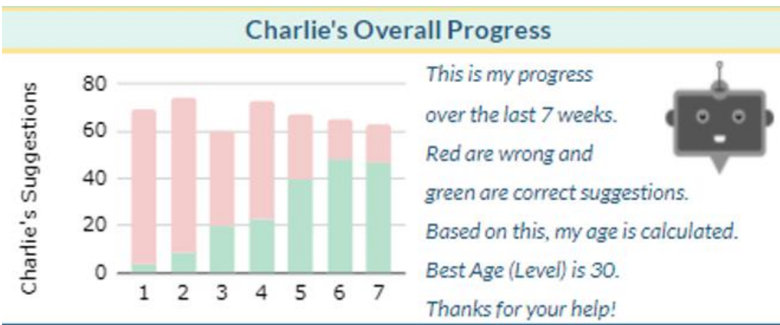

Fig. 7. Overall progress. (Color figure online)

2. *Progress Bars.* In our prototype, the user can assess the progress of given and needed feedback through an individual progress bar as well as a team's progress bar (see Fig. 8). The individual bar progresses with points received by giving feedback to the AI. As soon as the progress bar is full, it will start from the beginning again, but constitutes a new level. The same applies to the team's progress bar which is affected by all individuals of the same team.

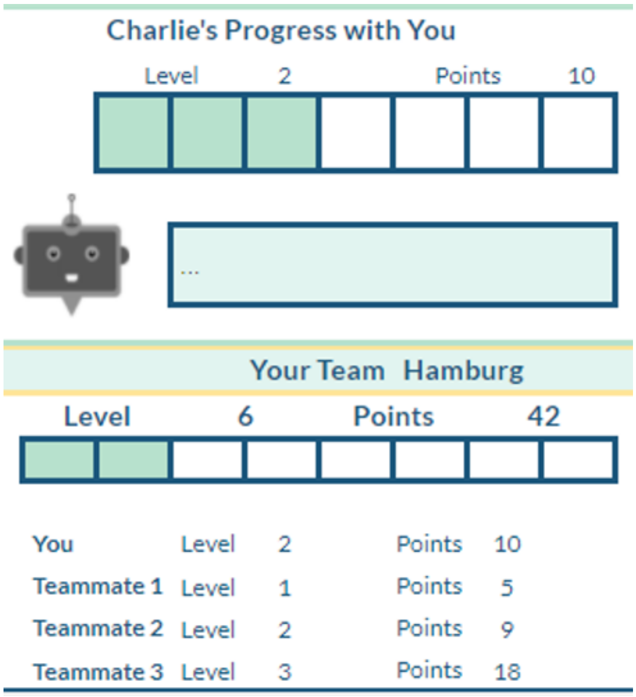

Fig. 8. Individual and team's progress bars

3. *Points and Levels.* To manage the progress, we make use of a points and levels system. Thus, for giving feedback, a user will receive points, which will increase the progress of the individual as well as the team's progress bar. When a progress bar is full, a new level is reached, and the points are further accumulated (see Fig. 8). Accordingly, also the AI progresses in levels through the feedback of all users.
4. *Competition.* The element of competition is implemented as a team competition. As the organization has several subsidiaries in different cities, the teams are defined per city. As the prototype implements a points and levels system, each user contributes with the individual points to a specific team. These points are then used to rank the teams in the context of the team competition. The ranking is visually represented as depicted in Fig. 9. Furthermore, the top three teams will be rewarded, e.g., with internal organizational benefits.

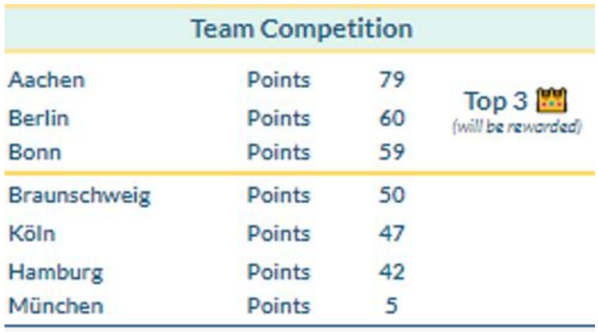

Fig. 9. Team competition.

5. *AI Personality*. The AI in our prototype is equipped with a personality including a name (Charlie), an avatar, and an age, which increases with its corresponding level. It also introduces itself as co-customer manager and asks for feedback to learn (see Fig. 6). Additionally, it also has the ability to thank users for feedback in a separate field below the individual progress bar (see Fig. 8). Hence, for provided feedback it will say "Thank you!". As an alternative, when the user does not provide feedback over a certain amount of time, the AI gives a reminder: "Please don't forget to feedback me". While the field is usually shaded in a greenish color, the latter statement will be highlighted with red color-coding.
6. *The better the feedback, the more points*. To ensure that feedback is not only good in quantitative numbers but also qualitatively valuable, we give users the opportunity to provide feedback in two different ways (see Fig. 10). For one thing, users can click a "helpful" or "not helpful" button based on their evaluation of the suggestion given by the AI. To make the feedback more valuable for the AI to learn, the users can also choose and select the correct FAQ according to the question asked by the customer. The field is populated with all relevant FAQ managed in a dropdown menu. When selecting an FAQ, a short answer to the question will appear in the field below. By clicking the "Send Feedback" button, the feedback will be submitted for the AI to learn by matching the customer question with the according FAQ. Both feedback versions can be used for helpful and for not helpful suggestions. To ensure that users will not only use the easier version of only clicking one "helpful/not helpful" button, we implement the rule that users will be rewarded for giving more qualitative feedback, i.e., they will receive two points instead of one.

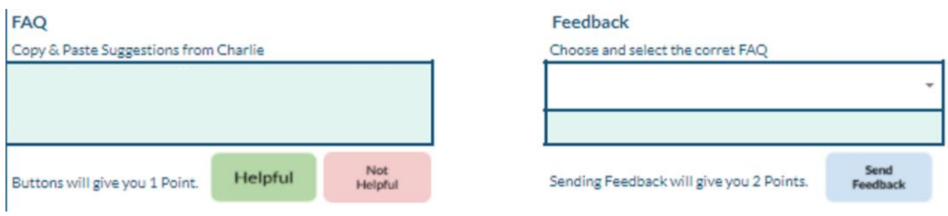

Fig. 10. Feedbacks

7. *Users are working in teams*. To encourage fellowship and enjoyment, users collect points together with other users within a team. As a team they further compete against

the other teams. As of the organizational structures, users are allocated to a team based on their location. Thus, the prototype is equipped with three fields for the users to set their name, location and level of experience, i.e., expert or novice. The selected team will then show up in the team progress section of the prototype (see Fig. 8). It shows each member of the team, their points and levels, and thus, how they contributed to their team's progress.

8. *Users are in charge - it's a collaboration, not an effort for replacement.* The prototype ensures the need of the users' participation in the collaboration. With this, the user remains in charge and thus is also free to choose when to give feedback. While using the prototype, users should be aware of the collaborative aspect, its purpose of making work more efficient and onboarding new employees. They should not fear any human replacement.

15.7 Evaluation

We assess our derived DPs through an evaluation of our developed prototype implementing the according gamification elements [45]. Therefore, we conduct user test simulations with subsequent semi-structured expert interviews [43]. The simulation included a thorough presentation of the prototype as well as exemplary interactions with the prototype. Each user test run including simulation and interview lasted from around 40 to 60 min. We selected eleven domain expert users (U1–U11) (4 female, 7 male; average age of 23.64, SD = 2.19) with experience in customer service from the organization providing our use case for demonstration at hand. The interview guideline was designed to specifically address and evaluate our DPs and gamification effects. Overall, ten of the eleven participants recognized and were mostly positive about the prototype concluding a successful implementation of the DPs and gamification elements. However, one participant was rather negative about the prototype coming from a general skepticism toward human-AI collaboration ("*I don't want to participate in my eventual re-placement*" (U7)). We accordingly considered further respective remarks with carefulness.

DP1 - Epic Meaning. The main motivation of the majority of the participants is of intrinsic nature. Thus, they give feedback because it is necessary and relevant for the AI to learn (U2, U3, U5, U8, U11), especially from many people (U5, U11), which enables better support and education for novice employees (U2, U4, U5, U7, U9, U10). The element of the Epic Meaning supports this intrinsic motivation with numbers of employees and novices participating (U1, U3–5, U9) as it is "*good as an overview [for] the user to understand also their part in the development of the AI*" (U8). Apart from that, the introduction is "*pretty motivating as well because it's inviting you to give feedback and it's done in a nice way*" (U10).

DP2 - AI Personality. The personality attributes of the AI were mostly perceived as very likable, e.g., the smiling of the avatar, politeness, eagerness to learn, friendliness and respectfulness (U2–4, U9). Furthermore, the introduction text and greeting of the AI positively influenced this perception (U3, U4, U10). Also, the interaction itself with the AI was appreciated due to the manner of asking and thanking for feedback (U3–5, U9, U10). All in all, "*it doesn't give you the feeling that you are just talking to an algorithm*" (U1), but to kind of a partner (U4, U10). However, although the majority considered the AI as "cute" (U2, U4, U6, U8–10), the participants suggested adapting the age or level accordingly, as it is and should not be a child or teenager you are talking to (U6, U9). Additionally, the age should not be "*random*" (U7) or "*related to [...] the number of feedback*" (U4) but related to the ML.

DP3 - Users are in Charge. Though one participant predicted a decreasing relevance of the collaboration with increasing AI capabilities (U6), the intention for human-AI collaboration instead of human replacement was recognized and confirmed by most of the participants (U1–5, U8–11), e.g., "*on the one hand the AI is helping me, but on the other hand I can also shape and support the AI*" (U1) and "*the emphasis is so strongly on giving feedback, which I think is a very essential part of collaborating and getting better together*" (U2). Apart from that, the most skeptical participant emphasized the impossibility to "*collaborate with a non-human being*" (U7).

DP4 - Users are Working in Teams. The teamwork setting contributes to team feeling, spirit, and dynamics (U5, U6, U9, U10). Additionally, team members can hold each other accountable (U1, U6, U8). With the according competition (DP5) this makes the feedback process more enjoyable, satisfying, and motivating (U1–3, U8). What is more, it encourages the users to include Charlie in the team, e.g., "*It makes Charlie not only a co-worker but also a member of the team that everyone can support and should help*" (U4) and "*I see Charlie as a good addition to the team*" (U5). One aspect, which should be further discussed is the transparency of the individuals' points. It might discourage the employee being the last in the hierarchy of the team (U11).

DP5 - Competition and the Better the Feedback, the More Points. Most participants stated that the competition element contributes most to their continuance intention to use the feedback system (U2, U5, U6, U9–11). Eventually, they have fun participating in a competition (U1, U2, U4, U5, U8) and like that more effort will be rewarded (U3, U4). Thus, the rule for different feedback "*makes sense*" (U2, U3, U5, U6, U8, U10, U11). In general, they like to have two options for giving feedback in terms of efficiency, flexibility, and quality (U1, U4, U5, U8–10). However, there are two aspects to consider for improving the competition element. First, the scores should be calculated in relation to team size for a fair comparison of differently sized teams (U6). Second, manipulations of the feedback system need to be prevented (U11), e.g., checking, if correct feedback is provided (U5). At last, participants raised awareness on the fact that not all people are motivated by competitions (U7, U10, U11).

DP6 - Progress Bars. To visually see the progress and one's own contribution to it was perceived as both motivating and satisfying (U1, U3–6, U9, U10), e.g., "*I really like it because you can see your own contribution for yourself and also toward the team. And therefore I'm more motivated to use [the prototype] on a regular basis and to put more effort in it*" (U4). Additionally, it positively contributes to the team spirit, dynamics, and accountability (U3, U9–11). As the team aspect of the prototype was highly recognized, it is further suggested to improve the prototype by limiting the representation of the personal progress bar: "*For me, my personal progress bar is a bit over-represented, I think the team's progress bar seems cool as I contribute to something bigger than me kind of*" (U2).

DP7 - Points and Levels. The system of gaining points to reach levels positively contributed to the motivating game-aspects of the prototype, e.g., "*I really like it. I think it's a cool incentive system in itself to level up*" (U6). Besides, it makes sense as it is easy and transparent to understand and to track progress (U2, U3, U8–10), as "*it gives you check points that you can strive towards*" (U4). For improvement, it was suggested to increase the difficulty of reaching higher levels, i.e., with each higher level, more points are required to reach the next level (U4, U6). Also, to further contribute to the team aspect, it could be relevant to not be able to train the AI alone to a higher level (U4).

Design. Eventually, regarding the prototype itself, the participants especially appreciated the aesthetics and design as it is well-structured and therefore easy to handle and to understand (U2, U3, U5, U6, U8–11). Still, the prototype is perceived as a little crowded with all elements (U2–4). Thus, it is suggested to hide or separate the competition element on another site (U2, U4) as this element shifts the focus from the customer experience to educating the AI the most (U2).

15.8 Discussion and Conclusion

Overall, we formulated seven DPs based on 16 MRs to gamify HITL in customer service. We conducted our study along the DSR process [27] and systematically aligned it to the framework for gamifying collaboration processes [24]. Hence, we first identified intended outcomes as objectives of a solution through expert interviews (Q1). Then we derived DPs from the MRs as gamification principles for collaboration process interactions (Q2). Based on these, we could eventually identify matching gamification elements to gamify HITL in customer service (Q3). By implementing the DPs with the according gamification elements in a prototype, we could evaluate them and their effects through user test simulations and following interviews with domain experts (Q1). Based on the results of our evaluation, the DPs were successfully implemented to elevate users' motivation and with this their continuance intention to use. However, the evaluation also disclosed that one gamification element alone would not have reached the intended effects over a long time and not with all users, i.e., gamification effects strongly depend on the users and how to individually motivate them [36]. Thus, a good mixture of elements is required to trigger the motivation of the various individual users. For instance, while for some participants the rewarding nature of the competition (DP5) is the main motivator to give feedback to the AI and with this participate in the HITL, others do not feel affected at all by such extrinsic motivation. They mostly do not see the value in such a competition, as it does not have a bigger meaning for what they do. Hence, the Epic Meaning (DP1) is the most important and powerful element triggering intrinsic motivation and continuance intention to use. Thus, users give feedback because it has purpose affecting their work, e.g., making work more efficient and educating novices with an improved AI. Consequently, as progress bars (DP6) make users' contribution to the AI's improvement visible, they are considered equally meaningful and of intrinsically motivating nature. Seeing progress is of uttermost importance when training AI, otherwise the extra effort will be perceived useless leading to a lower continuance intention. The effects of the points and levels system to manage the progress (DP7) are twofold. For one thing, as it complements the progress visualization, it supports intrinsic motivation. For another thing, it is equally extrinsically motivating as the competition is based on the calculation of points for each team. Regarding the gamification rules, they positively supported the gamification affordances. First, it makes sense that users gain more points for more qualitative feedback (DP6). Second, as employees are still responsible for communicating with the customer and can choose, if and what feedback they want to provide to the AI, DP3 is successfully implemented to foster collaboration instead of human replacement. However, some people might hesitate to call it a "collaboration". They would rather call it a tool or support. This could come from a general disapproval, fear or inexperience toward AI and should be considered carefully. The third rule stating that users work in a team (DP4) positively impacts both the competition as well as the meaning behind the system as users are working together with their fellow colleagues to win as well as to improve the system together. At last, regarding the AI personality (DP2), we successfully confirm the application of personality attributes to an AI as gamification object toward a more enjoyable experience. All in all, our results show potential for applying gamification in HITL learning. Thus, our study contributes to HIS research [12] combining the field of gamification with ML [22] in an augmentation scenario [31]. We therefore provide prescriptive design knowledge toward a theory of design and action in the form of seven DPs for gamifying HITL learning in customer service as well as an evaluated prototype implementing gamification elements following the DPs [25,26,45]. Additionally, we can confirm applicability of the framework for gamifying collaboration processes to human-AI collaboration processes [24]. We encourage both practitioners and researchers to draw on our findings to gamify HITL learning, as well as align with the framework to systematically gamify other collaboration process scenarios. Besides the promising results of this research, there are a few limitations to consider. First, the implementation for demonstration as well as evaluation is limited to only one organization. Additionally, we only performed user test simulations. Thus, we encourage future research to evaluate our DPs in various naturalistic settings. Second, we only conducted a qualitative situational evaluation. It would be valuable to quantitatively measure gamification effects on satisfaction, motivation, and continuance intention over a longer period of time using the system. Eventually, our DPs and prototype can serve as a fundament for combining gamification and ML as well as a complement to HIS research toward advancing HITL learning.

15.9 References

1. Statistik der Woche. https://www.heise.de/news/Statistik-der-Woche-Wo-KI-in-der-deutschen-Wirtschaft-boomt-6172430.html. Accessed 8 Feb 2021
2. Huang, M.-H., Rust, R.T.: Artificial intelligence in service. J. Serv. Res. **21**(2), 155–172 (2018)
3. de Keyser, A., K¨ocher, S., Alkire, L., Verbeeck, C., Kandampully, J.: Frontline service technology infusion: conceptual archetypes and future research directions. J. Serv. Manag. **30**(1), 156–183 (2019)
4. Robinson, S., et al.: Frontline encounters of the AI kind: an evolved service encounter framework. J. Bus. Res. **116**, 366–376 (2020)
5. Xu, Y., Shieh, C.-H., van Esch, P., Ling, I.-L.: AI customer service: task complexity, problem-solving ability, and usage intention. Australas. Mark. J. **28**(4), 189–199 (2020)
6. Holzinger, A.: Interactive Machine Learning (iML). Informatik Spektrum **39**(1), 64–68 (2016)
7. Holzinger, A., Plass, M., Holzinger, K., Cri, san, G.C., Pintea, C.-M., Palade, V.: Towards interactive Machine Learning (iML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: Buccafurri, F., Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CDARES 2016. LNCS, vol. 9817, pp. 81–95. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45507-5_6
8. Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: the role of humans in interactive machine learning. AI Mag. **35**(4), 105–120 (2014)
9. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. **3**, 119–131 (2016)
10. Holzinger, A., Plass, M., Holzinger, K., Crisan, G.C., Pintea, C.-M., Palade, V.: A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop. arXiv:1708.01104v1 (2017)
11. Dellermann, D., Calma, A., Lipusch, N.,Weber, T.,Weigel, S., Ebel, P.: The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS), Hawaii, USA (2019)
12. Dellermann, D., Ebel, P., S¨ollner, M., Leimeister, J.M.: Hybrid intelligence. Bus. Inf. Syst. Eng. **61**, 637–643 (2019)

13. Bittner, E., Oeste-Reis, S., Leimeister, J.M.: Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS), Hawaii, USA (2019)
14. Frick, W.: When your boss wears metal pants. Harvard Bus. Rev. **93**, 84–89 (2015)
15. Seeber, I., et al.: Machines as teammates: a research agenda on AI in team collaboration. Inf. Manag. **57**, 103174 (2020)
16. Martínez, M.A.M., Nadj, M., Maedche, A.: Towards an integrative theoretical framework of interactive machine learning systems. In: Proceedings of the 27th European Conference on Information Systems, Stockholm-Uppsala, Sweden (2019)
17. Dudley, J.J., Kristensson, P.O.: A review of user interface design for interactive machine learning. ACM Trans. Interact. Intell. Syst. **8**(1), 1–37 (2018)
18. Lowry, P.B., Gaskin, J.E., Moody, G.D.: Proposing the Multi-Motive Information Systems Continuance Model (MISC) to better explain end-user system evaluations and continuance intentions. J. Assoc. Inf. Syst. **16**(7), 515–579 (2015)
19. Bhattacherjee, A., Premkumar, G.: Understanding changes in belief and attitude toward information technology usage. A theoretical model and longitudinal test. Manag. Inf. Syst. Q. **28**(2), 229–254 (2004)
20. Darejeh, A., Salim, S.S.: Gamification solutions to enhance software user engagement - a systematic review. Int. J. Hum.-Comput. Interact. **32**(8), 613–642 (2016)
21. Steffens, F., Marczak, S., Figueira Filho, F., Treude, C., Singer, L., Redmiles Ban Al-Ani, D.: Using gamification as a collaboration motivator for software development teams. A preliminary framework. In: Brazilian Symposium on Collaborative Systems, Salvador, Brazil (2015)
22. Khakpour, A., Colomo-Palacios, R.: Convergence of gamification and machine learning: a systematic literature review. Technol. Learn. **26**, 597–636 (2021)
23. Meske, C., Brockmann, T., Wilms, K., Stieglitz, S.: Gamify employee collaboration - a critical review of gamification elements in social software. In: Australasian Conference on Information Systems, Adelaide (2016)
24. Wiethof, C., Tavanapour, N., Bittner, E.: Designing and evaluating a collaborative writing process with gamification elements: toward a framework for gamifying collaboration processes. AIS Trans. Hum.-Comput. Interact. **13**(1), 38–61 (2021)
25. Gregor, S.: The nature of theory in information systems. MIS Q. **30**(3), 611–642 (2006)
26. Gregor, S., Hevner, A.: Positioning and presenting design science research for maximum impact. MIS Q. **37**(2), 337–355 (2013)
27. Peffers, K., et al.: The design science research process. A model for producing and presenting information systems research. In: Proceedings of the International Conference on Design Science Research in Information Systems and Technology, Claremont, CA (2006)

