Information Systems 120 (2024) 102311

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

A Value Co-Creation Perspective on Data Labeling in Hybrid Intelligence Systems: A Design Study

Mahei Manhai Li [a,*], Philipp Reinhard [a], Christoph Peters [a,b], Sarah Oeste-Reiss [a], Jan Marco Leimeister [a,b]

[a] *University of Kassel, Germany*
[b] *University of St.Gallen, Switzerland*

ARTICLE INFO

Keywords:
Hybrid intelligence
artificial intelligence
Value Co-creation
IT Service Management (ITSM)
IT support
machine learning
Human-in-the-loop
Interactive labeling

ABSTRACT

The adoption of innovative technologies confronts IT-Service-Management (ITSM) with an increasing volume and variety of requests. Artificial intelligence (AI) possesses the potential to augment customer service employees. However, the training data for AI systems are annotated by domain experts with little interest in labeling correctly due to their limited perceived value. Ultimately, insufficient labeled data leads to diminishing returns in AI performance. Following a design science research approach, we provide a novel human-in-the-loop (HIL) design for ITSM support ticket recommendations by incorporating a value co-creation perspective. The design incentivizes ITSM agents to provide labels during their everyday ticket-handling procedures. We develop a functional prototype based on 17,120 support tickets provided by a pilot partner as an instantiation and evaluate the design through accuracy metrics and user evaluations. Our evaluation revealed that recommendations after label improvement showed increased user ratings, and users are willing to contribute their domain knowledge. We demonstrate that our design benefits for both human agent and AI systems in the form of hybrid intelligence service systems. Overall, our results emphasize agents' need for value-in-use by providing better results if they improve the labeling of support tickets pre-labeled by AI. Thus, we provide prescriptive knowledge of a novel HIL design that enables efficient and interactive labeling in the context of diverse applications of reinforcement learning systems.

1. Introduction

With new artificial intelligence (AI) technologies and digitalization projects gaining popularity, the IT landscape in businesses has become increasingly more complex and heterogeneous. The IT services market has reached $57 billion in 2021 and is projected to reach $82 billion in 2027 [59]. Thus, IT service management (ITSM) and its frontline support agents face higher customer expectations and a rapidly increasing number of heterogeneous customer requests [35,73–75]. Recent research in frontline service technologies has drawn upon the technological advances made in AI and particularly hybrid intelligence (HI) [16] to augment and empower support agents in their problem-solving activities [35,50,51,76].

HI systems often rely on high-quality annotated data [16] and integrate the human user to leverage their expert domain knowledge into the learning mechanism as a so-called human-in-the-loop (HIL) [13]. Relevant and new data needs to be continuously validated [32] and audited during model initialization and system use [28]. The creation and maintenance of the underlying data for AI systems in terms of labeling the ground truth and incorporating domain experts remain a key challenge for support organizations that want to leverage the potential of AI [3,77,78]. The work arrangements of actors that label tickets and actors that reap the benefit of using systems that rely on those labeled tickets are called HIL configurations [5]. Prior research on HI systems often focused on improving the upfront labeling activities, e.g., through increasing user engagement [67], semi-automating labeling processes [63], or gamifying the annotation tasks [68]. They focused on making labeling more efficient and interactive through a process perspective. Still, prior approaches have not considered the underlying cause of the mentioned data labeling challenge. Despite the importance of internal employees [79], current HIL configurations often do not provide support agents with an outlook on the importance of their work. With support agents being overworked and subject to high turnover rates, they are left with little incentive to do labeling task annotators are either only

[*] Corresponding author.
E-mail address: mahei.li@uni-kassel.de (M.M. Li).

https://doi.org/10.1016/j.is.2023.102311
Received 27 August 2023; Received in revised form 17 October 2023; Accepted 18 October 2023
Available online 30 October 2023

incentivized extrinsically or are rewarded with a delay by receiving better recommendations in the future.

To overcome the challenges revolving around annotated training data and to facilitate the co-creation of value in human-AI interaction via labeling activities, we incorporate a value co-creation perspective and a service-dominant logic [27,64,80] as a theoretical lens. We propose that the lack of suitable incentives refers to a lack of value-in-use agents expected in human-AI interactions for annotating data. Therefore, our research goal is to design an HI system incorporating an interactive HIL-based labeling mechanism to provide immediate value-in-use to support agents in exchange for data labeling. Therefore, we state the following research question (RQ): *How can domain experts be integrated into the data labeling process using a HI system? What are the design characteristics (*i.e.*, design requirements, design principles, and design features) for a human-in-the-loop configuration for data labeling that provides immediate value-in-use for domain experts during use?*

Following the Design Science Research (DSR) process, we propose, implement, and test a novel solution for value co-creation-based labeling. With the configured pipeline and the derived design principle, we aim to emphasize and strengthen humans' role in AI's socially desirable development and application [72]. Therefore, we focus primarily on the interaction between the AI-based system and the end user.

2. Theoretical foundation

With Service-Dominant Logic (SDL) as a theoretical lens on hybrid intelligence, the interaction between humans and AI can be seen as a value co-creation where both actors provide and integrate resources [53, 64]. Human users or annotators contribute domain-specific knowledge and offer feedback, while the prediction model learns patterns, provides recommendations, and augments the human workplace. A key aspect of SDL is the co-creation of value, which emphasizes the collaborative creation of value between actors and entities through the mutually beneficial integration of resources [8,47,56,81]. Following the theories on value co-creation, value recipients – here support agents – are not only consuming a service, but they are also active participants in the value creation process through interaction [64,82] by which the user is made better off in some way [25] – for instance by reducing the time for finding a solution or by receiving high-quality recommendations. The unique nature of SDL is the focus on value-in-use [65], which refers to the individually perceived value when using an AI service instead of only consuming a recommended solution [26,64]. Value creation in interaction appears in a joint sphere of human and machine intelligence (Fig. 1). Thus, following a value co-creation and value-in-use perspective, we derive design knowledge for a value-driven HIL configuration for reinforcement learning-based AI systems and interactive labeling that implies value co-creation at a single point of interaction rather than at different points in time [83].

3. Related work

3.1. Hybrid intelligence in ITSM

Machine learning (ML) in complex environments, such as ITSM, demands high-quality domain-specific user input [49]. Meza Martínez et al. [40] differentiate between two possible ML strategies to overcome the lack of domain-specific knowledge. The first approach suggests that ML practitioners learn from domain experts when developing ML models [49]. However, as users are not involved in the development, there is a lack of user engagement and trust because the systems are considered a "black box" [6,31]. Another approach, which represents the underlying foundation for this research project, is hybrid intelligence [16], sometimes called interactive ML [30,31]. HI systems stress the often overlooked limitations of AI systems (e.g., data quality, availability) and the importance of humans as both value recipients and data validators [16,28]. Furthermore, HI systems rely on the combination of AI and human intelligence and the interaction between AI systems and human collaborators on how to build adaptive systems [4]. Thereby HI induces a hybrid of human and machine intelligence by joining both – humans and machines – in a learning mechanism. This perspective calls for more user-centricity and value on the human site and counteracts the issues of autonomous agents [57], which includes providing explanations to human users [1].

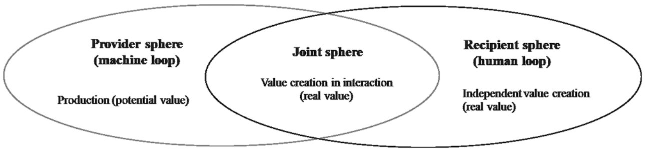


Fig. 1. Value creation spheres according to Grönroos and Voima [27].

3.2. Human-in-the-Loop configurations

A common practice among developers involves employing HIL configurations to ensure that users are directly involved and that the models learn continuously as users provide domain-specific input [6]. Meza Martínez et al. [40] distinguish between supervised, active, and reinforcement learning as HIL mechanisms. This paper focuses only on a ticket recommender system that relies on reinforcement learning [2]. Reinforcement learning is characterized by a self-learning mechanism based on rewards and punishments during use [33]. The HIL learning mechanisms require labeled data. Therefore, labeling is an important part of building high-performance HI systems [43]. Interactive labeling constitutes a field of practice and research in which the role of the human is especially prevalent for providing domain-specific knowledge for training the models. At the same time, other approaches see the role of domain experts in developing the functionalities, for example, in low-code development projects [21]. Most of the prior work on interactive labeling concern the number of high-quality labels as the main objective. Therefore, optimizing labeling processes and motivating human users to contribute labeled data has relevance for practice and within the ITSM literature [12,31,35,43].

4. Research design

This paper follows a design science research (DSR) approach [46] (Fig. 2) and is contextualized in cooperation with three ITSM service providers. The authors regularly meet with 18 stakeholders, including support Fagents, management and work council representatives, and developers in roundtable workshop settings (every 8 weeks) for 3,5 years to review newly developed tools (Appendix A). Based on participant observations and 13 interviews, a recurring pattern of problems that arose during the design of the intelligent frontline support system with all three pilot partners were a) how can we ensure that the system is learning and b) how can we motivate agents to provide labels continuously? The interviewees' information is listed in Table 1. Thus, the research design is a problem-centered initialization [46].

