Augmenting Frontline Service Employee Onboarding via Hybrid Intelligence: Examining the Effects of Different Degrees of Human-GenAI Interaction

Philipp Reinhard[1[0000–0003–3207–666X]], Nicolas Neis[2[0009–0004–1206–3575]], Lisa Kolb[3[0000–0002–2807–7777]], Dennis Wischer[4[0009–0000–4191–881X]], Mahei Manhai Li[1[0000–0002–7171–9130]], Axel Winkelmann[2[0000–0002–2896–566X]], Frank Teuteberg[4[0000–0003–3870–6642]], Ulrike Lechner[3[0000–0002–4286–3184]], and Jan Marco Leimeister[1[0000–0002–1990–2894]]

[1] University of Kassel, Pfannkuchstr. 1, 34121, Germany
{philipp.reinhard, mahei.li,leimeister}@uni-kassel.de
[2] University of Würzburg, Würzburg, 97070, Germany
{nicolas.neis, axel.winkelmann}@uni-wuerzburg.de
[3] University of the Bundeswehr Munich, Neubiberg, 85579, Germany
{lisa.kolb, ulrike.lechner}@unibw.de
[4] University of Osnabrück, Osnabrück, 49074, Germany
dennis.wischer@caballito.io, frank.teuteberg@uni-osnabrueck.de

Abstract. High turnover rates within help desks, caused by excessive workloads, make the efficient onboarding of novices a persistent and recurring challenge. Generative artificial intelligence (GenAI) possesses the potential to augment novice frontline service employees (FSE) during their onboarding phase. However, there is a lack of knowledge on how to design the interaction of FSE and GenAI. Thus, following design science research (DSR), we propose a conversational agent - called co-agent - that leverages the capabilities of large language models and the concept of hybrid intelligence to augment novice FSE. To examine the interaction between novices and GenAI given various task difficulties, we derive and instantiate two hybrid intelligence meta-designs - a *supporter* that provides recommendations and a *collaborator* that allows for prompting the coagent. The results from an online experiment with 75 laypeople show that novices interact with GenAI more frequently and show a higher engagement, especially in difficult tasks. Overall, we uncover a paradox: Despite an increased interaction and a greater time investment, FSEs experience a lower perceived workload with a GenAI-based *collaborator*. From that, we derive implications for designing employee-facing co-agents in customer services.

Keywords: Generative AI, Hybrid Intelligence, Customer Service.

1 Introduction

Onboarding new employees in customer service poses a recurring challenge, especially given the high turnover rate of up to 70% for new hires in customer support [48]. Furthermore, the situation is complicated by the high time-to-performance of new call center agents which make up more than six months on average [3] due to the high workload initially experienced by novices. The new FSE in customer service help desks are typically tasked with problem-solving shortly after their initial formal training. Generally, they handle a spectrum of customer issues ranging from simple customer inquiries to more intricate issues and situations that surpass the agent's capacity, necessitating escalation to an expert. Simple inquiries may often be resolved through referencing FAQs. On the other hand, the resolution of more technical queries typically involves consulting detailed descriptions in ticket documentation or exploring discussions in online forums. After the formal training, novices can handle most support requests they have learned during their first weeks, all without direct customer interaction. However, when faced with unfamiliar issues, novices often spend considerable time resolving them, must escalate cases to dedicated experts, and are under great pressure [56].

To reduce novices' workload while preventing longer upfront training and deteriorating performance, leveraging artificial intelligence (AI) to support novices during real-time customer interactions is recommended [54]. Such an AI-assisted approach can address the limitations of traditional upfront training, potentially reducing resignation rates and improving job satisfaction. While research already aimed at providing initial insights about the impact and potentials of AI coaches [36,22] and assistants [21], we observe a lack of research on the application and design of employee-facing AI-based co-agents and co-pilots [62] to augment novices' problem-solving skills and experiences during their customer conversations. The current challenge involves analyzing the extent of collaboration between humans and AI in human-AI hybrids [17]. To investigate the degree of interaction between novices and AI across diverse task complexities, we establish two hybrid intelligence meta-designs: (1) a *supporter* offering advice and (2) a *collaborator* enabling prompting the co-agent. Through an online experiment involving 75 laypeople, we analyze the impact of these interaction levels and the effect of task difficulty on interaction with AI, task performance, and workload. Hence, we aim to address the following research question: *How does the degree of interaction with emerging GenAI systems influence the workload of novice frontline service employees?*