28. Cheng, X., Yin, G., Azadegan, A., Kolfschoten, G.: Trust evolvement in hybrid team collaboration: a longitudinal case study. Group Decis. Negot. **25**(2), 267–288 (2015). https://doi.org/10.1007/s10726-015-9442-x
29. Przybilla, L., Baar, L., Wiesche, M., Krcmar, H.: Machines as teammates in creative teams: digital facilitation of the dual pathway to creativity. In: Proceedings of SIGMIS-CPR, Nashville, TN, USA (2019)
30. Wiethof, C., Tavanapour, N., Bittner, E.: Implementing an intelligent collaborative agent as teammate in collaborative writing: toward a synergy of humans and AI. In: Proceedings of the 54th Hawaii International Conference on System Sciences (2021)
31. Benbya, H., Pachidi, S., Jarvenpaa, S.L.: Artificial intelligence in organizations: implications for information systems research. J. Assoc. Inf. Syst. **22**(2), 281–303 (2021)
32. Larivière, B., et al.: “Service Encounter 2.0”: an investigation into the roles of technology, employees and customers. J. Bus. Res. **79**, 238–246 (2017)
33. Wiethof, C., Bittner, E.: Hybrid intelligence - combining the human in the loop with the computer in the loop: a systematic literature review. In: Forty-Second International Conference on Information Systems, Austin (2021)
34. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L.: A survey of human-in-the-loop for machine learning. arXiv:2108.00941v1 (2021)
35. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining “gamification”. In: MindTrek 2011, Tampere, Finland (2011)
36. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work? - A literature review of empirical studies on gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences, Hawaii, USA (2014)
37. Marczak, S., Filho, F.F., Singer, L, Treude, C., Steffens, F., Redmiles Ban Al-Ani, D.: Studying gamification as a collaboration motivator for virtual software teams. Social issues, cultural issues, and research methods. In: Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work and Social Computing, Vancouver, BC, Canada (2015)
38. Hunicke, R., LeBlanc, M., Zubek, R.: MDA: a formal approach to game design and game research. In: Proceedings of the Challenges in Games AI Workshop, Nineteenth National Conference of Artificial Intelligence, AAAI, San Jose, USA, pp. 1–5 (2004)
39. Liu, D., Santhanam, R., Webster, J.: Toward meaningful engagement: a framework for design and research of gamified information systems. Manag. Inf. Syst. Q. **41**(4), 1011–1034 (2017)
40. Suh, A., Cheung, C.M.K., Ahuja, M., Wagner, C.: Gamification in the workplace: the central role of the aesthetic experience. J. Manag. Inf. Syst. **34**(1), 268–305 (2017)
41. Tseng, S.-L., Sun, H.: Playful design elements and stages of player experience in gamification. In: Proceedings of the 38th International Conference on Information Systems, Seoul (2017)

42. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. J. Soc. Issues **56**(1), 81–103 (2000)
43. Meuser, M., Nagel, U.: ExpertInneninterviews - vielfach erprobt, wenig bedacht. VS Verlag f¨ur Sozialwissenschaften, Wiesbaden (2002). https://doi.org/10.1007/978-3-322-93270-9_3
44. Chandra, L., Gregor, S., Seidel, S.: Prescriptive knowledge in IS research. Conceptualizing design principles in terms of materiality, action, and boundary conditions. In: Proceedings of the 48th Hawaii International Conference on System Sciences, Kauai, Hawaii, USA (2015)
45. Venable, J., Pries-Heje, J., Baskerville, R.: A comprehensive framework for evaluation in design science research. In: Peffers, K., Rothenberger, M., Kuechler, B. (eds.) DESRIST 2012. LNCS, vol. 7286, pp. 423–438. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29863-9_31

16 Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service

Wiethof, C., Poser, M., and Bittner, E. A. C. 2022a. "Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service," in *Pacific Asia Conference on Information Systems (PACIS)*, Virtual Event.

Abstract. Conversational agents (CAs) are increasingly deployed to automate online customer service encounters. Hence, researchers and practitioners have so far predominantly addressed attributes and features of customer-facing CAs toward more efficient customer request processing. However, as CAs still regularly fail to answer complex issues, the concept of Hybrid Intelligence (HI) suggests combining artificial with human intelligence in a Hybrid Intelligence System to overcome the weaknesses of CAs and service employees (SEs) and promote their strengths leading to enhanced performance results and collaborative learning through mutual augmentation. Thus, following a Design Science Research approach, we formulate design principles (DPs) to develop an employee-facing CA for augmenting SEs simultaneously to their customer interaction. We implement a CA prototype and evaluate it with 21 participants in a user test. We found that the DPs were successfully implemented. Thereby, we contribute to practice, customer service, and HI research and provide avenues for future research.

Keywords: Conversational Agents, Augmentation, Customer Service, Hybrid Intelligence

16.1 Introduction

In recent years, companies are increasingly exploring the potential of infusing information technology (IT) into online customer service to improve operational efficiency (Bitner et al. 2000; Glushko and Nomorosa 2013). Following this trend, service providers in different domains (e.g., finance, e-commerce, IT) are deploying instant messaging platforms to enable customers to interact with service employees (SEs) in real-time via chat (McLean and Osei-Frimpong 2017; McLean and Wilson 2016). This enables the execution of engaging and personalized service encounters to support customers instantly and in an individualized manner (Canhoto and Clear 2020; Huang and Rust 2018; Wirtz et al. 2018).

To further increase operational efficiency of online customer service, conversational agents (CAs) powered by artificial intelligence (AI) are progressively deployed to automate frontstage interactions with customers (Følstad and Skjuve 2019). CAs are defined as intelligent software systems that interact with users through spoken (e.g., voice-based assistants) or textual (e.g., chatbots) natural language (Bittner et al. 2019; McTear et al. 2016). Based on their analytical capabilities to quickly search large amounts of data, standardized customer requests can be processed reliably, which increases the accessibility and speed of companies' service delivery (Adam et al. 2021). However, CAs still have problems comprehending complex customer requests frequently causing service failure, which deteriorates service quality (Poser et al. 2021).

To prevent this common pitfall, augmentation strategies are currently being investigated to promote the strengths of AI and humans and compensate for the limitations of the other (Benbya et al. 2021; Jain et al. 2021; Østerlund et al. 2021). By combining their intelligences to a Hybrid Intelligence (HI), a Hybrid Intelligence System (HIS) enables AI and humans to achieve better results together than each could alone while ensuring continuous improvement through hybrid learning (Dellermann et al. 2019b; Dellermann et al. 2019a; Jain et al. 2021). Leveraging associated benefits, HIS have to be designed and developed addressing aspects of human-computer interaction to meet the requirements of input and output channels to facilitate hybrid task processing (Limerick et al. 2014; Pinhanez 2020; Rzepka and Berger 2018). With reference to research on AI, especially Machine Learning (ML), various interface modalities enabling human-computer interaction have been investigated (Amershi et al. 2014; Pinhanez 2020). For example, Dubey et al. (2020) developed a framework for human-AI collaboration to build a dashboard prototype with several AI-related functionalities to augment SEs' knowledge during customer interaction. Besides integrated AI solutions, CAs - representing AI-based agents - are another prominent class of interfaces for human-computer interaction (Glikson and Woolley 2020). Initial research has shown that CAs have a positive impact on employees' performance across various digital workplaces allowing intuitive dyadic, dialog-based interaction to receive output from information systems (IS) and to provide input and commands as well as feedback to improve the AI (Feng and Buxmann 2020; Meyer von Wolff et al. 2019b; Zheng et al. 2022). While capabilities of CAs have been extensively exploited in online customer service as customer-facing service channels, there is a lack of research on the design of employee-facing CAs to support human-human interaction (Hohenstein and Jung 2018; Meyer von Wolff et al. 2019a; Seering et al. 2019; Zheng et al. 2022). Referring to the concept of HI, combining human intelligence and AI can lead to superior results based on humans' intuitive capabilities, e.g., empathy, creativity, and flexibility and AI's analytical skills, e.g., consistency, speed, and efficiency (Dellermann et al. 2019b). Hence, augmenting the capabilities of either SE or CA through artificial or human intelligence respectively could advance SE-customer interaction (McLean and Osei-Frimpong 2017; McLean and Wilson 2016) as well as CA-customer interaction in chat-based service encounters in efficiency, speed, and individualization (Adam et al. 2021; Følstad and Skjuve 2019; Janssen et al. 2020; Rapp et al. 2021). In terms of HI, mutual augmentation then ensures both the augmentation of the SE through the CA and vice versa. Therefore, addressing the described knowledge gap, we gather requirements and insights into the feasibility of a CA that presents information to and collects feedback from a SE in parallel to the SE-customer interaction. With the SE as the center of the CA-SE-customer interaction, we firstly focus on the augmentation of the SE while ensuring the augmentation of the CA. Consequently, we formulate the following research question (RQ):

RQ: *How can an employee-facing CA be designed and developed to augment SEs during customer interaction within HIS?*

To answer the research question, we design and develop a HIS for online customer service with a conversational interface. Therefore, the paper is structured as follows: In the Related Work, we present the conceptual background and the current state of research concerning CAs and HIS in online customer service. Subsequently, we introduce the Research Approach by outlining the Design Science Research (DSR) procedure and the applied methods. In Objectives of a Solution, we introduce meta-requirements (MRs) derived from literature and expert interviews. Thereafter, we define design principles (DPs) and describe their instantiation via design features (DFs) in a technical prototype. Next, the results of the evaluation of the implemented prototype via user test are presented. In the Discussion, we outline our findings, address the limitations of the study, and identify avenues for future research. The paper ends with a conclusion.

16.2 Related Work

16.2.1 CAs in Online Customer Service

With organizations increasingly adopting AI-based technologies, CAs particularly gained in popularity among both practitioners and researchers (Benbya et al. 2021; Klopfenstein et al. 2017; Schuetzler et al. 2021). Hence, research on CAs is widespread in the field of IS and can be conducted, organized, and grouped along various perspectives, e.g., tasks, application areas, and objectives (Meyer von Wolff et al. 2019a; Meyer von Wolff et al. 2019b). For a common understanding, we define CAs referring to Diederich et al. (2022, p. 4) as “technological artifacts with which users interact through natural language, both in written and spoken form”. For our research, we limit the definition of CAs to their written form, as we specifically study CAs in the form of text-based chatbots. Eventually, CAs provide an alternative interface to graphical user interfaces (UIs) for accessing IS in a dialogical fashion via natural language (Feng and Buxmann 2020; Følstad and Brandtzaeg 2017; Følstad and Skjuve 2019; Klopfenstein et al. 2017). Hence, comprehensive research has been conducted to provide conceptual foundations, categories, design guidelines, and potential avenues for future research on CAs (Diederich et al. 2022; Feng and Buxmann 2020; Janssen et al. 2020; Meyer von Wolff et al. 2019a; Meyer von Wolff et al. 2019b). For instance, Janssen et al. (2020) established a taxonomy of design elements for domain-specific CAs along the perspectives of context, intelligence, and interaction. In terms of context, one prominent application domain is customer service (Diederich et al. 2022; Janssen et al. 2020; Meyer von Wolff et al. 2019a; Meyer von Wolff et al. 2019b). As organizations strive for automation, they exploit the potential of CAs to replace SEs with self-service solutions (Huang and Rust 2018; Meyer von Wolff et al. 2019b; Robinson et al. 2020), i.e., answering simple customer requests (Dwivedi et al. 2019; Xu et al. 2020) to provide instant customer support (Svenningsson and Faraon 2019). With this, organizations aim to increase their productivity covering efficiency and cost reduction (Brandtzaeg and Følstad 2017; Janssen et al. 2020; Meyer von Wolff et al. 2019a). Research on CAs in customer service predominantly focuses on this endeavor, i.e., the design and development of suitable customer-facing CAs, e.g., with social cues like language style or typing indicators (Gnewuch et al. 2017; Gnewuch et al. 2018; Gnewuch et al. 2020), empathy (Xu et al. 2017), human-likeness (Svenningsson and Faraon 2019), verbal anthropomorphic design (Adam et al. 2021), and considering task complexity as well as usage intention (Xu et al. 2020).

Despite research on advancing CA usage in customer service, CAs' ability to provide adequate service and support is of uttermost importance (Følstad and Skjuve 2019). However, as of now, technological advancements in AI have not yet reached a general intelligence to properly understand natural language in its full diversity. This impairs understanding and processing of, as well as reacting to customer requests and emotions, which leads to CA failure (Brandtzaeg and Følstad 2017; Dellermann et al. 2019b; Følstad and Skjuve 2019). Hence, researchers suggest involving SEs for request escalation through CA-SE handovers (Følstad and Skjuve 2019; Poser et al. 2021). While such handover scenarios enable a sequential combination of CA and SE, recent research disclosed a new perspective of combining both AI and SE's intelligence to process requests simultaneously through mutual augmentation (Keyser et al. 2019; Larivière et al. 2017; Poser et al. 2022b). The bold arrows in Figure 1 depict how this concept allows for (1) human-human interaction, i.e., SE-customer, while (2) AI augments the SE invisibly for the customer. However, as most research focuses on customer-facing CAs, there is a lack of research on employee-facing CAs (Meyer von Wolff et al. 2019a).

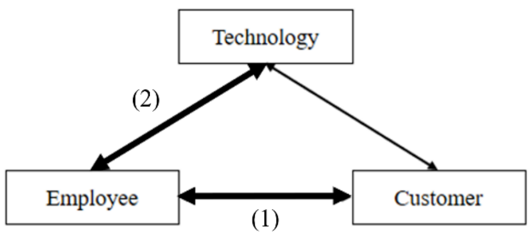

Figure 1. Frontline service technology infusion, augmentation scenario (Keyser et al. 2019), adapted

16.2.2 HIS in Online Customer Service

When it comes to augmentation of both artificial and human intelligence, researchers increasingly design and develop HIS as they "*have the ability to accomplish complex goals* *by combining human and AI to collectively achieve superior results than each of them could have done in separation and continuously improve by learning from each other*" (Dellermann et al. 2019a, p. 3). Thereby, HI specifies a collaboration between humans and machines, i.e., a task is collaboratively solved by humans and machines within a HIS (Dellermann et al. 2019a). Investigating collaborative agents as well as human-machine-teaming, the concept of HI emerged from research on human-machine-collaboration (Bittner et al. 2019; Norman 2017; Seeber et al. 2020; Strohmann et al. 2019; Wiethof et al. 2021; Yu et al. 2019). This concept focuses on mutual augmentation mechanisms toward artificial and human intelligence leading to better results and continuous collaborative learning (Dellermann et al. 2019b; Wiethof and Bittner 2021).

So far, AI is implemented through UI features in HIS, e.g., dashboards (Dubey et al. 2020; Poser et al. 2022a; Wiethof and Bittner 2022) in terms of dedicated applications (Følstad and Brandtzaeg 2017). Still, due to their intuitive and engaging nature, CA interfaces are likely to become the preferred UI (Følstad and Brandtzaeg 2017; Klopfenstein et al. 2017). Nevertheless, there is a lack of research on CA involvement in human-human conversation. For instance, while Feine et al. (2020a, 2020b) investigated CA development systems focusing on the development process through interaction between domain expert and CA, Gao and Jiang (2021) and Hohenstein and Jung (2018) provided starting points for HIS with an agent focusing on the human-human-interaction. Thereby, they examine, among others, usage, efficiency, and quality of suggestions provided by the CA. These insights provide a valuable starting point to extend the study of HIS with a CA. Accordingly, Gao and Jiang (2021) call for future research on the use and evaluation of domain-specific CAs in the context of real-world tasks and natural human-human interaction. In this context, valuable results could be obtained by ensuring the motivation and goal orientation of human participants (Brandtzaeg and Følstad 2017; Janssen et al. 2020). With our paper, we address the research gap on employee-facing CAs in online customer service and adopt findings from team research toward CA-SE collaboration. Based on this, we design a human-centered HIS with a conversational interface contributing to research on HI.

16.3 Research Approach

We adopt the DSR approach to develop a socio-technical solution for a prevalent real-world problem (Gregor and Hevner 2013; Hevner et al. 2004). To structure the process of generating prescriptive design knowledge in the form of DPs for the construction of a HIS with an employee-facing CA, we follow the six steps of the established DSR method of Peffers et al. (2007) (see Figure 2).

DSR Steps	Activities
(1) Problem Identification	Identification of research gap on employee-facing CAs to augment SEs during customer interaction
(2) Objectives of a Solution	Identification of MRs based on a SLR and expert interviews
(3) Design & Development	Derivation of DPs and development of an employee-facing CA
(4) Demonstration	Implementation of a CA prototype
(5) Evaluation	User test and mixed-method evaluation of CA prototype to provide a proof-of-concept
(6) Communication	This publication

Figure 2. Research Approach with DSR Steps

The first step refers to *Problem Identification*. By reviewing core publications in Sections 1 (Introduction) and 2 (Related Work), the current real-world challenge of companies was identified. Currently, organizations face the challenge to ensure an efficient application of AI-based CAs for interaction with customers in the online customer service frontstage by accommodating their current limitations.

As part of the second step, *Objectives of a Solution*, we utilized results of a structured literature search (SLR) according to Webster and Watson (2002) and vom Brocke et al. (2015). This SLR was conducted in the context of a preceding study about characteristics of AI in (online) customer service that are structured along the dimensions of (1) **service context**, (2) **capabilities**, (3) **deliverables**, (4) **integration**, and (5) **intelligence** (Poser et al. 2022b). Using these insights, we identified MRs that define the scope, capabilities, task types, and deliverables of an employee-facing CA. Furthermore, the appearance and behavior as well as the interaction with customers and SEs, and the intelligence (data basis and its processing) are determined. To supplement these literature-based MRs with insights from experts (E1-5) in the application domain, five semi-structured interviews according to Myers and Newman (2007) were conducted with SEs from one cooperating company.

For steps three and four (*Design and Development; Demonstration*), the previously identified MRs were used to define DPs. More precisely, following the taxonomy of Möller et al. (2020), supportive DPs were developed in one iteration based on the previously derived MRs. To this end, two researchers identified thematic commonalities across MRs and formulated materiality-and-action-oriented DPs according to Chandra et al. (2015). These DPs describe how the artifact should be produced and what it should contain. In addition, a data set for the development of the CA prototype was created based on frequently asked questions (FAQs) and matching answers from the cooperating company. To instantiate the DPs, DFs were defined to guide the development and situated implementation of the prototype.

In step five *Evaluation*, applying a mixed-method approach, the CA prototype was assessed. For a user test, 21 participants were recruited via a university email distribution list. The sample consists of individuals (five female, 16 male) between the ages of 18 and 40. Eight of these participants have pertinent work experience in customer service. 19

participants indicated that they do not have experience in using CAs or use them infrequently. The user test followed a standardized procedure. First, participants were introduced to the task and functionalities of the CA prototype by members of the research team. Second, participants engaged in a customer interaction to test the prototype. Third, following the user test, participants completed a questionnaire and participated in a semi-structured interview. In this semi-naturalistic evaluation setting, customer requests during the prototype test were simulated by the research team with predefined scripts. Utilizing an interview guide, participants were asked to assess the (1) applicability and completeness of the DPs, (2) usage characteristics, performance, and (3) impact of the CA on work practices. To supplement these qualitative data, a questionnaire was used to obtain insights on (1) perceived humanness, (2) perceived usefulness, and (3) continuance intention to use. Validated scales were used for this purpose. Perceived usefulness was measured with four items based on Davis (1989) with a 5-point Likert scale. Continuance intention to use was assessed with three items according to Bhattacherjee et al. (2008) with a 5-point Likert scale. Perceived humanness was measured with six items based on a 9-point semantic differential scale according to Holtgraves and Han (2007). In addition, quantitative measures of usage behavior (click and typing behavior) based on the usage logs, were collected and analyzed. As part of step six *Communication*, we present the activities of the described DSR steps in this paper.

16.4 Objectives of a Solution

The identification of MRs is guided by the concept of HI (Dellermann et al. 2019b; Dellermann et al. 2019a). By integrating an AI-based CA into online customer service, we strive to combine both artificial and human intelligence leading to mutual augmentation. In this hybrid collaboration scenario, an employee-facing CA, as UI of a HIS, represents an artificial teammate for SEs to conduct customer interactions. According to Social Response Theory, humans perceive computers with social cues as social actors (Nass and Moon 2000). As CAs have anthropomorphic characteristics, individuals unconsciously apply social rules and develop expectations toward human-like behavior of CAs in accordance with the social context (Araujo 2018; Feine et al. 2019). Serving the role of an artificial teammate, the employee-facing CA should therefore meet SEs' expectations toward humanoid team behavior (Poser and Bittner 2020). Therefore, to structure the derivation of MRs from literature and expert interviews, we use the established input-process-output model from team research. With this model, relevant capabilities, activities, and performance outcomes for hybrid collaboration between an employee-facing CA and a SE can be assigned based on the three dimensions. In this model, (I) **inputs** refer to the capabilities of involved agents to process a given task, (II) **processes** include activities performed by involved agents to achieve a joint task goal, and (III) **outputs** concern the evaluation of the team performance and fulfillment of team members' needs (Kozlowski and Bell 2006).

Inputs address capabilities that enable the CA to behave as an artificial teammate. To serve augmentation, the CA should be deployed in the frontstage to propose suitable response suggestions to the SE for customer requests (Dellermann et al. 2019a; Ostrom et al. 2019) (**MR1**). Depending on the use case, the CA should be able to suggest responses to simple, routine requests or more demanding problems involving the processing of data and information with or without the analysis of customers' emotions (Huang and Rust 2018; Wirtz et al. 2018). As the majority of requests are simple, the CA should provide support for these customer issues (E1-4). Hence, the CA should be able to at least process FAQs and deliver adequate responses (**MR2**). To do so, the CA has to be equipped with a database comprising a vast set of problem-solution pairs to provide suitable suggestions to the SE (Krogh 2018; Schuetzler et al. 2021) (**MR3**). To reliably identify corresponding answers to a request, the CA should be able to recognize customers' intentions irrespective of the phrasing (E5) (Følstad and Brandtzaeg 2017; Hill et al. 2015; Mallios and Bourbakis 2016) (**MR4**). Considering the nature of a HIS (Dellermann et al. 2019b; Dellermann et al. 2019a), the CA should allow SEs to provide feedback during interaction. Thereby, the CA can be augmented and the knowledge base can continuously evolve after a number of customer interaction sessions (**MR5**).

