Please quote as: Reinhard, P., Li, M. M., Peters, C. & Leimeister, J. M. (2024). LET EMPLOYEES TRAIN THEIR OWN CHATBOTS: DESIGN OF GENERATIVE AI-ENABLED DELEGATION SYSTEMS. *European Conference on Information Systems (ECIS)*, Paphos, Cyprus.

LET EMPLOYEES TRAIN THEIR OWN CHATBOTS: DESIGN OF GENERATIVE AI-ENABLED DELEGATION SYSTEMS

Completed Research Paper

Philipp Reinhard, University of Kassel, Kassel, Germany, philipp.reinhard@uni-kassel.de

Mahei Manhai Li, University of Kassel, Kassel, Germany, mahei.li@uni-kassel.de

Christoph Peters, University of Kassel/University of St.Gallen, Kassel/St.Gallen, Germany/Switzerland, christoph.peters@uni-kassel.de

Jan Marco Leimeister, University of Kassel/University of St.Gallen, Kassel/St.Gallen, Germany/Switzerland, janmarco.leimeister@unisg.ch

Abstract

While chatbots can be implemented with very little effort, scaling and maintaining chatbots remains a challenge. This is crucial in knowledge-intensive customer service like IT support, where domain knowledge must stay current with the evolving IT landscape. Following design science research, we derive design principles for a generative AI (GPT4) enabled textual training data creation and curation system (T^2C^2) as part of a new class of systems – bot delegation systems. For the design of T^2C^2, chatbot and domain expert viewpoints are integrated. We evaluate two instances of T^2C^2, each with distinct degrees of human-ai delegation where employees act both as creators and curators of training data. The paper's theoretical contribution is two-fold: (1) we present a novel kernel theory that represents the material characteristics of bot delegation systems by contextualizing the IS delegation framework to the self-determination theory; (2) the design and evaluation of T^2C^2 as the built-and-evaluated artifact.

Keywords: Chatbot; generative AI, customer service, IS delegation, self-determination theory

1 Introduction

While the market for IT services is continuously increasing, the shortage of IT professionals is expected to reach 85.2 million by 2030 (Korn Ferry Institute 2018). To maintain top-quality support services for digital products, service organizations turn to leverage the potential of intelligent conversational agents (CA), often instantiated as chatbots, for automating customer service work (Følstad and Brandtzæg 2017; Gnewuch et al. 2017; Haupt et al. 2023). Figures show that the global market for chatbots will increase by 23,9 % per year until 2032 (Statista 2024). Accordingly, there has been a considerable amount of research on initial setups and design decisions for improving customer experience and human-chatbot interaction from a customer perspective (Chaves and Gerosa 2021; Følstad and Brandtzæg 2017; Nißen et al. 2022; Yu et al. 2024). Overall, chatbot development is made easier. For instance, even non-programmer domain experts can design intent-based chatbots with low-code platforms (Li et al. 2022). Intents refer to the classes of requests that the chatbot should recognize and perform. Typically, intent-based chatbots are first trained on a small set of training data (i.e., chatbot response, intents, intent example sentences). Therefore, the question of how chatbot training data is maintained (e.g., revising example sentences and intents and adding example sentences) after their initial roll-out remains largely neglected in research and practice (Janssen et al. 2022; Khan 2017). Similar to other AI applications, it is crucial to include domain expertise in training the systems (Grønsund and Aanestad 2020; Li et al.

2024; Reinhard et al. 2023b). Nevertheless, the integration of frontline service employees (FSE) with limited technological expertise in chatbot development remains a challenging aspect (Feine et al. 2020). While chatbot development teams usually closely work with service management, adding new data to chatbots is referred to as both time-consuming and difficult for FSE. While more intelligent CAs can use training data to continuously learn, in reality, the training data are often initialized during the initial CA implementation project, making it quickly and partially obsolete (Gao et al. 2021). This results in low motivation to contribute knowledge to keep the bot up to date as known from data annotation tasks in AI research (Nowak and Rüger 2010).

We aim to create a textual training data creation and curation system (T^2C^2) for chatbots that streamlines textual training data generation and motivates FSEs to act as both data curators (i.e., improving training data, labeling data, etc.) and data creators (i.e., creating intents, replies, and examples). Extending the research on interactive chatbot development (Feine et al. 2020), our solution is integrated into the problem-solving routines of FSEs and is enabled by recent technological advances of generative AI (GenAI) such as GPT4 (Ouyang et al. 2022). By doing so, employee's workplaces are changed and influenced by AI in two overarching directions (Larivière et al. 2017): (a) the trained chatbots partly substitute employees given the selected cases and b) GenAI augments the creation of chatbot training data during post-implementation. As FSEs are demanded to delegate the role of AI agent (here: GenAI) as well as the chatbot itself, we refer to the IS delegation framework (Baird and Maruping 2021) to ensure cost-efficient allocation of tasks and the human's willingness to delegate AI. Accordingly, we aim at increasing the employees' intrinsic motivation to maintain chatbot training data continuously. Therefore, self-determination theory (SDT) (Deci and Ryan 2012) acts as another theoretical lens for incorporating the perspective and resources of affected FSEs in delegating chatbots. SDT has already been applied to chatbot research (Jiménez-Barreto et al. 2021; Nguyen et al. 2022) and AI assistance (Vreede et al. 2021) and can inform the design of hybrid intelligence systems (Poser et al. 2022). Thus, we raise the following research question: *How to design GenAI-enabled training data creation and curation systems that motivate frontline service employees to maintain chatbots?*

The paper is structured as follows. First, we introduce the theoretical foundation spanning the IS delegation framework (Baird and Maruping 2021) and SDT (Deci and Ryan 2012). Afterward, the related work regarding chatbot development and maintenance as well as GenAI is outlined. After describing our research methodology, the paper communicates the design and development of T^2C^2. Finally, we discuss the implications for practice and theory.

2 Theoretical Foundation

The domain of Information Systems (IS) has conventionally perceived the interplay between humans and technology as a scenario where humans hold the upper hand while technology serves as a submissive instrument. More recent frameworks emphasize that technology is also an active agent in this relationship (Leonardi 2011). Baird and Maruping (2021) put forth a conceptual structure for the delegation of IS, which is inspired by decision-making scholarship that centers on how authority, obligations, and synchronization are assigned among actors (Akinola et al. 2018; Ribes et al. 2013). This structure outlines three elements for human and artificial agents, promoting a shift to a delegation-focused approach: (a) preferences, (b) endowments, and (c) roles. We apply the IS delegation framework to design training data creation and curation systems using hybrid intelligence (Dellermann et al. 2019), aiming for enhanced productivity. The delegation framework (Baird and Maruping 2021) ensures an efficient and outcome-maximizing allocation of tasks and selective automation and augmentation within the creation of training data.

Self-determination theory (Deci and Ryan 2012) captures three psychological needs ensuring that people feel intrinsically motivated and enjoy being productive: autonomy, competence, and relatedness. Prior research in information systems has shown that the three basic needs reliably predict continuous usage (Rezvani et al. 2017) and user motivation (Menard et al. 2017). SDT has been applied to multiple research on chatbots (Ballou et al. 2022)– primarily from a user experience perspective (Nguyen et al.

2022). However, SDT is particularly applicable to AI integration into employees' workplaces and the guidance of design in the direction of desired goals such as engagement, motivation, and well-being (Vreede et al. 2021). Multiple studies considered SDT as a theoretical lens in DSR projects (Lohrenz et al. 2022; Poser et al. 2022; Vreede et al. 2021). We incorporate self-determination theory (SDT) (Deci and Ryan 2012) as a kernel theory and thereby meet the three fundamental needs.

