Towards the Design of Hybrid Intelligence Frontline Service Technologies – A Novel Human-in-the-Loop Configuration for Human-Machine Interactions

Mahei Manhei Li
University of Kassel
Mahei.li@uni-kassel.de

Denise Löfflad
University of Kassel
Denise.loefflad@uni-kassel.de

Cornelius Reh
University of Kassel
Cornelius.reh@wi-kassel.de

Sarah Oeste-Reiß
University of Kassel
Oeste-reiss@uni-kassel.de

Abstract

The rapid adoption of innovative technologies confronts IT-Service-Management (ITSM) with incoming support requests of increasing complexity. As a consequence, job demands and turnover rates of ITSM support agents increase. Recent technological advances have introduced assistance systems that rely on hybrid intelligence to provide support agents with contextually suitable historical solutions to help them solve customer requests. Hybrid intelligence systems rely on human input to provide high-quality data to train their underlying AI models. Yet, most agents have little incentives to label their data, lowering data quality and leading to diminishing returns of AI systems due to concept drifts. Following a design science research approach, we provide a novel Human-in-the-Loop design and hybrid intelligence system for ITSM support ticket recommendations, which incentivize agents to provide high-quality labels. Specifically, we leverage agents' need for instant gratification by simultaneously providing better results if they improve labeling automatically labeled support tickets.

Keywords: Hybrid Intelligence, ITSM, IT support, AI, human-in-the-loop

1. Introduction

The IT services market has reached $57 billion in 2021 and is projected to reach $82 billion in 2027 (Statista, 2022). With new AI technologies and digitalization projects gaining popularity, the IT landscape in businesses has become increasingly more complex and heterogeneous. Thus, IT service management (ITSM) and its frontline support agents face higher customer expectations and a rapidly increasing number of more complex and heterogeneous customer requests (Kubiak & Rass, 2018). Support agents are at the forefront of service provider-customer touchpoints. They provide frontline services (Keyser, Köcher, Alkire, Verbeeck, & Kandampully, 2019) to ensure the "availability of IT services and flawless business operations" (Kubiak & Rass, 2018, p. 63664).

Providing high-quality services has become a critical success factor for IT service providers (Pentland, 1992). Thus, it falls upon support agents to provide frontline service to a company's customers. They do so by drawing upon their experience or information material that they use as reference points to solve incoming customer problems (Das, 2003).

Recent research in frontline service technologies has drawn upon the technological advances made in artificial intelligence and particularly hybrid intelligence (HI) (Dellermann, Ebel, Söllner, & Leimeister, 2019) to augment support agents in their problem-solving activities (Kubiak & Rass, 2018; Poser et al., 2022; Poser & Bitner, 2021; Schmidt, Li, Weigel, & Peters, 2021). HI systems are defined by combination of complementary capabilities of human (e.g.: flexibility, empathy, common sense, creativity) and artificial intelligence (e.g.: pattern recognition, probabilistic, speed, consistency) to achieve complex goals (Dellermann et al., 2019). Similar to adaptive automation (AA) systems (Scerbo, 2007), HI systems consist of an automation-based subsystem and an adaptive part that is influenced by humans. However, HI system research focuses explicitly on AI based systems, while simultaneously drawing on coinciding AA concepts, such as human-centered approach to reduce cognitive load and workload of human agents (Kaber, Riley, Tan, & Endsley, 2001). When confronted with incoming support tickets, which is a digital record of any customer issue or problem, existing (artificial) intelligent systems (Bailey & Barley, 2020) rely on pre-trained models and data repositories to provide information to support agents,

URI: https://hdl.handle.net/10125/102669
978-0-9981331-6-4

who can use it to make better decisions. These systems aid them in solving customer issues (Poser & Bitner, 2021). The integration of intelligent frontline IT-service systems that augment the work processes of support agents have been identified as important (Pentland, 1992; Poser & Bitner, 2021; Schmidt et al. 2021). Mechanisms of providing high-quality input data for the system during operations have recently also gained research interest (Grønsund & Aanestad, 2020).

HI systems often rely on high-quality data to train their prediction models (Dellermann et al., 2019). Relevant and new data needs to be continuously validated (Jiang, Gradus, & Rosellini, 2020) and audited by domain experts, such as service agents not just during model initialization, but continuously during system use (Grønsund & Aanestad, 2020). Yet, human-in-the-loop (HiL) configurations for support agents, who are already over-worked and subject to high turnover rates, provide little incentive to them other than an outlook that the work is important. Support agents might ask themselves - why should we label data on top of our already stressful daily business? Thus, our research goal is to *design a HI system that incorporates a support-agent friendly HiL mechanism to provide immediate utilitarian value to support agents in exchange for high-quality data labeling*. Moreover, we want to leverage the human need for instant gratification and utilitarian value to motivate support agents to diligently label tickets continuously during operations. Thus, we address the following research question:

RQ*: How do we design a HiL-interaction point that simultaneously provides immediate value for the human actor and long-term benefits for the HI-System*?