Processes refer to activities performed by the CA to collaborate with the SE to provide customer service, e.g., solving customer requests. For a hybrid handling of requests in the frontstage with the joint objective to answer customer questions, the goals of SE and CA should be aligned (Elshan and Ebel 2020; Nguyen et al. 2021) (**MR6**). As the SE is responsible for the execution and control of the customer interaction, the CA performs the subordinate goal to follow the chat-based, real-time interaction between SE and customer in order to suggest appropriate responses to the SE (E1-4). In doing so, the SE should perceive the availableness of the CA for interaction while allowing its monitoring of actions (Bulu 2012; Goel et al. 2013) (**MR7**). This form of support is helpful for SEs to perform the customer interaction (E1-4). Serving the augmentation role, the involvement of the CA should not be transparent to the customer, limiting the interaction to the SE (Ostrom et al. 2019; Robinson et al. 2020) (**MR8**). During performing the joint task of processing and solving customer requests, SE and CA have to establish a shared focus and mutual understanding via interacting with one another (Nguyen et al. 2021). To adapt augmentation to the conditions of a fast-paced, synchronous SE-customer interaction, the CA should behave reactively by displaying suggestions instantly and in sync with messages from the customer to avoid delays (McLean and Wilson 2016; Portela and Granell-Canut 2017; Song and Zinkhan 2008) (**MR9**). To provide the means to process requests in a goal-oriented fashion, the CA should present multiple response options that allow SEs to choose from but do not overwhelm them (E1) (Følstad and Taylor 2020) (**MR10**). In terms of SEs' effectiveness and efficiency to complete the task, i.e., resolving a customer request, the CA should allow SEs to effortlessly use and/or adapt the provided suggestions and provide feedback (Diederich et al. 2019) (**MR11**).

Outputs refer to requirements addressing the evaluation of the results of the hybrid team by customers. As customers overestimate waiting times leading to a negative service experience, the problem should be solved in short sessions by the hybrid team (Cheong et al. 2008; McLean and Osei-Frimpong 2017). In doing so, the quality of the answers should be adequate and fulfill the need of customers for an engaging and personalized service interaction (Canhoto and Clear 2020; Turel and Connelly 2013). In this context, empathy is important to build a relationship with the customer (Paluch and Wirtz 2020). To support SEs to comprehend the emotional situation and respond to the customer's needs, the CA should offer different tonality in response suggestions for the SE (Medhi Thies et al; Xu et al. 2020) (**MR12**). Given that the nature of hybrid collaboration is relevant in addition to the outcome, a natural interaction between CA and SE should be promoted to positively influence the continuance intention of CA use by SEs (Bhattacherjee 2008; Følstad and Brandtzaeg 2017). Therefore, the CA should be designed to be user-friendly (E3), engaging, and equipped with personality traits, e.g., an icon and name for the CA (E2) (Jenkins et al. 2007; Nass et al. 1994; Nass and Moon 2000) (**MR13**). For the establishment of a personal connection, the proposed suggestions should be introduced by short and understandable messages from the CA to the SE (Nguyen et al. 2021) (**MR14**).

16.5 Artifact Design, Development, and Demonstration

16.5.1 Artifact Design

To design and develop a HIS, we considered both humans (SEs as in online customer service) and AI (CAs as with conversational interface) for collaboration toward mutual augmentation (Dellermann et al. 2019b). Therefore, based on the 14 MRs, we derived five DPs that constitute prescriptive knowledge and define what aspects should be considered (design) and how (action) an employee-facing CA can be created for a HIS in online customer service (Gregor 2006; Gregor et al. 2020) (see Figure 3). Thereby, the DPs include distinct mechanisms involving SEs and CAs as enactors and/or users (Gregor et al. 2020). Following the framework of Wache et al. (2022), we present DPs that have a balanced level of abstraction and density of concepts.

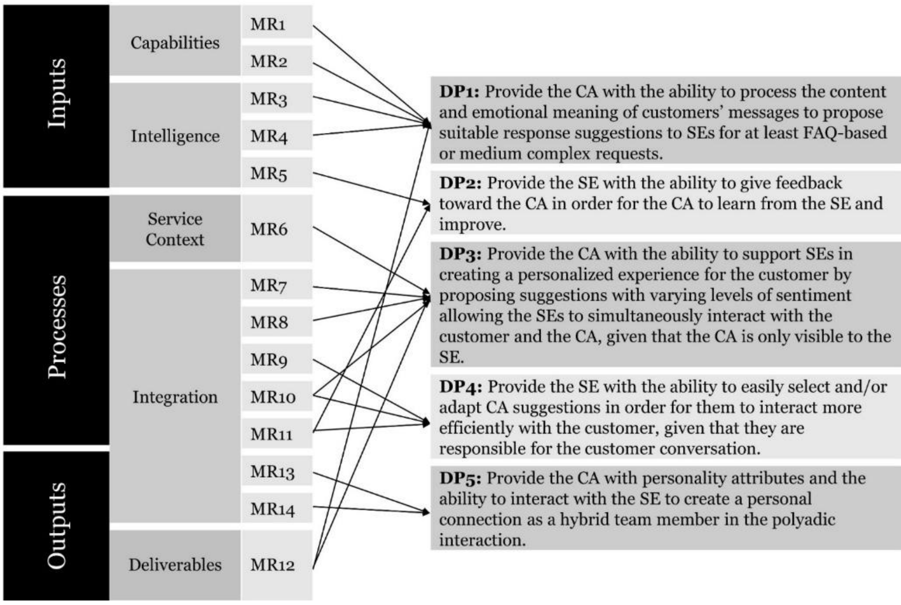

Figure 3. DPs based on MRs

The CA's capability in the form of processing incoming messages from the customer in terms of content and emotional tone is a relevant input for hybrid teamwork with the SE to propose appropriate response suggestions (**DP1**). An additional feature that is relevant in terms of a HIS allowing mutual augmentation and continuous learning is the ability of the CA to learn based on SE feedback for proposed responses (**DP2**). During the hybrid processing of a customer request, the CA, invisible to the customer, should follow the SE-customer interaction simultaneously. For the submission of suggestions, the CA should act reactively to only suggest answers for incoming customer messages, if suitable solutions are available. In terms of a positive service experience as an output of the hybrid teamwork between CA and SE, the SE should be facilitated to create a personalized interaction with the customer. Therefore, the CA should offer several response suggestions with different sentiment levels to match customers' emotional states (**DP3**). In addition, for effective hybrid teamwork, the CA should enable the SE to easily use and customize the suggested responses to support prompt reactions to customer questions (**DP4**). To enable satisfactory hybrid teamwork from the SE's perspective, the CA should establish a personal connection with the SE through a virtual identity and messages (**DP5**).

16.5.2 Artifact Development and Demonstration

To create a situated instantiation of these DPs in the form of a technical proof-of-concept prototype, we defined a set of DFs. These DFs refer to attributes and capabilities of the artifact to address the DPs (Meth et al. 2015) (see Figure 4). To allow an effortless utilization, the CA is integrated into the customer chat window and displays a text message, visible only to the SE, with two response suggestions as an immediate reaction to a customer message, if a solution for the request is available. The suggestions have the same content with different wording: (1) neutral-factual and (2) personalized (**DF1**: *DP1, DP3*). To provide SEs with the means to effortlessly use one of the suggestions, the two response options each have a button function. Once a suggestion is clicked, it is directly sent as a message to the customer (**DF2**: *DP1*, *DP4*). The learning function of the CA is enabled by SE's selection behavior. A thumb-down button can be used as feedback for unsuitable suggestions. The use of an answer by clicking on it represents positive feedback for a suggested response (**DF3**: *DP2, DP4*). An integrated edit button allows the modification of suggested responses in the text entry field before sending (**DF4**: *DP1*, *DP4*). The CA prototype is presented with a virtual identity via an avatar and its name "Sam". Furthermore, the CA proposes suggestions along with a message to the SE (**DF5**: *DP3, DP5*).

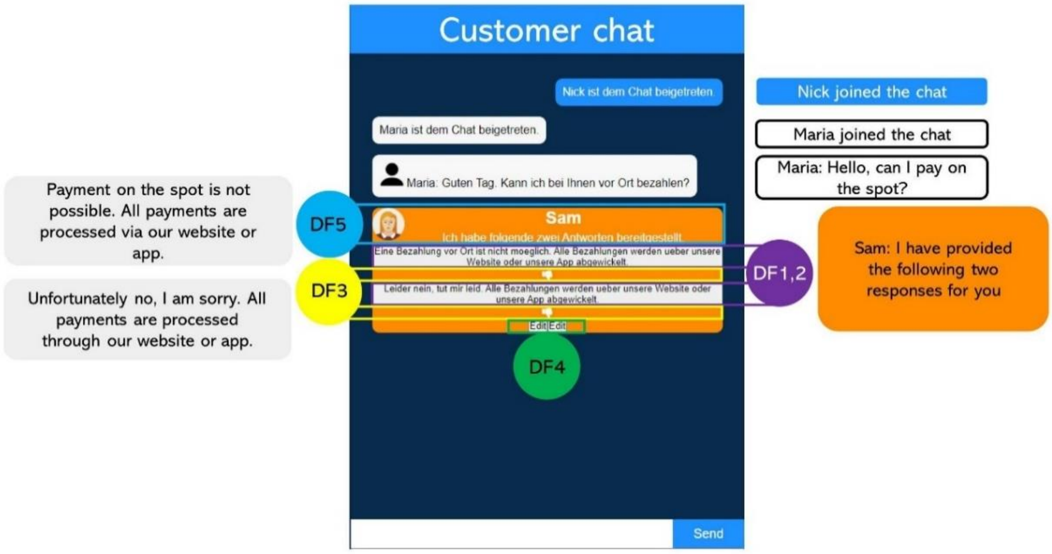

Figure 4. Web-based CA prototype with DFs

The implementation of these DFs is illustrated in Figure 4. Before development, a decision between an ML-based and rule-based architecture was made based on the availability of data. As only a small number of problem-solution pairs to customer requests were available in the dataset from the cooperating company, a rule-based architecture was used to implement a proof-of-concept CA. This rule-based prototype allows better anticipation of the CA's behavior for the user test to prevent negative influences from ML that may not yet be ideal during the user test. To propose factual and personalized CA response suggestions, a database consisting of intents with predefined tags, recognition patterns as well as corresponding responses was created based on the FAQ. A web application was developed for the integration of the CA. To incorporate the CA into the application, the web framework Flask was used. HTML was utilized for the design of the graphical UI. During operation, input in the form of incoming messages of a customer is broken down into segments and processed according to specified rules and recognition patterns. To retrieve suitable responses, intents and response specifications are generated. The identified responses are then displayed in a message from the CA within the interface via Flask. If the input does not match the predefined recognition patterns, no action is performed.

16.6 Evaluation

We assessed the deployment of an employee-facing CA augmenting SEs during customer interaction by conducting a user test with 21 participants (P1-21) (approximately 30 minutes each) and analyzed their usage logs, i.e., the interactions with the CA suggestions. To address the successful implementation of the DPs, we supplement the usage results with quantitative measures of perceived humanness, perceived usefulness, and continuance intention to use via questionnaires, and qualitative insights by conducting semi-structured interviews.

The usage logs comprise a total of 209 interactions each encompassing one customer question and the subsequent SE activities. 100 interactions show the usage of the CA's factual response suggestion while 90 reveal the usage of the CA's personalized response suggestion. Only eight response suggestions were edited, four of them personalized and four factual suggestions. The remaining 11 interactions show rejections of the suggestions, i.e., the answer was formulated without CA augmentation. Thus ~95 % of all interactions prove the successful usage of CA suggestions. Apart from that, the feedback functionality toward one suggestion was used in 25 interactions, 18 for factual and seven for personalized suggestions.

The usage log data give a first impression of the CA involvement in the customer interaction and SEs' usage of CA suggestions and thereby account for the successful implementation of the DPs. The analysis of the interview transcripts and questionnaire measures further complement the results with valuable insights as follows.

DP1: Participants like the response suggestions specifically due to their preciseness, as they are adequate, on point, and match the customer questions (P3, P7, P11, P13-15, P17-21). Derived from FAQs, the suggestions help to structure and align customer interactions (P13) as well as reduce errors (P8, P11, P19). As participants appreciate the CA for its fast access to all required information (P5, P6, P8, P10, P12, P14), its application for such a use case is considered meaningful and useful (P1, P14, P16).

DP2: The limited usage of the feedback functionality can, for one thing, be ascribed to a comprehension difficulty. More specifically, participants did not know what would happen when the button is clicked (P1, P2, P6). For another thing, it was criticized that there is only the option to explicitly give negative feedback but not positive feedback (P2, P12). Nevertheless, in general, a feedback functionality was judged to be a useful feature (P9, P15, P18, P21). To further improve the functionality, it is not only necessary to include the option to give positive feedback, but also to specify feedback (P15). This could be realized by giving access to data structures, e.g., the FAQ set (P15).

DP3: Although the response suggestion feature was highly appreciated by the participants, some responses should have been better formulated, e.g., they were lacking courtesy, were too casual, or too emotional (P4-6, P11, P12, P15, P16, P19-21). Thus, it was beneficial to have different options (P10, P11, P16-18). Still, participants recommended having more suggestions (P2, P7, P12, P16), e.g., at least a third option combining the more factual with the more personalized suggestion (P1, P15). However, some also wished for less redundancy as sometimes the response suggestions were very similar (P3, P8, P13-15). For further improvement, the CA should also provide adequate conversation beginnings and endings, e.g., greetings and goodbyes (P3, P4, P6, P11, P12, P19). At last, it was positively emphasized that the direct suggestions were provided fast and clearly (P9, P17, P18, P20, P21) allowing the SEs to simultaneously interact with the customer and the CA.

DP4: Due to its fast information processing and delivery pace, the CA increased the efficiency (P2, P3, P5, P6, P8, P10, P11, P13, P16, P19) and simplicity (P3-8, P10-15, P17-21) of work activities by allowing faster customer request processing (P1, P3-8, P11-21). This was further supported by the intuitive UI (P3-10, P12, P14, P15, P17-21) enabling an easy selection of the response suggestions (P2, P4, P10, P19) as well as an easy adaptation in terms of adding, emphasizing, individualizing, or personalizing responses (P1, P3, P6-8, P10, P11, P14, P15, P18, P19, P21). For improvement, the "edit" buttons could be better positioned, e.g., directly attached to each response suggestion (P1, P3, P7, P9, P10, P16, P17).

DP5: Regarding the perceived humanness of the CA, the analysis of questionnaire measures delivered mixed results (*Mean*: 6.67, *SD*: 1.56, *Median*: 7.00) that are supplemented by further evidence in the interview transcripts. While some reasoned for the CA's humanness, e.g., predominantly because of its language style and suggestions (P1, P3, P7, P8, P15, P18-21), the lack of appropriate greetings, goodbyes, and empathetic filler words as well as the limited response suggestions made the CA less human-like for others (P2, P5, P6, P12, P14). Nevertheless, none of the participants pointed out an impact of perceived humanness on their satisfaction with the CA.

Eventually, the high continuance intention to use (*Mean*: 3.87, *SD*: 1.13, *Median*: 4.00), based on satisfaction, and the perceived usefulness (*Mean*: 4.27, *SD*: 0.63, *Median*: 4.33) strongly prove the successful implementation of the DPs.

16.7 Discussion

The results of our work contribute to research in online customer service, HI, and employee-facing CAs by addressing hybrid teamwork between SE and CA. By investigating an employee-facing CA as an interface of a HIS, we present one possible way to create an augmentation scenario in the frontstage of online customer service. To this end, 14 MRs were identified based on findings from literature and practice to define five overarching DPs, whose implementation could provide employee-facing CAs with the capability to handle customer requests simultaneously to and as support for SEs as part of a HIS.

To gain insights into the applicability and completeness of our DPs, we developed a proof-of-concept CA. The deployment of this CA for a user test showed that the instantiated DPs supported participants in performing a customer interaction with provided response suggestions. This is reflected by the high number of utilized suggestions by participants (198 used out of 209 suggestions) and supported by their high usefulness rating of the CA along with a strong intention to work with it again in the future. With DP1 and DP4, the combination of the CA capabilities to propose suitable response suggestions and the opportunity for SEs to effortlessly use or adapt them, SEs are enabled to efficiently handle customer requests in real-time by utilizing the provided information probably causing fewer errors. The demonstrated usefulness of the CA's ability to quickly propose suggestions to the SE in sync with the customer interaction is consistent with research findings indicating that customers expect prompt responses in online customer service contexts (McLean and Osei-Frimpong 2017; McLean and Wilson 2016). Therefore, it is conceivable that SE's use of responses might have a positive effect on the customers' service experience. An additional factor that can contribute to this perception of customers is the adaptation of responses to their current situation (Canhoto and Clear 2020; Wirtz et al. 2018).

To individualize the interaction, DP3 addresses the proposal of different response suggestions by the CA in relation to the current level of sentiment of customer messages. The interview data suggest that participants perceived the provided number and content of responses to be beneficial. This finding is supported by the usage logs, which show that participants heavily relied on the suggestions using factual and personalized responses to the same extent. Despite the positive assessment of DP3, the interviews revealed that user test participants would have preferred multiple response options. This result is inconsistent with the practice-based requirements, as experts reported that they would want two suggestions. This discrepancy could be explained by the fact that suggestions by the implemented CA are too context-independent due to its rule-based architecture.

In terms of HI research, HIS do not only support the augmentation of SEs through CA suggestions but should also enable CA augmentation through the SE (Dellermann et al. 2019b; Dellermann et al. 2019a). This is predominantly realized by implementing feedback functionalities (Abdel-Karim et al. 2020; Kulesza et al. 2015; Lees et al. 2011; Oliveira et al. 2020; Schneider and Handali 2019). With DP2, the SEs are provided with the option to critically reinforce the CA by using a "thumb-down" button. Even though the analysis of the interviews revealed that this feature is important confirming mutual augmentation intentions, the user test showed that the instantiation of DP2 via the feedback button was not distinct nor fully comprehensible. Thus, many hesitated to use the feedback button not knowing its utility or impact on the interaction with either the customer or the CA. Another reason for the minor usage of the feedback button (25 of 209 interactions) could also be the high satisfaction with the suggestions. As ~ 95 % of the suggestions were used or adapted, many participants did not see the need in giving critical feedback. Research also indicates that people are more likely to be polite and rather positive toward a computer (Nass et al. 1994; Nass and Moon 2000). Hence, for further developments, DP2 could be enhanced by, e.g., integrating a "thumb-up" button for positive feedback. Additionally, future research could specifically focus on this learning feature, i.e., how to best augment the CA.

At last, DP5 was implemented to support the CA's role as a team member ensuring a satisfying hybrid collaboration between SE and CA. Following Social Response Theory (Nass and Moon 2000), the CA was instantiated with a human-like appearance equipped with social cues and anthropomorphic attributes. In general, the evaluation reveals a human-like perception of the CA among the SEs. However, this is tightly connected with the successful implementation of DP3 providing the CA with the ability to propose more personalized suggestions. Hence, the employees could personalize the response options toward the customer. As opposed to CA criticism addressing unnatural interactions (Grudin and Jacques 2019) or response limitations (Amershi et al. 2014; Harms et al. 2019), most participants were surprisingly positive about the variety and use of language. Even though they knew that the nature of the CA is artificial, most reasoned in favor of the perceived humanness. For the ones reasoning against perceived humanness, the CA could be improved by an increased usage of empathetic words as well as conversation beginnings and endings. Nevertheless, to specifically facilitate and nurture the team connection of SE and CA, future research could focus on their distinct interactions, e.g., increased personalization. Additionally, to better examine the impact of perceived humanness on the SEs' satisfaction with the CA, future research could conduct experiments including a baseline condition without a virtual identity of the CA.

Besides the promising results of this research, there are a few limitations to consider. First, we assessed the DPs by deploying our prototype in an artificial evaluation setting as the customers were simulated by the research team. This might have also influenced the high response acceptance by the SEs, i.e., technical performance in terms of typing errors or incomprehensible questions was not in the scope of this study. Second, even though we could confirm a successful application of a rule-based CA, we call for further research to examine if an ML-based CA would have achieved better or different results. This also applies to the amount and form of response suggestions. Additionally, future research should implement a prototype in a more naturalistic setting to evaluate, whether the results remain positive, specifically in terms of SE support and enhanced operational performance. Besides performance, service quality is a criterion that should be addressed in future research by evaluating customer satisfaction with the service provided. At last, our design knowledge for employee-facing CAs could be combined with research on CA-SE handovers (Følstad and Skjuve 2019; Poser et al. 2021), i.e., integrating sequential with simultaneous customer request handling.

16.8 Conclusion

Our paper investigates a conversational interface for HIS with an employee-facing CA toward augmentation and hybrid teamwork between CA and SE enabling their simultaneous handling of customer requests in real-time customer conversations. Thereby, we address the real-world challenge of companies to use AI-based CAs in an efficient way in the frontstage of online customer service, considering their current weaknesses as well as strengths. Following a DSR approach, we present prescriptive knowledge about design and action in the form of MRs and higher-order DPs for an employee-facing CA that, invisible to the customer, supports SEs to process customer requests in real-time. Our findings contribute to research and practice. The results have implications for research about employee-facing CAs, as we provide and extend existing design knowledge. In addition, our investigation of an employee-facing CA contributes to research on HI, as we explore a conversational interface for HIS. Our results serve as promising first insights that can be used to design mutual augmentation and learning via natural language interaction. Furthermore, we adopted findings from team research in order to design a human-centered HIS. The inclusion of this research stream can serve as a starting point for future research. The investigation of the use of an employee-facing CA also has implications for online customer service research, as it offers a solution to find a suitable balance between efficiency (automation) and personalization (human-touch) to conduct service encounters. Besides research-related aspects, we also contribute to practice. The presentation of implementable design knowledge enables companies to deploy a CA in their online customer service. Furthermore, the evaluation of the CA indicates advantages of using this CA in combination with SEs. These insights are helpful for companies to decide on the application of an employee-facing CA in order to increase their operational efficiency.