3 Related Work

Chatbots are progressively being integrated into frontline service encounters for answering simple FAQs or even helping with complex customer requests (Janssen et al. 2022). Customer service chatbots are task-oriented dialogue systems that support customers with domain-specific requests (Xiao et al. 2019). They predominantly rely on text-based interactions, utilize AI solely for user intention (intent) classification, and exhibit restricted conversational capabilities and scope of topics (Luo et al. 2022). For the trained intents such intent-based CAs perform very well, provide a compliant customer experience, and can support highly domain-specific and complex processes (Luo et al. 2022). Natural language understanding (NLU) techniques such as machine learning approaches are applied to classify the intents, where the model is trained on a set of training examples for given intents. In case the intent cannot be predicted with certainty, the request is documented and forwarded to human FSE (Keyser et al. 2019). Thus, classifying a large number of intents correctly is key for frontline service quality and technology acceptance. Misclassification will lead to service breakdowns (Benner et al. 2021) and will congest already overloaded FSE with repetitive and monotone tasks.

According to Gao et al. (2021), many long-term chatbot projects fail due to the overwhelming efforts to maintain chatbot training data. Maintenance of chatbots however remains a rather neglected field of research. Deriving the main steps of managing chatbot projects from prior work (Janssen et al. 2022; Meyer Von Wolff et al. 2022) reveals that customer service organizations have to manage the phases (1) planning, (2) design, (3) development, and (4) maintenance (Figure 1). However, research on chatbots in the IS community is primarily concerned with the first three phases and often neglects the challenges of long-term scaling and maintenance. However, DSR approaches to solving the issue of chatbot performance and scalability for task-oriented bots after the development phase have been a matter of prior studies (Feine et al. 2020; Schloß et al. 2022). Feine et al. (2020) developed an interactive chatbot development system through which users can directly adjust chatbots' dialogues. While Schloß et al. (2022) aim to mine chatbot conversations to improve the model. Research from the field of computer science currently pursues to find technical solutions for new intent discovery utilizing deep learning (Zhang et al. 2022). However, regarding extending the chatbot, service departments are currently required to provide accurate data. Maintenance involves several key activities, including revising example sentences, refining intents, and augmenting the dataset by adding new example sentences. These tasks are vital for the chatbot's accuracy and ability to address diverse inquiries.

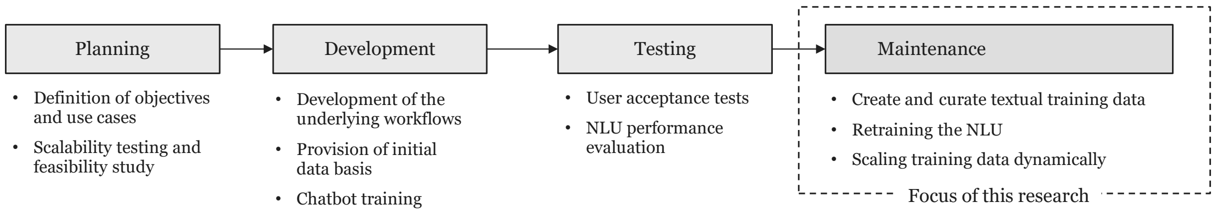

Figure 1. Chatbot setup based on Meyer Von Wolff et al. (2022) and Janssen et al. (2022)

In recent times, transformer-based large language models (LLM) have gained attention as deep learning foundation models for processing and fine-tuning natural language tasks, such as summarization, classification, and sentiment analysis (Reinhard et al. 2024). These models are often referred to as GenAI because they can both predict and create text. LLMs (e.g., GPT4.0, LLaMA) are suitable for general questions and open-domain question-answering (Petroni et al. 2019). For domain-specific service requests that require the integration of internal sources and high factual certainty, utilizing a compliant

NLU in the form of intent detection remains more powerful and reliant (Luo et al. 2022). Despite its limitations, GenAI has the potential to transform chatbot development and operations for frontline support services. GenAI can be applied as a method for data augmentation (Bayer et al. 2022) that can impact the quality of intent classifiers and thereby improve NLU.

4 Research Approach

To answer the beforementioned research question, we conduct a DSR project and follow a process similar to Kuechler and Vaishnavi (2008) and Meth et al. (2015) (Figure 1).

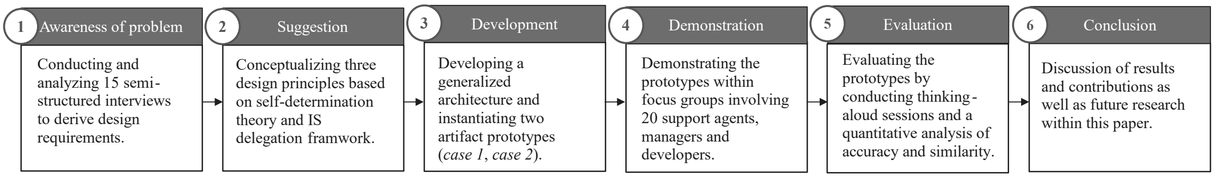

Figure 2. Design science approach based on Kuechler and Vaishnavi (2008)

To gain a better understanding of the problem (*Awareness of problem*) of continuous chatbot training data maintenance, 15 interviews with a development team and conversational AI experts were conducted. After reviewing the existing literature as well as interviewing the involved stakeholders we then conceptualized design requirements (DRs). Within the next phase (*Suggestion*), we derive design principles (DPs) from the proposed DRs by applying SDT (Deci and Ryan 2012) and the IS delegation framework (Baird and Maruping 2021) as justificatory knowledge (Gregor and Jones 2007). In alignment with theories on design principles (Chandra et al. 2015; Gregor et al. 2020), our DPs were constituted by material properties, boundary conditions, and rationales. Then we concretized the abstracted design principles into an abstracted architecture of T^2C^2 and concretized two cases of different GenAI-enabled T^2C^2 (*Development*). We instantiate one architecture as a prototype for generating FAQs and translating FAQs into chatbot training data by utilizing GenAI. Additionally, we present a low-fidelity prototype that integrates T^2C^2 into an existing ticket recommender system and incorporates GenAI, which augments the agent's problem-solving activities. As our research team meets regularly every 6-8 weeks, we demonstrated the prototypes to our pilot partner, a small and medium-sized enterprise (SME) IT service provider within a focus group session and presented our prototypes (*Demonstration*). Furthermore, we showed the derived FAQs and the trained chatbot to two chatbot developers from the pilot partner. Afterward (*Evaluation*), we conducted a user evaluation through semi-structured interviews with four external experts as well as employees from the pilot partners as a summative and naturalistic evaluation (Venable et al. 2016). Furthermore, we demonstrated the prototype to five chatbot experts, conducted thinking-aloud interviews, and evaluated the effectiveness of generated training data in terms of intent accuracy and similarity. The results aim to reveal training data generation efficiency, engagement in chatbot maintenance, and evolving design principles throughout development, demonstration, and evaluation. In the conclusion, we discuss findings and future research directions.

5 Textual Training Data Creation and Curation System (T^2C^2)

5.1 Awareness of problem

In sum, we conducted 15 interviews (Table 1) with employees from a chatbot project team of an SME and external experts in the field of conversational AI to understand the relevant challenges in maintaining chatbot data from different perspectives (including SME and large companies). We asked interviewees to elaborate on the chatbot data maintenance process, any related challenges, the potential for optimization, and the role of employees in maintaining chatbots and analyzed the qualitative content systematically (Mayring 2004). Currently, chatbot development is mainly directed by in-house developers who, in conjunction with support agents, manually update a substantial spreadsheet with intent-solution matches (I1.1, I2.1, I2.2). Unfortunately, this method has led to diminishing engagement, declining NLU performance, and data quality issues (I4.1, I7, Freire et al. 2024). Despite advancements in GenAI, experts predict that intent-based chatbots will continue to play a significant role, offering resource-efficient and highly compliant responses for specific use cases (Luo et al. 2022). Experts propose a hybrid approach involving NLU for closed-domain scenarios and LLMs for open-domain and context-dependent cases (I6, I8).