We first motivate the underlying design challenges of HI systems (*Phase 1*) extracted from a review of relevant literature and 13 semi-structured interviews with support agents [42]. Afterward, we derive three design requirements (DRs) as part of defining the objectives of our solution (*Phase 2*). In the phase of design and development (*Phase 3*), we outline the design of the system pipeline and explain how the system initializes its underlying model and works during operations. Then, our research demonstrates how the system integrates a HIL design to address all design requirements and presents its design principle by translating design features (DFs) into a prototypical instantiation (*Phase 4*). The paper concludes with five formative and summative evaluations [66] by following the pattern lifecycle for design patterns evaluation according to Petter et al. [48] (*Phase 5*). In the development phase, we assessed the

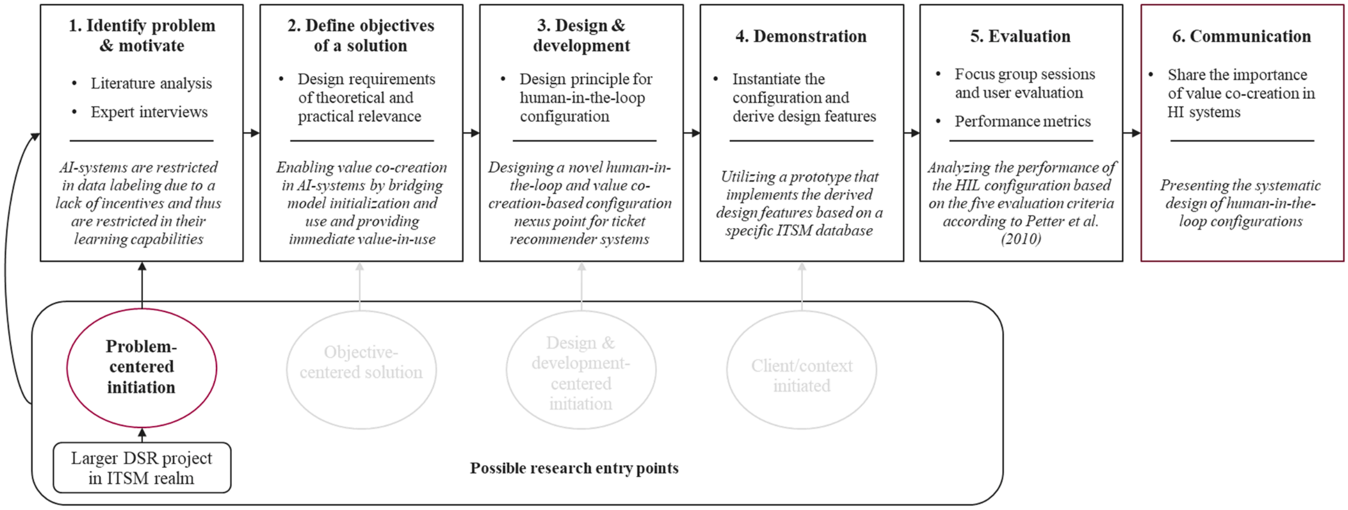

1. Identify problem & motivate
• Literature analysis
• Expert interviews
AI-systems are restricted in data labeling due to a lack of incentives and thus are restricted in their learning capabilities
2. Define objectives of a solution
• Design requirements of theoretical and practical relevance
Enabling value co-creation in AI-systems by bridging model initialization and use and providing immediate value-in-use
3. Design & development
• Design principle for human-in-the-loop configuration
Designing a novel human-in-the-loop and value co-creation-based configuration nexus point for ticket recommender systems
4. Demonstration
• Instantiate the configuration and derive design features
Utilizing a prototype that implements the derived design features based on a specific ITSM database
5. Evaluation
• Focus group sessions and user evaluation
• Performance metrics
Analyzing the performance of the HIL configuration based on the five evaluation criteria according to Petter et al. (2010)
6. Communication
• Share the importance of value co-creation in HI systems
Presenting the systematic design of human-in-the-loop configurations
Problem-centered initiation
Objective-centered solution
Design & development-centered initiation
Client/context initiated
Larger DSR project in ITSM realm
Possible research entry points

Fig. 2. Design Science Research Approach according to Peffers et al. [46].

Table 1
Interview series with support managers and agents.

ID	Role description	Gender	Support type	Company size	Domain	Duration in min
E1	IT support manager	Male	External IT support	Small and medium-sized enterprises (SME)	IT provider	19:20
E2	IT support agent	Male	External IT support	SME	IT provider	29:33
E3	IT support agent	Male	External IT support	SME	IT provider	35:00
E4	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	28:19
E5	IT support agent	Female	Internal IT support	Large enterprise	Healthcare	33:00
E6	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	30:26
E7	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	29:44
E8	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	26:16
E9	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	27:45
E10	IT support agent	Female	External IT support	SME	IT provider	32:13
E11	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	25:24
E12	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	23:48
E13	IT support agent	Male	Internal IT support	Large enterprise	Healthcare	29.23

reliability and plausibility of our design. We conducted a Wizard-of-Oz experiment and then conducted focus groups with domain experts, including agents and managers. Finally, we evaluated the feasibility during the deployment phase by instantiating a prototype.

Additionally, we perform a simulation to show the learning capabilities of a multi-armed bandit prediction model and present the performance of the automated labeling system to ensure predictivity. Finally, during use, we evaluate perceived value-in-use by utilizing domain expert evaluations of relabeling tickets to determine whether suggestions improved and demonstrated effectiveness. The paper concludes with a discussion and outlook on future work by sharing the importance of value co-creation in HI systems (*Phase 6*).

5. Problem identification and objectives

HI systems employ HIL mechanisms to leverage intuition and the real-world knowledge of domain experts to augment work [16]. From a perspective of IS identity, AI systems in workplaces also have to cope with a lack of control [41]. Therefore, service employees should experience more interaction with AI, where they act as supervisors [9] and verify machine outcomes [16,29]. Especially in complex problem-solving tasks like IT support, trust in AI is limited by the restricted performance and reliance on recommender systems and the difficulty of recommending optimal solutions [36,84]. For example, interviewees E1 and E5 state that they *typically need much time to find a certain ticket in the database and thus rely on colleagues instead of using an AI-based system*. To improve the algorithms, humans must audit data and use the HI system to produce more data. Both forms of data editing are required to continue training the models [28]. Because of the increasing number of incoming tickets for IT support, new data is generated rapidly during use. Therefore "*it [the knowledge base] is very high-maintenance, it is outdated, and you cannot find anything*" (E4). Model operators must monitor the performance and initiate new labeling phases to improve the model and adapt to data drift continuously [38]. The mentioned challenges motivate the need for value co-creation in HI systems during both model initialization and operations phases bridging the provider and recipient spheres [27].

> **DR1:** Intelligent frontline support technologies should utilize and leverage HIL value co-creation configurations to bridge their model initialization and operation phase.

Labeling data by domain experts, such as support agents in the case of IT support, is associated with human effort. *Typically retaining knowledge is time-intensive and costly* (E2; E3, E4, E6) [12,70]. However, it is essential for training and testing the models [49]. An improved labeling process for model initialization and operations should reduce the required volume of ground truth data during the initial incorporation of domain-specific knowledge and generate labeled data during operation. The upfront labeling efforts are not supported by the traditional labeling processes, and users act as an oracle [43]. Typically, samples for labeling are selected by the development team and then forwarded to the annotators uncured., which hampers labeling efficiency, discourages annotators in the long term, and restricts the underlying training data. In practice, a *knowledge manager provides domain knowledge* (E4, E7), and "*agents are already under large pressure*" (E6). An improved process should therefore aim at augmenting and semi-automating at least a part of the labeling tasks [18] and ensure self-efficacy and competence in labeling the data [71]. From a service perspective, this mediates the integration of the user's resources and ensures accumulating value throughout the user's value-creation process [64]. Therefore, we propose following design requirement:

> **DR2:** *HIL configurations of intelligent frontline support technologies should augment the labeling activities.*

Typically, annotators do not benefit from labeling the data [11]. As such, the relationship between the task giver and the annotator can be described by the principle-agent theory, where both actors strive for different goals [20]. Accordingly, an improved labeling process should ensure that the HTIL takes over both – the role of a value recipient and a value enabler by auditing data [28] to ensure a form of intrinsic motivation to – "*easily get the tickets from a specific time period on a certain topic*" (E1, E7). However, even if dedicated experts are involved in the labeling process upfront, they only experience delayed effects of their resource contribution to the prediction model. Overall, there is a lack of incentives to label data [35]. In the context of IT frontline services, this task falls on support agents, who are under immense operational pressure [55]. Even without the advantages of intelligent frontline support technologies, the data quality of tickets has thus been rather challenging (E3, E5, E6, E7) [54]. "*So lots of colleagues, [and] I also understand why they do it, because […] sometimes we have lots of tickets and you have no time to describe how you would solve [the ticket]. Yet, if you find the ticket and you want to [use it] to solve it [the problem], the resolution most of the time does not help you.* "(E5). Conventional systems provide no direct incentive for the person to label the tickets [35]. Thus, labeling tickets should directly be linked to some form of value-in-use [27] and cognitive involvement [22], which means that the agents need to realize that their action leads to utility not only for others but for themselves [58]. Therefore, we propose the following design requirement:

> **DR3:** *HIL interactions of intelligent frontline support technologies should consider the operational context of IT support agents and provide immediate value-in-use for the agent to motivate the ticket labeling activity.*

6. Hybrid Intelligence System Design & Development

The following two subsections provide a processual perspective of the model pipeline (Fig. 3), which is used to accommodate the ITSM system context and mitigate shortcomings of real-world data. We refined the pipeline after conducting a first iteration [85]. Lastly, we derive the design principle utilizing abstraction [52].

