6-18-2022

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

Christina Wiethof
Information Systems, wiethof@informatik.uni-hamburg.de

Eva A. C. Bittner
University of Hamburg, bittner@informatik.uni-hamburg.de

TOWARD A HYBRID INTELLIGENCE SYSTEM IN CUSTOMER SERVICE: COLLABORATIVE LEARNING OF HUMAN AND AI

Research Paper

Christina Wiethof, Universität Hamburg, Hamburg, Germany, christina.wiethof@uni-hamburg.de

Eva A. C. Bittner, Universität Hamburg, Hamburg, Germany, eva.bittner@uni-hamburg.de

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.

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.

2 Related Work

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.

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.

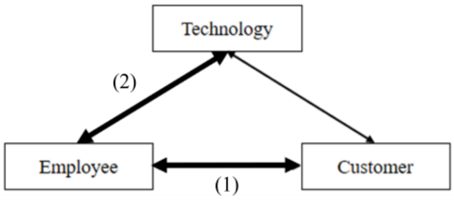

Figure 1. Frontline service technology infusion, augmentation scenario highlighted (Keyser et al., 2019), adapted.

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.

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).

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).

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).

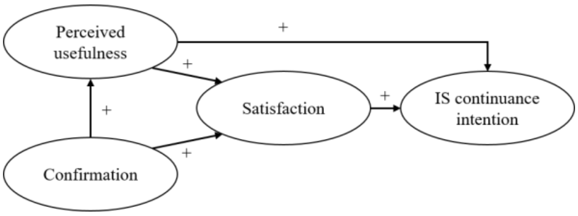

Figure 2. A post-acceptance model of IS continuance (Bhattacherjee, 2001), adapted.

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).

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.

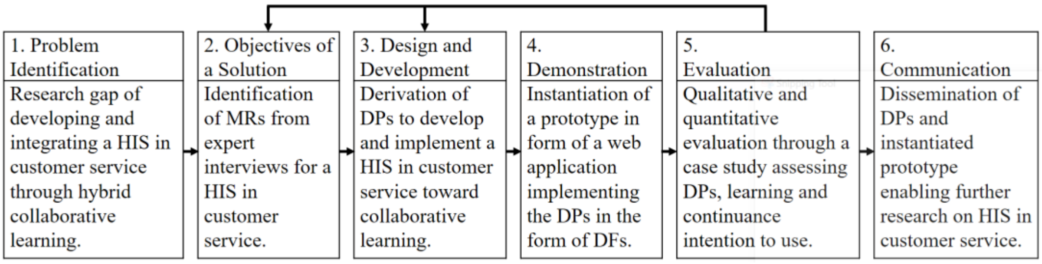

Figure 3. Structure along the DSR process.

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).

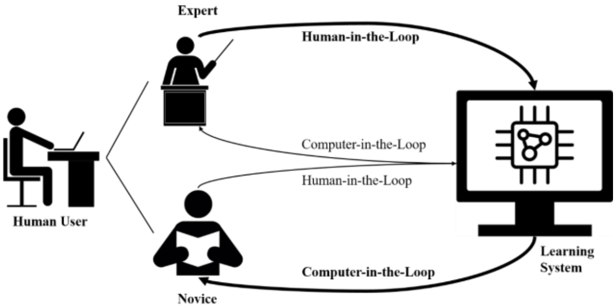

Figure 4. Hybrid collaborative learning differentiating between experts and novices.

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)

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.

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)

Table 1. DPs with according MRs.

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.

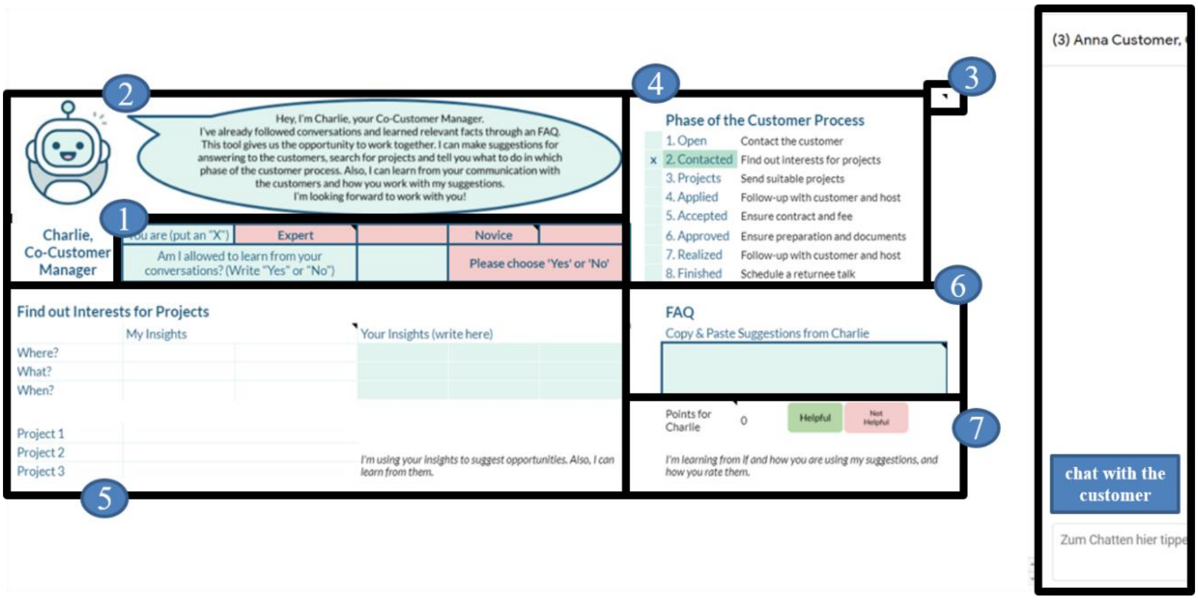

Figure 5. Prototype – user interface of the web application with DFs 1-7[1].

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).

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

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.

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.

7.1 Quantitative results

To identify differences between novices and experts, we first tested for normal distribution. As the sample sizes are relatively small ($n_{novices} = 30$ and $n_{experts} = 31$), 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.

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	

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.
[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.

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.

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.

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.

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.