ID	Role	Gender	Company	Related Phase	Min
I1.1	Chatbot Developer	Female	SME	Status Quo	52:17
I1.2	Chatbot Developer	Female	SME	Awareness & Evaluation	71:49
I2.1	Chatbot Developer	Male	SME	Status Quo	39:31
I2.2	Chatbot Developer	Male	SME	Awareness & Evaluation	71:41
I3	Chatbot Developer	Male	SME	Status Quo	80:44
I4.1	Manager IT-Support	Male	SME	Status Quo	34:35
I4.2	Manager IT-Support	Male	SME	Awareness & Evaluation	55:48
I5	Conversational AI Specialist	Male	Large enterprise	Awareness	46:45
I6	Conversational AI Specialist	Male	Conversational AI Provider	Awareness	55:56
I7	Chatbot Developer	Male	Conversational AI Provider	Awareness & Evaluation	54:37
I8	Chatbot Developer	Female	Startup for chatbot training data	Awareness	52:19
I9	Chatbot Developer	Female	Large enterprise	Awareness	26:57
I10	Conversational AI Specialist	Male	Conversational AI Provider	Awareness	39:52
I11	Conversational AI Specialist	Male	Conversational AI Provider	Awareness	26:45
I12	Chatbot Developer	Female	SME	Awareness & Evaluation	59:34

Table 1. Overview of conducted interviews

From a practical as well as theoretical perspective, it is crucial to incorporate these individuals because they possess firsthand information about customer preferences and how they express their queries, which they acquire through their everyday interactions (I5, I8, I9, I10, Feine et al. 2020, Li et al. 2024) (*DR1: Domain expertise*): "*Ideally, you have individuals who are most knowledgeable about the subject, which in this case means having agents who deal with user communication on a daily basis. We possess the best knowledge, potentially even undocumented knowledge, which can still be harnessed effectively."* (I5). DR1 is above motivated by SDT as chatbots are influencing the FSE's workplaces and thus chatbots should be in the hands of the employees themselves. As such, decisions about use cases that should be automated by the chatbot should lie with the FSE who are mainly influenced by the AI to increase autonomy (Rohde et al. 2024) (*DR2: Use case control).* While the current process is flawed by manual efforts and non-interactive spreadsheet use, a to-be process should be augmented to improve efficiency and reduce the load of already overworked support agents through data augmentation (I8, I12) (*DR5: Data augmentation*): "*Otherwise, it's crucial to instruct the AI that the sentences and questions should be extremely simple, encompassing a variety of formats and all possible forms.*" (I8). DR5 ensures that despite the high pressure, FSEs are capable in a sense of competence to maintain chatbot training data. Still, FSE should possess control over the training data generation and should be able to regenerate and adjust the data interactively from a perspective of autonomy and competence (Wang et al. 2023) (*DR3: Content control*): "*In our case, the agents who are within the bot itself have the capability to provide feedback through their frontend interface. For instance, they can do so when suggesting a new response*" (I9). A key requirement for making the data creation process more efficient is the provision and access to solution material and examples from knowledge bases to inform the GenAI unit and support the agent's creation and labeling process (I5, I9, I10) (*DR4: Knowledge access*): "*It would be a significant relief for us if we had a centralized data source for all services, from which relevant responses could be generated, essentially as building blocks. Unfortunately, we do not have this. While the service does have a knowledge database, it is not integrated with our chatbot."* (I9) Since the chatbot answers with predefined solution templates, the system should support the solution creation and documentation through GenAI (*DR6: Solution support*). Again, satisfying competence is key to engaging FSE in training data creation activities in the long term. Because FSE are no experts in training NLU, they require feedback and recommendations for optimizing the data (I8, I11, I12) (*DR7: Feedback*): "*So, we do indeed measure to some extent. Whether such an intent would be truly relevant for the upcoming inquiries, whether we anticipate that there will be more requests related to that specific internet or not. This way, the customer also has a statistical basis to determine its actual relevance.*" (I11) As chatbots act as coworkers (Wang and Yuan 2022) it is required to induce a sense of ownership (*DR8: Ownership*) and integrate the data creation and curation into existing routines and processes (I9) (*DR9: Routines*) to satisfy the need for relatedness.

5.2 Suggestion

Given the literature on chatbot development and the derived design requirements of practical relevance, we further integrated SDT and IS delegation framework as kernel theories to conceptualize design principles for bot delegation systems (Gregor and Jones 2007). In plain language, a bot delegation system comprises all activities to manage and maintain the data of a chatbot while combining human and AI agencies along the maintenance process. The bot delegation system defines the rules a chatbot uses to answer questions and have conversations with users. Our bot delegation system, T²C², combines all three design principles and augments the maintenance of chatbot data. Our overall T²C² system is being instantiated within two cases given the underlying task and situation – here the underlying data quality (Reinhard et al. 2023a). Case 1 shows how a bot delegation system based on curated knowledge repositories (e.g., product documentation) can look like, while case 2 demonstrates a system that leverages low data quality and new cases on the job. Our conceptualization of design principles places a strong emphasis on defining the activities that T²C² should facilitate in both scenarios (Figure 3). This is intended to boost employee motivation for chatbot maintenance and to ensure a cost-efficient allocation of tasks between humans and AI, aligning with the IS delegation framework (Baird and Maruping 2021).

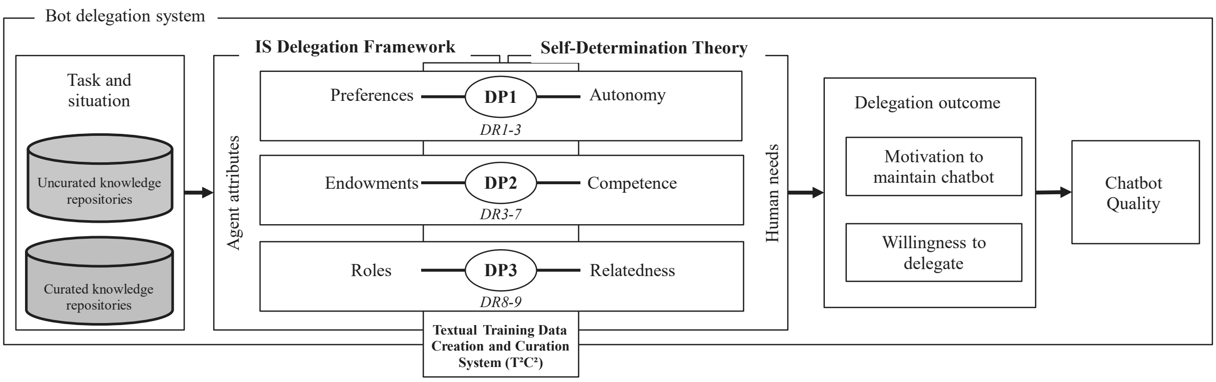

Figure 3. Conceptualization of design principles for a bot delegation system

DP1 – Principle of preferences and autonomy: According to Baird and Maruping (2021), human agents' *preferences* refer to their motivations for pursuing specific goals and shifting their focus between previously mentioned pairs of actors or entities (Bandura 2006; Enfield and Kockelman 2017; Schanze 1987; Shapiro 2005). As a result, the preferences that motivate agents to enter delegation dyads - whether human or artificial - are the determining factors in how and why such interactions occur (Baird and Maruping 2021). Taking into account FSEs' preferences is part of ensuring autonomy (Deci and Ryan 2012). *Autonomy* refers to providing employees with a choice and control of their behavior according to their decision models and goals. A sense of control is becoming more important in times of AI infusion (Calisto et al. 2021; Dietvorst et al. 2018) as AI is taking over (sub-)tasks and augmenting employees. Two levels of autonomy can be distinguished concerning chatbot training data maintenance. Integrating support agents as domain experts in the selection of automation use cases for scaling the chatbot during post-implementation refers to the first level of autonomy (DR1, DR2). T²C² should enable FSE to delegate mundane and repetitive tasks to an AI agent (Vassilakopoulou et al. 2023). Another level of autonomy involves data control and chatbot training data. New automation-appropriate requests should be curated by domain experts (DR3). *T²C² should provide FSE with a choice of automation use cases and provide employees control on data curation and creation given the underlying data and their preferences to satisfy the need for autonomy.*