6.1. Model Initialization

The proposed pipeline consists of a model training phase and an operations phase, where the latter includes HIL access points. The initial training phase is necessary to set up the model before using it in an organizational context in which the system will adapt to the new and context-specific data by learning based on agents' feedback. In the first step, manually labeled customer tickets are used to train an automatic pre-labeling model, for example, using BERT [19]. With labeling, we refer to the annotation of key entities [17] in tickets, including phrases describing the system, fault description, trigger, and service requests. This label classification model is later used to propose automatic labeling of new incoming tickets, which the support agent can edit. Finally, the manually labeled tickets are simultaneously clustered based on clustering approaches. Based on this grouping, the system determines a reward rating scale between zero and one for the further reinforcement-learning process: The closer a suggested ticket is to an incoming ticket regarding the cluster, the higher the reward for the system.

The next step in the pipeline is the initialization and self-learning training of the prediction model. This model predicts possible suggestions of historical tickets and is based on a multi-armed contextual bandit. Multi-armed bandits are a type of reinforcement learning which uses an exploration-exploitation mechanism to decide the best prediction within a closed set of possible solutions. Such models are frequently used in online recommendation settings such as news websites where different articles are presented to the reader based on expected reading preferences [39]. Once the reward scale is set, the prediction model can be initialized. Next, the prediction model enters a self-learning phase, predicting possible historical solutions and using the rewarding scale for immediate feedback to adapt and predict new solutions. This phase leads to a prediction model that can be utilized to create a HIL value co-creation configuration that can be deployed within an organizational context.

6.2. Model Operations

The pipeline on the operational level consists of two main blocks: the machine learning loop and the human loop. The machine loop starts with the labeling and prediction models built during the initialization phase. First, an incoming customer ticket is automatically labeled to determine the problem. The prediction model then uses this labeling to select a set of four ticket suggestions based on the initial reward scale. This set of historical tickets is then presented to the support agent.

Here, the pipeline creates a configuration of value co-creation: the agent can choose to edit the ticket labels and correct or improve possibly faulty labels. If the labels are edited, the prediction model compares the

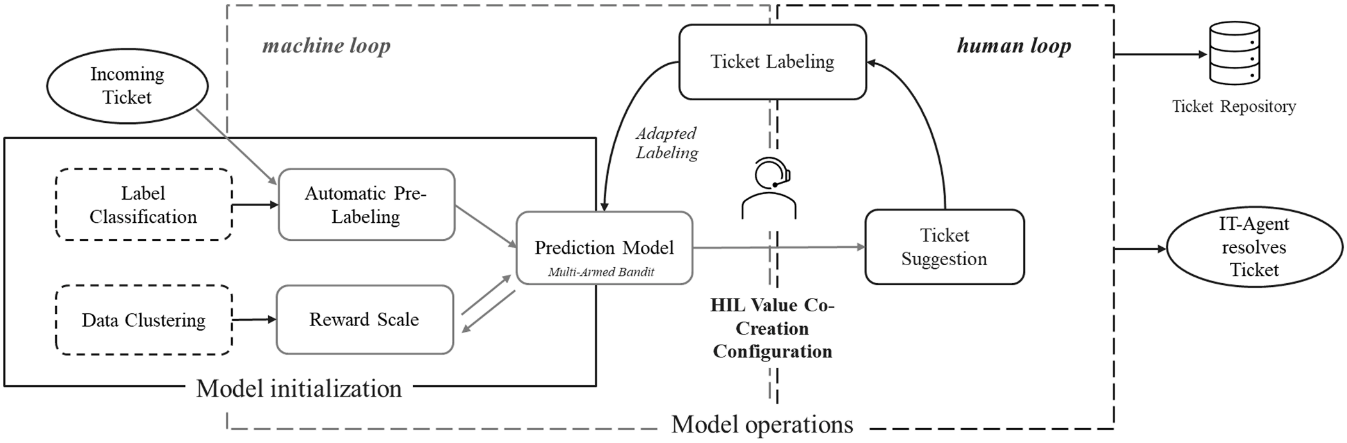


Fig. 3. Pipeline including HIL value co-creation configuration.

newly labeled ticket to historical tickets again and presents the support agent with a new set of ticket suggestions. This loop can be repeated until the support agent is satisfied with the result and can solve the customer ticket. This HIL value co-creation configuration has the advantage of incentivizing the support agent to edit the labels because it results in better ticket suggestions. Additionally, support agents can evaluate the helpfulness of the suggested tickets and give direct feedback to the prediction model. The rating is converted to fit and adapt the rating scale and thus influences the reward used by the prediction model. Finally, the rating function is applied to examine the perceived value of the value co-creation configuration.

6.3. *Design principle - a value co-creation-based HIL configuration*

Given the design requirements, the aim is to design an intelligent system that augments support agent problem-solving capabilities, which keeps receiving domain knowledge after its initial development (DR1), offers interactive and augmented labeling (DR2), and focuses on the touchpoints of support agents to provide an incentive to engage in the HIL activities (DR3). Such incentives can be better decision support or more personalized recommendations and must be provided immediately after contributing high-quality domain knowledge as part of the value-in-use [64]. To accommodate all three design requirements, we propose the design of a HIL value co-creation configuration within the context of the operations pipeline as a novel design principle. We argue that the HIL-enabled act of labeling is strongly tied to several key features of the HI system. It ties together model improvement of the HI system, maintaining training data and ensuring ticket labeling over time and model suggestions, where the HIL takes on the simultaneous role as both value recipient and data auditor [28], shown in Fig. 3. We call the different touchpoints in the HIL a nexus, where the system converges process- and role-wise and simultaneously spans data, prediction model, time, and user needs. Thus, we formulate our design principle as follows:

> **DP - HIL Value Co-Creation Configuration:** *Intelligent ITSM-frontline technologies should design a human-in-the-loop interaction point that provides support agents with immediate value-in-use for labeling activities through value co-creation during model operations.*

7. Demonstration

For this paper, one of our partners gave us access to 17,120 real-world support tickets. As expected with real-world data [10], data quality was poor and had to be cleaned accordingly. Therefore, the tickets were subjected to an initial data cleansing (e.g., empty tickets, non-requests, etc.), resulting in 10,494 and manual filtering of 1st-level frontline tickets, leaving us with 2835 tuples. For more details on the processing steps, refer to Appendix C. We continued to work with cleansed support tickets with ID, title, problem description, solution text field, and an answer history.

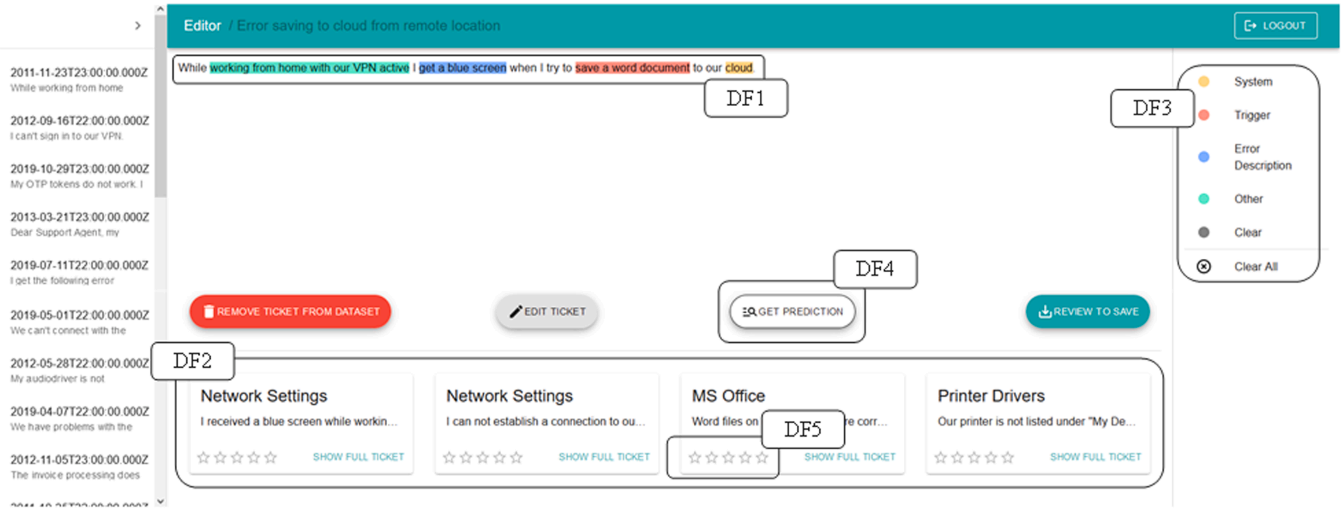
This section demonstrates how we used the design principle to guide us in the instantiation of our system [24]. Thus, we present our resulting five design features (DF1 – DF5) in Fig. 4 [52] to indicate how support agents would interact with our proposed HIL configuration [28]. Our system's support agents first see pre-labeled support tickets (DF1: Pre-labeling) and initial solution suggestions (DF2:Automatic suggestions). However, the suggested tickets might not be ideal because the automatic labeling might be inaccurate. Thus, the support agent can relabel the initial ticket (DF3 Label editing). The label categories were defined previously by support agents. Next, the system provides revised solutions based on the newly highlighted tickets (DF4: Immediately updated suggestions) with better results. Lastly, the historical ticket suggestions provide necessary information for the agents to find a solution, solve the customer request, and save its feedback, as indicated by the introduced 5-star ranking (DF5: User feedback). An overview of the design requirements, design principle, and design features can be found in Fig. 5.