2. Design science research approach

We followed the design science research (DSR) approach by (Peffers et al., 2007) for designing and evaluating design requirements and principles, see **Figure *1***. We thus formulate practice-based design requirements that describe the problem-space and derive appropriate design principles (Chandra, Seidel, & Gregor, 2015). Based on this, we developed various design features which were then used to implement a fully operational system, which we tested in a two-fold study.

This paper is embedded in a larger DSR project setting that has been ongoing for over two years. The overall purpose of the DSR project is to develop intelligent systems for ITSM to aid its IT support agents. The project includes two research institutes, two software implementation businesses, and three pilot partners that provide system requirements from practice and users to test the resulting system, conducting consortium research (Österle & Otto, 2010). The three pilot partners all provide professional IT support services. The authors regularly meet with over 24 stakeholders (a floating number between 15-39 people), including support agents, management and work council representatives, developers, and researchers in roundtable workshop settings (every 6-8 weeks) to review newly developed tools.

Recurring 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" b) "how can we motivate agents to continuously provide labels?" and c) "how can we support agents with a system that helps them solve the ticket". All three questions have unanimously been agreed upon by both management and support agents and are the basis for the design science research at hand. Thus, the research design is categorized as problem-centered initialization (Peffers et al., 2007) and each initially identified problem will be explored further in this paper in the next section.

Technical support augmentations have been redesigned at a business process level (Poser & Bittner, 2021). This paper focuses on the context of IT support services and the system design to provide support agents with historical support tickets based on the incoming support request data. The matched

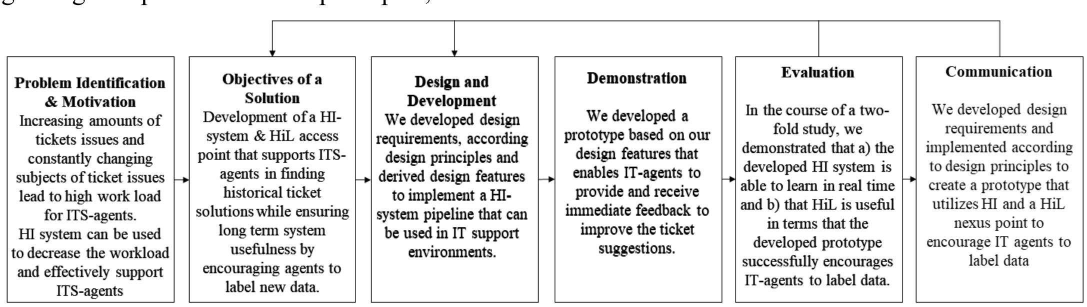

Figure 1. DSR Approach adapted from (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

support ticket functions as reference material to provide agents with input for solving the incoming customer issue. Overall our design science research approach was based on questions in a domain of interest and characterized by "creating/obtaining sources of data germane to relevant phenomena in the domain and cleansing, extracting, annotating data streams" (Maass, Parsons, Purao, Storey, & Woo, 2018, p. 1253). For this paper, we gained access to 17120 real-world historical support tickets from 2019 to 2020 by one of our partners. As expected with real-world data (Cai & Zhu, 2015), data quality was poor and needed to be cleansed accordingly. The tickets were subjected to an initial data cleansing (e.g., empty tickets, non-requests, etc.), resulting in 10494, and manual filtering of 1st-level frontline tickets leaving us with 2835 tuples. We worked with support tickets that had an ID, title, problem description and solution text field, and answer history. Thus, we used our cleanse semi-structured data set as training for our HI system to create predictions, following an algorithmic modeling approach (Breiman, 2001; Shmueli, 2010).

Hence, this paper is structured as follows: Following the here presented initial motivation, section 3 sheds light on three main design requirements derived from extant research and used to inform design principles (Gregor, Kruse, & Seidel, 2020). We thus follow a theory-driven approach (Drechsler & Hevner, 2016). Section 4 depicts the iterative machine learning steps, the so-called system pipeline, employed to solve the problem at hand and further explains how the system initializes its underlying model and works during operations. Since the system output is essentially a prediction, the pipeline is an ensemble of different methods and models that simultaneously represents a kernel predictive model (Prat, Akoka, Comyn-Wattiau, & Storey, 2022).