DP2 – Principle of endowments and competence: According to the interviews, creating training data for chatbots requires not only effort but also skills to efficiently train NLU. Current AI has strengths in mechanical and analytical intelligence (Huang and Rust 2022) given the availability and access to highly structured data. As such, the infusion of GenAI and large language models in T²C² should be aligned with human and AI agent *endowments*. Endowments are either resources (e.g., knowledge) or capabilities (e.g. thinking abilities). Skills and knowledge are attributed to the relationship between human agents and artificial agents (Schmitt et al. 2023). For example, an artificial agent could have the ability to access and capability to analyze data, whereas the human agent provides awareness of its resulting meaningfulness and usefulness. On the other hand, both agents require enough differences that justify any costs of the delegation from human to artificial agents (Baird and Maruping 2021). Therefore, the AI agent has to provide sufficient benefits to the human agents (e.g., time-savings, inspiration). SDT emphasizes the human need for competence, underlining the importance of feeling capable and assurance of effectively interacting with the environment and achieving desired outcomes. To instill confidence in employees tasked with creating and curating chatbot training data, it is essential to enhance the process's efficiency through GenAI and simplified data editing and regeneration options, ensuring control (DR2, DR3). GenAI should play a role in enhancing data, reducing labeling efforts (DR5), and aiding in solution creation and documentation tasks (DR6), which include handling large volumes of text and extracting usable data, such as FAQs. Furthermore, existing studies mentioned the challenge of information overload and overreliance by novices (Seeber et al. 2020), while experts might possess algorithm aversion (Calisto et al. 2022). Thus, the system should accommodate varying employee skill levels. In the case of T²C², a differentiation of chatbot developers and FSE should be considered to maximize engagement. *T²C² system should ensure a shared fit of resources and capabilities by augmenting the data generation process given the underlying data, enabling FSE to interact with GenAI, and providing feedback given human and artificial agents' endowments to satisfy the need for competence.*

DP3 – Principle of Roles and Relatedness: *Roles* are constituted by a set of rights and responsibilities that agents either have or transfer to one another (Baird and Maruping 2021). In hybrid intelligence systems, human employees take over new roles and tasks such as annotators. To ensure a secure, reliable, and at the same time continuously improving AI system, employees take on new responsibilities, such as *modifying* (input and model), *supervising* (e.g. intervening), *sustaining* (e.g. ensuring proper and safe function), *explaining* (the outcomes) or *training the* AI systems (e.g., labeling, retraining) (Burton et al. 2020; Dellermann et al. 2021; Volkmar et al. 2022). The roles and responsibilities should support the need for relatedness. *Relatedness* refers to the need to build a relationship with others and with information systems such as AI agents (Deci and Ryan 2012; Poser et al. 2022). Thus, this principle encompasses configurations that fulfill the need for relatedness, both with the augmentation tool and the chatbot. Therefore, the design should promote a sense of ownership (DR8), with the tool seamlessly integrated into support agents' workflows and routines. To further enhance relatedness, support agents should be offered dashboards showcasing their contributions and be involved in testing their proposed use cases. The third design principle should enable human agents to delegate rights and responsibilities to the GenAI unit, while still retaining certain responsibilities in supervising or modifying the data. *T²C² should be integrated into work processes and provide users a high degree of interactivity by considering the underlying data as well as new roles and responsibilities to satisfy the need for relatedness.*

5.3 Development

We translated the three design principles into a generalized T²C² architecture that can be applied to multiple data sources and domains and allows for integrating several discrete functionalities. Given the data and the characteristics of one of our pilot partners, we instantiated two types of T²C² systems. The first case concerns the creation of FAQs and corresponding chatbot training data based on well-documented product documentation. Therefore, case one shows a higher degree of AI agency and involves FSE as data curators. The second case is integrated into the problem-solving routines of FSE that resolve new requests and create data from available ticket documentation. The latter involves FSE as data creators.

5.3.1 T²C² architecture

We designed a T²C² architecture as an overall infrastructure for generating chatbot training data based on the approach of service-oriented architecture (SOA). Thus, the architecture consists of several distinct units that provide certain functionality (Figure 4). The approach ensures that the creation of data can be augmented for different constellations along the post-implementation phase (see case 1, case 2). Therefore, our architecture can rely on multiple data sources and can be utilized for transferring existing knowledge bases into chatbot data as well as supporting employees to create new data on the job. First, the architecture integrates multiple data sources such as support tickets, product documentation, or knowledge base articles and ensures adequate data preprocessing for each type. Given the preference of input knowledge, the data must be limited to the size of one intent (e.g., FAQ, incident, service request). This extraction of results can be performed automatically through GenAI or manually through search for example. Afterward, the selected results inform the GenAI unit that generates an intent title, multiple example utterances, and a response utilizing state-of-the-art LLMs. The core of the GenAI unit refers to prompt engineering. According to SOA, multiple prompt patterns are provided to efficiently generate data depending on the data type and the source of data. To decide whether a chatbot case might already be part of the initial training set, a similarity check helps to detect equal intents and allows curating existing intents. The architecture ensures access to the set of intents and the underlying training data. After creating or curating the data the NLU test unit allows to build and test an NLU classifier given the new data. The NLU test unit provides the user with feedback services and recommendations to improve the data. Finally, the data is transferred to the chatbot development team, which is responsible for deploying the data and updating the chatbot.

5.3.2 Case 1: High AI agency vs. low human agency (FAQ bot)

In *case 1*, FSEs take over the role and responsibilities of curators by reviewing, testing, and adjusting training data that is initially generated by a GenAI (DP3). The starting point are well-structured and digital product documentation that are regularly published with each new product release. Given the characteristics of the data and the clarity of questions that are incorporated within the data, FSEs welcome automating possible questions with a self-service chatbot that rises with each new product release cycle. That in turn implies lowering the degree of FSE autonomy in selecting chatbot use cases (DP1). Much human effort was already put into writing and updating the product documentation. Hence, FSEs are incentivized to delegate data creation tasks to the AI agent. Furthermore, a higher degree of AI agency is suitable to largely automate the knowledge extraction and data creation phase (DP2). The delegation between human and AI agents in developing chatbot data is highlighted in Figure 4. Given the reduced efforts of data creation implied by the high AI agency, we expect that FSEs are motivated to curate the data and test the chatbot accordingly.