2 Foundation

Despite the widespread use of AI and primarily conversational agents in customer service, numerous challenges remain (e.g., [2,41]). AI can reduce costs and streamline processes, but personalized service, as demanded by customers, can only be provided to a limited extent [31]. This leads to a trade-off for companies between service efficiency and quality [2]. The remedy is a combination in which humans perform personal customer interaction, but AI augments them in problem-solving. Combining humans and AI creates a socio-technological ensemble that can serve customers' needs [13,49]. This hybrid intelligence can achieve complex goals with better results than individual intelligence [12]. AI system developers must coordinate the novice's interaction with the AI to maximize synergies in hybrid intelligence [13]. The AI-based co-agent we introduce in this paper intervenes in an ongoing interaction between humans and supports novices. The intervention should be designed so that the FSE is not interrupted but receives support in solving the problem [39,58]. The aim is a synergetic integration of humans and AI with a reduced human workload [49]. The hybrid approach facilitates the generation of new knowledge in challenging domains. This enables the FSEs to learn from the AI, and the knowledge of experienced experts can be provided to novices without the need for extensive onboarding [13,61].

The evolving landscape of frontline service, shaped significantly by emerging intelligent technologies, is a central area of investigation across numerous studies [55,11,34]. In particular, AI-driven approaches, such as the regulation of collaborative AI [30] or AI-assisted interpersonal emotion regulation [26], demonstrate research in this area. Beyond the purely technical perspective of using AI-based augmentation, human collaboration with technologies like AI [50] is also an evolving area of research. Whether investigating service quality [8], the effects on user compliance [2], or specifying the role of AI bots in teams [6,4], the focus is on human-machine interaction. Other IS scholars are investigating the combination of the intelligence of machines and humans through hybrid intelligence [49,61]. This involves specifying areas of application [15], determining requirements [56], and deriving design principles [14]. Research in hybrid intelligence combines humans and AI and focuses on fundamental aspects, such as factors influencing cooperation between humans and AI [25,35]. The use of AI does not only have positive effects; depending on previous experience, but divergent effects on performance can also be observed [64]. In addition, several studies have found a negative effect of AI enhancements on users' mental workload [7,57]. Due to its invasive properties, the interaction between humans and AI can lead to stressful situations. The study extends previous research on the design principles of hybrid intelligence systems and focuses on collaboration via prompts. It sheds light on the optimal collaboration between humans and AI and emphasizes the impact on human workload in the context of hybrid intelligence.

3 Research Approach

A design-oriented approach [23] is chosen to conduct an experimental study on the impact of different degrees of human-AI interaction. Hevner et al. (2008) [27] serve as the general work foundation. To identify the relationship between the degree of human-AI interaction and the mental workload, we follow the method proposed by Kuechler and Vaishnavi (2008) [33], shown in Fig.1.

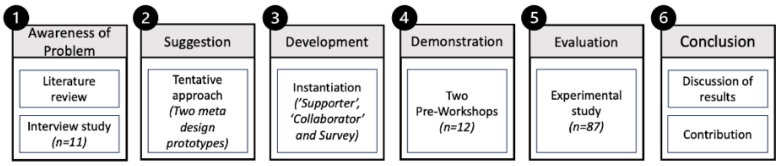

Fig.1. Research Design (cf. [33])

The study's needs, goals, stakeholders, and requirements [37] are identified through a literature review on hybrid intelligence and a previous study on AI augmentation, human-AI interaction, and the FSE journey [51]. Semi-structured interviews are conducted to deepen the findings and identify further aspects. In the suggestion phase, a precise picture of our research design is created within the ISDT framework [42]. The development phase involves designing and implementing two prototypes based on GenAI for an online experiment on human-AI interaction in customer service [18]. We demonstrate and test the prototypes at two workshops to gather input from the field of IT support and improve the high-fidelity prototypes. Afterward, we analyze the impact of two types of human-AI hybrids (*supporter* and *collaborator*) and their influence on the workload of an FSE in an experimental study with n=75 participants. Additionally, a survey is conducted to gather insights into participants' perceptions of the GenAI tool [32]. The final operationalization phase, including findings, contributions, limitations, and further research, encompassing the study results [24].