Then, section 5 demonstrates how the system integrates a human-in-the-loop (HiL) design to address all design requirements and presents its design principles. The paper concludes with tentative formative and summative evaluations (Venable, Pries-Heje, & Baskerville, 2016). First, we conduct an artificial evaluation in section 6 using simulation to show the learning capabilities of a multi-armed bandit prediction model. Second, we evaluate the HiL perceived utility using domain expert evaluations of relabeling 45 tickets to determine whether suggestions improved. Lastly, in section 7, we conducted a relevance check with system stakeholders in a workshop setting with 16 potential users. The paper concludes with a discussion and outlook on future work.

3. Theoretical foundation and problem identification

Based on the problem-initialized issues mentioned by our pilot partners, we moved towards a theory-driven design and identified relevant literature that provides the grounding for deriving intelligent frontline support technology design requirements, (Prat et al., 2022).

3.1 HI systems and HiL interaction

Hybrid intelligence 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 (Dellermann et al., 2019; Grønsund & Aanestad, 2020). While AI systems are better at pattern recognition (Akata et al., 2020), they often lack real-world and domain knowledge, ultimately still leaving problem-solving decisions to human actors (Grønsund & Aanestad, 2020). 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 (Akata et al., 2020).

Hybrid intelligence systems employ human-in-the-loop mechanisms to leverage intuition and real-world knowledge of domain experts to eventually augment work (Dellermann et al., 2019). Recent IS research calls for analyzing and designing novel HiL configurations. To improve the algorithms, humans need to audit data and use the HI system to produce more data, both forms of data editing are required to continue training the models (Grønsund & Aanestad, 2020). This coincides with the initial training of a predictive model and model use, with the latter coinciding with our system use (Shmueli, 2010). Thus, for our HI support system, we derive the following design requirement:

DR1-*HiL*: *Intelligent frontline support technologies should utilize and leverage human-in-the-loop mechanisms to bridge both their model training and system use phase.*

3.2 IT-frontline technology and HI systems

Once a customer encounters an issue with an IT product, they can turn to IT frontline service providers for help (Kajko-Mattsson, 2004). Customer issues or problems are externalized as support tickets, which are assigned to IT, support agents. Based on ***Technical Support Work Theory***, support agents engage in problem-solving activities based on the experience of the individual support agent, their available resources, and the problem type (Das, 2003). To solve an incoming customer request, the support agent can either *deflect* and route the ticket to a more suitable expert or attempt to solve it, for which Das (2003) differentiates between three forms of activities. 1) *Locate and retrieve information* - the support agent finds an almost identical problem within their available knowledge database and solves the ticket with little to no modification to the retrieved solution. 2) *Adapt solution suggestions* – the support ticket is similar to a previously solved issue and the support agent can retrieve information on how the historic issue was solved. However, the suggested solution is considered a reference solution and needs to be adapted by the support agent. The historical solution can also help the agent abductively diagnose the issue. 3) *Generate new solutions*. If the issue is completely new, the support agent needs to find a novel and innovative solution to a new problem by experimenting and reasoning (Das, 2003; Poser & Bitner, 2021). The first two are the subject of this paper.

To improve frontline service encounters between IT service agents and customers, the infusion of frontline service technologies has become quasi-omnipresent (Giebelhausen, Robinson, Sirianni, & Brady, 2014; Keyser et al., 2019; van Doorn et al., 2017). In light of recent advancements in artificial and hybrid intelligence-based (HI) systems (Dellermann et al., 2019), many IT frontline encounters are augmented (Keyser et al., 2019) to improve operations (Kubiak & Rass, 2018; Poser & Bitner, 2021). They augment support agents in information retrieval endeavors and provide more relevant solution suggestions to augment agents in their *problem-solving activities* (Das, 2003). Thus, we derive the following design requirement for our HI support system:

DR2-*TSWT*: *Intelligent frontline support technologies should augment the problem-solving activities of support agents.*