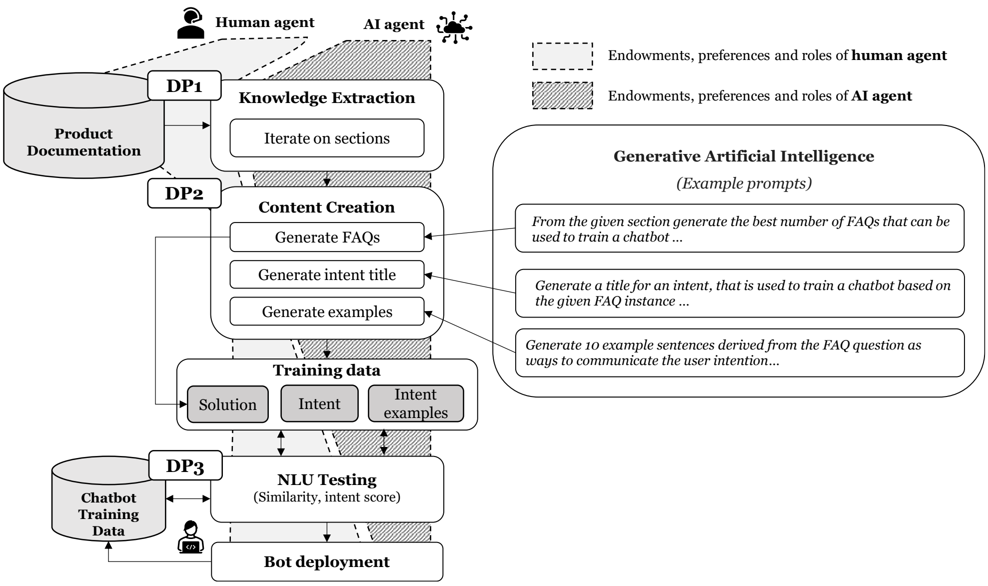

Figure 4. T²C² with high AI agency and low agency based on product documentation

Therefore, we transferred the PDF files provided by the pilot partner into a string format and preprocessed the data accordingly. Given the sections of the documents, we utilized GenAI to develop possible FAQs (i.e., question and solution) for each section. The GenAI unit is based on a recent GPT4.0 model that can be accessed via an API. Then the system iterates on each instance of the FAQ and applies question and answer as input for the GenAI unit to generate an intent title and intent examples. As the generated FAQs include questions and answers, the answers can be used as solutions. After iterating on all FAQs, similarity scores are calculated to identify equal or similar intents. The scores support the employee's competence in selecting and merging intents. Finally, the data is being used to train a test chatbot. The instantiation includes FSE as curators by reviewing the FAQs and the training data and thereby satisfies the need for autonomy and relatedness. The process and prototype of the FAQ bot are presented in Figure 5.

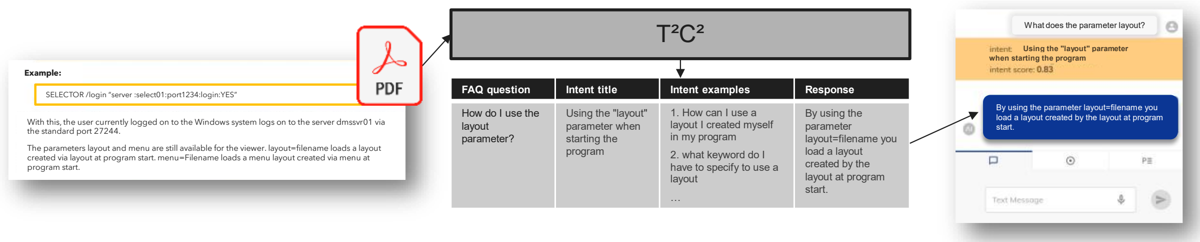

Figure 5. Process and prototype for T²C² with high AI agency

5.3.3 Case 2: High human agency vs. low AI agency (ticket bot)

In *case 2*, FSEs take over the role of data creators due to the novelty and complexity of customer cases and the lack of curated data. Many IT support organizations possess a large repository of solved tickets and are creating new tickets continuously. However, the data quality of such tickets and subsequent automation potential as shown in *case 1* remains limited. While the FAQ-based development approach ensures that simple questions can be answered, other requests and incidents that are not covered within the documentation are neglected. In that case, the chatbot forwards the request to a human FSE (Keyser et al. 2019). Thus, in case 2 we facilitate the FSE's willingness to delegate a new and more complex request to a chatbot. With the motivation to maintain the bot, we refer to the FSE's motivation to create training data and curate data provided by the GenAI. Meanwhile, the FSE should be empowered to delegate certain data creation steps to the AI agent.

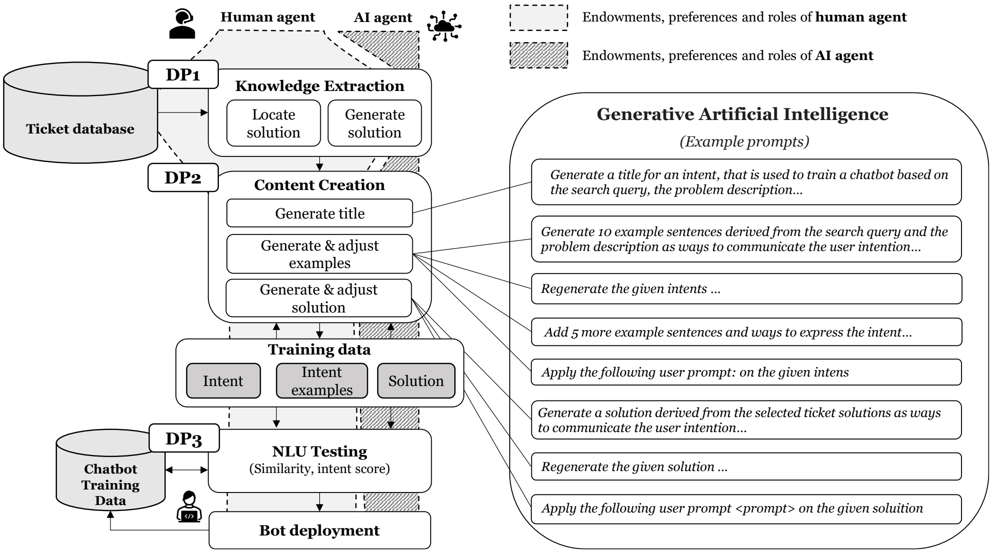

Figure 6. T²C² with high human agency and low AI agency based on support tickets

Thus, we present a system for continuous chatbot training data creation with a high degree of human agency in chatbot training data creation (Figure 6). While solving the request, the FSE searches for solutions in ticket recommendation systems or other data sources or generates a new solution (Das 2003) and decides to automate the given case. If the agent identifies a suitable solution or generates a new one, those can be selected and used to inform the prompts within the GenAI component. The generated data can be adjusted by regenerating, adding intent examples, or applying individual prompts to modify the data. Similar intents will be presented during data creation to prevent inconsistencies within the chatbot training data. The quality of the created data can be tested through an intent score by the NLU test unit. Finally, the data can be saved and transferred to the development team. In comparison to the prior outlined FAQ-based bot development, *case 2* requires a higher integration of FSEs into the data creation process as highlighted in Figure 6. We argue that for new requests that cannot be solved via simple FAQs domain knowledge is required to not only curate the automation case but to create the data. The process and prototype of the graphical user interface are presented in Figure 7. The green-colored arrows and numbers mark the added steps of data creation integrated into the problem-solving routines.