4 Designing a Co-Agent for Novice Frontline Service Employees

4.1 Awareness

The study utilized semi-structured interviews with support staff and experts to comprehend the needs and challenges faced by practitioners in customer service activities [37]. Along with different FSE routines (e.g., assign, transfer, locate, adapt, generate, retain) derived from technical support theory [9,47], we elaborated on how AI can overcome persistent challenges. We interviewed 11 experts and analyzed the qualitative data through content coding [38] and found several ways to integrate generative AI into support routines in a human-centered way. The interviews showed that generative AI can be used in different ways to process customer requests. Expert 3 explained: "*[...] [that] the tickets [are] not always [...] helpful in the search because they're too long or poorly documented.*" The literature states that the greatest potential lies in decision support and human-AI collaboration [51]. In this way, FSEs retain control of the customer conversation but are accompanied by an AI-based co-agent. This interaction between the FSE and the AI is intended to maximize effectiveness while reducing workload when solving complex tasks [10]. In particular, when a human interacts with a generative AI, increased productivity and an improvement in overall task performance can be observed [43]. Expert 7 describes the following scenario: "*I took some sample tickets [...] and simply changed the ticket description [...]. And then an email was sent directly to the customer [...]. Even though I can't technically assess whether everything in there is correct, I thought it sounded very good*". We have, therefore, identified the problem that the FSE may only receive support from the GenAI to the extent that the FSE must actively process the task to ensure that the response to the customer is correct.

4.2 Suggestion

Our study proposes two GenAI-based co-agents to enhance the FSE onboarding process. We focus on two problem-solving routines, locating and adapting, that require high cognitive effort during customer interactions [9]. This approach is shown in Fig. 2 using the IS design theory (ISDT) approach to connect meta-design, meta-requirements, and constructs in an inner and outer model for our study [42]. For the meta-design, we distinguish between two degrees of interaction: *supporter* and *collaborator*. The *supporter* provides recommendations with minimal interaction, while the *collaborator* enables the FSE to prompt the GenAI-based system.

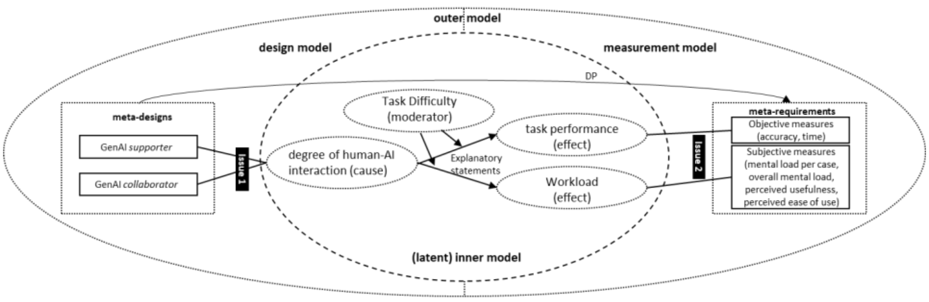

Fig.2. Research design in the ISDT framework based on [42]

The meta-requirements are based on literature and interview results clustered according to key elements. These key elements are integrated into the experiment using a survey to measure the construct's usefulness, ease of use, and mental load (subjective measurements), as well as task time, the accuracy of correctly solved cases, and interaction counts with the GenAI-based support systems based on participants' responses during the experiment (objective measurements). The core relation (inner model) we analyze through the experiment is the influence of the degree of human-AI interaction (cause) on FSEs' mental workload and task performance (effect). *Task difficulty* acts as a moderator with three difficulty levels. Based on this ISDT research model design, we propose a design principle [14], that guides this study: *To augment novice FSE's problem-solving capabilities in real-time customer interactions with various degrees of difficulty, a higher degree of human-GenAI interaction should be enabled because of the complementary strengths of humans and AI in the form of hybrid intelligence.*

4.3 Development

Next, we developed two different prototypes as instantiations of our meta-designs. Participants interact with the prototypes via a web-based chatbot interface developed with the frontend framework Gradio. The prototypes include a GenAI-based customer bot with a specified customer request and a solution reference. Above that, the prototypes are extended with a ”listening” and advice-giving GenAI bot - the co-agent. Additionally, both instantiations contain a search engine based on GPT embeddings. Additionally, we built a backend with a knowledge base and a database to log usage data and messages from the participants and the LLM-generated responses for the customer and the co-agent. We used publicly available data from a mobile provider and generated a data set linked with the co-agent and search engine.