3.3. Support agents and immediate gratification

Frontline support technologies typically rely on IT support tickets to train their underlying models (Kubiak & Rass, 2018), to ultimately augment customer interaction (van Doorn et al., 2017). Thus, support ticket quality is crucial for support agents, especially for supervised machine learning approaches that require properly labeled data (Jiang et al., 2020). HI systems often turn to Human-in-the-Loop (HiL) mechanisms asking employees to label their data during the initial training phase and operations (Wiethof & Bittner, 2021). The latter is particularly important to allow incremental learning to avoid data-drift, which describes a discrepancy between past training data and future test data (Mallick, Hsieh, Arzani, & Joshi, 2022; Tsymbal, 2004). In the context of IT frontline services, this task falls on support agents in hopes to leverage their domain knowledge. Yet, support agents are under immense operational pressure (Schmidt et al., 2021). Even without the advantages of intelligent frontline support technologies, the data quality of tickets has thus been rather challenging (Salah, Maciá-Fernández, Díaz-Verdejo, & Sánchez-Casado, 2016) and conventional systems provide no direct incentive for the person, who is supposed to label the tickets (Kubiak & Rass, 2018). Thus, labeling tickets should directly be linked to some form of ***utilitarian value***, which means that the agents need to realize that their action leads to utility not only for others but for themselves (Shaw, 1994). Furthermore, people have a certainty bias (Kahneman & Tversky, 1979; Weber & Chapman, 2005), which leads to a preference for ***instant gratification*** (Wolfe & Patel, 2017). Thus, for our HI support system we derive the following design requirement:

DR3-UtilGrat: *Human-in-the-loop interactions of intelligent frontline support technologies should consider the context of IT support agents and provide immediate and utilitarian value for the agent to motivate the ticket labeling activity.*

While others have focused on individual data-driven techniques to analyze IT services (Kubiak & Rass, 2018), this paper focuses on a) the particular design of HI service support systems and provides a reference pipeline to guide future implementations and b) an IT-enabled HiL mechanism to address the data-drift conundrum of supervised machine learning approaches.

4. Hybrid intelligence system design & development

The following two subsections provide a processual perspective of the AI model pipeline, which is used to accommodate the ITSM system context and mitigate shortcomings of real-world data. Lastly, we derive design principles by means of abstraction (Prat et al., 2022). The proposed pipeline consists of a model training phase, described in Section 4.1, and operations phase, described in Section

4.2, where the latter includes its 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 IT-Agent feedback.

4.1. Model initialization

Figure *2* presents the model initialization pipeline based solely on AI mechanisms. In a very first step, a database of 200 manually labeled customer tickets is created, which is then used to train an automatic labeling model using a bi-directional Long Short-Term Memory (LSTM) neural network (Yu, Si, Hu, & Zhang, 2019). This label classification model is later used to propose automatic labeling of new incoming tickets which can be edited by the support agent.

During the model initialization, the manually labeled tickets are simultaneously used for the LSTM classification training and clustered based on the OPTICS clustering algorithm (Ankerst, Breunig, Kriegel, & Sander, 1999) to determine textually close tickets.

Next, based on the previously created grouping, the self-learning initial system determines a system reward scale between zero and one for further reinforcement-learning purposes: The closer a suggested ticket is to an incoming ticket with regards to the cluster, the higher the reward for the system. Based on this scale, the system initializes a prediction system that determines a set of possible historical ticket solutions. This set is immediately evaluated using the system rewarding scale, providing timely feedback for the system. Thus, the second step in the pipeline is a self-learning phase in which the prediction system is constantly adapted. Once this cycle hits a sufficient threshold, a final prediction model and an automatic labeling model are created by the system. These models are then utilized in the HI-system pipeline, described in the following section.

The prediction system is based on a multi-armed bandit, which is a type of reinforcement learning. Multi-armed bandits use an exploration-exploitation mechanism to decide the best prediction within a closed set of possible solutions. Such multi-armed bandit models, are frequently used in online-learning recommendation contexts (Mary, Gaudel, & Preux, 2015). In contrast to offline learning, online learning allows for the system to constantly adapt itself to new incoming data (Li, Zhou, & Cao, 2021). As ITSM is constantly exposed to new problems and therefore new and unseen data, we deemed online learning to be adequate

for this task and therefore opted for a multi-armed bandit algorithm. This allows us to provide immediate online feedback predictions to the user, depending on their labelling input.

4.2. Operations

The pipeline on the operational level, depicted in **Figure *3***, implements the HiL-touchpoint and consists of two main blocks: the machine learning loop and the human loop. The machine loop starts with the labeling and prediction models that were built during the initialization phase, described in section 4.1. In a first step, an incoming customer ticket is automatically labeled to determine the problem at hand. This labeling is then used by the prediction model to select a set of four ticket suggestions based on the initial system reward scale. This set of historical tickets is then presented to the support agent.