8. Evaluation

We extensively evaluated the HIL value co-creation configuration based on multi-dimensional evaluation criteria according to [48]. As Table 2 illustrates, our evaluation procedure ensures that the design principle is plausible, effective, feasible, predictive, and reliable. In the following, the different evaluation phases and results will be summarized.

8.1. *User evaluation and interviews (eval 1)*

First, by conducting a Wizard of Oz-based user evaluation [14,60], we aim to abstract the design principle from its technological fundamentals to evaluate the **reliability** within the development phase. We introduced the design to people with a technical background and only a rudimentary understanding of IT support, conducted a user test, and interviewed 11 participants. Thus, we could ensure that the value co-creation principle shows its effect reliability regardless of the development or deployment approaches [48]. Solving the tickets took an average of 20 min. The interview partners were, on average, 26 years



Editor / Error saving to cloud from remote location

While working from home with our VPN active I get a blue screen when I try to save a word document to our cloud.

2011-11-23T23:00:00.000Z
While working from home

2012-09-16T22:00:00.000Z
I can't sign in to our VPN.

2019-10-29T23:00:00.000Z
My OTP tokens do not work. I

2013-03-21T23:00:00.000Z
Dear Support Agent, my

2019-07-11T22:00:00.000Z
I get the following error

2019-05-01T22:00:00.000Z
We can't connect with the

2012-05-28T22:00:00.000Z
My audiodriver is not

2019-04-07T22:00:00.000Z
We have problems with the

2012-11-05T23:00:00.000Z
The invoice processing does

Network Settings
I received a blue screen while workin...

Network Settings
I can not establish a connection to ou...

MS Office
Word files on ... re corr...

Printer Drivers
Our printer is not listed under "My De...

Fig. 4. Demonstration of HIL instantiation with design features df1-df5.

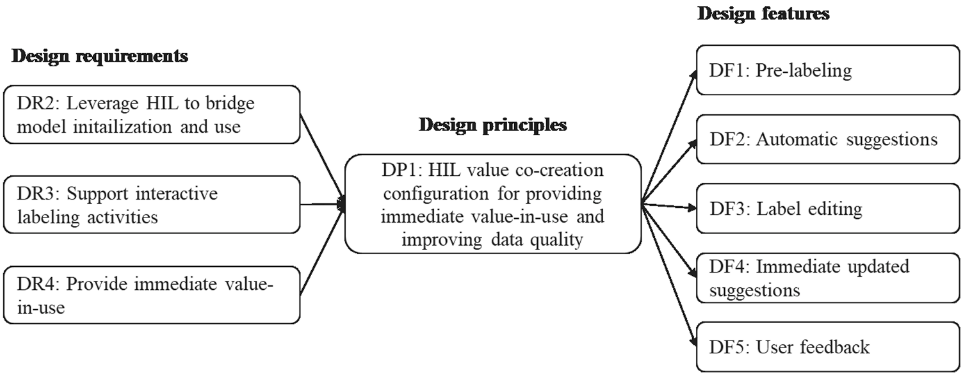

Design requirements
DR2: Leverage HIL to bridge model initaillization and use
DR3: Support interactive labeling activities
DR4: Provide immediate value-in-use
Design principles
DP1: HIL value co-creation configuration for providing immediate value-in-use and improving data quality
Design features
DF1: Pre-labeling
DF2: Automatic suggestions
DF3: Label editing
DF4: Immediate updated suggestions
DF5: User feedback

Fig. 5. Overall design framework.

Table 2
Overview of the evaluation according to [48].

Evaluation criteria	Evaluation	Characteristics	Results
Reliability: The principle yields consistent outcomes, regardless of implementation.	**Eval 1:** Summative artificial user evaluation	Qualitative Wizard-of-Oz user evaluation with 11 professionals with a rudimentary understanding of IT support.	Confirmation of the general mechanisms of the design principle regardless of the technological backend and domain knowledge
Feasibility: The principle can be operationalized or implemented as described.	**Eval 2:** Demonstration of a fully functional prototype	Prototypical instantiation and deployment based on a BERT-labeling model and multi-armed contextual bandit system given 17,120 real-world support tickets.	Transformer-based language models can reliably predict labels based on a small initial data set, while the bandit provides reliable results.
Predictivity: The principle produces the expected result.	**Eval 3:** Formative evaluation of prediction model	Evaluating the warm start of a multi-armed contextual bandit (epsilon 0,25; 59 training tickets, 15 test tickets).	Accuracy of 66,67% and rewards showing increasing growth with decreasing marginal returns.
Plausibility: The principle seems reasonable based on current domain knowledge.	**Eval 4:** Summative focus group evaluation [62]	Three focus group sessions with, on average, 17 real-world ITSM experts, including support agents, managers, and developers.	Agents' willingness to use the system and confirmation of system functionality, with minor concerns relating to the user interface.
Effectiveness: The suggested principle addresses the underlying causes of the problem.	**Eval 5:** Summative naturalistic user evaluation	Two annotators with expert domain knowledge and 90 real-world tickets were annotated and evaluated.	Tickets showed a mean average rating increase of 0.9, while for the highest slot, the mean average rating increase was 2.23.

old, and all possess a technical background and experience with IT. Regarding the augmentation of the labeling process (DR2), the interviewed participants stated that the tool intuitively supported the labeling, and annotating the data was perceived as straightforward. In addition, interviewees mentioned that the labeling supported their cognitive processing of the presented problem cases and understanding of the recommended solutions. The automated labeling could also be extended to the presented solutions to enable an easier matching of problem-solution pairs: "*I think next time I would label first and then read the tickets, to be able to match the problem with the recommended tickets easier*" (I3). Overall, the participants did not perceive labeling the data as effortful, unnecessary, or meaningless. Interestingly, the willingness to contribute to higher data quality through interactive labeling was broadly confirmed and reasoned by the benefit of receiving better recommendations (DR3): "*Because then I noticed that with correct labeling I also get immediately meaningful solution possibilities*" (I4).

Furthermore, the users understood that labeling the data supported the AI to "*narrow down the problem request*" (I1) or "*filter based on important phrases*" (I2). An interviewee compared the labeling to providing prompts to ChatGPT: "*You have to specify the input to the AI so that it can answer your question exactly – that is similar to this ChatGPT*" (I1). In conclusion, the system motivates users to input their knowledge and justifies the effort of labeling the text. The Wizard of Oz system can be accessed via the provided link in Appendix D.

8.2. Demonstration of a fully functional prototype (Eval 2)

To validate the **feasibility**, we instantiated a reinforcement learning system based on the proposed design principle during the deployment phase. We applied clustering to generate the reward and a BERT-based transformer model for our pre-labeling automation. For more details regarding the complete data pipeline for setting up the contextual bandit system refer to Appendix E. The core of the HI system represents a multi-armed contextual bandit. A BERT-based approach achieved viable accuracy scores for pre-labeling data for our underlying database. The results (Accuracy: 0.667; Precision: 0.578; Recall: 0.581; F1-Score: 0.573) suggested that transformer-based machine learning tools like BERT can provide annotators with helpful suggestions for labeling the tickets based on a small database. Although the pre-labeling model has a comparatively low performance due to the unstructured and informal character of most of the problem descriptions, users considered the recommendations useful (see *Eval 5*). Overall, the results reveal that transformer-based large language models can augment the interactive labeling mechanisms (DR2) integrated into the value co-creation configuration and thus bridge model initialization and operations (DR1). The reinforcement system achieved an accuracy of 0.667 and confirmed that the initial set-up with a small amount of manually labeled data performs sufficiently well to build a self-learning system. The evaluation confirms the principle's feasibility as it can be operationalized and implemented as described [48].