16.9 Acknowledgment

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, 02L18A111).

16.10 References

Abdel-Karim, B. M., Pfeuffer, N., Rohde, G., and Hinz, O. 2020. “How and What Can Humans Learn from Being in the Loop?” *KI - Künstliche Intelligenz* (34), pp. 199-207.

Adam, M., Wessel, M., and Benlian, A. 2021. “AI-based chatbots in customer service and their effects on user compliance,” *Electronic Markets* (31:4), pp. 427-445.

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. 2014. “Power to the People: The Role of Humans in Interactive Machine Learning,” *AI Magazine* (35:4), pp. 105-120.

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

Benbya, H., Pachidi, S., and Jarvenpaa, S. 2021. "Artificial Intelligence in Organizations: Implications for Information Systems Research," *Journal of the Association for Information Systems* (22:2), pp. 1-25.

Bhattacherjee, A. 2008. "Understanding Information Systems Continuance: A Theoretic Extension and Empirical Test," *Journal of Computer Information Systems* (49:1), pp. 17-26.

Bhattacherjee, A., Perols, J., and Sanford, C. 2008. "Information Technology Continuance: A Theoretic Extension and Empirical Test," *Journal of Computer Information Systems* (49:1), pp. 17-26.

Bitner, M. J., Brown, S. W., and Meuter, M. L. 2000. "Technology Infusion in Service Encounters," *J. of the Acad. Mark. Sci. (Journal of the Academy of Marketing Science)* (28:1), pp. 138-149.

Bittner, E., Oeste-Reiß, S., and Leimeister, J. M. 2019. "Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work," in *52nd Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA.

Brandtzaeg, P. B., and Følstad, A. 2017. "Why people use chatbots," in *4th International Conference on Internet Science*, Thessaloniki, Greece.

Bulu, S. T. 2012. "Place Presence, Social Presence, Co-presence, and Satisfaction in Virtual Worlds," *Computers & Education* (58:1), pp. 154-161.

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

Chandra, L., Seidel, S., and Gregor, S. 2015. "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," in *48th Hawaii International Conference on System Sciences (HICSS)*, Kauai, USA.

Cheong, K., Kim, J., and So, S. 2008. "A Study of Strategic Call Center Management: Relationship between Key Performance Indicators and Customer Satisfaction," *European Journal of Social Sciences* (6:2), pp. 286-276.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319-340.

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., and Ebel, P. 2019a. "The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems," in *52nd Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA.

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. 2019b. "Hybrid Intelligence," *Business & Information Systems Engineering* (61), pp. 637-643.

Diederich, S., Brendel, A. B., Lichtenberg, S., and Kolbe, L. 2019. "Design for Fast Request Fulfillment or Natural Interaction? Insights from an Experiment with a Conversational Agent," in *27th European Conference on Information Systems (ECIS)*, Stockholm-Uppsala, Sweden.

Diederich, S., Brendel, A. B., Morana, S., and 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 (JAIS)* (23:1), pp. 96-138.

Dubey, A., Abhinav, K., Jain, S., Arora, V., and Puttaveerana, A. 2020. "HACO: A Framework for Developing Human-AI Teaming," in *13th Innovations in Software Engineering Conference (ISEC)*, Jabalpur, India.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Janssen, M., and More Authors 2019. "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management* (57).

Elshan, E., and Ebel, P. 2020. "Let's Team Up: Designing Conversational Agents as Teammates," in *41st International Conference on Information Systems (ICIS)*, virtual.

Feine, J., Gnewuch, U., Morana, S., and Maedche, A. 2019. "A Taxonomy of Social Cues for Conversational Agents," *International Journal of Human-Computer Studies* (132), pp. 138-161.

Feine, J., Morana, S., and Maedche, A. 2020a. "A Chatbot Response Generation System," in *MuC*, Magdeburg, Germany.

Feine, J., Morana, S., and Maedche, A. 2020b. "Designing Interactive Chatbot Development Systems," in *41st International Conference on Information Systems (ICIS)*, virtual.

Feng, S., and Buxmann, P. 2020. "My Virtual Colleague: A State-of-the-Art Analysis of Conversational Agents for the Workplace," in *53rd Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA.

Følstad, A., and Brandtzaeg, P. B. 2017. "Chatbots and the New World of HCI," *interactions* (24:4), pp. 38-42.

Følstad, A., and Skjuve, M. 2019. "Chatbots for Customer Service: User Experience and Motivation," in *1st International Conference on Conversational User Interfaces*, Dublin, Ireland.

Følstad, A., and Taylor, C. 2020. "Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives," in *Chatbot Research and Design*, A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger and P. B. Brandtzaeg (eds.), Cham: Springer International Publishing, pp. 201-214.

Gao, Z., and Jiang, J. 2021. "Evaluating Human-AI Hybrid Conversational Systems with Chatbot Message Suggestions," in *CIKM*, virtual, QLD, Australia.

Glikson, E., and Woolley, A. W. 2020. “Human Trust in Artificial Intelligence: Review of Empirical Research,” *Academy of Management Annals* (14:2), pp. 627-660.

Glushko, R. J., and Nomorosa, K. J. 2013. “Substituting Information for Interaction: A Framework for Personalization in Service Encounters and Service Systems,” *Journal of Service Research* (16:1), pp. 21-38.

Gnewuch, U., Morana, S., Adam, M. T. P., and Maedche, A. 2018. ““The Chatbot is typing …” – The Role of Typing Indicators in Human-Chatbot Interaction,” in *16th Annual Pre-ICIS Workshop on HCI Research in MIS*, San Francisco, USA.

Gnewuch, U., Morana, S., and Maedche, A. 2017. “Towards Designing Cooperative and Social Conversational Agents for Customer Service,” in *38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.

Gnewuch, U., Yu, M., and Maedche, A. 2020. “The Effect of Perceived Similarity in Dominance on Customer Self-Disclosure to Chatbots in Conversational Commerce,” in *28th European Conference on Information Systems (ECIS)*, Marrakesh, Morocco.

Goel, L., Johnson, N., Junglas, I., and Ives, B. 2013. “Predicting Users' Return to Virtual Worlds: A Social Perspective,” *Information Systems Journal* (23:1), pp. 35-63.

Gregor, S. 2006. “The Nature of Theory in Information Systems,” *MIS Quarterly* (30), pp. 611-642.

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

Gregor, S., Kruse, L., and Seidel, S. 2020. “Research Perspectives: The Anatomy of a Design Principle,” *Journal of the Association for Information Systems* (21:6), pp. 1622-1652.

Grudin, J., and Jacques, R. 2019. “Chatbots, Humbots, and the Quest for Artificial General Intelligence,” in *CHI*, Glasgow, Scotland, UK.

Harms, J.-G., Kucherbaev, P., Bozzon, A., and Houben, G.-J. 2019. “Approaches for Dialog Management in Conversational Agents,” *IEEE Internet Computing* (23:2), pp. 13-22.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. “Design Science in Information Systems Research,” *MIS Quarterly* (28), pp. 75-105.

Hill, J., Randolph Ford, W., and Farreras, I. G. 2015. “Real Conversations with Artificial Intelligence: A Comparison between Human–human Online Conversations and Human–chatbot Conversations,” *Computers in Human Behavior* (49), pp. 245-250.

Hohenstein, J., and Jung, M. 2018. “AI-Supported Messaging: An Investigation of Human-Human Text Conversation with AI Support,” in *CHI*, Montreal, QC, Canada.

Holtgraves, T., and Han, T.-L. 2007. “A Procedure for Studying Online Conversational Processing Using a Chatbot,” *Behavior Research Methods* (39:1), pp. 156-163.

Huang, M.-H., and Rust, R. T. 2018. “Artificial Intelligence in Service,” *Journal of Service Research* (21:2), pp. 155-172.

Jain, H., Padmanabhan, B., Pavlou, P. A., and Raghu, T. S. 2021. “Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society,” *Information Systems Research* (32:3), pp. 675–687.

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

Jenkins, M.-C., Churchill, R., Cox, S., and Smith, D. 2007. “Analysis of User Interaction with Service Oriented Chatbot Systems,” in *HCII 2007, LNCS 4552*, Berlin Heidelberg: Springer-Verlag, pp. 76-83.

Keyser, A. de, Köcher, S., Alkire, L., Verbeeck, C., and Kandampully, J. 2019. “Frontline Service Technology infusion: conceptual archetypes and future research directions,” *Journal of Service Management* (30:1), pp. 156-183.

Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. 2017. “The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms,” in *DIS*, Edinburgh, United Kingdom.

Kozlowski, S. W. J., and Bell, B. S. 2006. “Work Groups and Teams in Organizations,” in *Handbook of Psychology*, I. B. Weiner (ed.), Hoboken, USA: Wiley, pp. 412-469.

Krogh, G. von 2018. “Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing,” *Academy of Management Discoveries* (4:4), pp. 404-409.

Kulesza, T., Burnett, M., Wong, W.-K., and Stumpf, S. 2015. “Principles of Explanatory Debugging to Personalize Interactive Machine Learning,” in *IUI*, Atlanta, GA, USA.

Larivière, B., Bowen, D., Andreassen, T. W., Sirianni, N. J., Voss, C., Wünderlich, N. V., and Keyser, A. de 2017. “"Service Encounter 2.0": An investigation into the roles of technology, employees and customers,” *Journal of Business Research* (79), pp. 238-246.

Lees, M., Evans, R., and Mareels, I. 2011. “Autonomous Performance Monitoring of an On-Line Real-Time Decision Support System,” in *18th World Congress The International Federation of Automatic Control*, Milano, Italy, pp. 13080-13085.

Limerick, H., Coyle, D., and Moore, J. W. 2014. “The Experience of Agency in Human-computer Interactions: A Review,” *Frontiers in Human Neuroscience* (8), pp. 1-10.

Mallios, S., and Bourbakis, N. 2016. “A Survey on Human Machine Dialogue Systems,” in *7th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Chalkidiki, Greece.

McLean, G., and Osei-Frimpong, K. 2017. “Examining Satisfaction with the Experience during a Live Chat Service Encounter- Implications for Website Providers,” *Computers in Human Behavior* (76), pp. 494-508.

McLean, G., and Wilson, A. 2016. “Evolving the Online Customer Experience … Is there a Role for Online Customer Support?” *Computers in Human Behavior* (60), pp. 602-610.

McTear, M., Callejas, Z., and Griol, D. 2016. *The Conversational Interface: Talking to Smart Devices,* Cham: Springer.

Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., and O'Neill, J. "How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians," in *Human-Computer Interaction - INTERACT 2017,* R. Bernhaupt, G. Dalvi, A. Joshi, K. Balkrishan, J. O'Neill and M. Winckler (eds.), Cham: Springer, pp. 441-459.

Meth, H., Mueller, B., and Maedche, A. 2015. "Designing a Requirement Mining System," *Journal of the Association for Information Systems* (16:9), pp. 799-837.

Meyer von Wolff, R., Hobert, S., and Schumann, M. 2019a. "How May I Help You? –State of the Art and Open Research Questions for Chatbots at the Digital Workplace," in *52nd Hawaii International Conference on System Sciences (HICSS),* Maui, Hawaii, USA.

Meyer von Wolff, R., Masuch, K., Hobert, S., and Schumann, M. 2019b. "What Do You Need Today? –An Empirical Systematization of Application Areas for Chatbots at Digital Workplaces," in *25th Americas Conference on Information Systems (AMCIS),* Cancun, Mexico.

Möller, F., Guggenberger, T., and Otto, B. 2020. "Towards a Method for Design Principle Development in Information Systems," in *15th International Conference on Design Science Research in Information Systems and Technology (DESRIST),* Kristiansand, Norway.

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

Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81-103.

Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers are Social Actors," in *CHI*, Boston, Massachusetts, USA.

Nguyen, T. H., Waizenegger, L., and Techatassanasoontorn, A. A. 2021. ""Don't Neglect the User!" – Identifying Types of Human-Chatbot Interactions and their Associated Characteristics," *Information Systems Frontiers*.

Norman, D. 2017. "Design, Business Models, and Human-Technology Teamwork," *Research-Technology Management* (60:1), pp. 26-30.

Oliveira, C. S. de, Sanin, C., and Szczerbicki 2020. "Human Feedback and Knowledge Discovery: Towards Cognitive Systems Optimization," *Procedia Computer Science* (176), pp. 3093-3102.

Østerlund, C., Jarrahi, M. H., Willis, M., Boyd, K., and Wolf, C. 2021. "Artificial Intelligence and the World of Work, a Co-constitutive Relationship," *Journal of the Association for Information Science and Technology* (72:1), pp. 128-135.

Ostrom, A. L., Fotheringham, D., and Bitner, M. J. 2019. "Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences," in *Handbook of Service Science, Volume II*, P. P. Maglio, C. A. Kieliszewski, J. C. Spohrer, K. Lyons,

L. Patrício and Y. Sawatani (eds.), Cham: Springer International Publishing, pp. 77-103.

Paluch, S., and Wirtz, J. 2020. "Artificial Intelligence and Robots in the Service Encounter," *Journal of Service Management Research* (4:1), pp. 3-8.

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

Pinhanez, C. S. 2020. "HCI Research Challenges of the Next Generation of Conversational Systems," in *2nd Conference on Conversational User Interfaces (CUI)*, New York, USA.

Portela, M., and Granell-Canut, C. 2017. "A New Friend in Our Smartphone? Observing Interactions with Chatbots in the Search of Emotional Engagement," in *International Conference on Human Computer Interaction*, Cancun, Mexico.

Poser, M., and Bittner, E. 2020. "Hybrid Teamwork: Consideration of Teamwork Concepts to Reach Naturalistic Interaction between Humans and Conversational Agents," in *15th International Conference on Wirtschaftsinformatik*, Potsdam, Germany.

Poser, M., Singh, S., and Bittner, E. 2021. "Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents," in *54th Hawaii International Conference*.

Poser, M., Wiethof, C., Banerjee, D., Subramanian, V. S., Paucar, R., and Bittner, E. A. C. 2022a. "Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service," in *17th International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, St. Petersburg, FL, USA.

Poser, M., Wiethof, C., and Bittner, E. 2022b. "Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions," in *30th European Conference on Information Systems (ECIS)*, Timisoara, Romania.

Rapp, A., Curti, L., and Boldi, A. 2021. "The Human Side of Human-Chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-based Chatbots," *International Journal of Human-Computer Studies* (151).

Robinson, S., Orsingher, C., Alkire, L., Keyser, A. de, Giebelhausen, M., Papamichail, K. N., Shams, P., and Temerak, M. S. 2020. "Frontline encounters of the AI kind: An evolved service encounter framework," *Journal of Business Research* (116), pp. 366-376.

Rzepka, C., and Berger, B. 2018. "User Interaction with AI-enabled Systems: A Systematic Review of IS Research," in *39th International Conference on Information Systems (ICIS)*, San Francisco, USA.

Schneider, J., and Handali, J. P. 2019. "Personalized Explanation for Machine Learning: A Conceptualization," in *27th European Conference on Information Systems (ECIS)*, Stockholm-Uppsala, Sweden.

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., and Rosser, H. K. 2021. "Deciding Whether and How to Deploy Chatbots," *MIS Quarterly Executive* (20), pp. 1-15.

Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. 2020. “Machines as teammates: A research agenda on AI in team collaboration,” *Information & Management* (57:2).

Seering, J., Luria, M., Kaufman, G., and Hammer, J. 2019. “Beyond Dyadic Interactions: Considering Chatbots as Community Members,” in *CHI*, Glasgow, UK.

Song, J. H., and Zinkhan, G. M. 2008. “Determinants of Perceived Web Site Interactivity,” *Journal of Marketing* (72:2), pp. 99-113.

Strohmann, T., Siemon, D., and Robra-Bissantz, S. 2019. “Introducing The Virtual Companion Canvas - Towards Designing Collaborative Agents,” in *Workshop on Designing User Assistance in Intelligent Systems*, Stockholm, Sweden.

Svenningsson, N., and Faraon, M. 2019. “Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots,” in *AICCC*, Kobe, Japan.

Turel, O., and Connelly, C. E. 2013. “Too Busy to Help: Antecedents and Outcomes of Interactional Justice in Web-based Service Encounters,” *International Journal of Information Management* (33:4), pp. 674-683.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. “Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research,” *Communications of the Association for Information Systems* (37:1), pp. 205-224.

Wache, H., Möller, F., Schoormann, T., Strobel, G., and Petrik, D. 2022. “Exploring the Abstraction Levels of Design Principles: The Case of Chatbots,” in *17th International Conference on Wirtschaftsinformatik (WI)*, Nürnberg, Germany.

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

Wiethof, C., and Bittner, E. 2021. “Hybrid intelligence - combining the human in the loop with the computer in the loop: a systematic literature review,” in *42nd International Conference on Information Systems (ICIS)*, Austin, USA.

Wiethof, C., and Bittner, E. 2022. “Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI,” in *30th European Conference on Information Systems (ECIS)*, Timisoara, Romania.

Wiethof, C., Tavanapour, N., and Bittner, E. 2021. “Implementing an intelligent collaborative agent as teammate in collaborative writing: toward a synergy of humans and AI,” in *54th Hawaii International Conference on System Sciences (HICSS)*, virtual.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., and Martins, A. 2018. “Brave New World: Service Robots in the Frontline,” *Journal of Service Management* (29:5), pp. 907-931.

Xu, A., Liu, Z., Guo, Y., Sinha, V., and Akkiraju, R. 2017. “A New Chatbot for Customer Service on Social Media,” in *CHI*, Denver, CO, USA.

Xu, Y., Shieh, C.-H., van Esch, P., and Ling, I.-L. 2020. “AI customer service: Task complexity, problem-solving ability, and usage intention,” *Australasian Marketing Journal* (28:4), pp. 189-199.

Yu, K., Berkovsky, S., Taib, R., Zhou, J., and Chen, F. 2019. “Do I Trust My Machine Teammate? An Investigation from Perception to Decision,” in *24th International Conference on Intelligent User Interfaces (IUI),* Marina del Rey, CA, USA.

Zheng, Q., Tang, Y., Liu, Y., Liu, W., and Huang, Y. 2022. “UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library,” (arXiv:2202.09895).

17 References

Abdel-Karim, B. M., Pfeuffer, N., Rohde, G., and Hinz, O. 2020. “How and What Can Humans Learn from Being in the Loop?” *Künstliche Intelligenz* (34), pp. 199-207 (doi: 10.1007/s13218-020-00638-x).

Adam, M., Wessel, M., and Benlian, A. 2021a. “AI-based chatbots in customer service and their effects on user compliance,” *Electronic Markets* (31), pp. 427-445 (doi: 10.1007/s12525-020-00414-7).

Adam, M. T. P., Gregor, S., Hevner, A., and Morana, S. 2021b. “Design Science Research Modes in Human-Computer Interaction Projects,” *AIS Transactions on Human-Computer Interaction* (13:1), pp. 1-11 (doi: 10.17705/1thci.00139).

Adams, T. 2015. “And the Pulitzer goes to…a computer,” *The Guardian*.

Ahuja, M., and Thatcher, J. B. 2005. “Moving Beyond Intentions and Towards the Theory of Trying: Effects of Work Environment and Gender on Post-Adoption Information Technology Use,” *MIS Quarterly* (29:3), 427-459 (doi: 10.2307/25148691).

Akata, Z., Balliet, D., Rijke, M. de, Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., van Hoof, H., van Riemsdijk, B., van Wynsberghe, A., Verbrugge, R., Verheij, B., Vossen, P., and Welling, M. 2020. “A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence,” *Computer* (53:8), pp. 18-28 (doi: 10.1109/MC.2020.2996587).

Ameen, N., Tarhini, A., Reppel, A., and Anand, A. 2021. “Customer experience in the age of artificial intelligence,” *Computers in Human Behavior* (114) (doi: 10.1016/j.chb.2020.106548).

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. 2014. “Power to the People: The Role of Humans in Interactive Machine Learning,” *AI Magazine* (35:4), pp. 105-120 (doi: 10.1609/aimag.v35i4.2513).

Amorim, M., Cohen, Y., Reis, J., and Rodrigues, M. 2019. “Exploring Opportunities for Artificial Emotional Intelligence in Service Production Systems,” *IFAC-PapersOnLine* (52:13), pp. 1145-1149 (doi: 10.1016/j.ifacol.2019.11.350).

Andras, P., Esterle, L., Guckert, M., Han, T. A., Lewis, P. R., Milanovic, K., Payne, T., Perret, C., Pitt, J., Powers, S. T., Urquhart, N., and Wells, S. 2018. “Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems,” *IEEE Technology and Society Magazine* (37:4), pp. 76-83 (doi: 10.1109/MTS.2018.2876107).

Aoki, N. 2021. “The importance of the assurance that “humans are still in the decision loop” for public trust in artificial intelligence: Evidence from an online experiment,” *Computers in Human Behavior* (114) (doi: 10.1016/j.chb.2020.106572).

Araujo, T. 2018. “Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions," *Computers in Human Behavior* (85), pp. 183-189 (doi: 10.1016/j.chb.2018.03.051).