Here, the pipeline creates two nexus points: On the one hand, support agents can evaluate the helpfulness of the suggested tickets and give direct feedback to the prediction model. On the other hand, the agent can choose to edit the ticket. If the labels are edited, the prediction model compares the 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-touchpoint has the short-term advantage that the support agent is incentivized to edit the labels because it results in better ticket suggestions. From a

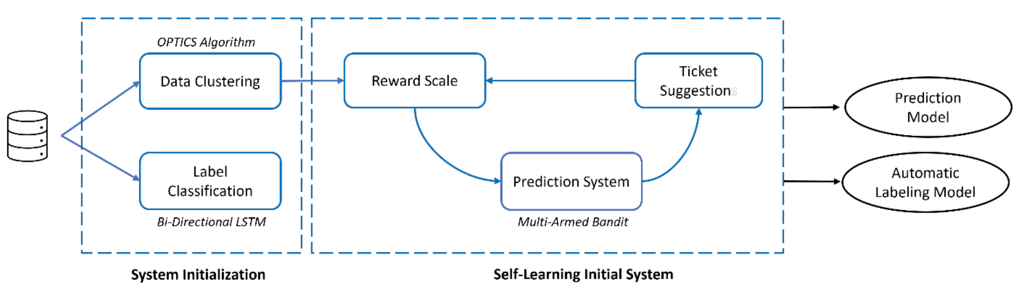

Figure 2: HIS Pipeline for Self-Learning Model Initialization to Build Prediction and Labeling Model.

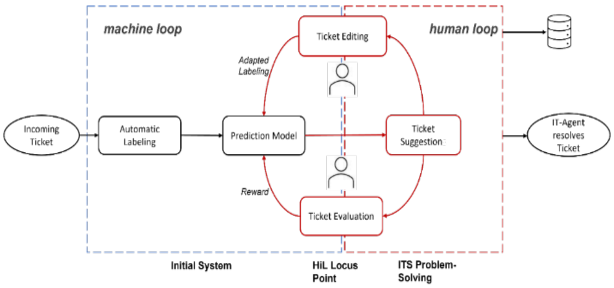

Figure 3. Operational Pipeline including HiL-touchpoint.

long-term perspective, the updated labelling leads to an updated data repository. Once a significant number of tickets has been re-labelled by expert support agents, the updated data repository is used to re-train the system, starting with the automatic labelling model, leading to more precise data clustering and a better multi-armed bandit model. This ensures a long-lasting higher-quality data repository that is constantly updated with new tickets and thus adapted to the dynamically changing requirements in IT-Support and counteracting potential data shift issues.

4.3. HiL-as-Nexus (solution space)

The worth of customer touchpoints has long been researched in service research, where customer journeys are constituted by service encounters, which are enabled by touchpoints (Hogan, Almquist, & Glynn, 2005; Kronqvist & Leinonen, 2019). From a HI perspective, recent attention has moved towards human-in-the-loop configuration, where experts are involved in augmenting algorithms (Grønsund & Aanestad, 2020; Wiethof & Bittner, 2021). We argue that touchpoints have to augment frontline service encounters (Keyser et al., 2019), the aim is to design an intelligent system that augments support agent problem-solving capabilities (DR2), which keeps on learning after its initial development (DR1) and focuses on the touchpoints of support agents to provide an incentive to engage in the HiL activities (DR3). To accommodate all three design requirements, we propose the design of a HiL agent touchpoint within the context of the operations pipeline as a novel design principle. Furthermore, we argue that the HiL Nexus Point acts as an enabler for humans to act as boundary spanners (Tushman, 1977). 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 *data quality* and ensuring ticket labeling *over time* and *model suggestions*, where the human-in-the-loop takes on the simultaneous role as both system beneficiary and data auditor (Grønsund & Aanestad, 2020). We call the different touchpoints that are present in the HiL a nexus point, 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-HilNxp: ***HiL Nexus Point****- Intelligent IT-frontline technologies should design human-in-the-loop interaction points as locus points that simultaneously provide support agents with immediate utilitarian value and improves labeling quality to ultimately improve the system models in the long term.*

The *HiL Nexus principle* thus integrates all three design requirements (DR1; DR2; DR3).

5. Human-in-the-loop demonstration

This section demonstrates how we used the design principles to guide us in the instantiation of our system (Gregor et al., 2020). Thus, we present our resulting 6 design features (df1 – df6) (Prat et al., 2022) to indicate how support agents would interact with our HiL nexus point (Grønsund & Aanestad, 2020), depicted in **Figure** *4*. During the design of the prototype and the interface, we made sure to