8.3. Formative and naturalistic evaluation of the prediction model (Eval 3)

To evaluate the learning mechanism of the prediction model and validate the **predictivity** [48], we use the average reward during the self-learning phase. As the prediction model is based on a multi-armed bandit algorithm, the average reward shows the average closeness of suggested tickets to the incoming ticket. This measure depicts whether the model is learning throughout the training process: If the average reward increases, the model can find and suggest historical tickets close to the incoming ticket. Fig. 6 presents the development of the average reward over the number of iterations as a weighted average reward over the four suggested slots and for the highest slot (best ticket suggestion). As seen in the plot in Fig. 6, both rewards show increasing growth with decreasing marginal returns over the number of iterations and thus indicate learning within the system during the initial training phase. It is to be noted that the average reward for the highest slot shows stronger growth, which indicates that one out of four slots is more similar to the incoming ticket. In the context of IT frontline support technologies, this is a satisfactory result because it indicates that our system can adapt and learn, as is intended, with multi-armed bandits, which is in line with proof of feasibility. We predict that in an organizational context, the system will rapidly adapt to new data as it will be fed with tickets labeled by support agents and receive an immediate evaluation of the suggested tickets.

8.4. Summative focus group evaluation (Eval 4)

Within three focus group sessions (S9–11), we demonstrated the pipeline and prototype to each case of 18, 15, and 19 members of our research project since all members are experts in ITSM. They included support agents, managers, support system developers, managing directors, and work council members across the research consortium. We opted to conduct a focus group review of our system to gain further insights by stimulating a discussion among our experts [61,62]. Primary concerns were satisfactorily addressed, and the HIL configuration design was accepted, with only minor concerns relating to the size of UI tiles. Most importantly, they confirmed that the HIL configuration and overall system design fully address the design requirements, confirming **plausibility** by indicating domain relevance and showing that the "concept is more than just a belief" [48]. Our prototype was limited in demonstrating additional design features, such as multi-category and hierarchical labeling (S9). Different categories were elaborated as the consortium included multiple pilot partners, showing adaptability to different ITSM environments. The focus groups emphasized the importance of reliable initial recommendations (S10). Finally, we discussed the integration of the proposed HIL value co-creation configuration in existing ticket recommendations from a user experience perspective (S11).

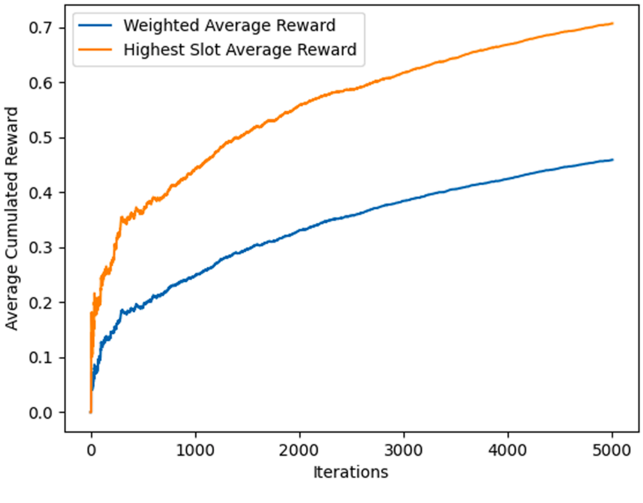


Fig. 6. Average reward of self-learning system over iterations.

8.4. Summative Naturalistic User Evaluation of the Ticket Recommender System (eval 5)

The HIL value co-creation touchpoint was evaluated concerning the ticket labeling and suggestion evaluation mechanisms for **effectiveness** during use [48] by simulating work environments during operations. Thus, two annotators with expert domain knowledge initiated a mock operations environment and evaluated the tickets in two steps: First, they evaluated the suggestions and whether they were helpful based on the automatic labeling only. In the second step, they edited the labeled tickets first, were presented with the new set of four suggestions, and evaluated the suggestions the system made based on the annotator-labeled ticket. A total of 90 tickets were annotated and evaluated. Fifteen tickets showed clear signs of improving the suggestions after round 1, with a mean average rating increase of 0.9. For the highest slot, the mean average rating increase was 2.23. This coincided with a post-evaluation interview, in which both annotators perceived that usually, only one out of the four ticket suggestions appeared useful, and only seldomly did they perceive two or more ticket suggestions as useful.

Furthermore, both annotators responded positively to the immediate feedback, whereas A1 states that "immediate suggestion makes labeling meaningful" and that "when I felt like the incoming ticket text was specific enough to be able to assign labels, […] the resulting suggestion was much more likely to be better" (A1). Moreover, they reported that the relabeling process is "quick" and intuitive. These factors led to a high intention to relabel and evaluate tickets, which were "not like the usual labeling tasks" (A2). We, therefore, expect a quick growth of expert annotated data once the system is used in an organizational context.

Eighteen tickets were not rated since one annotator found no new ticket suggestion relevant, and 4 ticket ratings have gotten lower ratings. On the other hand, 13 tickets remained the same. During post-evaluation interviews, the annotators explained that many ticket suggestions did not fit the new support request and suspected that the original 200 labeled tickets did not include many relevant support requests. Although functional, the annotators did not choose to add any newly annotated ticket into the repository, even though the data set includes its suggestions. Although the total number of evaluated instances is comparably low, the results indicate that manual relabeling by support agents can positively impact the quality of suggested historical tickets. Our results suggest that the newly labeled tickets can improve suggestion quality, but due to the breadth of different customer request types, the number of high-quality tickets needs to be carefully annotated. Table II summarizes the complete evaluation procedure.

9. Discussion and conclusion

We expect that our design principle *HIL value-co-creation configuration* can be applied to different labeling tasks and learning mechanisms. Furthermore, our evaluation suggests that the perceived immediate value-in-use can stimulate the willingness to co-create value in HIL configurations. Thus, we provide novel insights into solving the challenge of data labeling in AI by incorporating the theories of SD-L and value co-creation [26,37]. With our design principle for novel HIL configuration for ITSM systems, we provide an improvement DSR contribution by developing new solutions for a known problem [23].

9.1. Theoretical contributions

The results of the research project show multiple contributions to literature and practice. For the literature on HI system design and particularly in the context of support services [7,35,50], we contribute to its body of design knowledge by suggesting a novel HIL configuration in which the labeling human-in-the-loop is simultaneously the value recipient. Our principle contributes to a nascent design theory by providing prescriptive knowledge as an "operational principle", whereas our artifact instantiation is a "situated implementation of an artifact" ([23], p. 342).

Specifically, we focus on knowledge for instantiations, presenting the rationale behind design requirements, design principles, and design features [52]. For IT support services, we contribute to the body of knowledge on frontline service technology infusions by providing a novel form of support agent integration [34] to augment their work-related problem-solving activities [15]. Our HIL configuration principle also contributes to the research on HIL design and configurations [28,69,86]. We argue that our HIL design provides individual and organizational benefits, complementing the HI system's main functionality. As one of the first papers in the literature stream on interactive machine learning [30], hybrid intelligence [16], and HIL configurations [28], we show how value co-creation can be the core of self-learning systems and how an integrated interactive labeling process removes the dichotomy between value recipients and value providers. By enabling value-in-use in terms of immediate perceived value and providing a sense of control, we contribute novel configurations to the knowledge of designing interactive labeling systems [43]. According to Grönroos and Voima [27], we differentiate between a (1) provider sphere, (2) recipient sphere, and (3) joint sphere. The machine loop represents the provider sphere as it supports the human user through recommendations and augments the labeling activities. The machine loop only generates a potential value-in-use that is activated within the joint sphere and realized by the user within the recipient sphere (Fig. 7). Overall, the mechanisms of SD-L [65] and value co-creation [26] should be at the core of knowledge-intensive HI systems. Given the theoretical contributions, we phrase the following proposition.

Proposition: A HIL value co-creation configuration that provides immediate value-in-use for HI system users can sustainably provide high-quality data for continuous prediction model improvement.

9.2. *Practical implications*

For practice, the paper provides several guidelines and insights for future ITSM labeling-based AI system designs. First, the paper is rooted in a real-world need elicited by practitioners cooperating in a research development project [45] and provides 17,120 real-world support tickets as a database. Second, our research reports detail design decisions of our system implementation, guiding future practitioners to design and develop a similar pipeline. This pipeline can be adapted to different data and used as the basis for future work on improving ITSM processes by supporting and facilitating human decision-making concerning IT support. Third, we further demonstrate how our novel HIL design should follow a key principle to allow for a suitable system that addresses all three design requirements. For more specific implementation instructions, our five design features guide how the principle can guide developers in instantiating an appropriate system. Fourth, from a practical perspective, the model can improve ML development and operations regarding efficient initialization, continuous usage, and model maintenance. In addition, the approach outlines a way to incentivize users to contribute domain-specific knowledge. Finally, the optimized interaction between humans and AI will lead to a higher prediction and subsequent service performance.