Bailey, K. D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*, SAGE Publications, Inc.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., and Rossi, M. 2018. "Design Science Research Contributions: Finding a Balance between Artifact and Theory," *Journal of the Association for Information Systems* (19:5), pp. 358-376 (doi: 10.17705/1jais.00495).

Bassano, C., Barile, S., Saviano, M. L., Cosimato, S., and Pietronudo, M. C. 2020. "AI technologies & value co-creation in a luxury goods context," in *53rd Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA, pp. 1618-1627 (doi: 10.24251/hicss.2020.199).

Benbya, H., Pachidi, S., and Jarvenpaa, S. 2021. "Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research," *Journal of the Association for Information Systems* (22:2), pp. 281-303 (doi: 10.17705/1jais.00662).

Benedikt, L., Joshi, C., Nolan, L., Henstra-Hill, R., Shaw, L., and Hook, S. 2020. "Human-in-the-loop AI in government: a case study," in *25th International Conference on Intelligent User Interfaces (IUI)*, Cagliari, Italy, pp. 488-497 (doi: 10.1145/3377325.3377489).

Benner, D., Elshan, E., Schöbel, S., and Janson, A. 2021. "What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents," in *42nd International Conference on Information Systems (ICIS)*, Austin, Texas, USA.

Berger, B., Adam, M., Rühr, A., and Benlian, A. 2021. "Watch Me Improve - Algorithm Aversion and Demonstrating the Ability to Learn," *Business & Information Systems Engineering* (63:1), pp. 55-68 (doi: 10.1007/s12599-020-00678-5).

Bhattacherjee, A. 2001. "Understanding Information Systems Continuance: An Expectation-Confirmation Model," *MIS Quarterly* (25:3), pp. 351-370 (doi: 10.2307/3250921).

Bhattacherjee, A., Perols, J., and Sanford, C. 2008. "Information Technology Continuance: A Theoretic Extension and Empirical Test," *Journal of Computer Information Systems* (49:1), pp. 17-26 (doi: 10.1080/08874417.2008.11645302).

Bhattacherjee, A., and Premkumar, G. 2004. "Understanding Changes in Belief and Attitude toward Information Technology Usage: A Theoretical Model and Longitudinal Test," *MIS Quarterly* (28:2), pp. 229-254 (doi: 10.2307/25148634).

Bitner, M. J., Brown, S. W., and Meuter, M. L. 2000. "Technology infusion in service encounters," *Journal of the Academy of Marketing Science* (28:1), pp. 138-149 (doi: 10.1177/0092070300281013).

Bittner, E. A. C., and Leimeister, J. M. 2014. "Creating Shared Understanding in Heterogeneous Work Groups: Why It Matters and How to Achieve It," *Journal of*

Management Information Systems (31:1), pp. 111-144 (doi: 10.2753/MIS0742-1222310106).

Bittner, E. A. C., Oeste-Reiß, S., Ebel, P. A., and Söllner, M. 2019a. "Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen," *HMD Praxis der Wirtschaftsinformatik* (56), pp. 34-49 (doi: 10.1365/s40702-018-00487-1).

Bittner, E. A. C., Oeste-Reiß, S., and Leimeister, J. M. 2019b. "Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work," in *52nd Hawaii International Conference on System Sciences (HICSS)*, Grand Wailea, Maui, Hawaii, USA, pp. 284-293 (doi: 10.24251/HICSS.2019.035).

Bock, D. E., Wolter, J. S., and Ferrell, O. C. 2020. "Artificial intelligence: disrupting what we know about services," *Journal of Services Marketing* (34:3), pp. 317-334 (doi: 10.1108/jsm-01-2019-0047).

Böttcher, B., and Nüttgens, M. 2013. "Überprüfung der Gebrauchstauglichkeit von Anwendungssoftware," *HMD Praxis der Wirtschaftsinformatik* (50), pp. 16-25 (doi: 10.1007/BF03342065).

Bowers, C. A., Pharmer, J. A., and Salas, E. 2000. "When Member Homogeneity is Needed in Work Teams: A Meta-Analysis," *Small Group Research* (31:3), pp. 305-327 (doi: 10.1177/104649640003100303).

Brady, M. K., Cronin, J. J., and Brand, R. R. 2002. "Performance-only measurement of service quality: a replication and extension," *Journal of Business Research* (55:1), pp. 17-31 (doi: 10.1016/S0148-2963(00)00171-5).

Brandt, M. 2021. "Statistik der Woche: Wo KI in der deutschen Wirtschaft boomt," available at https://www.heise.de/news/Statistik-der-Woche-Wo-KI-in-der-deutschen-Wirtschaft-boomt-6172430.html, accessed on Aug 9 2021.

Brandtzaeg, P. B., and Følstad, A. 2017. "Why People Use Chatbots," in *Internet Science. INSCI 2017. Lecture Notes in Computer Science, vol 10673*, I. Kompatsiaris, J. Cave, A. Satsiou, G. Carle, A. Passani, E. Kontopoulos, S. Diplaris and D. McMillan (eds.), Springer, Cham, pp. 377-392 (doi: 10.1007/978-3-319-70284-1_30).

Brandtzaeg, P. B., and Følstad, A. 2018. "Chatbots: changing user needs and motivation," *Interactions* (25:5), pp. 38-43 (doi: 10.1145/3236669).

Briggs, R. O., Kolfschoten, G. L., Vreede, G.-J. de, Albrecht, C. C., and Lukosch, S. G. 2010. "Facilitator in a Box: Computer Assisted Collaboration Engineering and Process Support Systems for Rapid Development of Collaborative Applications for High-Value Tasks," in *43rd Hawaii International Conference on System Sciences (HICSS)*, Koloa, Kauai, Hawaii, USA, pp. 1-10 (doi: 10.1109/HICSS.2010.206).

Briggs, R. O., Kolfschoten, G. L., Vreede, G.-J. de, Lukosch, S. G., and Albrecht, C. C. 2013. "Facilitator-in-a-Box: Process Support Applications to Help Practitioners Realize the Potential of Collaboration Technology," *Journal of Management Information Systems* (29:4), pp. 159-194 (doi: 10.2753/MIS0742-1222290406).

Briggs, R. O., and Vreede, G.-J. de. 2009. *ThinkLets: Building Blocks for Concerted Collaboration*, Omaha, Nebraska: University of Nebraska, Center for Collaboration Science.

Brunken, R., Plass, J. L., and Leutner, D. 2003. "Direct Measurement of Cognitive Load in Multimedia Learning," *Educational Psychologist* (38:1), pp. 53-61 (doi: 10.1207/S15326985EP3801_7).

Bulu, S. T. 2012. "Place presence, social presence, co-presence, and satisfaction in virtual worlds," *Computers & Education* (58:1), pp. 154-161 (doi: 10.1016/j.compedu.2011.08.024).

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G. S., Stumpe, M. C., and Terry, M. 2019. "Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making," in *Conference on Human Factors in Computing Systems (CHI '19)*, Glasgow, Scotland, UK (doi: 10.1145/3290605.3300234).

Campbell, C., Sands, S., Ferraro, C., Tsao, H.-Y., and Mavrommatis, A. 2020. "From data to action: How marketers can leverage AI," *Business Horizons* (63:2), pp. 227-243 (doi: 10.1016/j.bushor.2019.12.002).

Canhoto, A. I., and Clear, F. 2020. "Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential," *Business Horizons* (63:2), pp. 183-193 (doi: 10.1016/j.bushor.2019.11.003).

Carlsson, S. A., Henningsson, S., Hrastinski, S., and Keller, C. 2011. "Socio-technical IS design science research: developing design theory for IS integration management," *Information Systems and e-Business Management* (9), pp. 109-131 (doi: 10.1007/s10257-010-0140-6).

Cerf, V. G. 2013. "Augmented Intelligence," *IEEE Internet Computing* (17:5), p. 96 (doi: 10.1109/MIC.2013.90).

Chandra, L., Seidel, S., and Gregor, S. 2015. "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," in *48th Hawaii International Conference on System Sciences (HICSS)*, Kauai, Hawaii, USA, pp. 4039-4048 (doi: 10.1109/HICSS.2015.485).

Charms, R. de. 1983. *Personal Causation: The Internal Affective Determinants of Behavior*, New York: Routledge.

Charters, E. 2003. "The Use of Think-aloud Methods in Qualitative Research: An Introduction to Think-aloud Methods," *Brock Education Journal* (12:2), pp. 68-82 (doi: 10.26522/brocked.v12i2.38).

Chen, M., Liu, F., and Lee, Y.-H. 2022. "My Tutor is an AI: The Effects of Involvement and Tutor Type on Perceived Quality, Perceived Credibility, and Use Intention," in *Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science, vol 13336*, H. Degen and S. Ntoa (eds.), Springer, Cham, pp. 232-244 (doi: 10.1007/978-3-031-05643-7_15).

Cheng, X., Yin, G., Azadegan, A., and Kolfschoten, G. L. 2016. "Trust Evolvement in Hybrid Team Collaboration: A Longitudinal Case Study," *Group Decision and Negotiation* (25), pp. 267-288 (doi: 10.1007/s10726-015-9442-x).

Cheong, K. J., Kim, J. J., and So, S. H. 2008. "A study of strategic call center management: Relationship between key performance indicators and customer satisfaction," *European Journal of Social Sciences* (6:2), pp. 268-276.

Cheung, C. F., Lee, W. B., Wang, W. M., Chu, K. F., and To, S. 2003. "A multi-perspective knowledge-based system for customer service management," *Expert Systems with Applications* (24:4), pp. 457-470 (doi: 10.1016/S0957-4174(02)00193-8).

Chevalier-Boisvert, M., Bahdanau, D., Lahlou, S., Willems, L., Saharia, C., Nguyen, T. H., and Bengio, Y. 2019. "BabyAI: A Platform to Study the Sample Efficiency of Grounded Language Learning," in *7th International Conference on Learning Representations (ICLR)*, New Orleans, Louisiana, USA (doi: 10.48550/arXiv.1810.08272).

Chhibber, N., and Law, E. 2019. "Using Conversational Agents To Support Learning By Teaching," in *Conference on Human Factors in Computing Systems (CHI '19)*, Glasgow, Scotland, UK (doi: 10.48550/arXiv.1909.13443).

Chromik, M., Fincke, F., and Butz, A. 2020. "Mind the (persuasion) gap: contrasting predictions of intelligent DSS with user beliefs to improve interpretability," in *Companion: Companion Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS)* (doi: 10.1145/3393672.3398491).

Colton, S., Goodwin, J., and Veale, T. 2012. "Full-FACE poetry generation," in *3rd International Conference on Computational Creativity (ICCC)*, Dublin, Ireland, pp. 95-102.

Cook, G. J. 1993. "An Empirical Investigation of Information Search Strategies with Implications for Decision Support System Design," *Decision Sciences* (24:3), pp. 683-693 (doi: 10.1111/j.1540-5915.1993.tb01298.x).

Curtis, D. D., and Lawson, M. J. 2001. "Exploring collaborative online learning," *Journal of Asynchronous Learning Networks* (5:1), pp. 21-34 (doi: 10.24059/OLJ.V5I1.1885).

Dahlbäck, N., and Jönsson, A. 1989. "Empirical studies of discourse representations for natural language interfaces," in *4th Conference on European chapter of the Association for Computational Linguistics (EACL)*, Manchester, England, pp. 291-298 (doi: 10.3115/976815.976854).

Darejeh, A., and Salim, S. S. 2016. "Gamification Solutions to Enhance Software User Engagement—A Systematic Review," *International Journal of Human–Computer Interaction* (32:8), pp. 613-642 (doi: 10.1080/10447318.2016.1183330).

Davenport, T., Guha, A., Grewal, D., and Bressgott, T. 2020. "How artificial intelligence will change the future of marketing," *Journal of the Academy of Marketing Science* (48), pp. 24-42 (doi: 10.1007/s11747-019-00696-0).

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319-340 (doi: 10.2307/249008).

Dean, D. L., Hender, J. M., Rodgers, T. L., and Santanen, E. L. 2006. “Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation,” *Journal of the Association for Information Systems* (7:10), pp. 646-699 (doi: 10.17705/1jais.00106).

Debs, M. B. 1991. “Recent Research on Collaborative Writing in Industry,” *Technical Communication* (38:4), pp. 476-484.

Deci, E. L., and Ryan, R. M. 2000. “The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior,” *Psychological Inquiry* (11:4), pp. 227-268 (doi: 10.1207/S15327965PLI1104_01).

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., and Ebel, P. 2019a. “The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems,” in *52nd Hawaii International Conference on System Sciences (HICSS)*, Grand Wailea, Maui, Hawaii, USA, pp. 274-283 (doi: 10.24251/HICSS.2019.034).

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. 2019b. “Hybrid Intelligence,” *Business & Information Systems Engineering* (61), pp. 637-643 (doi: 10.1007/s12599-019-00595-2).

Dellermann, D., Lipusch, N., Ebel, P., and Leimeister, J. M. 2019c. “Design principles for a hybrid intelligence decision support system for business model validation,” *Electronic Markets* (29), pp. 423-441 (doi: 10.1007/s12525-018-0309-2).

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., and Leimeister, J. M. 2017. “Finding the Unicorn: Predicting Early Stage Startup Success Through a Hybrid Intelligence Method,” in *38th International Conference on Information Systems (ICIS)*, Seoul, South Korea (doi: 10.2139/ssrn.3159123).

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. 2011a. “From game design elements to gamefulness: defining “gamification”,” in *15th International Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere, Finland, pp. 9-15 (doi: 10.1145/2181037.2181040).

Deterding, S., Khaled, R., Nacke, L. E., and Dixon, D. 2011b. “Gamification: Toward a definition,” in *Conference on Human Factors in Computing Systems (CHI ‘11)*, Vancouver, BC, Canada.

DeVellis, R. F. 2003. *Scale Development: Theory and Applications*, Thousand Oaks, CA: SAGE Publications, Inc.

Diederich, S., Brendel, A. B., Lichtenberg, S., and Kolbe, L. 2019. “Design For Fast Request Fulfillment Or Natural Interaction? Insights From An Experiment With A Conversational Agent,” in *27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Diederich, S., Brendel, A. B., Morana, S., and 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), pp. 96-138 (doi: 10.17705/1jais.00724).

Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K. E., and Dugan, C. 2019. "Explaining models: an empirical study of how explanations impact fairness judgment," in *24th International Conference on Intelligent User Interfaces (IUI)*, Marina del Rey, CA, USA, pp. 275-285 (doi: 10.1145/3301275.3302310).

Dubey, A., Abhinav, K., Jain, S., Arora, V., and Puttaveerana, A. 2020. "HACO: A Framework for Developing Human-AI Teaming," in *13th Innovations in Software Engineering Conference on Formerly known as India Software Engineering Conference (ISEC)*, Jabalpur, India (doi: 10.1145/3385032.3385044).

Dudley, J. J., and Kristensson, P. O. 2018. "A Review of User Interface Design for Interactive Machine Learning," *ACM Transactions on Interactive Intelligent Systems* (8:2) (doi: 10.1145/3185517).

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Le Meunier-FitzHugh, K., Le Meunier-FitzHugh, L., Misra, S., Mogaji, E., Sharma, S. K., Singh, J. B., Raghavan, V., Raman, R., Rana, N. P., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., and Williams, M. D. 2021. "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management* (57) (doi: 10.1016/j.ijinfomgt.2019.08.002).

Ebel, P., Söllner, M., Leimeister, J. M., Crowston, K., and Vreede, G.-J. de. 2021. "Hybrid intelligence in business networks," *Electronic Markets* (31), pp. 313-318 (doi: 10.1007/s12525-021-00481-4).

Edwards, W. 1962. "Dynamic Decision Theory and Probabilistic Information Processings," *Human Factors* (4:2), pp. 59-74 (doi: 10.1177/001872086200400201).

Eick, S. G. 1994. "Graphically Displaying Text," *Journal of Computational and Graphical Statistics* (3:2), pp. 127-142 (doi: 10.2307/1390665).

Elshan, E., and Ebel, P. 2020. "Let's Team Up: Designing Conversational Agents as Teammates," in *41st International Conference on Information Systems (ICIS)*, Hyderabad, India.

Elson, J. S., Derrick, D. C., and Ligon, G. S. 2018. "Examining Trust and Reliance in Collaborations between Humans and Automated Agents," in *51st Hawaii International Conference on System Sciences (HICSS)*, Waikoloa Village, Hawaii, USA, pp. 430-439 (doi: 10.24251/HICSS.2018.056).

Epstein, S. L. 2015. "Wanted: Collaborative intelligence," *Artificial Intelligence* (221), pp. 36-45 (doi: 10.1016/j.artint.2014.12.006).

Erbe, H.-H. 2001. "Multi-Agent Systems Assisting Operators to Mastering Automation Systems," *IFAC Proceedings Volumes* (34:16), pp. 357-361 (doi: 10.1016/S1474-6670(17)41550-3).

Feine, J., Gnewuch, U., Morana, S., and Maedche, A. 2019. "A Taxonomy of Social Cues for Conversational Agents," *International Journal of Human-Computer Studies* (132), pp. 138-161 (doi: 10.1016/j.ijhcs.2019.07.009).

Feine, J., Morana, S., and Maedche, A. 2020a. "A chatbot response generation system," in *Proceedings of the Conference on Mensch und Computer (MuC)*, Magdeburg, Germany, pp. 333-341 (doi: 10.1145/3404983.3405508).

Feine, J., Morana, S., and Maedche, A. 2020b. "Designing Interactive Chatbot Development Systems," in *41st International Conference on Information Systems (ICIS)*, Hyderabad, India.

Feng, B., and MacGeorge, E. L. 2010. "The Influences of Message and Source Factors on Advice Outcomes," *Communication Research* (37:4), pp. 553-575 (doi: 10.1177/0093650210368258).

Feng, S., and Buxmann, P. 2020. "My Virtual Colleague: A State-of-the-Art Analysis of Conversational Agents for the Workplace," in *53rd Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA, pp. 156-165 (doi: 10.24251/hicss.2020.020).

Fingerle, T., Schmitt, S., and Tautz, C. 2002. "Customer Requirements Elicitation in Intelligent Electronic Sales Support Systems," in *15th Bled Electronic Commerce Conference (BLED)*, Bled, Slovenia, pp. 244-262.

Følstad, A., and Brandtzaeg, P. B. 2017. "Chatbots and the new world of HCI," *Interactions* (24:4), pp. 38-42 (doi: 10.1145/3085558).

Følstad, A., and Skjuve, M. 2019. "Chatbots for customer service: user experience and motivation," in *1st International Conference on Conversational User Interfaces (CUI)*, Dublin, Ireland, pp. 1-9 (doi: 10.1145/3342775.3342784).

Følstad, A., and Taylor, C. 2020. "Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives," in *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science, vol 11970*, A. Følstad, T. Araujo, S. Papadopoulos, E. Lai-Chong Law, O.-C. Granmo, E. Luger and P. B. Brandtzaeg (eds.), Springer, Cham, pp. 201-214 (doi: 10.1007/978-3-030-39540-7_14).

Forman, J. 1991. "Collaborative Business Writing: A Burkean Perspective for Future Research," *The Journal of Business Communication* (28:3), pp. 233-257 (doi: 10.1177/002194369102800304).

Forster, J. 1978. *Teams und Teamarbeit in der Unternehmung: eine gesamtheitliche Darstellung mit Meinungen und Beispielen aus der betrieblichen Praxis*, Zurich: Haupt.

Fortin, D. R., and Dholakia, R. R. 2005. "Interactivity and vividness effects on social presence and involvement with a web-based advertisement," *Journal of Business Research* (58:3), pp. 387-396 (doi: 10.1016/S0148-2963(03)00106-1).

Fortune Media. 2019. "Global 500," available at https://fortune.com/global500/2019/search/, accessed on Nov 1 2021.

Frick, W. 2015. “When Your Boss Wears Metal Pants,” *Harvard Business Review*, pp. 84-89.

Friedrich, J., Becker, M., Kramer, F., Wirth, M., and Schneider, M. 2020. “Incentive design and gamification for knowledge management,” *Journal of Business Research* (106), pp. 341-352 (doi: 10.1016/j.jbusres.2019.02.009).

Froehle, C. M. 2006. “Service Personnel, Technology, and Their Interaction in Influencing Customer Satisfaction,” *Decision Sciences* (37:1), pp. 5-38 (doi: 10.1111/j.1540-5414.2006.00108.x).

Froehle, C. M., and Roth, A. V. 2004. “New measurement scales for evaluating perceptions of the technology-mediated customer service experience,” *Journal of Operations Management* (22:1), pp. 1-21 (doi: 10.1016/j.jom.2003.12.004).

Gao, Z., and Jiang, J. 2021. “Evaluating Human-AI Hybrid Conversational Systems with Chatbot Message Suggestions,” in *30th ACM International Conference on Information & Knowledge Management (CIKM)*, Virtual Event, pp. 534-544 (doi: 10.1145/3459637.3482340).

Gartner. 2019. “How to Manage Customer Service Technology Innovation,” available at https://www.gartner.com/smarterwithgartner/27297-2/, accessed on Oct 3 2021.

Gartner. 2020. “Gartner Magic Quadrant for the CRM Customer Engagement Center,” available at https://www.gartner.com/en/documents/3985947/magic-quadrant-for-the-crm-customer-engagement-center, accessed on Nov 16 2021.

Gaurav, T. 2016. “On Interactive Machine Learning,” available at https://www.trivedigaurav.com/blog/on-interactive-machine-learning/, accessed on Oct 21 2020.

Gavriushenko, M., Kaikova, O., and Terziyan, V. 2020. “Bridging human and machine learning for the needs of collective intelligence development,” *Procedia Manufacturing* (42), pp. 302-306 (doi: 10.1016/j.promfg.2020.02.092).