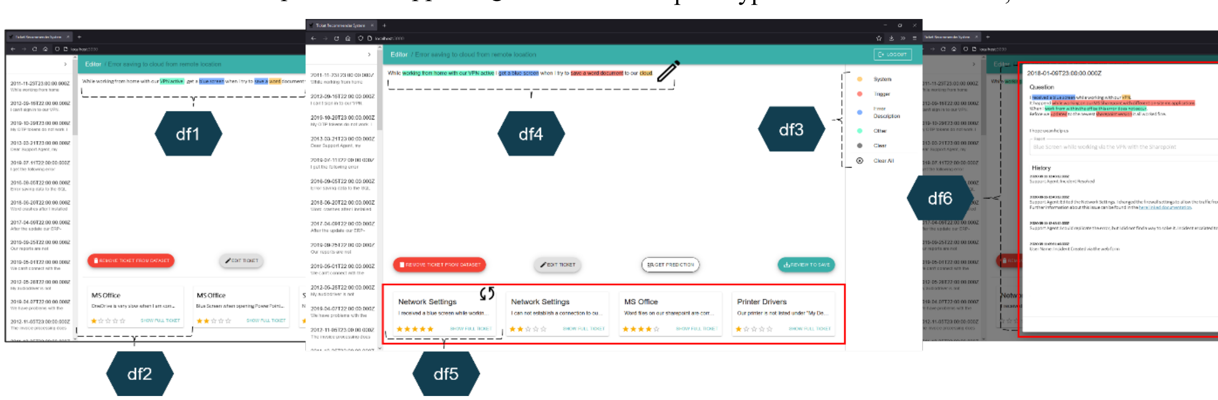

Figure 4. Demonstration of HiL instantiation with design features *df1-df6.*

implement an intuitive and user-friendly system to avoid unnecessary workload increase (Kaber et al., 2001).

The first things support agents see within our system are 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 is incentivized to indicate its 1-star rating uses the appropriate pre-selected label highlights (df3:*label_choice*) to re-label the initial ticket (df4:*label_editing*). The label categories were defined previously by support agents. Next, the system provides revised solutions based on the newly highlighted tickets (df5:*immediate_feedback*) with better results. Lastly, the historical ticket suggestions provide necessary information for the agents to find a solution and solve the customer request, and save its feedback, as indicated by the newly gained 5-star ranking (df6:*solution_materials*).

The design features allow the support agent to be incentivized to check and relabel faulty support tickets and be rewarded with immediate better suggestions (DR3). Furthermore, high-quality solution materials help support agents in their problem-solving activities (DR2), more precisely in 1) locating useful information and 2) providing several suggestions that can be potentially adapted to find a solution for the customer request (Das, 2003). Finally, the HiL nexus point is deeply integrated into the inner workings of both model training and 4 system operations (DR1). Thus, the HiL touchpoint satisfied all three design requirements.

6. Evaluation

To evaluate the learning mechanism of the prediction model, we use the average system 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 is able to find and suggest historical tickets that are close to the incoming ticket. **Figure 5** presents the development of the average system 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 **Figure 5**, both system 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 note 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 preliminary satisfactory result, because it indicates that our system is able to adapt and learn, as is intended with multi-armed bandits, which is in line with an initial proof-of-concept. This result shows that the initial set-up with a small amount of manually labeled data performs sufficiently well to build a self-learning system. 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.

The HiL touchpoint was evaluated with regards to the ticket editing and suggestion evaluation mechanisms, by simulating work environments during operations. Thus, two annotators with expert domain knowledge initiated in 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 a second step, they edited the labeled tickets first, were presented the new set of four suggestions, and evaluated the suggestions the system made based on the annotator-labeled ticket. A total of 45 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 the four ticket suggestions appeared to be useful, and only seldomly did they perceive two or more ticket suggestions 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

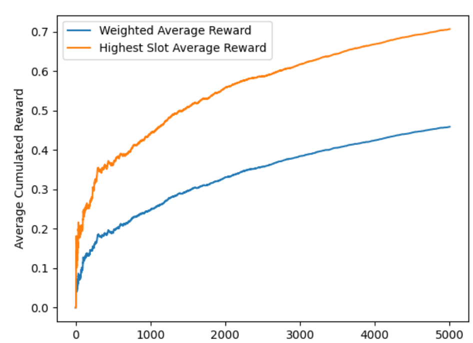

Figure 5. Learning rates over iterations.

likely to be better" (A1). Moreover, they reported that the re-labeling process is "quick" and intuitive. These factors lead to a high motivation to re-label 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.

18 tickets were not rated, since one annotator did not find any new ticket suggestion relevant, and 4 ticket ratings have gotten lower ratings. 13 tickets remained the same. During post-evaluation interviews, the annotators explained that they felt that many of the ticket suggestions did not fit the new support request and suspect 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. Albeit the total number of evaluated instances is comparably low, the results indicate that manual re-labeling by support agents can have a positive impact on 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.