9.3. *Limitations and future research*

Our research comes with limitations and provides room for future research. For example, given the scope of this research, the role of operators and how they are integrated into the pipeline remains neglected. Future research should examine the role of ML operators in our proposed pipelines and conduct a naturalistic evaluation design of a long time to test the systems in day to day business of different organizations as a means for a proof-of-value [44] and as an insight into the long-term impacts on data quality and ML operations. In addition, our evaluation does not consider labeling quality as metrics. As Appendix E includes the entire instantiated pipeline, which encompasses 12 process steps, we have opted for a summative evaluation. Individually optimizing each step would also contribute to an overall improvement of system output. This indicates potential for both future research and a system of reference for practitioners.

Another limitation refers to the selected machine learning type. For example, our technological implementation relies on a multi-armed bandit reinforcement learning system. However, considering aspects of other types, such as active learning, can provide additional insights into value co-creation-based labeling. For example, a system could only request new labels when the data is needed to improve the model or the results. Thereby researchers and practitioners could further reduce the demand on users to label data [43]. Simultaneously, additional multi-armed bandits and parameterizations could be implemented to ensure stronger reliability. Furthermore, the potential of generative AI and large language models can be utilized to augment routines of support agents in multi-faceted ways (Reinhard et al. 2024), e.g., by improving the highlighting mechanism. Also, future research could apply the proposed design to train generative AI models via reinforcement learning [86].

Overall, further research has to be conducted concerning our proposition to ensure an accumulation of value and value co-creation in the long term by evaluating the performance of the automated labeling model after adding manual labels and simultaneously validating the

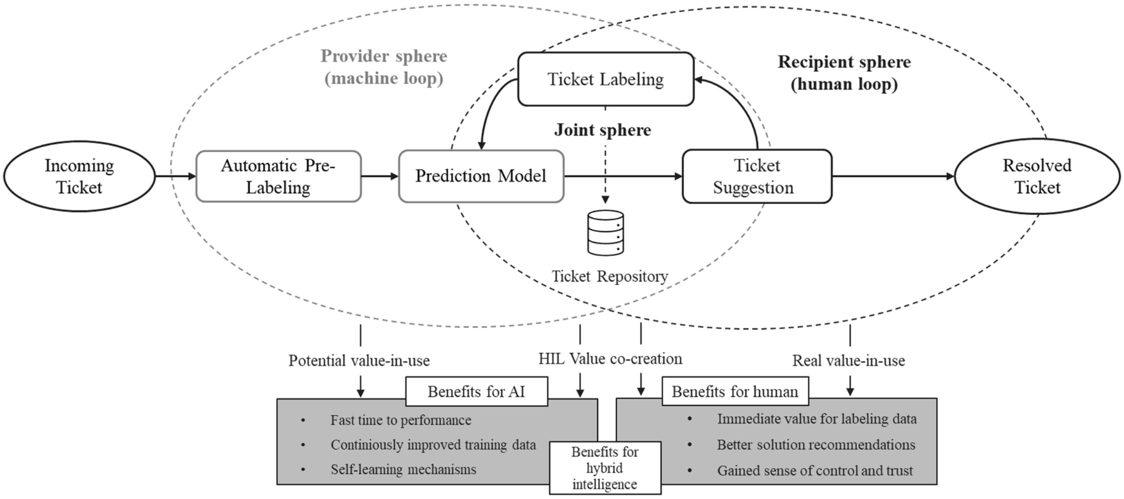


Fig. 7. Conceptualization of value co-creation-based HIL configurations.

performance improvement of the prediction model. Furthermore, a future large-scale experiment should aim for quantitative analysis that could underline the effects mentioned by the interviewees and focus groups. An experiment could also reveal whether our HITL mechanism provides additional benefits for users such as increasing explainability and thereby could contribute to explainable AI (XAI) literature as well. Nonetheless, our paper provides insights into the reasoning behind our HI support system and innovative HIL design and paves the way for future research endeavors.

Research data

Due to the sensitive nature and criticality of the data included in the IT support tickets within our demonstration, the partner company was assured that raw data would remain confidential and would not be shared. To provide an intuition on the underlying data, the performed processing steps as well as the contextual bandit system, we manually anonymized single data instances for exemplary demonstration. The available data is part of the appendix.

Disclosures

During the preparation of this work the author did not use any tool/service, such as generative AI. The author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication. The work was not published and is not under review in any other journal.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Appendix A. Workshop Sessions with pilot partners

ID	Events	Content	Participants	Date
S1	Workshop "Status Quo"	Investigation of the status quo at one of our pilot partners and potential for improvement.	5	04.02.2020
S2	Workshop "Vision"	Developing a vision for the future of AI-augmented IT support.	21	20.05.2020
S3	Focus group session 1	Elaborating basic functionalities of a ticket recommender system.	20	19.01.2021
S4	Focus group session 2	Developing and reviewing a basic ticket recommender system without value co-creation HIL configuration based on matching algorithms and an ontology.	20	21.04.2021
S5	Focus group session 3		23	16.06.2021
S6	Focus group session 4		16	25.08.2021
S7	Focus group session 5		16	19.01.2022
S8	Focus group session 6		14	20.04.2022
S9	Focus group session 7	Presentation of a new prototype, including the HIL value co-creation configuration and discussing the user experience.	18	05.07.2022
S10	Focus group session 8	Presentation of another version of the prototype and discussion of additional features for the ticket recommender system.	15	06.09.2022
S11	Focus group session 9	Reviewing an introduction of the HIL value co-creation configuration into a basic ticket recommender system from the point of usability.	19	02.11.2022

Appendix B. Overview of interviewees for Wizard of Oz user evaluation

ID	Expertise	Educational background	Gender	Age
I1	Digital professional	Advanced technical college	Male	21
I2	Digital professional	Bachelor	Male	24
I3	Digital professional	Bachelor	Female	23
I4	Developer	General university entrance qualification	Male	31
I5	Engineer	Master (or Diploma)	Male	25
I6	Research Associate	Master (or Diploma)	Male	30
I7	Doctor	General university entrance qualification:	Male	27
I8	Civil engineer	Bachelor	Male	29
I9	Research Associate	Master (or Diploma)	Male	24
I10	Purchasing Manager	Bachelor	Male	30
I11	Digital professional	Advanced technical college entrance qualification	Male	19

Appendix C. Anonymized exemplary data

The following table outlines the original data structure provided by the partner:

Processing Step	Description	Example
Number	Ticket number for identifying the ticket	*INC00000182522*
Opened_By	Agent who opened the ticket	*[agent name 1]*
Short_Description	Title or short description of the problem	*Unable to login because it was expired*
Priority	Priority of the problem	*3 – Standard*
Assignment_Group	Group that was first assigned to the incident	*Global IT Support Help Desk*
Assigned_To	Agent who was assigned to the incident	*[agent name 1]*
Main_Category	Main category	*SAP*
Subcategory_1	First subcategory	*General*
Subcategory_2	Second subcategory	*All*
Subcategory_3	Third subcategory	–
Resolver_Group	Group that resolved the ticket	*Global IT Support Help Desk*
Country	Country abbreviation	*UK*
Type	Type of ticket	*Incident*
Regio	Region abbreviation	*EU*
Contact_Type	Type of initial contact	*Phone*
Resolved_At	Date of resolving ticket	*2021–01–01 03:28:21*
Closed_At	Date of closing the ticket	*2021–01–06 04:00:04*
Assignment_Group_History	History of all assigned groups	*Global IT Support Help Desk*
Work_Notes	Agent's notes on solving the incident	*2021–01–01 03:24:54 - Agent (Work notes)* *Who called:User* *Ext Number: +00,000,000,000* *Locations: EU* *Short Description: User cannot login to SAP because her employment contact was set as 2020–12–31.* *Issue(s): The user called and mentioned that this user has already extended the work contact until 2021–12–31. He also confirmed that it is better to put the new due date as 2021–12,031 because he cannot know what will happen.* *My action taken: I extend the SAP*
Comments	Any additional comments	–
Close_Notes	Notes on closing the ticket and answering the user	*2021–01–01 03:28:22 - Agent (Close notes (Customer visible))* *Dear User,* *As per conversation on the phone, you mentioned that you would like to extend the due date for SAP on user_name.* *Thank you*
Assignee_History	History of all assigned agents	*[agent name 1]*
Reassignment_count_assignee	Number of reassigned agents	*1*
Reassignment_count	Number of reassigned groups	*1*

Appendix D. Wizard of Oz System

The following link provides access to the Wizard of Oz System to experience the user interface of the contextual bandit system. To access the system, copy the link and replace the placeholder "name" with an arbitrary input: https://hybrid-intelligence.herokuapp.com/artificial-intelligence-bandit-hiss/name

Appendix E. Anonymized pipeline

The following is a manually anonymized instance of the overall data pipeline for setting up the contextual bandit system. It provides exemplary input and output data for each processing step. In addition, it includes the program code for application in similar cases. However, log files as well as temporary data (such as pickle-files) are excluded. https://anonymous.4open.science/r/Bandit-Backend-Clean-7506/

The pipeline spans the following processing steps:

N.	Processing Step	Details
01	Language detection	Filtering English tickets and separating data points with mixed languages
02	Ticket anonymization	Pseudonymizing sensitive data
03	Pre-processing	Additional data cleansing and aggregating multiple data fields
04	Topic modeling	Identifying recurring topics over the complete data set
05	Topic clustering	Aggregating multiple related topics via a hierarchical clustering approach
06	Topic cluster labeling	Identifying useful ticket clusters by incorporating domain experts
07	Ticket quality labeling	Preparing labelled data for training a scoring model and filtering high quality tickets
08	Highlighting labeling	Preparing labelled data for training the subsequent highlighting model
09	Highlighting model	Highlighting keywords and phrases within the problem description
10	Feature extraction	Extracting most relevant features for subsequent scoring model training
11	Scoring model	Predicting quality labels and filtering high quality tickets
12	Contextual bandit training	Training the recommendation system

References

[1] B. Abedin, Managing the tension between opposing effects of explainability of artificial intelligence: a contingency theory perspective, Int. Res. 32 (2) (2022) 425–453. No.