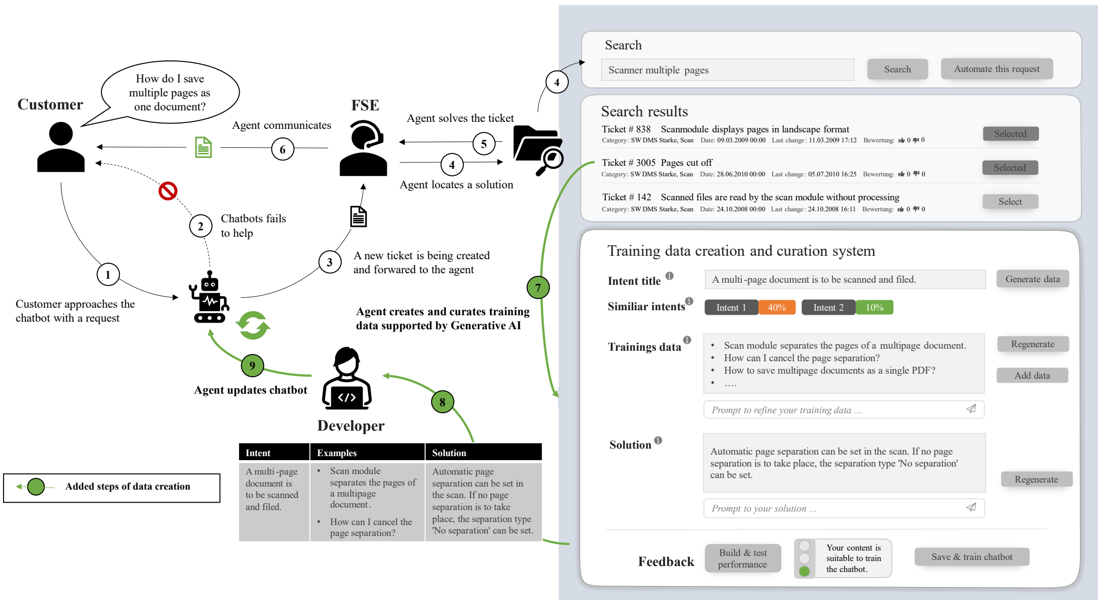

Figure 7. Process and prototype for T²C² with high human agency

5.4 Demonstration and evaluation

Our overarching goal is to improve delegation outcomes to increase subsequent chatbot quality as depicted in Figure 3. We demonstrate and evaluate the two instantiations within a summative evaluation (Venable et al. 2016). Hence, in the following, we describe our two-fold evaluation of the design's impact on the willingness to delegate and the motivation to maintain chatbots given each case.

5.4.1 Case 1: High AI agency vs. low human agency (FAQ bot)

We initiate and evaluate *Case 1* as a fully functional prototype based on provided product documentation within a summative setting (Venable et al. 2016). In that context, we focus on the willingness to delegate the FAQs to a chatbot and content creation to an AI agent as well as on the motivation to curate the proposed data. We presented the prototype to five expert chatbot developers (I1.2, I2.2, I4.2, I7, I12) and discussed our results. Both developers stressed the effectiveness and efficiency of GenAI as a tool to extract knowledge from well-documented and structured knowledge repositories. Developing and curating a set of FAQs and translating the data into processible training data is associated with high effort for both FSE and chatbot developers. Hence, the developers confirmed their willingness to delegate the data creation to the GenAI and emphasized the importance and motivation of curating the generated training data. FSE and developers should act as a quality gate before incorporating the data (I1.2). Thereby the design not only reduces manual effort and increases the motivation of employees but especially leverages the chatbot due to the wide coverage of questions. The generated chatbot data will be further utilized to maintain the current chatbot data by the FSE to ensure relatedness. Above the thinking-aloud interviews, we evaluated the FAQ bot through performance measures to validate the subsequent quality of the chatbot. The main objective is improving and maintaining the NLU. Therefore, we show that GenAI can support the creation of training data both quantitatively and qualitatively. Quantitatively, we refer to efficiently generating a high number of intent sentence samples that enable effective training of an NLU. Out of 33 pages of product documentation PDFs, we generated 192 intents and correspondingly 1960 intent example sentences to train a FAQ bot. Despite the large number of intents and without further preprocessing, we achieved a comparatively high intent accuracy of 0.85 by training a conversational AI NLU. Filtering 141 intents of low quality and similarities with other intents thus resulted only in a small increase to an intent accuracy of 0.91. The results suggest a high initial quality of generated intent example sentences to train an NLU. An analysis of the cosine similarity of FAQs revealed that the generated FAQs differ greatly (average similarity = 0.29), suggesting that the intents incorporate exclusive topics. Two chatbot developers (I1.2, I12) tested the bot and confirmed that the chatbot produced reasonable questions that "could have been asked by novice customers" (I12). Additionally, we calculated the pairwise cosine similarity of the concatenated intent examples. The results show that only 46 pairs have a high similarity of greater than 0.9. Overall, the evaluation results suggest that the architecture produces reasonable FAQs and valuable training data and thus ensures both human and AI competence. FSEs can hence reliably delegate FAQ generation towards the AI agent and are motivated to act as a curator given the initial quality of the created training data.

5.4.2 Case 2: High human agency vs. low AI agency (ticket bot)

In *case 2* analysis, we evaluated the tool's effectiveness in promoting user willingness to assign complex tasks and data creation to AI. Additionally, we considered the motivation to add new data continuously on the job. As proposed by Meth et al. (2015), we conducted a focus group session as part of the demonstration phase. The focus group includes 20 participants from three different service organizations including support agents, managers, developers, and work council members. Within the session which lasted about 1.5 hours, the prototype was presented and then discussed regarding the delegation outcomes outlined in chapter 5.2. Additionally, the groups evaluated the architecture and prototype concerning user friendliness and self-efficacy. Overall, the participants confirmed the practical relevance of such systems as current documentation of knowledge base articles, templates, and FAQs require a lot of time. Correspondingly all pilot partners emphasized the applicability of their automation endeavours and service processes and the willingness to delegate parts of the data creation to AI agents as a form of augmentation. The idea of hybrid and GenAI-enabled data creation was further transferred to the pressing issue of data quality. The design principles hence could improve the documentation of knowledge in the long-term and are expected to increase the motivation to maintain bots. In addition to the focus group session, we evaluated our architecture by conducting semi-structured interviews with five chatbot experts as indicated in Table 1 to assess the impact of our design on delegation outcomes. The interview partners were presented with the concept and the architecture incorporating the design principles. In comparison to the demonstration phase, the experts were able to provide more detailed feedback regarding the functionalities and less regarding visual elements. Especially, they took over the perspective of developers who are the recipients of the generated data as they finally deploy the intents proposed by the human and AI agents. Covering several data sources is important as most organizations provide chatbot developers with different data formats and data quality (I4.1, I7). To ensure traceability it is crucial to include certain meta-data such as author, ticket number, reference to documentation, or creation date (I1.2). Additionally, they confirm the need for a distinction between new and old intents (I1.2, I7). Typically, the sets of intents are available and need to be mapped with new intents (I7). If employees want to adjust existing intents, they should be provided with the existing training data and use the interface with the initialized data (I1.2). Otherwise, FSE would not be able and willing to delegate. Multiple interviewees stated that a dashboard for developers where all open chatbot cases are listed and accessible and where chatbot developers themselves can utilize the system to fine-tune the data would be important (I1.2, I2.2, I7). That in turn fosters the willingness to delegate new tickets to the chatbot. Additionally, it was confirmed that the design provided FSE with control: "*Having direct guidance on hand is quite beneficial. This way, you can immediately say, "No, that doesn't work," and make further adjustments, even with a prompt."* (I12). Thereby, the design ensures motivation to curate the generated data and maintain the data in an interactive manner. Despite the potential of the proposed design, experts claim, that additional quality assurance is required (I1.2): "*And my concern would be that what comes out down there, I mean, I would place a person in front of this NLU test unit again, just to check if it makes sense in terms of content."*.