Gelbrich, K., Hagel, J., and Orsingher, C. 2021. “Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence,” *International Journal of Research in Marketing* (38:1), pp. 176-193 (doi: 10.1016/j.ijresmar.2020.06.004).

Ghahramani, Z. 2015. “Probabilistic machine learning and artificial intelligence,” *Nature* (521), pp. 452-459 (doi: 10.1038/nature14541).

Ghosh, B., Wilson, H. J., Burden, A., and Daugherty, P. R. 2019. “Taking a Systems Approach to Adopting AI,” available at https://hbr.org/2019/05/taking-a-systems-approach-to-adopting-ai, accessed on Aug 8 2022.

Ghuman, R., and Kumari, R. 2013. “Narrative Science: A Review,” *International Journal of Science and Research* (2:9), pp. 205-207.

Giraffa, L. M. M., and Viccari, R. M. 1998. “The use of agents techniques on intelligent tutoring systems,” in *18th International Conference of the Chilean Society of Computer Science*, Antofogasta, Chile, pp. 76-83 (doi: 10.1109/SCCC.1998.730785).

Glikson, E., and Woolley, A. W. 2020. "Human Trust in Artificial Intelligence: Review of Empirical Research," *Academy of Management Annals* (14:2) (doi: 10.5465/annals.2018.0057).

Glushko, R. J., and Nomorosa, K. J. 2013. "Substituting Information for Interaction: A Framework for Personalization in Service Encounters and Service Systems," *Journal of Service Research* (16:1), pp. 21-38 (doi: 10.1177/1094670512463967).

Glushko, R. J., and Tabas, L. 2008. "Bridging the "Front Stage" and "Back Stage" in Service System Design," in *41st Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, Big Island, Hawaii, USA, p. 106 (doi: 10.1109/HICSS.2008.77).

Gnewuch, U., Adam, M. T. P., Morana, S., and Maedche, A. 2018. ""The Chatbot is typing …" - The Role of Typing Indicators in Human-Chatbot Interaction," in *17th Annual Pre-ICIS Workshop on HCI Research in MIS*, San Francisco, California, USA.

Gnewuch, U., Morana, S., and Maedche, A. 2017. "Towards Designing Cooperative and Social Conversational Agents for Customer Service," in *38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.

Gnewuch, U., Yu, M., and Maedche, A. 2020. "The Effect of Perceived Similarity in Dominance on Customer Self-Disclosure to Chatbots in Conversational Commerce," in *28th European Conference on Information Systems (ECIS)*, Virtual Event.

Goel, L., Johnson, N., Junglas, I., and Ives, B. 2013. "Predicting users' return to virtual worlds: a social perspective," *Information Systems Journal* (23:1), pp. 35-63 (doi: 10.1111/j.1365-2575.2011.00396.x).

Göker, M. H., and Roth-Berghofer, T. 1999. "The development and utilization of the case-based help-desk support system HOMER," *Engineering Applications of Artificial Intelligence* (12:6), pp. 665-680 (doi: 10.1016/S0952-1976(99)00037-8).

Gonzalez, C. 2005. "Decision support for real-time, dynamic decision-making tasks," *Organizational Behavior and Human Decision Processes* (96:2), pp. 142-154 (doi: 10.1016/j.obhdp.2004.11.002).

Graef, R., Klier, M., Kluge, K., and Zolitschka, J. F. 2021. "Human-machine collaboration in online customer service – a long-term feedback-based approach," *Electronic Markets* (31), pp. 319-341 (doi: 10.1007/s12525-020-00420-9).

Gregor, S. 2006. "The Nature of Theory in Information Systems," *MIS Quarterly* (30:3), pp. 611-642 (doi: 10.2307/25148742).

Gregor, S., Chandra, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the Association for Information Systems* (21:6), pp. 1622-1652 (doi: 10.17705/1jais.00649).

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355 (doi: 10.25300/MISQ/2013/37.2.01).

Grudin, J., and Jacques, R. 2019. "Chatbots, Humbots, and the Quest for Artificial General Intelligence," in *Conference on Human Factors in Computing Systems (CHI '19)*, Glasgow, Scotland, UK (doi: 10.1145/3290605.3300439).

Gulati, S., Sousa, S., and Lamas, D. 2018. "Modelling trust in human-like technologies," in *9th Indian Conference on Human Computer Interaction*, Bangalore, India (doi: 10.1145/3297121.3297124).

Guntzviller, L. M., Liao, D., Pulido, M. D., Butkowski, C. P., and Campbell, A. D. 2020. "Extending advice response theory to the advisor: Similarities, differences, and partner-effects in advisor and recipient advice evaluations," *Communication Monographs* (87:1), pp. 114-135 (doi: 10.1080/03637751.2019.1643060).

Gupta, A. 2021. "4 Key Tech Trends in Customer Service to Watch," available at https://www.gartner.com/smarterwithgartner/4-key-tech-trends-in-customer-service-to-watch, accessed on Aug 8 2022.

Hamari, J., Koivisto, J., and Sarsa, H. 2014. "Does Gamification Work? - A Literature Review of Empirical Studies on Gamification," in *47th Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, Hawaii, USA, pp. 3025-3034 (doi: 10.1109/HICSS.2014.377).

Hanika, T., Herde, M., Kuhn, J., Leimeister, J. M., Lukowicz, P., Oeste-Reiß, S., Schmidt, A., Sick, B., Stumme, G., Tomforde, S., and Zweig, K. A. 2019. "Collaborative Interactive Learning - A clarification of terms and a differentiation from other research fields," available at https://arxiv.org/abs/1905.07264.

Harms, J.-G., Kucherbaev, P., Bozzon, A., and Houben, G.-J. 2019. "Approaches for Dialog Management in Conversational Agents," *IEEE Internet Computing* (23:2), pp. 13-22 (doi: 10.1109/MIC.2018.2881519).

Hassanein, K., and Head, M. 2005. "The Impact of Infusing Social Presence in the Web Interface: An Investigation Across Product Types," *International Journal of Electronic Commerce* (10:2), pp. 31-55 (doi: 10.2753/JEC1086-4415100202).

Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., Maron, R. C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-Letz, T., Izar, B., Kalle, C. von, Fröhling, S., and Brinker, T. J. 2019. "Superior skin cancer classification by the combination of human and artificial intelligence," *European Journal of Cancer* (120), pp. 114-121 (doi: 10.1016/j.ejca.2019.07.019).

Hevner, A. R. 2007. "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems* (19:2).

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105 (doi: 10.2307/25148625).

Hill, J., Ford, W. R., and Farreras, I. G. 2015. "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations," *Computers in Human Behavior* (49), pp. 245-250 (doi: 10.1016/j.chb.2015.02.026).

Hjorth, A. 2021. "NaturalLanguageProcesing4All: - A Constructionist NLP tool for Scaffolding Students' Exploration of Text," in *17th ACM Conference on International Computing Education Research (ICER)*, Virtual Event, pp. 347-354 (doi: 10.1145/3446871.3469749).

Hohenstein, J., and Jung, M. 2018. “AI-Supported Messaging: An Investigation of Human-Human Text Conversation with AI Support,” in *Conference on Human Factors in Computing Systems (CHI ‘18)*, Montreal, QC, Canada (doi: 10.1145/3170427.3188487).

Holtgraves, T., and Han, T.-L. 2007. “A procedure for studying online conversational processing using a chat bot,” *Behavior Research Methods* (39:1), pp. 156-163 (doi: 10.3758/BF03192855).

Holzinger, A. 2016a. “Interactive Machine Learning (iML),” *Informatik Spektrum* (39), pp. 64-68 (doi: 10.1007/s00287-015-0941-6).

Holzinger, A. 2016b. “Interactive machine learning for health informatics: when do we need the human-in-the-loop?” *Brain Informatics* (3), pp. 119-131 (doi: 10.1007/s40708-016-0042-6).

Holzinger, A., Malle, B., Saranti, A., and Pfeifer, B. 2021. “Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI,” *Information Fusion* (71), pp. 28-37 (doi: 10.1016/j.inffus.2021.01.008).

Holzinger, A., Plass, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. 2016. “Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to Solve the Traveling Salesman Problem with the Human-in-the-Loop Approach,” in *Availability, Reliability, and Security in Information Systems. CD-ARES 2016. Lecture Notes in Computer Science, vol 9817*, F. Buccafurri, A. Holzinger, P. Kieseberg, A. Tjoa and E. Weippl (eds.), Springer, Cham, pp. 81-95 (doi: 10.1007/978-3-319-45507-5_6).

Holzinger, A., Plass, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. 2017. “A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop,” available at https://arxiv.org/abs/1708.01104.

Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. 2019. “Interactive machine learning: experimental evidence for the human in the algorithmic loop,” *Applied Intelligence* (49), pp. 2401-2414 (doi: 10.1007/s10489-018-1361-5).

Hong, W., Thong, J. Y. L., and Tam, K. Y. 2004. “The Effects of Information Format and Shopping Task on Consumers’ Online Shopping Behavior: A Cognitive Fit Perspective,” *Journal of Management Information Systems* (21:3), pp. 149-184 (doi: 10.1080/07421222.2004.11045812).

Hu, W.-L., Rivetta, C., MacDonald, E., and Chassin, D. P. 2019. “Optimal Operator Training Reference Models for Human-in-the-loop Systems,” in *52nd Hawaii International Conference on System Sciences (HICSS)*, Grand Wailea, Maui, Hawaii, USA, pp. 3524-3531.

Huang, M.-H., and Rust, R. T. 2018. “Artificial Intelligence in Service,” *Journal of Service Research* (21:2), pp. 155-172 (doi: 10.1177/1094670517752459).

Hudec, M., Mináriková, E., Mesiar, R., Saranti, A., and Holzinger, A. 2021. “Classification by ordinal sums of conjunctive and disjunctive functions for explainable AI and interpretable machine learning solutions," *Knowledge-Based Systems* (220) (doi: 10.1016/j.knosys.2021.106916).

Hun Lee, M., Siewiorek, D. P., Smailagic, A., Bernardino, A., and Bermúdez i Badia, S. 2021. "A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment," in *Conference on Human Factors in Computing Systems (CHI '21),* Virtual Event (doi: 10.1145/3411764.3445472).

Hung, S.-Y., and Liang, T.-P. 2001. "Effect of computer self-efficacy on the use of executive support systems," *Industrial Management & Data Systems* (101:5), pp. 227-237 (doi: 10.1108/02635570110394626).

Hunicke, R., LeBlanc, M., and Zubek, R. 2004. "MDA: A Formal Approach to Game Design and Game Research," in *Challenges in Games AI Workshop, 19th National Conference of Artificial Intelligence*, San Jose, California, USA.

Ishii, K., and Markman, K. M. 2016. "Online customer service and emotional labor: An exploratory study," *Computers in Human Behavior* (62), pp. 658-665 (doi: 10.1016/j.chb.2016.04.037).

Jacobs, I., Hong, D., Dawson, S., Sjoblom, S., and Lynch, D. 2019. "The Forrester New Wave: Conversational AI for Customer Service," available at https://www.forrester.com/report/The+Forrester+New+Wave+Conversational+AI+For+Customer+Service+Q2+2019/-/E-RES144416, accessed on Nov 8 2021.

Jain, H., Padmanabhan, B., Pavlou, P. A., and Raghu, T. S. 2021. "Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society," *Information Systems Research* (32:3), pp. 675-687 (doi: 10.1287/isre.2021.1046).

Janssen, A., Passlick, J., Rodríguez Cardona, D., and Breitner, M. H. 2020. "Virtual Assistance in Any Context - A Taxonomy of Design Elements for Domain-Specific Chatbots," *Business & Information Systems Engineering* (62:3), pp. 211-225 (doi: 10.1007/s12599-020-00644-1).

Jenkins, M.-C., Churchill, R., Cox, S., and Smith, D. 2007. "Analysis of User Interaction with Service Oriented Chatbot Systems," in *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments. HCI 2007. Lecture Notes in Computer Science, vol 4552*, J. A. Jacko (ed.), Springer, pp. 76-83 (doi: 10.1007/978-3-540-73110-8_9).

Johns, R. 2010. "Likert items and scales," available at https://ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf, accessed on Aug 8 2022.

Johnson, D. W., and Johnson R. T. 2004. "Cooperation and the Use of Technology," in *Handbook of Research on Educational Communications and Technology*, D. H. Jonassen (ed.), Routledge, pp. 785-811.

Johnson, E. J., and Payne, J. W. 1985. "Effort and Accuracy in Choice," *Management Science* (31:4), pp. 395-414 (doi: 10.1287/mnsc.31.4.395).

Jones, D., and Gregor, S. 2007. "The Anatomy of a Design Theory," *Journal of the Association for Information Systems* (8:5), pp. 312-335 (doi: 10.17705/1jais.00129).

Joy-Matthews, J., Megginson, D., and Surtees, M. 2004. *Human Resource Development*, Londong and Sterling: Kogan Page.

Kahai, S. S., and Cooper, R. B. 2003. "Exploring the Core Concepts of Media Richness Theory: The Impact of Cue Multiplicity and Feedback Immediacy on Decision Quality," *Journal of Management Information Systems* (20:1), pp. 263-299 (doi: 10.1080/07421222.2003.11045754).

Kankanhalli, A., Lee, O.-K., and Lim, K. H. 2011. "Knowledge reuse through electronic repositories: A study in the context of customer service support," *Information & Management* (48:2-3), pp. 106-113 (doi: 10.1016/j.im.2011.02.002).

Kaplan, A., and 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), pp. 15-25 (doi: 10.1016/j.bushor.2018.08.004).

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W. 2020. "Dense Passage Retrieval for Open-Domain Question Answering," available at https://arxiv.org/abs/2004.04906.

Kartoun, U. 2017. "Text nailing: an efficient human-in-the-loop text-processing method," *Interactions* (24:6), pp. 44-49 (doi: 10.1145/3139488).

Keyser, A. de, Köcher, S., Alkire, L., Verbeeck, C., and Kandampully, J. 2019. "Frontline Service Technology infusion: conceptual archetypes and future research directions," *Journal of Service Management* (30:1), pp. 156-183 (doi: 10.1108/JOSM-03-2018-0082).

Khakpour, A., and Colomo-Palacios, R. 2021. "Convergence of Gamification and Machine Learning: A Systematic Literature Review," *Technology, Knowledge and Learning* (26), pp. 597-636 (doi: 10.1007/s10758-020-09456-4).

Kiefer, S. 2022. "CaSE: Explaining Text Classifications by Fusion of Local Surrogate Explanation Models with Contextual and Semantic Knowledge," *Information Fusion* (77), pp. 184-195 (doi: 10.1016/j.inffus.2021.07.014).

Kirchkamp, O., and Strobel, C. 2019. "Sharing responsibility with a machine," *Journal of Behavioral and Experimental Economics* (80), pp. 25-33 (doi: 10.1016/j.socec.2019.02.010).

Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. 2017. "The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms," in *Conference on Designing Interactive Systems (DIS '17)*, Edinburgh, United Kingdom, pp. 555-565 (doi: 10.1145/3064663.3064672).

Koch, M., and Gross, T. 2006. "Computer-Supported Cooperative Work - Concepts and Trends," in *AIM 2006 – Information Systems and Collaboration: State of the Art and Perspectives (AIM)*, F. Feltz, B. Otjacques, A. Oberweis and N. Poussing (eds.), Bonn: Koellen, pp. 165-172.

Kolfschoten, G. L., and Vreede, G.-J. de. 2009. "A Design Approach for Collaboration Processes: A Multimethod Design Science Study in Collaboration Engineering," *Journal of Management Information Systems* (26:1), pp. 225-256 (doi: 10.2753/MIS0742-1222260109).

Koppenhagen, N., Gaß, O., and Müller, B. 2012. "Design Science Research in Action - Anatomy of Success Critical Activities for Rigor and Relevance," in *20th European Conference on Information Systems (ECIS)*, Barcelona, Spain (doi: 10.5445/IR/1000055012).

Kozlowski, S. W. J., and Bell, B. S. 2003. "Work groups and teams in organizations," in *Handbook of Psychology (Vol. 12)*, W. C. Borman, D. R. Ilgen and R. J. Klimoski (eds.), New York: Wiley-Blackwell, pp. 333-375 (doi: 10.1002/0471264385.wei1214).

Krannich, D. 2010. *Mobile Usability Testing. Ein toolbasiertes Vorgehensmodell zum Rapid-Prototyping und Usability-Testing von Mobilen Systemen im originären Benutzungskontext*, Dissertation: Uni-Bremen.

Krening, S., and Feigh, K. M. 2018. "Interaction Algorithm Effect on Human Experience with Reinforcement Learning," *ACM Transactions on Human-Robot Interaction* (7:2) (doi: 10.1145/3277904).

Krogh, G. von. 2018. "Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing," *Academy of Management Discoveries* (4:4), pp. 404-409 (doi: 10.5465/amd.2018.0084).

Kuechler, W., and Vaishnavi, V. 2012. "A Framework for Theory Development in Design Science Research: Multiple Perspectives," *Journal of the Association for Information Systems* (13:6), pp. 395-423 (doi: 10.17705/1jais.00300).

Kühl, N., Goutier, M., Hirt, R., and Satzger, G. 2019. "Machine Learning in Artificial Intelligence: Towards a Common Understanding," in *52nd Hawaii International Conference on System Sciences (HICSS)*, Grand Wailea, Maui, Hawaii, USA, pp. 5236-5245.

Kulesza, T., Burnett, M., Wong, W.-K., and Stumpf, S. 2015. "Principles of Explanatory Debugging to Personalize Interactive Machine Learning," in *20th International Conference on Intelligent User Interfaces (IUI)*, Atlanta, Georgia, USA, pp. 126-137 (doi: 10.1145/2678025.2701399).

Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., and Szopinski, D. 2021. "An Update for Taxonomy Designers. Methodological Guidance from Information Systems Research," *Business & Information Systems Engineering* (doi: 10.1007/s12599-021-00723-x).

Langan-Fox, J., Wilson, J. R., and Anglim, J. 2004. "Mental models, team mental models, and performance: Process, development, and future directions," *Human Factors and Ergonomics in Manufacturing & Service Industries* (14:4), pp. 331-352 (doi: 10.1002/hfm.20004).

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wünderlich, N. V., and Keyser, A. de. 2017. ""Service Encounter 2.0": An investigation into the roles of technology, employees and customers," *Journal of Business Research* (79), pp. 238-246 (doi: 10.1016/j.jbusres.2017.03.008).

Lee, T. Y., Smith, A., Seppi, K., Elmqvist, N., Boyd-Graber, J., and Findlater, L. 2017. "The human touch: How non-expert users perceive, interpret, and fix topic models," *International Journal of Human-Computer Studies* (105), pp. 28-42 (doi: 10.1016/j.ijhcs.2017.03.007).

Lees, M., Evans, R., and Mareels, I. 2011. "Autonomous Performance Monitoring of an On-Line Real-Time Decision Support System," *IFAC Proceedings Volumes* (44:1), pp. 13080-13085 (doi: 10.3182/20110828-6-IT-1002.00421).

Leimeister, J. M. 2014. *Collaboration Engineering: IT-gestützte Zusammenarbeitsprozesse systematisch entwickeln und durchführen*, Berlin, Heidelberg: Springer Gabler.

Lerch, F. J., and Harter, D. E. 2001. "Cognitive Support for Real-Time Dynamic Decision Making," *Information Systems Research* (12:1), pp. 63-82 (doi: 10.1287/isre.12.1.63.9717).

Levy, F. 2018. "Computers and populism: artificial intelligence, jobs, and politics in the near term," *Oxford Review of Economic Policy* (34:3), pp. 393-417 (doi: 10.1093/oxrep/gry004).

Libai, B., Bart, Y., Gensler, S., Hofacker, C. F., Kaplan, A., Kötterheinrich, K., and Kroll, E. B. 2020. "Brave New World? On AI and the Management of Customer Relationships," *Journal of Interactive Marketing* (51:1), pp. 44-56 (doi: 10.1016/j.intmar.2020.04.002).

Likert, R. 1932. "A technique for the measurement of attitudes," *Archives of Psychology* (22:140).

Limerick, H., Coyle, D., and Moore, J. W. 2014. "The experience of agency in human-computer interactions: a review," *Frontiers in Human Neuroscience* (8) (doi: 10.3389/fnhum.2014.00643).

Lindvall, M., Lundström, C., and Löwgren, J. 2021. "Rapid Assisted Visual Search: Supporting Digital Pathologists with Imperfect AI," in *26th International Conference on Intelligent User Interfaces (IUI)*, College Station, Texas, USA, pp. 504-513 (doi: 10.1145/3397481.3450681).

Liu, D., Santhanam, R., and Webster, J. 2017. "Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems," *MIS Quarterly* (41:4), pp. 1011-1034 (doi: 10.25300/MISQ/2017/41.4.01).

Liu, J., Wilson, A., and Gunning, D. 2014. "Workflow-based Human-in-the-Loop Data Analytics," in *Workshop on Human Centered Big Data Research (HCBDR ,14)*, Raleigh, North Carolina, USA, pp. 49-52 (doi: 10.1145/2609876.2609888).

Liu, M., Ding, Q., Zhang, Y., Zhao, G., Hu, C., Gong, J., Xu, P., Zhang, L., and Wang, Q. 2020. "Cold Comfort Matters - How Channel-Wise Emotional Strategies Help in a Customer Service Chatbot," in *Conference on Human Factors in Computing Systems (CHI ,20)*, Honolulu, Hawaii, USA (doi: 10.1145/3334480.3382905).

Lloyd, E., Huang, S., and Tognoli, E. 2017. "Improving Human-in-the-Loop Adaptive Systems Using Brain-Computer Interaction," in *IEEE/ACM 12th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS),* Buenos Aires, Argentina, pp. 163-174 (doi: 10.1109/SEAMS.2017.1).