[2] M.M. Afsar, T. Crump, B. Far, Reinforcement learning based recommender systems: a survey, CoRR (2021). No. abs/2101.06286.

[3] A.K. Ahmad, A. Jafar, K. Aljoumaa, Customer churn prediction in telecom using machine learning in big data platform, J. Big Data 6 (1) (2019) 1–24. No.

[4] Z. Akata, D. Balliet, M.de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerincx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen, M Welling, A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence, Computer (Long Beach Calif) 53 (8) (2020) 18–28. No.

[5] S. Alter, Work system theory: overview of core concepts, extensions, and challenges for the future, J. Assoc. Inf. Syst. (2013) 72.

[6] S. Amershi, M. Cakmak, W.B. Knox, T. Kulesza, Power to the people: the role of humans in interactive machine learning, AI Magazine 35 (4) (2015) 105–120. No.

[7] D.E. Bailey, S.R. Barley, Beyond design and use: how scholars should study intelligent technologies, Inf. Org. 30 (2) (2020), 100286. No.

[8] M. Blaschke, U. Riss, K. Haki, S. Aier, Design principles for digital value co-creation networks: a service-dominant logic perspective, Electron. Markets 29 (3) (2019) 443–472. No.

[9] M. Braun, M. Greve, J. Riquel, A.B. Brendel, L. Kolbe, Meet your new colle(ai)gue – exploring the impact of human-ai interaction designs on user performance, ECIS (2022).

[10] L. Cai, Y. Zhu, The challenges of data quality and data quality assessment in the big data era, Data Sci. J. 14 (0) (2015) 2. No.

[11] H.-A. Cao, T.K. Wijaya, K. Aberer, N. Nunes, A collaborative framework for annotating energy datasets, in: 2015 IEEE International Conference on Big Data (Big Data), IEEE, 2015.

[12] M. Choi, C. Park, S. Yang, Y. Kim, J. Choo, S.R. Hong, AILA: attentive interactive labeling assistant for document classification through attention-based deep neural networks, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, ACM, New York, NY, USA, 2019.

[13] L.F. Cranor, A framework for reasoning about the human in the loop, in: UPSec'08: Proceedings of the 1st conference on usability, psychology and security, Berkeley, 2008.

[14] Dahlbäck, N., Jönsson, A. and Ahrenberg, L. (1993), "Wizard of Oz studies: why and how", available at: https://dl.acm.org/doi/pdf/10.1145/169891.169968.

[15] A. Das, Knowledge and productivity in technical support work, Manage. Sci. 49 (4) (2003) 417–431. No.

[16] D. Dellermann, P. Ebel, M. Söllner, J.M. Leimeister, Hybrid Intelligence, Bus. Inf. Syst. Eng. 61 (5) (2019) 637–643. No.

[17] Dernoncourt, F., Lee, J.Y. and Szolovits, P. (2017), *NeuroNER: an easy-to-use program for named-entity recognition based on neural networks*.

[18] M. Desmond, E. Duesterwald, K. Brimijoin, M. Brachman, Q. Pan, Semi-automated data labeling, NeurIPS 2020 Comp. Demonstr. Track (2021) 156–169.

[19] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (Eds.) (2019), *BERT: pre-training of deep bidirectional transformers for language understanding*.

[20] K.M. Eisenhardt, Agency theory: an assessment and review, Acad. Manage. Rev. 14 (1) (1989) 57–74. No.

[21] E. Elshan, P.A. Ebel, M. Söllner, J.M. Leimeister, Leveraging low code development of smart personal assistants: an integrated design approach with the SPADE Method, J. Manag. Inf. Syst. (JMIS) 40 (1) (2022) 96–129. No.

[22] C.M.N. Faisal, D. Fernandez-Lanvin, J.de Andrés, M Gonzalez-Rodriguez, Design quality in building behavioral intention through affective and cognitive involvement for e-learning on smartphones, Int. Res. 30 (6) (2020) 1631–1663. No.

[23] S. Gregor, A.R. Hevner, Positioning and presenting design science research for maximum impact, MIS Quarterly 37 (2) (2013) 337–355. No.

[24] S. Gregor, L. Kruse, S. Seidel, Research perspectives: the anatomy of a design principle, J. Assoc. Inf. Syst. 21 (2020) 1622–1652.

[25] C. Grönroos, Service logic revisited: who creates value? And who co-creates? Eur. Bus. Rev. 20 (4) (2008) 298–314. No.

[26] C. Grönroos, Value co-creation in service logic: a critical analysis, Mark. Theory 11 (3) (2011) 279–301. No.

[27] C. Grönroos, P. Voima, Critical service logic: making sense of value creation and co-creation, J. Acad. Mark. Sci. 41 (2) (2013) 133–150. No.

[28] T. Grønsund, M. Aanestad, Augmenting the algorithm: emerging human-in-the-loop work configurations, J. Strat. Inf. Syst. 29 (2) (2020), 101614. No.

[29] P. Hemmer, M. Schemmer, L. Riefle, N. Rosellen, M. Vössing, N. Kühl, Factors that influence the adoption of human-AI collaboration in clinical decision-making, ECIS (2022).

[30] A. Holzinger, Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inf. 3 (2) (2016) 119–131. No.

[31] L. Jiang, S. Liu, C. Chen, Recent research advances on interactive machine learning, J. Visualization 22 (2) (2019) 401–417. No.

[32] T. Jiang, J.L. Gradus, A.J. Rosellini, Supervised Machine Learning: a Brief Primer, Behav. Ther. 51 (5) (2020) 675–687. No.

[33] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement Learning: a Survey, J. Artificial Intelligence Res. 4 (1996) 237–285.

[34] A.de Keyser, S. Köcher, L. Alkire, C. Verbeeck, J Kandampully, Frontline service technology infusion: conceptual archetypes and future research directions, J. Service Manag. 30 (1) (2019) 156–183. No.

[35] P. Kubiak, S. Rass, An overview of data-driven techniques for IT-service-management, IEEE Access 6 (2018) 63664–63688.

[36] S. Lockey, N. Gillespie, D. Holm, I.A. Someh, A review of trust in artificial intelligence: challenges, Vulnerab. Future Directions (2021).

[37] R.F. Lusch, S.L. Vargo, Service-dominant logic: reactions, reflections and refinements, Market. Theory 6 (3) (2006) 281–288. No.

[38] A. Mallick, K. Hsieh, B. Arzani, G. Joshi, Matchmaker: data drift mitigation in machine learning for large-scale systems, in: Proceedings of Machine Learning and Systems, 2022, pp. 77–94. Vol. 4.

[39] J. Mary, R. Gaudel, P Preux, Bandits and recommender systems, in: P. Pardalos, M. Pavone, G.M. Farinella, V. Cutello (Eds.), Machine Learning, Optimization, and Big Data, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2015, pp. 325–336. Vol. 9432.

[40] M.A. Meza Martínez, M. Nadj, A. Maedche, Towards an integrative theoretical framework of interactive machine learning systems, ECIS (2019).

[41] M. Mirbabaie, F. Brünker, Möllmann Frick, R.J. Nicholas, S. Stieglitz, The rise of artificial intelligence – understanding the AI identity threat at the workplace", Electronic Markets 32 (1) (2022) 73–99. No.

[42] M.D. Myers, M. Newman, The qualitative interview in IS research: examining the craft, Inf. Org. 17 (1) (2007) 2–26. No.

[43] M. Nadj, M. Knaeble, M.X. Li, A. Maedche, Power to the oracle?, in: Design Principles for Interactive Labeling Systems in Machine Learning, 34 KI - Künstliche Intelligenz, 2020, pp. 131–142. No.

[44] J.F. Nunamaker, R.O. Briggs, D.C. Derrick, G. Schwabe, The last research mile. Achieving both rigor and relevance in information systems research, J. Manag. Inf. Syst. 32 (3) (2015) 10–47. No.

[45] H. Österle, B. Otto, Consortium research, Bus. Inf. Syst. Eng. 2 (5) (2010) 283–293. No.

[46] K. Peffers, T. Tuuanen, M.A. Rotheberger, S. Chatterjee, T. Tuunanen, M. A. Rothenberger, A design science research methodology for information systems research, J. Manag. Inf. Syst. 24 (3) (2007) 45–77. No.