6 Discussion, Limitations, and Future Research

This paper presents insights into how chatbot data maintenance can be improved by a novel approach to integrating GenAI and FSE. We build a textual training data creation and curation system (T^2C^2) instantiated as two cases and evaluate the results in a summative approach concerning delegation outcomes. Our results show that our DPs facilitate the ability to train a chatbot and thus increase the willingness to delegate. Based on our design, FSE are more motivated to maintain the chatbot. Our research is distinctive with regards to combining both the IS delegation framework as well as SDT as the underlying kernel theory. Thus, we provide a nascent design theory and action according to Gregor and Hevner (2013). This represents a DSR contribution type of improvement as it is a new solution to a known problem. Furthermore, we apply the IS delegation framework to the specific context of GenAI-powered IT support and show its adequacy for this future-relevant field. The underlying agency concept of the IS delegation framework is used to realize two instantiations of the T^2C^2 system. One instantiation is realized with a high(er) AI agency, and one instantiation is realized with a high(er) human agency given the underlying data quality and the AI and human attributes. Thereby, not only the validity of the DSR artifact as described above is increased (as instantiations from different parts of the agency spectrum are covered), but additionally, the adequacy of the IS delegation framework for the important context of GenAI-powered information systems is realized. Moreover, we apply SDT in the context of GenAI-powered IT support and demonstrate how its underlying principles impact the design of information systems in this context. Within both instantiations (high human agency and high AI agency) of our designed T^2C^2, we use the three basic principles of autonomy, competence, and relatedness of SDT to investigate how these affect the motivation to maintain chatbot data. Our results suggest that considering both IS delegation and self-determination ensures motivation to maintain chatbots and willingness to delegate GenAI agents.

Our research comes with certain limitations. First, it is based on a single pilot partner, and two specific instantiations, and lacks the integration of FSE's perspectives. Expanding the demonstration of design principles and architecture to various service departments and knowledge repositories could yield more generalized insights, design adjustments, and impacts on delegation outcomes. Second, while our demonstration and evaluation serve as proof of concept, further research is needed to establish proof of value. Prototypes should be integrated into pilot partners' workplaces and undergo a long-term evaluation to observe FSE's experiences and intentions regarding data delegation and creation on the job. A quantitative study comparing the proposed design to the baseline spreadsheet model would enhance the findings. Additionally, we only conducted a summative evaluation, hence future research has to evaluate the individual design principles in terms of autonomy, competence, and relatedness. Third, our DSR project did not explore different delegation mechanisms within the data creation and curation process. Future research should investigate human interaction with GenAI for chatbot data generation and knowledge repository enhancement. Additional design cycles are planned to incorporate evaluation feedback, create a dedicated FAQ database, and conduct further assessments.

From a DSR perspective, we contribute to the emerging IS paradigm of systems that augment employees to delegate tasks and functions to autonomous bots (Baird and Maruping 2021). We introduce a novel class of bot delegation systems that help users maintain chatbot data efficiently. With the advent of GenAI and large language models, we expect that bot delegation systems will continue to become increasingly relevant. Furthermore, we offer a novel perspective by synthesizing a kernel theory, which amalgamates the IS Delegation Framework and SDT. This synergy between two fundamental theories offers a deeper and more comprehensive insight into the inner workings of bot delegation systems and contributes to the IS design theory (Niehaves and Ortbach 2016) by demonstrating the interplay of two kernel theories within the (latent) inner model of an explanatory IS design theory (Baskerville and Pries-Heje 2010). Finally, our contribution involves the refinement of design principles for bot delegation systems and their structural framework, drawing upon the works of Chandra et al. (2015) and Gregor et al. (2020). We achieve this through the creation of an adapted knowledge representation. This representation aligns the material agency attributes of bot delegation systems with either human agents or the agentic attributes of the IS artifact, as proposed by Baird and Maruping (2021). The ultimate goal of this alignment is to address the crucial requirements of bot delegation systems, in line with the needs of self-determination, as elucidated by Deci and Ryan (2012). Our proposed structure for design principles for bot delegation systems can be broken down into several key components: material properties, boundary conditions, and rationales. Material properties play a crucial role in defining the functionalities of the system within specified boundary conditions, which encompass attributes of both humans and AI agents, including preferences, resources, and roles. These boundary conditions also involve fundamental considerations related to the quality of underlying data. Additionally, the principles guiding the system are grounded in core rationales, which are deeply rooted in fundamental human needs according to SDT. In practice, these insights have relevance for various IT support stakeholders, including IT support managers, support agents, business developers, and technology officers. On a practical level, our work demonstrates how new technologies can be leveraged in a human-centered, employee-empowering manner. This approach, where employees (in this case, support agents) determine how and where GenAI is integrated to enhance their daily work, is particularly important in a context where there is increased concern about job displacement due to these technologies (Huang and Rust 2018). By actively involving employees in shaping the use of such technologies in service triads of customers, employees, and GenAI, we move from passive acceptance to proactive influence on their consequences.

References

Baird, A., and Maruping, L. M. 2021. "The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts," *MIS Quarterly* (45:1), pp. 315-341.

Ballou, N., Deterding, S., Tyack, A., Mekler, E. D., Calvo, R. A., Peters, D., Villalobos-Zúñiga, G., and Turkay, S. 2022. "Self-Determination Theory in HCI: Shaping a Research Agenda," in *CHI*

Conference on Human Factors in Computing Systems Extended Abstracts, New York, NY, USA, ACM (doi: 10.1145/3491101.3503702).

Baskerville, R., and Pries-Heje, J. 2010. “Explanatory Design Theory,” *Business & Information Systems Engineering* (2:5), pp. 271-282.

Bayer, M., Kaufhold, M.-A., and Reuter, C. 2022. “A survey on data augmentation for text classification,” *ACM Computing Surveys* (55:7), pp. 1-39.

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 *International Conference on Information Systems 2021*.

Burton, J. W., Stein, M.-K., and Jensen, T. B. 2020. “A systematic review of algorithm aversion in augmented decision making,” *Journal of Behavioral Decision Making* (33:2), pp. 220-239.

Calisto, F. M., Nunes, N. J., and Nascimento, J. C. 2022. “Modeling Adoption of Intelligent Agents in Medical Imaging,” *SSRN Electronic Journal*.

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 *2015 48th hawaii international conference on system sciences*, pp. 4039-4048.

Chaves, A. P., and Gerosa, M. A. 2021. “How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design,” *International Journal of Human–Computer Interaction* (37:8), pp. 729-758.

Das, A. 2003. “Knowledge and Productivity in Technical Support Work,” *Management Science* (49:4), pp. 416-431.

Deci, E. L., and Ryan, R. M. 2012. “Self-Determination Theory,” in *Handbook of Theories of Social Psychology: Volume 1*, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, pp. 416-437 (doi: 10.4135/9781446249215.n21).

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., and Ebel, P. 2021. *The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems*.

Feine, J., Morana, S., and Maedche, A. (eds.). 2020. *Designing Interactive Chatbot Development Systems*.

Følstad, A., and Brandtzæg, P. B. 2017. “Chatbots and the new world of HCI,” *Interactions* (24:4), pp. 38-42.

Freire, S. K., Wang, C., and Niforatos, E. 2024. “Chatbots in Knowledge-Intensive Contexts: Comparing Intent and LLM-Based Systems,” *arXiv preprint arXiv:2402.04955 [Titel anhand dieser ArXiv-ID in Citavi-Projekt übernehmen]*.

Gao, M., Liu, X., Xu, A., and Akkiraju, R. 2021. “Chatbot or Chat-Blocker: Predicting Chatbot Popularity before Deployment,” in *Designing Interactive Systems Conference 2021*, New York, NY, USA, New York, NY, USA: ACM.

Gnewuch, U., Stefan Morana, and Alexander Maedche (eds.). 2017. *Towards Designing Cooperative and Social Conversational Agents for Customer Service*.

Gregor, S., Chandra Kruse, L., Seidel, S., and others. 2020. “Research perspectives: the anatomy of a design principle,”

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., and Jones, D. 2007. “The Anatomy of a Design Theory,” *1536-9323*.

Grønsund, T., and Aanestad, M. 2020. “Augmenting the algorithm: Emerging human-in-the-loop work configurations,” *The Journal of Strategic Information Systems* (29:2), p. 101614.