Löffler, C., Hoffmann, J., and Bodendorf, F. 2010. *Forschungsmethoden in der Wirtschaftsinformatik,* Nürnberg: Lehrstuhl Wirtschaftsinformatik II.

Lösser, B., Oberländer, A. M., and Rau, D. 2019. "Taxonomy Research in Information Systems: A Systematic Assessment," in *27th European Conference on Information Systems (ECIS),* Stockholm & Uppsala, Sweden.

Lounis, S., Doukidis, S., Papastamatiou, N., and Doukoulos, T. 2016. "Gamification of Authoring Interactive E-Books for Children: The Q-Tales Ecosystem," in *10th Mediterranean Conference on Information Systems (MCIS),* Paphos, Cyprus.

Lowry, P. B., Gaskin, J. E., and Moody, G. D. 2015. "Proposing the Multi-Motive Information Systems Continuance Model (MISC) to Better Explain End-User System Evaluations and Continuance Intentions," *Journal of the Association for Information Systems* (16:7), pp. 515-579 (doi: 10.17705/1jais.00403).

Luger, E., and Sellen, A. 2016. ""Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents," in *Conference on Human Factors in Computing Systems (CHI ,16),* San Jose, California, USA, pp. 5286-5297 (doi: 10.1145/2858036.2858288).

Luong, A., Kumar, N., and Lang, K. 2019. "Man-Machine Collaboration in Organizational Decision-Making: An Experimental Study Using Loan Application Evaluations," in *Pre-ICIS SIGDSA Symposium on Inspiring mindset for Innovation with Business Analytics and Data Science,* Munich, Germany.

MacGeorge, E. L., Feng, B., and Guntzviller, L. M. 2016. "Advice: Expanding the Communication Paradigm," *Annals of the International Communication Association* (40:1), pp. 213-243 (doi: 10.1080/23808985.2015.11735261).

MacGeorge, E. L., and van Swol, L. M. 2018. "Advice across disciplines and contexts," in *The Oxford Handbook of Advice,* E. L. MacGeorge and L. M. van Swol (eds.), New York: Oxford University Press, pp. 3-17.

Macnish, K., and Fernandez Inguanzo, A. 2019. "Case Study - Customer Relation Management, Smart Information Systems and Ethics," *The ORBIT Journal* (2:2), pp. 1-24 (doi: 10.29297/orbit.v2i2.114).

Maedche, A., Gregor, S., Morana, S., and Feine, J. 2019. "Conceptualization of the Problem Space in Design Science Research," in *14th International Conference on Design Science Research in Information Systems and Technology (DESRIST). Lecture Notes in Computer Science,* Worcester, Massachusetts, USA: Springer, Cham, pp. 18-31 (doi: 10.1007/978-3-030-19504-5_2).

Maglio, P. P., and Spohrer, J. 2008. "Fundamentals of service science," *Journal of the Academy of Marketing Science* (36), pp. 18-20 (doi: 10.1007/s11747-007-0058-9).

Mallios, S., and Bourbakis, N. 2016. “A survey on human machine dialogue systems,” in *7th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Chalkidiki, Greece (doi: 10.1109/IISA.2016.7785371).

Manjavacas, E., Karsdorp, F., Burtenshaw, B., and Kestemont, M. 2017. “Synthetic Literature: Writing Science Fiction in a Co-Creative Process,” in *Workshop on Computational Creativity in Natural Language Generation (CC-NLG)*, Santiago de Compostela, Spain, pp. 29-37 (doi: 10.18653/v1/W17-3904).

March, S. T., and Smith, G. F. 1995. “Design and natural science research on information technology,” *Decision Support Systems* (15:4), pp. 251-266 (doi: 10.1016/0167-9236(94)00041-2).

Marczak, S., Figueira Filho, F., Singer, L., Treude, C., Steffens, F., and Redmiles Ban Al-Ani, D. 2015. “Studying Gamification as a Collaboration Motivator for Virtual Software Teams: Social Issues, Cultural Issues, and Research Methods,” in *18th ACM Conference Companion on Computer-Supported Cooperative Work and Social Computing (CSCW)*, Vancouver, Canada.

Martínez, M. A. M., Nadj, M., and Maedche, A. 2019. “Towards An Integrative Theoretical Framework Of Interactive Machine Learning Systems,” in *27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Mayring, P. 2001. “Combination and Integration of Qualitative and Quantitative Analysis,” *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* (2:1) (doi: 10.17169/fqs-2.1.967).

Mayring, P. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*, Klagenfurt, Austria.

McCamish, B., Termehchy, A., and Touri, B. 2017. “A Game-theoretic Approach to Data Interaction: A Progress Report,” in *2nd Workshop on Human-In-the-Loop Data Analytics*, Chicago, Illinois, USA (doi: 10.1145/3077257.3077270).

McLean, G., and Osei-Frimpong, K. 2017. “Examining satisfaction with the experience during a live chat service encounter-implications for website providers,” *Computers in Human Behavior* (76), pp. 494-508 (doi: 10.1016/j.chb.2017.08.005).

McLean, G., and Wilson, A. 2016. “Evolving the online customer experience … is there a role for online customer support?” *Computers in Human Behavior* (60), pp. 602-610 (doi: 10.1016/j.chb.2016.02.084).

McTear, M., Callejas, Z., and Griol, D. 2016. *The Conversational Interface: Talking to Smart Devices*, Springer, Cham.

Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., and O‘Neill, J. 2017. “How Do You Want Your Chatbot? An Explanatory Wizard-of-Oz Study with Young, Urban Indians,” in *Human-Computer Interaction - INTERACT 2017. INTERACT 2017. Lecture Notes in Computer Science, vol 10513*, R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O‘Neill and M. Winckler (eds.), Springer, Cham, pp. 441-459 (doi: 10.1007/978-3-319-67744-6_28).

Mele, C., Russo Spena, T., Kaartemo, V., and Marzullo, M. L. 2021. “Smart nudging: How cognitive technologies enable choice architectures for value co-creation,” *Journal of Business Research* (129), pp. 949-960 (doi: 10.1016/j.jbusres.2020.09.004).

Mero, J. 2018. “The effects of two-way communication and chat service usage on consumer attitudes in the e-commerce retailing sector,” *Electronic Markets* (28), pp. 205-217 (doi: 10.1007/s12525-017-0281-2).

Meske, C., Brockmann, T., Wilms, K., and Stieglitz, S. 2015. “Gamify Employee Collaboration - A Critical Review of Gamification Elements in Social Software,” in *Australasian Conference on Information Systems*, Adelaide, Australia (doi: 10.48550/arXiv.1606.01351).

Meth, H., Mueller, B., and Maedche, A. 2015. “Designing a Requirement Mining System,” *Journal of the Association for Information Systems* (16:9), pp. 799-837 (doi: 10.17705/1jais.00408).

Meuser, M., and Nagel, U. 1991. “ExpertInneninterviews - vielfach erprobt, wenig bedacht: ein Beitrag zur qualitativen Methodendiskussion,” in *Qualitativ-empirische Sozialforschung : Konzepte, Methoden, Analysen*, D. Garz and K. Kraimer (eds.), Opladen: Westdeutscher Verlag, pp. 441-471.

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., and Bitner, M. J. 2000. “Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters,” *Journal of Marketing* (64:3), pp. 50-64 (doi: 10.1509/jmkg.64.3.50.18024).

Meyer von Wolff, R., Hobert, S., and Schumann, M. 2019a. “How May I Help You? - State of the Art and Open Research Questions for Chatbots at the Digital Workplace,” in *52nd Hawaii International Conference on System Sciences (HICSS)*, Grand Wailea, Maui, Hawaii, USA, pp. 95-104 (doi: 10.24251/HICSS.2019.013).

Meyer von Wolff, R., Masuch, K., Hobert, S., and Schumann, M. 2019b. “What Do You Need Today? - An Empirical Systematization of Application Areas for Chatbots at Digital Workplaces,” in *25th Americas Conference on Information Systems (AMCIS)*, Cancún, Mexico.

Mikołajczyk, K. 2022. “Changes in the approach to employee development in organisations as a result of the COVID-19 pandemic,” *European Journal of Training and Development* (45:5/6), pp. 544-562 (doi: 10.1108/EJTD-12-2020-0171).

Mingers, J. 2001. “Combining IS Research Methods: Towards a Pluralist Methodology,” *Information Systems Research* (12:3), pp. 240-259 (doi: 10.1287/isre.12.3.240.9709).

Molino, P., Zheng, H., and Wang, Y.-C. 2018. “COTA: Improving the Speed and Accuracy of Customer Support through Ranking and Deep Networks,” in *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, United Kingdom, pp. 586-595 (doi: 10.1145/3219819.3219851).

Möller, F., Guggenberger, T. M., and Otto, B. 2020. “Towards a Method for Design Principle Development in Information Systems,” in *Designing for Digital Transformation. Co-Creating Services with Citizens and Industry. DESRIST 2020.*

Lecture Notes in Computer Science, vol 12388, S. Hofmann, O. Müller and M. Rossi (eds.), Springer, Cham, pp. 208-220 (doi: 10.1007/978-3-030-64823-7_20).

Monteith, K., Martinez, T., and Ventura, D. 2012. "Automatic Generation of Melodic Accompaniments for Lyrics," in *3rd International Conference on Computational Creativity (ICCC)*, Dublin, Ireland, pp. 87-94.

Mori, M., MacDorman, K. F., and Kageki, N. 2012. "The Uncanny Valley," *IEEE Robotics & Automation Magazine* (19:2), pp. 98-100 (doi: 10.1109/MRA.2012.2192811).

Morris, A., and Ulieru, M. 2011. "FRIEND: A human-aware BDI agent architecture," in *IEEE International Conference on Systems, Man, and Cybernetics*, Anchorage, Alaska, USA, pp. 2413-2418 (doi: 10.1109/ICSMC.2011.6084039).

Mullins, R., Fouse, A., Ganberg, G., and Schurr, N. 2020. "Practice Makes Perfect: Lesson Learned from Five Years of Trial and Error Building Context-Aware Systems," in *53rd Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA, pp. 279-288 (doi: 10.24251/hicss.2020.035).

Münscher, R., Vetter, M., and Scheuerle, T. 2015. "A Review and Taxonomy of Choice Architecture Techniques," *Journal of Behavioral Decision Making* (29:5), pp. 511-524 (doi: 10.1002/bdm.1897).

Myers, M. D., and Newman, M. 2007. "The qualitative interview in IS research: Examining the craft," *Information and Organization* (17:1), pp. 2-26 (doi: 10.1016/j.infoandorg.2006.11.001).

Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81-103 (doi: 10.1111/0022-4537.00153).

Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers are social actors," in *Conference on Human Factors in Computing Systems (CHI ‚94)*, Boston, Massachusetts, USA, pp. 72-78 (doi: 10.1145/191666.191703).

Nguyen, T. H., Waizenegger, L., and Techatassanasoontorn, A. A. 2021. "„Don‘t Neglect the User!“ - Identifying Types of Human-Chatbot Interactions and their Associated Characteristics," *Information Systems Frontiers* (doi: 10.1007/s10796-021-10212-x).

Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A method for taxonomy development and its application in information systems," *European Journal of Information Systems* (22:3), pp. 336-359 (doi: 10.1057/ejis.2012.26).

Norman, D. 2017. "Design, Business Models, and Human-Technology Teamwork," *Research-Technology Management* (60:1), pp. 26-30 (doi: 10.1080/08956308.2017.1255051).

Oliveira, C. S. de, Sanin, C., and Szczerbicki, E. 2020. "Human Feedback and Knowledge Discovery: Towards Cognitive Systems Optimization," *Procedia Computer Science* (176), pp. 3093-3102 (doi: 10.1016/j.procs.2020.09.179).

Oliver, R. L. 1980. "A Cognitive Model for the Antecedents and Consequences of Satisfation Decisions," *Journal of Marketing Research* (17:4), pp. 460-469 (doi: 10.2307/3150499).

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., and Sinz, E. J. 2011. "Memorandum on design-oriented information systems research," *European Journal of Information Systems* (20:1), pp. 7-10 (doi: 10.1057/ejis.2010.55).

Østerlund, C., Jarrahi, M. H., Willis, M., Boyd, K., and Wolf, C. T. 2021. "Artificial intelligence and the world of work, a co-constitutive relationship," *Journal of the Association for Information Science and Technology* (72:1), pp. 128-135 (doi: 10.1002/asi.24388).

Ostheimer, J., Chowdhury, S., and Iqbal, S. 2021. "An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles," *Technology in Society* (66) (doi: 10.1016/j.techsoc.2021.101647).

Ostrom, A. L., Fotheringham, D., and Bitner, M. J. 2019. "Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences," in *Handbook of Service Science, Volume II*, P. P. Maglio, C. A. Kieliszewski, J. Spohrer, K. Lyons, L. Patrício and Y. Sawatani (eds.), Springer, Cham, pp. 77-103 (doi: 10.1007/978-3-319-98512-1_5).

Ozkara, B. Y., Ozmen, M., and Kim, J. W. 2016. "Exploring the relationship between information satisfaction and flow in the context of consumers' online search," *Computers in Human Behavior* (63), pp. 844-859 (doi: 10.1016/j.chb.2016.06.038).

Paas, F., Renkl, A., and Sweller, J. 2004. "Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture," *Instructional Science* (32:1/2), pp. 1-8 (doi: 10.1023/B:TRUC.0000021806.17516.d0).

Paluch, S., and Wirtz, J. 2020. "Artificial Intelligence and Robots in the Service Encounter," *Journal of Service Management Research* (4:1), pp. 3-8 (doi: 10.15358/2511-8676-2020-1-3).

Paschen, J., Wilson, M., and Ferreira, J. J. 2020. "Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel," *Business Horizons* (63:3), pp. 403-414 (doi: 10.1016/j.bushor.2020.01.003).

Pavlou, P. A., Liang, H., and Xue, Y. 2007. "Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective," *MIS Quarterly* (31:1), pp. 105-136 (doi: 10.2307/25148783).

Peeters, M. M. M., van Diggelen, J., van den Bosch, K., Bronkhorst, A., Neerincx, M. A., Schraagen, J. M., and Raaijmakers, S. 2021. "Hybrid collective intelligence in a human–AI society," *AI & Society* (36), pp. 217-238 (doi: 10.1007/s00146-020-01005-y).

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., and Bragge, J. 2006. "The Design Science Research Process : A Model for Producing and Presenting Information Systems Research," in *1st International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, Claremont, California, USA, pp. 83-106 (doi: 10.48550/arXiv.2006.02763).

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77 (doi: 10.2753/MIS0742-1222240302).

Pereira, M. M., and Paulovich, F. V. 2020. "RankViz: A visualization framework to assist interpretation of Learning to Rank algorithms," *Computers & Graphics* (93), pp. 25-38 (doi: 10.1016/j.cag.2020.09.017).

Pinhanez, C. S. 2020. "HCI Research Challenges of the Next Generation of Conversational Systems," in *2nd Conference on Conversational User Interfaces (CUI)*, Bilbao, Spain (doi: 10.1145/3405755.3406153).

Portela, M., and Granell-Canut, C. 2017. "A new friend in our smartphone?: observing interactions with chatbots in the search of emotional engagement," in *18th International Conference on Human Computer Interaction*, Cancun, Mexico (doi: 10.1145/3123818.3123826).

Poser, M., and Bittner, E. A. C. 2020. "Hybrid Teamwork: Consideration of Teamwork Concepts to Reach Naturalistic Interaction between Humans and Conversational Agents," in *15th International Conference on Wirtschaftsinformatik*, Potsdam, Germany (doi: 10.30844/wi_2020_a6-poser).

Poser, M., Singh, S., and Bittner, E. A. C. 2021. "Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents," in *54th Hawaii International Conference on System Sciences (HICSS)*, Virtual Event, pp. 1181-1190 (doi: 10.24251/HICSS.2021.144).

Poser, M., Wiethof, C., Banerjee, D., Subramanian, V. S., Paucar, R., and Bittner, E. A. C. 2022a. "Let‘s Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service," in *The Transdisciplinary Reach of Design Science Research. DESRIST 2022. Lecture Notes in Computer Science, vol 13229*, A. Drechsler, A. Gerber and A. Hevner (eds.), Springer, Cham, pp. 142-153 (doi: 10.1007/978-3-031-06516-3_11).

Poser, M., Wiethof, C., and Bittner, E. A. C. 2022b. "Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions," in *30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Prentice, C., and Nguyen, M. 2020. "Engaging and retaining customers with AI and employee service," *Journal of Retailing and Consumer Services* (56) (doi: 10.1016/j.jretconser.2020.102186).

Pries-Heje, J., Baskerville, R., and Venable, J. 2008. "Strategies for Design Science Research Evaluation," in *16th European Conference on Information Systems (ECIS)*, Galway, Ireland.

Przybilla, L., Baar, L., Wiesche, M., and Krcmar, H. 2019. "Machines as Teammates in Creative Teams: Digital Facilitation of the Dual Pathways to Creativity," in *Computers and People Research Conference*, Nashville, Tennessee, USA, pp. 94-102 (doi: 10.1145/3322385.3322402).

Püschel, L., Röglinger, M., and Schlott, H. 2016. “What‘s in a Smart Thing? Development of a Multi-layer Taxonomy,” in *37th International Conference on Information Systems (ICIS)*, Dublin, Ireland.

Raisch, S., and Krakowski, S. 2021. “Artificial Intelligence and Management: The Automation-Augmentation Paradox,” *Academy of Management Review* (46:1), pp. 192-210 (doi: 10.5465/amr.2018.0072).

Raj, M., and Seamans, R. 2019. “Primer on artificial intelligence and robotics,” *Journal of Organization Design* (8) (doi: 10.1186/s41469-019-0050-0).

Rank, S., Hoffmann, S., Struck, H.-G., Spierling, U., and Petta, P. 2012. “Creativity in Configuring Affective Agents for Interactive Storytelling,” in *3rd International Conference on Computational Creativity (ICCC)*, Dublin, Ireland, pp. 165-169.

Rao, S., Chitranshi, J., and Punjabi, N. 2020. “Role of Artificial Intelligence in Employee Engagement and Retention,” *Journal of Applied Management* (12:2).

Rapp, A., Curti, L., and Boldi, A. 2021. “The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots,” *International Journal of Human-Computer Studies* (151) (doi: 10.1016/j.ijhcs.2021.102630).

Reani, M., Peek, N., and Jay, C. 2019. “How different visualizations affect human reasoning about uncertainty: An analysis of visual behaviour,” *Computers in Human Behavior* (92), pp. 55-64 (doi: 10.1016/j.chb.2018.10.033).

Rhodes, M. 1961. “An Analysis of Creativity,” *The Phi Delta Kappan* (42:7), pp. 305-310.

Richter, A., Heinrich, P., Stocker, A., and Schwabe, G. 2018. “Digital Work Design - The Interplay of Human and Computer in Future Work Practices as an Interdisciplinary (Grand) Challenge,” *Business & Information Systems Engineering* (60:3), pp. 259-264 (doi: 10.1007/s12599-018-0534-4).

Riek, L. D. 2012. “Wizard of Oz studies in HRI: a systematic review and new reporting guidelines,” *Journal of Human-Robot Interaction* (1:1), pp. 119-136 (doi: 10.5898/JHRI.1.1.Riek).

Robb, A., and Lok, B. 2014. “Social presence in mixed agency interactions,” in *IEEE Virtual Reality (VR)*, Minneapolis, Minnesota, USA, pp. 111-112 (doi: 10.1109/VR.2014.6802076).

Robinson, S., Orsingher, C., Alkire, L., Keyser, A. de, Giebelhausen, M., Papamichail, K. N., Shams, P., and Temerak, M. S. 2020. “Frontline encounters of the AI kind: An evolved service encounter framework,” *Journal of Business Research* (116), pp. 366-376 (doi: 10.1016/j.jbusres.2019.08.038).

Rothe, H., Wessel, L., and Barquet, A. P. 2020. “Accumulating Design Knowledge: A Mechanisms-Based Approach,” *Journal of the Association for Information Systems* (21:3), pp. 771-810 (doi: 10.17705/1jais.00619).

Rundo, L., Pirrone, R., Vitabile, S., Sala, E., and Gambino, O. 2020. “Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine,” *Journal of Biomedical Informatics* (108) (doi: 10.1016/j.jbi.2020.103479).

Russell, S. J., and Norvig, P. 2016. *Artificial Intelligence: A Modern Appoach*, Pearson.

Rust, R. T., and Huang, M.-H. 2014. “The Service Revolution and the Transformation of Marketing Science,” *Marketing Science* (33:2), pp. 206-221 (doi: 10.1287/mksc.2013.0836).

Rzepka, C., and Berger, B. 2018. “User Interaction with AI-enabled Systems: A Systematic Review of IS Research,” in *39th International Conference on Information Systems (ICIS)*, San Francisco, California, USA.

Sajeev, S., Huang, J., Karampatziakis, N., Hall, M., Kochman, S., and Chen, W. 2021. “Contextual Bandit Applications in a Customer Support Bot,” in *27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event, pp. 3522-3530 (doi: 10.1145/3447548.3467165).

Salam, M. A., Koone, M. E., Thirumuruganathan, S., Das, G., and Roy, S. B. 2019. “A Human-in-the-loop Attribute Design Framework for Classification,” in *The World Wide Web Conference (WWW)*, San Francisco, California, USA, pp. 1612-1622 (doi: 10.1145/3308558.3313547).

Salber, D., and Coutaz, J. 1993. “Applying the Wizard of Oz technique to the study of multimodal systems,” in *Human-Computer Interaction. EWHCI 1993. Lecture Notes in Computer Science, vol 753*, L. J. Bass, J. Gornostaev and C. Unger (eds.), Springer, pp. 219-230 (doi: 10.1007/3-540-57433-6_51).