Finally, as a relevance check, we demonstrated both pipeline and prototype to 14 members of our research project since all members are experts in IT frontline services. They included 6 support agents, 3 managers, 3 support system developers, 1 managing director, and 1 work council member across the research consortium. To gain further insights by stimulating a discussion between our experts we opted for conducting a focus group review of our system (Stewart & Shamdasani, 1990). Primary concerns 1) - 3) were satisfactorily addressed and the HiL-nexus design was accepted with only minor concerns relating to the size of UI tiles. Most importantly, they confirmed that the HiL mechanism and the overall system design fully address the design requirements, which indicates a relevance check.

7. Conclusion

In summary, the paper designs two solution artifacts (Prat et al., 2022). First is the overall hybrid intelligence IT-frontline technology to augment frontline services. The second artifact is the human-in-the-loop touchpoint that is nested as a subsystem within the HI system.

For the literature on HI system design and particularly in the context of support services (Bailey & Barley, 2020; Kubiak & Rass, 2018; Poser & Bitner, 2021; Schmidt et al., 2021), we contribute to its body of design knowledge. Specifically, we focus on knowledge for instantiations, presenting the rationale behind design requirements, design principles, and design features (Prat et al., 2022).

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 (Keyser et al., 2019) to augment their work-related problem-solving activities (Das, 2003).

Our *HiL Nexus principle* also contributes to the body of knowledge on HiL design and configurations (Grønsund & Aanestad, 2020; Wiethof & Bittner, 2021). Our design knowledge provides justificatory insights into our HiL configuration. We argue that our HiL design provides both individual and organizational benefits, complementing the HI system's main functionality.

For practice, the paper provides several guidelines and insights. The paper is rooted in a real-world need as elicited by practitioners cooperating in a research development project (Österle & Otto, 2010). All three initial questions 1-3) were addressed. by the design of an integrated HI system, while also reporting details on design decisions of our system implementation, guiding future practitioners deciding 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. 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 six design features guide how the principles can guide developers in instantiating an appropriate system.

Future work should focus on further exploring our initial results, since our three evaluation constitutes a very first design cycle of an ongoing research endeavour (Hevner, 2007; Österle & Otto, 2010) as an initial proof-of-concept (Nunamaker, Briggs, Derrick, & Schwabe, 2015). A formative evaluation (Venable et al., 2016) of each pipeline activity would be subject to future research. More importantly, our limited evaluation needs to be expanded upon, preferably with studies following experimental designs. Thus, we plan to test different types of multi-armed bandits and parameterizations and improve data labeling both for initialization and simulating operations, as a first summative evaluation with more support tickets and support agents as annotators. Additionally, we plan for a naturalistic evaluation design of one year time to test our systems in day to day business of our pilot partners, as means for a proof-of-value (Nunamaker et al., 2015). Furthermore, research on adaptive automation, specifically its adaptive strategies could also provide novel insights into HI system design (Scerbo, 2007). Nonetheless, our paper provides insights into the reasoning behind our HI support system and innovative HiL design and paves the way for future research endeavours.

8. References

Akata, Z., Balliet, D., Rijke, M. de, Dignum, F., Dignum, V., Eiben, G., . . . Welling, M. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer*, *53*(8), 18–28.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). Optics. *ACM SIGMOD Record*, *28*(2), 49–60.

Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, *30*(2), 100286.

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3).

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*(0), 2.

Chandra, L., Seidel, S., & Gregor, S. (2015). *Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditionsNA2015 48th Hawaii International Conference on System Sciences; no H-Index found SJR: NA; VHB: NA; CCF: NA. 2015 48th hawaii international conference on system sciences*. IEEE.

Das, A. (2003). Knowledge and Productivity in Technical Support Work. *Management Science*, *49*(4), 417-431.

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

Drechsler, A., & Hevner, A. (2016). A four-cycle model of IS design science research: capturing the dynamic nature of IS artifact design. *International Conference on Design Science Research in Information Systems and Technology (DESRIST)*.

Giebelhausen, M., Robinson, S. G., Sirianni, N. J., & Brady, M. K. (2014). Touch versus Tech: When Technology Functions as a Barrier or a Benefit to Service Encounters. *Journal of Marketing*, *78*(113-124).

Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, *21*, 1622–1652.

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

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

Hogan, S., Almquist, E., & Glynn, S. E. (2005). Brand-building: finding the touchpoints that count. *Journal of Business Strategy*, *26*(2), 11–18.

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behavior Therapy*, *51*(5).

Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics*, *5*(1), 37–57.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292.