Haupt, M., Rozumowski, A., Freidank, J., and Haas, A. 2023. “Seeking empathy or suggesting a solution? Effects of chatbot messages on service failure recovery,” *Electronic Markets* (33:1), p. 56.

Huang, M.-H., and Rust, R. T. 2018. “Artificial intelligence in service,” *Journal of service research* (21:2), pp. 155-172.

Huang, M.-H., and Rust, R. T. 2022. “A Framework for Collaborative Artificial Intelligence in Marketing,” *Journal of Retailing* (98:2), pp. 209-223.

Janssen, A., Cardona, D. R., Passlick, J., and Breitner, M. H. 2022. "How to Make chatbots productive–A user-oriented implementation framework," *International Journal of Human-Computer Studies* (168), p. 102921.
Jiménez-Barreto, J., Rubio, N., and Molinillo, S. 2021. ""Find a flight for me, Oscar!" Motivational customer experiences with chatbots," *International Journal of Contemporary Hospitality Management* (33:11), pp. 3860-3882.
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.
Khan, R. 2017. "Standardized Architecture for Conversational Agents a.k.a. ChatBots," *International Journal of Computer Trends and Technology* (50:2), pp. 114-121.
Korn Ferry Institute. 2018. "The $8.5 Trillion Talent Shortage," available at https://www.kornferry.com/content/dam/kornferry/docs/article-migration/FOWTalentCrunchFinal_Spring2018.pdf.
Kuechler, W., and Vaishnavi, V. 2008. "On theory development in design science research: anatomy of a research project," *European Journal of Information Systems*.
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.
Li, M. M., Peters, C., Poser, M., Eilers, K., and Elshan, E. 2022. "ICT-enabled job crafting: How Business Unit Developers use Low-code Development Platforms to craft jobs," in *International Conference on Information Systems 2022,* Copenhagen, Denmark. 9-14th December.
Li, M. M., Reinhard, P., Peters, C., Oeste-Reiss, S., and Leimeister, J. M. 2024. "A Value Co-Creation Perspective on Data Labeling in Hybrid Intelligence Systems: A Design Study," *Information Systems* (120), p. 102311.
Lohrenz, L., Michalke, S., Robra-Bissantz, S., and Lattemann, C. 2022. "Mechanisms to foster Self-Determination on Engagement Platforms-An Online Experiment," in *HICSS*.
Luo, B., Lau, R. Y. K., Li, C., and Si, Y.-W. 2022. "A critical review of state-of-the-art chatbot designs and applications," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (12:1), e1434.
Mayring, P. 2004. "Qualitative content analysis," *A companion to qualitative research* (1:2), pp. 159-176.
Meth, H., undefined, 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. 2022. *Chatbot Introduction and Operation in Enterprises – A Design Science Research-based Structured Procedure Model for Chatbot Projects*.
Nguyen, Q. N., Sidorova, A., and Torres, R. 2022. "User interactions with chatbot interfaces vs. Menu-based interfaces: An empirical study," *Computers in Human Behavior* (128), p. 107093.
Niehaves, B., and Ortbach, K. 2016. "The inner and the outer model in explanatory design theory: the case of designing electronic feedback systems," *European Journal of Information Systems* (25), pp. 303-316.
Nißen, M., Selimi, D., Janssen, A., Cardona, D. R., Breitner, M. H., Kowatsch, T., and Wangenheim, F. von. 2022. "See you soon again, chatbot? A design taxonomy to characterize user-chatbot relationships with different time horizons," *Computers in Human Behavior* (127), p. 107043.
Nowak, S., and Rüger, S. 2010. "How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation," in *Proceedings of the international conference on Multimedia information retrieval*, pp. 557-566.
Ouyang, L., Wu, J., Xu, J., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. 2022. "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems* (35), pp. 27730-27744.

Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A. H., and Riedel, S. 2019. *Language Models as Knowledge Bases?*

Poser, M., Wiethof, C., Banerjee, D., Shankar Subramanian, V., Paucar, R., and Bittner, E. A. C. 2022. "Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service," Springer, Cham, pp. 142-153 (doi: 10.1007/978-3-031-06516-3_11).

Reinhard, P., Li, M., Dickhaut, E., Peters, C., and Leimeister, J. M. 2023a. "Empowering Recommender Systems in ITSM: A Pipeline Reference Model for AI-based Textual Data Quality Enrichment," in *International Conference on Design Science Research (DESRIST),* Pretoria, South Africa. 31 May - 02 Jun 2023, Pretoria, South Africa.

Reinhard, P., Li, M., Dickhaut, E., Reh, C., Peters, C., and Leimeister, J. M. 2023b. "A Conceptual Model for Labeling in Reinforcement Learning Systems: A Value Co-Creation Perspective," in *International Conference on Design Science Research (DESRIST),* Pretoria, South Africa. 31 May - 02 Jun 2023, Pretoria, South Africa.

Reinhard, P., Li, M. M., Peters, C., and Leimeister, J. M. 2024. "Generative AI in Customer Support Services: A Framework for Augmenting the Routines of Frontline Service Employees," in *HICSS 2024,* Waikiki, Hawaii, USA.

Rohde, F., Wagner, J., Meyer, A., Reinhard, P., Voss, M., Petschow, U., and Mollen, A. 2024. "Broadening the perspective for sustainable artificial intelligence: sustainability criteria and indicators for Artificial Intelligence systems," *Current Opinion in Environmental Sustainability* (66), p. 101411.

Schloß, D., Gnewuch, U., and Maedche, A. 2022. "Towards Designing a Conversation Mining System for Customer Service Chatbots," in *International Conference on Information Systems 2022,* Copenhagen, Denmark. 9-14th December.

Schmitt, A., Zierau, N., Janson, A., and Leimeister, J. M. 2023. "The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution," *Journal of the Association for Information Systems (JAIS)*.

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), p. 103174.

Statista. 2024. "Global market volume of chatbots in 2022 and forecast until 2032," available at https://de.statista.com/statistik/daten/studie/1373729/umfrage/weltweites-marktvolumen-chatbots/.

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.

Volkmar, G., Fischer, P. M., and Reinecke, S. 2022. "Artificial Intelligence and Machine Learning: Exploring drivers, barriers, and future developments in marketing management," *Journal of Business Research* (149), pp. 599-614.

Vreede, T. de, Raghavan, M., and Vreede, G.-J. de. 2021. "Design Foundations for AI Assisted Decision Making: A Self Determination Theory Approach," in *HICSS 2021* (doi: 10.24251/hicss.2021.019).

Wang, W., Chen, L., Xiong, M., and Wang, Y. 2023. "Accelerating AI adoption with responsible AI signals and employee engagement mechanisms in health care," *Information Systems Frontiers* (25:6), pp. 2239-2256.

Wang, W., and Yuan, Y. 2022. "Employees' Attitude towards a Digital Teammate-Will AI-enabled Chatbot Lead to Enhancing Employees' Job Identity?" in *International Conference on Information Systems 2022,* Copenhagen, Denmark. 9-14th December.

Xiao, Z., Zhou, M. X., and Fu, W.-T. 2019. "Who should be my teammates," in *Proceedings of the 24th International Conference on Intelligent User Interfaces,* New York, NY, USA, ACM (doi: 10.1145/3301275.3302264).

Yu, S., Xiong, J., and Shen, H. 2024. "The rise of chatbots: The effect of using chatbot agents on consumers' responses to request rejection," *Journal of Consumer Psychology* (34:1), pp. 35-48.

Zhang, Y., Zhang, H., Zhan, L.-M., Wu, X.-M., and Lam, A. Y. S. 2022. *New Intent Discovery with Pre-training and Contrastive Learning*.