Samarasinghe, S. M., and Tretiakov, A. 2009. “A multi-dimensional measure of e-learning systems success,” in *Australasian Society for Computers in Learning in Tertiary Education (ASCILITE)*, Auckland, New Zealand, pp. 908-910.

Sampson, S. E., and Froehle, C. M. 2006. “Foundations and Implications of a Proposed Unified Services Theory,” *Production and Operations Management* (15:2), pp. 329-343 (doi: 10.1111/j.1937-5956.2006.tb00248.x).

Sarker, I. H. 2021. “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Computer Science* (2) (doi: 10.1007/s42979-021-00592-x).

Scherer, A., Wünderlich, N. V., and Wangenheim, F. von. 2015. “The Value of Self-Service: Long-Term Effects of Technology-Based Self-Service Usage on Customer Retention,” *MIS Quarterly* (39:1), pp. 177-200 (doi: 10.25300/MISQ/2015/39.1.08).

Schneider, J. 2020. “Humans learn too: Better Human-AI Interaction using Optimized Human Inputs,” available at https://arxiv.org/abs/2009.09266v1.

Schneider, J., and Handali, J. P. 2019. “Personalized explanation in machine learning: A conceptualization,” in *27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden (doi: 10.48550/arXiv.1901.00770).

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., and Rosser, H. K. 2021. “Deciding Whether and How to Deploy Chatbots,” *MIS Quarterly Executive* (20:1) (doi: 10.17705/2msqe.00039).

Seeber, I., Bittner, E. A. C., Briggs, R. O., Vreede, G.-J. de, Vreede, T. de, Druckenmiller, D., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. 2018. “Machines as Teammates: A Collaboration Research Agenda,” in *51st Hawaii*

International Conference on System Sciences (HICSS), Waikoloa Village, Hawaii, USA, pp. 420-429 (doi: 10.24251/HICSS.2018.055).

Seeber, I., Bittner, E. A. C., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. 2020. "Machines as teammates: A research agenda on AI in team collaboration," *Information & Management* (57:2) (doi: 10.1016/j.im.2019.103174).

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., and Lowry, P. B. 2019. "Reinventing Collaboration with Autonomous Technology-Based Agents," in *27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Seering, J., Luria, M., Kaufman, G., and Hammer, J. 2019. "Beyond Dyadic Interactions: Considering Chatbots as Community Members," in *Conference on Human Factors in Computing Systems (CHI ,19)*, Glasgow, Scotland, UK (doi: 10.1145/3290605.3300680).

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. 2011. "Action Design Research," *MIS Quarterly* (35:1), pp. 37-56 (doi: 10.2307/23043488).

Servion. 2018. "What Makes Emerging Technologies The Future Of Customer Experience?" available at https://servion.com/blog/what-emerging-technologies-future-customer-experience/, accessed on Aug 13 2022.

Shneiderman, B. 2020. "Human-Centered Artificial Intelligence: Three Fresh Ideas," *AIS Transactions on Human-Computer Interaction* (12:3), pp. 109-124 (doi: 10.17705/1thci.00131).

Short, J. C., Ketchen, D. J., and Palmer, T. B. 2002. "The Role of Sampling in Strategic Management Research on Performance: A Two-Study Analysis," *Journal of Management* (28:3), pp. 363-385 (doi: 10.1016/S0149-2063(02)00132-0).

Siemon, D., Eckardt, L., and Robra-Bissantz, S. 2015. "Tracking Down the Negative Group Creativity Effects with the Help of an Artificial Intelligence-Like Support System," in *48th Hawaii International Conference on System Sciences (HICSS)*, Kauai, Hawaii, USA, pp. 236-243 (doi: 10.1109/HICSS.2015.37).

Simmert, B., Ebel, P. A., Peters, C., Bittner, E. A. C., and Leimeister, J. M. 2019. "Conquering the Challenge of Continuous Business Model Improvement - Design of a Repeatable Process," *Business & Information Systems Engineering* (61:4), pp. 451-468 (doi: 10.1007/s12599-018-0556-y).

Smith, A., Kumar, V., Boyd-Graber, J., Seppi, K., and Findlater, L. 2018. "Closing the Loop: User-Centered Design and Evaluation of a Human-in-the-Loop Topic Modeling System," in *23rd International Conference on Intelligent User Interfaces (IUI)*, Tokyo, Japan, pp. 293-304 (doi: 10.1145/3172944.3172965).

Smith-Renner, A., Kumar, V., Boyd-Graber, J., Seppi, K., and Findlater, L. 2020. "Digging into user control: perceptions of adherence and instability in transparent models," in *25th International Conference on Intelligent User Interfaces (IUI)*, Cagliari, Italy, pp. 519-530 (doi: 10.1145/3377325.3377491).

Sniezek, J. A., and Buckley, T. 1995. "Cueing and Cognitive Conflict in Judge-Advisor Decision Making," *Organizational Behavior and Human Decision Processes* (62:2), pp. 159-174 (doi: 10.1006/obhd.1995.1040).

Song, J. H., and Zinkhan, G. M. 2008. "Determinants of Perceived Web Site Interactivity," *Journal of Marketing* (72:2), pp. 99-113 (doi: 10.1509/jmkg.72.2.99).

Sonnenberg, C., and vom Brocke, J. 2012. "Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research," in *Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012. Lecture Notes in Computer Science, vol 7286*, K. Peffers, M. Rothenberger and B. Kuechler (eds.), Springer, pp. 381-397 (doi: 10.1007/978-3-642-29863-9_28).

Sousa, D. N., Brito, M. A., and Argainha, C. 2019. "Virtual customer service: building your chatbot," in *3rd International Conference on Business and Information Management (ICBIM)*, Paris, France, pp. 174-179 (doi: 10.1145/3361785.3361805).

Steenwinckel, B., Paepe, D. de, Vanden Hautte, S., Heyvaert, P., Bentefrit, M., Moens, P., Dimou, A., van den Bossche, B., Turck, F. de, van Hoecke, S., and Ongenae, F. 2021. "FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning," *Future Generation Computer Systems* (116), pp. 30-48 (doi: 10.1016/j.future.2020.10.015).

Steffens, F., Marczak, S., Figueira Filho, F., Treude, C., Singer, L., and Redmiles Ban Al-Ani, D. 2015. "Using Gamification as a Collaboration Motivator for Software Development Teams: A Preliminary Framework," in *Brazilian Symposium on Collaborative Systems (SBSC)*, Salvador, BA, Brazil.

Storch, N. 2011. "Collaborative Writing in L2 Contexts: Processes, Outcomes, and Future Directions," *Annual Review of Applied Linguistics* (31), pp. 275-288 (doi: 10.1017/S0267190511000079).

Strohmann, T., Siemon, D., and Robra-Bissantz, S. 2019. "Introducing The Virtual Companion Canvas - Towards Designing Collaborative Agents," in *Workshop on Designing User Assistance in Intelligent Systems*, Stockholm, Sweden, pp. 4-6 (doi: 10.5445/IR/1000095222).

Subramaniam, S., Aggarwal, P., Dasgupta, G. B., and Paradkar, A. 2018. "COBOTS - A Cognitive Multi-Bot Conversational Framework for Technical Support," in *17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, Stockholm, Sweden, pp. 597-604.

Suh, A., Cheung, C. M. K., Ahuja, M., and Wagner, C. 2017. "Gamification in the Workplace: The Central Role of the Aesthetic Experience," *Journal of Management Information Systems* (34:1), pp. 268-305 (doi: 10.1080/07421222.2017.1297642).

Sujata, J., Aniket, D., and Mahasingh, M. 2019. "Artificial Intelligence Tools for Enhancing Customer Experience," *International Journal of Recent Technology and Engineering* (8:2S3), pp. 700-706 (doi: 10.35940/ijrte.B1130.0782S319).

Sun, W., Zhang, S., Balog, K., Ren, Z., Ren, P., Chen, Z., and Rijke, M. de. 2021. "Simulating User Satisfaction for the Evaluation of Task-oriented Dialogue Systems," in *44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, pp. 2499-2506 (doi: 10.1145/3404835.3463241).

Sutherland, J., and Canwell, D. 2004. *Key Concepts in Human Resource Management*, London: Red Globe Press.

Suwantarathip, O., and Wichadee, S. 2014. "The Effects of Collaborative Writing Activity Using Google Docs on Students' Writing Abilities," *The Turkish Online Journal of Educational Technology* (13:2), pp. 148-156.

Svenningsson, N., and Faraon, M. 2019. "Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots," in *2nd Artificial Intelligence and Cloud Computing Conference (AICCC)*, Kobe, Japan, pp. 151-161 (doi: 10.1145/3375959.3375973).

Sweller, J. 1988. "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science* (12:2), pp. 257-285 (doi: 10.1207/s15516709cog1202_4).

Sweller, J., Ayres, P., and Kalyuga, S. 2011. *Cognitive Load Theory*, New York: Springer.

Szopinski, D., Schoormann, T., and Kundisch, D. 2019. "Because your taxonomy is worth it: Towards a framework for taxonomy evaluation," in *27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden.

Tamagnini, P., Krause, J., Dasgupta, A., and Bertini, E. 2017. "Interpreting Black-Box Classifiers Using Instance-Level Visual Explanations," in *2nd Workshop on Human-In-the-Loop Data Analytics*, Chicago, Illinois, USA (doi: 10.1145/3077257.3077260).

Tavanapour, N., and Bittner, E. A. C. 2017. "Collaboration among Crowdsourcees: Towards a Design Theory for Collaboration Process Design," in *50th Hawaii International Conference on System Sciences (HICSS)*, Waikoloa Village, Hawaii, USA, pp. 333-342 (doi: 10.24251/HICSS.2017.040).

Thiebes, S., Lins, S., and Basten, D. 2014. "Gamifying Information Systems - A Synthesis of Gamification Mechanics and Dynamics," in *22nd European Conference on Information Systems (ECIS)*, Tel Aviv.

Thuan, N. H., Drechsler, A., and Antunes, P. 2019. "Construction of Design Science Research Questions," *Communications of the Association for Information Systems* (44), pp. 332-363 (doi: 10.17705/1CAIS.04420).

Traumer, F., Oeste-Reiß, S., and Leimeister, J. M. 2017. "Towards a Future Reallocation of Work between Humans and Machines - Taxonomy of Tasks and Interaction Types in the Context of Machine Learning," in *38th International Conference on Information Systems (ICIS)*, Seoul, South Korea (doi: 10.2139/ssrn.3159131).

Trujillo, A. C., Gregory, I. M., and Ackerman, K. A. 2019. "Evolving Relationships between Humans and Machines," *IFAC-PapersOnLine* (51:34), pp. 366-371 (doi: 10.1016/j.ifacol.2019.01.015).

Tseng, S.-L., and Sun, H. 2017. "Playful Design Elements and Stages of Player Experience in Gamification," in *38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.

Turel, O., and Connelly, C. E. 2013. "Too busy to help: Antecedents and outcomes of interactional justice in web-based service encounters," *International Journal of Information Management* (33:4), pp. 674-683 (doi: 10.1016/j.ijinfomgt.2013.03.005).

Ulfert, A.-S., Antoni, C. H., and Ellwart, T. 2022. "The role of agent autonomy in using decision support systems at work," *Computers in Human Behavior* (126) (doi: 10.1016/j.chb.2021.106987).

van der El, K., Barendswaard, S., Pool, D. M., and Mulder, M. 2016. "Effects of Preview Time on Human Control Behavior in Rate Tracking Tasks," *IFAC-PapersOnLine* (49:19), pp. 108-113 (doi: 10.1016/j.ifacol.2016.10.470).

Vanduhe, V. Z., Nat, M., and Hasan, H. F. 2020. "Continuance Intentions to Use Gamification for Training in Higher Education: Integrating the Technology Acceptance Model (TAM), Social Motivation, and Task Technology Fit (TTF)," *IEEE Access* (8), pp. 21473-21484 (doi: 10.1109/ACCESS.2020.2966179).

Vargo, S. L., Maglio, P. P., and Akaka, M. A. 2008. "On value and value co-creation: A service systems and service logic perspective," *European Management Journal* (26:3), pp. 145-152 (doi: 10.1016/j.emj.2008.04.003).

Venable, J., Pries-Heje, J., and Baskerville, R. 2012. "A Comprehensive Framework for Evaluation in Design Science Research," in *Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012. Lecture Notes in Computer Science, vol 7286*, K. Peffers, M. Rothenberger and B. Kuechler (eds.), Springer, pp. 423-438 (doi: 10.1007/978-3-642-29863-9_31).

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: a Framework for Evaluation in Design Science Research," *European Journal of Information Systems* (25:1), pp. 77-89 (doi: 10.1057/ejis.2014.36).

Venable, J. R. 2006. "The Role of Theory and Theorising in Design Science Research," in *1st International Conference on Design Science Research in Information Systems and Technology (DESRIST),* Claremont, California, USA.

Verdenius, F. 1995. "Managing Product Inherent Variance During Treatment," *IFAC Proceedings Volumes* (28:4), pp. 121-126 (doi: 10.1016/S1474-6670(17)45551-0).

Verhagen, T., van Nes, J., Feldberg, F., and van Dolen, W. 2014. "Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters," *Journal of Computer-Mediated Communication* (19:3), pp. 529-545 (doi: 10.1111/jcc4.12066).

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in *17th European Conference on Information Systems (ECIS),* Verona, Italy.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the Shoulder of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," *Communications of the Association for Information Systems* (37), pp. 205-224 (doi: 10.17705/1CAIS.03709).

vom Brocke, J., Winter, R., Hevner, A., and Maedche, A. 2020. "Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space," *Journal of the Association for Information Systems* (21:3), pp. 520-544 (doi: 10.17705/1jais.00611).

Wache, H., Möller, F., Schoormann, T., Strobel, G., and Petrik, D. 2022. "Exploring the Abstraction Levels of Design Principles: The Case of Chatbots," in *17th International Conference on Wirtschaftsinformatik (WI)*, Nürnberg, Germany.

Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. 1992. "Building an Information Systems Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36-59 (doi: 10.1287/isre.3.1.36).

Wambsganss, T., Küng, T., Söllner, M., and Leimeister, J. M. 2021. "ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills," in *Conference on Human Factors in Computing Systems (CHI ,21)*, Virtual Event (doi: 10.1145/3411764.3445781).

Wang, G. 2019. "Humans in the Loop: The Design of Interactive AI Systems," available at https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems, accessed on Nov 13 2020.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. 13-23.

Wegge, J., Roth, C., Neubach, B., Schmidt, K.-H., and Kanfer, R. 2008. "Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size," *Journal of Applied Psychology* (93:6), pp. 1301-1313 (doi: 10.1037/a0012680).

Weinmann, M., Schneider, C., and vom Brocke, J. 2016. "Digital Nudging," *Business & Information Systems Engineering* (58:6), pp. 433-436 (doi: 10.1007/s12599-016-0453-1).

Wiethof, C., and Bittner, E. A. C. 2021. "Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review," in *42nd International Conference on Information Systems (ICIS)*, Austin, Texas, USA.

Wiethof, C., and Bittner, E. A. C. 2022. "Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI," in *30th European Conference on Information Systems (ECIS)*, Timișoara, Romania.

Wiethof, C., Poser, M., and Bittner, E. A. C. 2022a. "Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service," in *Pacific Asia Conference on Information Systems (PACIS)*, Virtual Event.

Wiethof, C., Roocks, T., and Bittner, E. A. C. 2022b. "Gamifying the Human-in-the-Loop: Toward Increased Motivation for Training AI in Customer Service," in *Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science, vol 13336*, H. Degen and S. Ntoa (eds.), Springer, Cham, pp. 100-117 (doi: 10.1007/978-3-031-05643-7_7).

Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021a. “Designing and Evaluating a Collaborative Writing Process with Gamification Elements: Toward a Framework for Gamifying Collaboration Processes,” *AIS Transactions on Human-Computer Interaction* (13:1), pp. 38-61 (doi: 10.17705/1thci.00141).

Wiethof, C., Tavanapour, N., and Bittner, E. A. C. 2021b. “Implementing an Intelligent Collaborative Agent as Teammate in Collaborative Writing: toward a Synergy of Humans and AI,” in *54th Hawaii International Conference on System Sciences (HICSS)*, Virtual Event, pp. 400-409 (doi: 10.24251/HICSS.2021.047).

Wilde, T., and Hess, T. 2006. “Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung, Arbeitsbericht, No. 2,” (doi: 10.5282/ubm/epub.14146).

Wilde, T., and Hess, T. 2007. “Forschungsmethoden der Wirtschaftsinformatik: Eine empirische Untersuchung,” *Wirtschaftsinformatik* (49), pp. 280-287 (doi: 10.1007/s11576-007-0064-z).

Wilson, H. J., and Daugherty, P. R. 2018. “Collaborative Intelligence: Humans and AI Are Joining Forces,” *Harvard Business Review*, pp. 114-123.

Wintersberger, P., Klotz, T., and Riener, A. 2020. “Tell Me More: Transparency and Time-Fillers to Optimize Chatbots‘ Waiting Time Experience,” in *11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI)*, Tallinn, Estonia (doi: 10.1145/3419249.3420170).

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

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., and He, L. 2022. “A survey of human-in-the-loop for machine learning,” *Future Generation Computer Systems* (135), pp. 364-381 (doi: 10.1016/j.future.2022.05.014).

Xu, A., Liu, Z., Guo, Y., Sinha, V., and Akkiraju, R. 2017. “A New Chatbot for Customer Service on Social Media,” in *Conference on Human Factors in Computing Systems (CHI ‚17)*, Denver, Colorado, USA, pp. 3506-3510 (doi: 10.1145/3025453.3025496).

Xu, W. 2019. “Toward human-centered AI: a perspective from human-computer interaction,” *Interactions* (26:4), pp. 42-46 (doi: 10.1145/3328485).

Xu, Y., Shieh, C.-H., van Esch, P., and Ling, I.-L. 2020. “AI Customer Service: Task Complexity, Problem-Solving Ability, and Usage Intention,” *Australasian Marketing Journal* (28:4), pp. 189-199 (doi: 10.1016/j.ausmj.2020.03.005).

Yimam, S. M., and Biemann, C. 2018. “Demonstrating Par4Sem - A Semantic Writing Aid with Adaptive Paraphrasing,” in *Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, Belgium, pp. 48-53 (doi: 10.18653/v1/D18-2009).

Yimam, S. M., Biemann, C., Majnaric, L., Šabanović, Š., and Holzinger, A. 2016. “An adaptive approach for biomedical entity and relation recognition,” *Brain Informatics* (3), pp. 157-168 (doi: 10.1007/s40708-016-0036-4).

Yu, K., Berkovsky, S., Taib, R., Zhou, J., and Chen, F. 2019. "Do I trust my machine teammate?: an investigation from perception to decision," in *24th International Conference on Intelligent User Interfaces (IUI)*, Marina del Rey, CA, USA, pp. 460-468 (doi: 10.1145/3301275.3302277).

Zeni, M., Zhang, W., Bignotti, E., Passerini, A., and Giunchiglia, F. 2019. "Fixing Mislabeling by Human Annotators Leveraging Conflict Resolution and Prior Knowledge," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (3:1) (doi: 10.1145/3314419).

Zhang, B., Lin, H., Zuo, S., Liu, H., Chen, Y., Li, L., Ouyang, H., and Yuan, B. 2020. "Research on Intelligent Robot Engine of Electric Power Online Customer Services Based on Knowledge Graph," in *4th International Conference on Innovation in Artificial Intelligence (ICIAI)*, Xiamen, China, pp. 216-221 (doi: 10.1145/3390557.3394312).

Zhang, P., and Li, N. 2005. "The Intellectual Development of Human-Computer Interaction Research: A Critical Assessment of the MIS Literature (1990-2002)," *Journal of the Association for Information Systems* (6:11), pp. 227-292 (doi: 10.17705/1jais.00070).

Zhang, X., Olfman, L., and Firpo, D. 2011. "An Information Systems Design Theory for Collaborative ePortfolio Systems," in *44th Hawaii International Conference on System Sciences (HICSS)*, Koloa, Kauai, Hawaii, USA (doi: 10.1109/HICSS.2011.56).

Zheng, Q., Tang, Y., Liu, Y., Liu, W., and Huang, Y. 2022. "UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library," in *Conference on Human Factors in Computing Systems (CHI ,22)*, New Orleans, Louisiana, USA (doi: 10.1145/3491102.3501855).

Zierau, N., Elshan, E., Visini, C., and Janson, A. 2020. "A Review of the Empirical Literature on Conversational Agents and Future Research Directions," in *41st International Conference on Information Systems (ICIS)*, Hyderabad, India.

Zschech, P., Walk, J., Heinrich, K., Vössing, M., and Kühl, N. 2021. "A Picture is Worth a Collaboration: Accumulating Design Knowledge for Computer-Vision-Based Hybrid Intelligence Systems," in *29th European Conference on Information Systems (ECIS)*, Virtual Event (doi: 10.48550/arXiv.2104.11600).

18 Eidesstaatliche Versicherung

Hiermit versichere ich, **Christina Wiethof**, geboren am 17. August 1996 in Rheda-Wiedenbrück, an Eides statt, dass ich die vorliegende Dissertationsschrift mit dem Titel „Hybrid Intelligence Systems – Designing Interactions for Continuous Mutual Augmentation of Humans and AI“ selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Hamburg, 30.08.2022 C. Wiethof

Ort, Datum Unterschrift

19 Versicherung zur Identität der elektronischen und gedruckten Dissertationsschrift

Ich versichere, dass das gebundene Exemplar der Dissertation und das in elektronischer Form eingereichte Dissertationsexemplar und das zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertationsschrift identisch sind.

Hamburg, 30.08.2022 C. Wiethoff

Ort, Datum Unterschrift