Kajko-Mattsson, M. (2004). Problems within front-end support. *Journal of Software Maintenance and Evolution: Research and Practice*, *16*(45), 309–329.

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

Kronqvist, J., & Leinonen, T. (2019). Redefining Touchpoints: An Integrated Approach for Implementing Omnichannel Service Concepts. In M. A. Pfannstiel & C. Rasche (Eds.), *Service Design and Service Thinking in Healthcare and Hospital Management* (pp. 279–288). Cham: Springer International Publishing.

Kubiak, P., & Rass, S. (2018). An Overview of Data-Driven Techniques for IT-Service-Management. *IEEE Access*, *6*, 63664–63688.

Li, G., Zhou, X., & Cao, L. (2021). Machine Learning for Databases. In *The First International Conference on AI-ML-Systems* (pp. 1–2). New York, NY, USA: ACM.

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-Driven Meets Theory-Driven Research in the Era of Big Data: Opportunities and Challenges for Information Systems Research. *Journal of the Association for Information Systems*, 1253–1273.

Mallick, A., Hsieh, K., Arzani, B., & Joshi, G. (2022). Matchmaker: Data Drift Mitigation in Machine Learning for Large-Scale Systems. *Proceedings of Machine Learning and Systems*, *4*, 77–94.

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

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research. *Journal of Management Information*

Systems, *32*(3), 10–47. https://doi.org/10.1080/07421222.2015.1094961

Österle, H., & Otto, B. (2010). Consortium Research. *Business & Information Systems Engineering*, *2*(5), 283–293.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Pentland, B. T. (1992). Organizing moves in software support hot lines. *Administrative Science Quarterly*.

Poser, M., & Bitner, E. (2021). (Re)Designing IT Support: How Embedded and Conversational AI Can Augment Technical Support Work. *International Conference on Information Systems*, *42*, 1–17.

Poser, M., Wiethof, C., Banerjee, D., Shankar Subramanian, V., Paucar, R., & Bittner, E. A. C. (2022). Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service. In A. Drechsler, A. Gerber, & A. Hevner (Eds.), *Lecture Notes in Computer Science. The Transdisciplinary Reach of Design Science Research* (Vol. 13229, pp. 142–153). Cham: Springer International Publishing.

Prat, N., Akoka, J., Comyn-Wattiau, I., & Storey, V. C. (2022). A Granular View of Knowledge Development in Design Science Research. In (pp. 363–375). Springer, Cham.

Salah, S., Maciá-Fernández, G., Díaz-Verdejo, J. E., & Sánchez-Casado, L. (2016). A Model for Incident Tickets Correlation in Network Management. *Journal of Network and Systems Management*, *24*(1), 57–91.

Scerbo, M. (2007). *Adaptive automation: Neuroergonomics: The brain at work. Oxford series in human-technology interaction*. Oxford (EE. UU): Oxford University Press.

Schmidt, S., Li, M. M., Weigel, S., & Peters, C. (2021). Knowledge is Power: Provide your IT-Support with Domain-Specific High-Quality Solution Material. *International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, 1–14.

Shaw, E. H. (1994). The Utility of the Four Utilities Concept. *Research in Marketing*, *6*(44-66).

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3).

Statista (2022). T-consulting and implementation services market revenue worldwide from 2016 to 2027. Retrieved from https://www.statista.com/forecasts/1079927/it-consulting-implementationservices-revenue

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (1. printing). *Applied social research methods series: Vol. 20*. Newbury Park, Calif.: Sage Publ. Retrieved from http://www.loc.gov/catdir/enhancements/fy0655/90008235-d.html

Tsymbal, A. (2004). The problem of concept drift: definitions and related work. *Computer Science Department, Trinity College Dublin*, *106*(2), 1–58.

Tushman, M. L. (1977). Special Boundary Roles in the Innovation Process. *Administrative Science Quarterly*, *22*(4), 587–605.

Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto. *Journal of Service Research*, *20*(1), 43–58.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89.

Weber, B. J., & Chapman, G. B. (2005). The combined effects of risk and time on choice: Does uncertainty eliminate the immediacy effect? Does delay eliminate the certainty effect? *Organizational Behavior and Human Decision Processes*, *96*(2), 104–118.

Wiethof, C., & Bittner, E. A. C. (2021). Hybrid Intelligence - Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review. *International Conference on Information Systems (ICIS)*, 1–17.

Wolfe, M. T., & Patel, P. C. (2017). Instant gratification: temporal discounting and self-employment. *Small Business Economics*, *48*(4), 861–882.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, *31*(7), 1235–1270.