[47] C. Peters, Designing Work and Service Systems, Habilitation Thesis, School of Management, University of St.Gallen, St.Gallen, Switzerland, 2020, p. 2020.

[48] S. Petter, D. Khazanchi, J.D. Murphy, A design science based evaluation framework for patterns, in: ACM SIGMIS Database: the DATABASE For Advances in Information Systems, 41, 2010, pp. 9–26. No.

[49] R.B. Porter, J.P. Theiler, D.R. Hush, Interactive machine learning in data exploitation, Comput. Sci. Eng. 15 (5) (2013) 12–20. No.

[50] M. Poser, E. Bittner, (*Re*)Designing IT Support: how embedded and conversational AI can augment technical support work", in: International Conference on Information Systems 42, 2021, pp. 1–17.

[51] M. Poser, C. Wiethof, D. Banerjee, V. Shankar Subramanian, R. Paucar, E.A. C. Bittner, Let's team up with AI! toward a hybrid intelligence system for online customer service, , in: A. Drechsler, A. Gerber, A. Hevner (Eds.), The Transdisciplinary Reach of Design Science Research, Lecture Notes in Computer Science, , Springer International Publishing, Cham, 2022, pp. 142–153. Vol. 13229.

[52] N. Prat, J. Akoka, I. Comyn-Wattiau, V.C. Storey, A granular view of knowledge development in design science research, in: A. Drechsler, A. Gerber, A. Hevner (Eds.), *The Transdisciplinary Reach of Design Science Research, Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2022, pp. 363–375. Vol. 13229.

[53] D. Priharsari, B. Abedin, E. Mastio, Value co-creation in firm sponsored online communities: what enables, constrains, and shapes value, Int. Res. 30 (3) (2020) 763–788. No.

[54] S. Salah, G. Maciá-Fernández, J.E. Díaz-Verdejo, L. Sánchez-Casado, A model for incident tickets correlation in network management, J. Network Syst. Manag. 24 (1) (2016) 57–91. No.

[55] S. Schmidt, M.M. Li, S. Weigel, C. Peters, Knowledge is power: provide your IT-support with domain-specific high-quality solution material, in: International Conference on Design Science Research in Information Systems and Technology (DESRIST), 2021, pp. 1–14.

[56] R. Schüritz, K. Farrell, B. Wixom, G. Satzger, Value Co-creation in data-driven services: towards a deeper understanding of the joint sphere, in: paper presented at Fortieth International Conference on Information Systems, 2019 available at: https://www.researchgate.net/profile/ronny-schueritz/publication/336280678_value_co-creation_in_data-driven_services_towards_a_deeper_understanding_of_the_joint_sphere.

[57] I. Seeber, L. Waizenegger, S. Seidel, S. Morana, I. Benbasat, P.B. Lowry, Collaborating with technology-based autonomous agents: issues and research opportunities, Int. Res. (2020).

[58] E.H. Shaw, The utility of the four utilities concept, Res. Mark. 6 (1994) 44–66. No.

[59] Statista (2022), "T-consulting and implementation services market revenue worldwide from 2016 to 2027″, available at: https://www.statista.com/forecasts/1079927/it-consulting-implementationservices-revenue (accessed 14 June 2022).

[60] A. Steinfeld, O.C. Jenkins, B. Scassellati, The oz of wizard, in: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, New York, NY, USA, ACM, 2009.

[61] D.W. Stewart, P.N. Shamdasani, Focus groups: Theory and Practice, Applied social Research Methods Series, 20, Sage Publ, Newbury Park, Calif, 1990, 1. printing.

[62] J. Stewart, Grounded theory and focus groups: reconciling methodologies in indigenous Australian education research, Australian J. Indigenous Educ. 36 (S1) (2007) 32–37. No.

[63] H.M. Trivedi, M. Panahiazar, A. Liang, D. Lituiev, P. Chang, J.H. Sohn, Y.-Y. Chen, B.L. Franc, B. Joe, D Hadley, Large scale semi-automated labeling of routine free-text clinical records for deep learning, J. Digit. Imaging 32 (1) (2019) 30–37. No.

[64] S.L. Vargo, R.F. Lusch, The four service marketing myths: remnants of a goods-based, manufacturing model, J. Serv. Res. 6 (4) (2004) 324–335. No.
[65] S.L. Vargo, R.F. Lusch, Service-dominant logic: continuing the evolution, J. Acad. Mark. Sci. 36 (1) (2008) 1–10. No.
[66] J. Venable, J. Pries-Heje, R. Baskerville, FEDS: a framework for evaluation in design science research, Eur. J. Inf. Syst. 25 (1) (2016) 77–89. No.
[67] L. Viana, E. Oliveira, T. Conte, An Interface design catalog for interactive labeling systems, in: Proceedings of the 23rd International Conference on Enterprise Information Systems, SCITEPRESS - Science and Technology Publications, 2021.
[68] S. Warsinsky, M. Schmidt-Kraepelin, S. Thiebes, M. Wagner, A. Sunyaev, Gamified Expert Annotation Systems: Meta-Requirements and Tentative Design, Springer, Cham, 2022, pp. 154–166.
[69] C. Wiethof, E.A.C. Bittner, Hybrid intelligence - combining the human in the loop with the computer in the loop: a systematic literature review, in: International Conference on Information Systems (ICIS), 2021, pp. 1–17.
[70] Jie Yan, Hauptmann Yang, Automatically labeling video data using multi-class active learning, in: Proceedings Ninth IEEE International Conference on Computer Vision, IEEE, 2003.
[71] Y. Yan, X. Zhang, X. Zha, T. Jiang, L. Qin, Z. Li, Decision quality and satisfaction: the effects of online information sources and self-efficacy, Int. Res. (2017).
[72] F.M. Zanzotto, Viewpoint: human-in-the-loop Artificial Intelligence, J. Artificial Intelligence Res. 64 (2019) 243–252.
[73] Schmidt S, Li M, Peters C (2022) Requirements for an IT Support System based on Hybrid Intelligence. In: HICSS.
[74] Li MM, Peters C, Leimeister JM (2017) Designing a Peer-Based Support System to Support Shakedown.
[75] Amrou, S., Semmann, M., & Böhmann, T. (2015). Enhancing transfer-of-training for corporate training services: Conceptualizing transfer-supporting IT components with theory-driven design.
[76] K. Eilers, C. Peters, J.M. Leimeister, Why the agile mindset matters. Technological Forecasting and Social Change, Ausgabe/Nummer: 121650, Vol. 179, Erscheinungsjahr/Year: 2022, 2022, pp. 1–14.
[77] K. Eilers, E. Elshan, P. Ebel, M. Söllner, J.M. Leimeister, Leveraging low code development of smart personal assistants: an integrated design approach with the SPADE method, J. Manag. Inf. Syst. 40 (1) (2023) 96–129.
[78] M.M. Li, C. Peters, M. Poser, K. Eilers, E. Elshan, ICT-enabled job crafting: How Business Unit Developers use Low-code Development Platforms to craft jobs. International Conference on Information Systems (ICIS), 2022. Copenhagen, Denmark. Link zu job crafting im Kontext des Papers (ggf. als future research).
[79] V. Mrass, C. Peters, J.M. Leimeister, How Companies Can Benefit from Interlinking External Crowds and Internal Employees. Management Information Systems Quarterly Executive (MISQE), Ausgabe/Nummer: 1, Vol. 20, Erscheinungsjahr/Year: 2021, 2021, pp. 17–38.
[80] R. Knote, A. Janson, M. Söllner, J.M. Leimeister, Value co-creation in smart services: a functional affordances perspective on smart personal assistants, J. Assoc. Inf. Syst. 2020 (2020) 418–458.
[81] M.M. Li, C. Peters, Reconceptualizing Service Systems Introducing Service System Graphs, Int. Conf. Inf. Commun. Syst. (2018).
[82] M.M. Li. Theorizing a Service Structure. A Hypergraph[HYPHEN]based Modeling Approach and Applications Vol. 22, kassel university press GmbH, 2021.
[83] E. Dickhaut, M.M. Li, A. Janson, J.M. Leimeister, The role of design patterns in the development and legal assessment of lawful technologies, Electron. Mark. 32 (4) (2022) 2311–2331.
[84] P. Reinhard, M.M. Li, E. Dickhaut, C. Peters, J.M. Leimeister, Empowering Recommender Systems in ITSM: A Pipeline Reference Model for AI[HYPHEN] Based Textual Data Quality Enrichment. International Conference on Design Science Research in Information Systems and Technology, Springer Nature Switzerland, Cham, 2023, pp. 279–293.
[85] M. Li, D. Löfflad, C. Reh, et al., Towards the Design of Hybrid Intelligence Frontline Service Technologies A Novel Human in the Loop Configuration for Human Machine Interactions. HICSS, 2023.
[86] P. Reinhard, M.M. Li, C. Peters, J.M. Leimeister, Generative AI in Customer Support Services: A Framework for Augmenting the Routines of Frontline Service Employees. Hawaii International Conference on System Sciences (HICSS), Waikiki, Hawaii, USA, 2024.