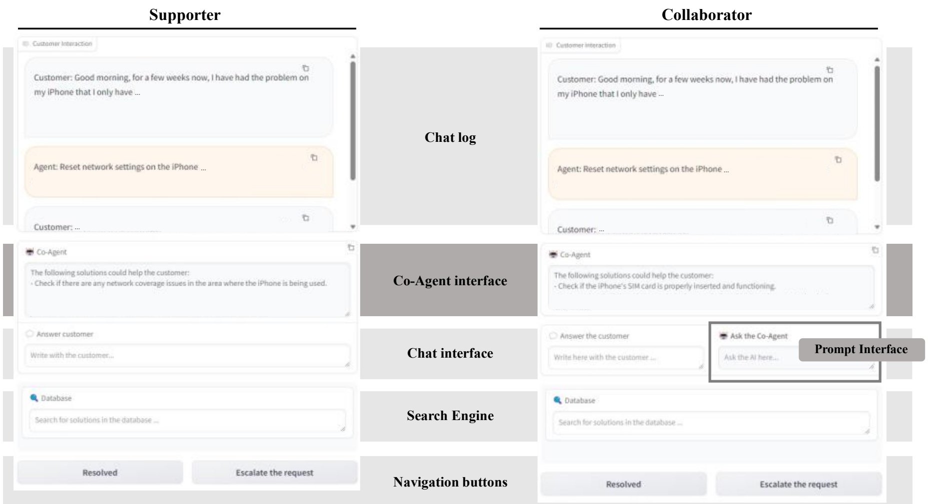

Fig.3. Side-by-side Comparison of Prototypes with Treatment

We used two prototypes to reflect the different degrees of interaction between humans and AI. Co-agent one serves as a *supporter* and co-agent two as a *collaborator*, allowing the FSE to interact with the agent for deeper insights or extra information, as shown in Figure 3. Both co-agent prototypes utilize a state-of-the-art LLM-based architecture [65] with an LLM [46] coupled with a retrieval augmented generation (RAG) system that was instantiated based on LlamaIndex- a framework for connecting GPT models to special knowledge bases and for developing chatbots. The RAG system was tested with a RAGAS evaluation framework [16]. The results ($faithfulness = 0.796$, $answer - relevancy = 0.840$, $context - precision = 0.658$, $context - recall = 0.784$) confirm the efficiency of the conversational agent and show that the co-agent provides reliable results and addresses user needs. For this purpose, we performed excessive prompt engineering. Participants receive a simulated customer request message and have to answer the same questions in the experiment. The tasks are randomly ordered with three levels of difficulty within the three customer requests that participants have to complete. Two cases can be directly solved by the FSE, while the third requires to *escalate* to a second-level agent. The customer request with *low difficulty* was extracted from a FAQ. The customer request with *high difficulty* is based on a forum post and requires more effort to find the correct solution. During the experiment, participants can use the ”solved” and ”escalate” buttons at the bottom of the interface. They can move on to the next case when they ”solved” the customer's request. If the request seems unsolvable, they can ”escalate” and move on to the next task.

4.4 Demonstration

Initially, we developed a low-fidelity prototype as a clickable mockup of the coagent [52]. We demonstrated the experimental study in front of 12 workshop participants, including help desk managers and FSEs. Then, we divided the participants into two focus groups. Each focus group was asked to discuss the general use cases for a co-agent in their organization, the potential for augmenting the onboarding phase, the augmentation interface and user experience, and additional configuration levels. In conclusion, the participants argued for a high potential of the presented use cases within the low-fidelity prototype. The participants stated that FSEs should still have direct access to ticket search engines to add specific details not mentioned during the conversation. Regarding the co-agent's interface, the group concluded that the recommendations should be easily transferred to the chat and, at the same time, revised efficiently. Thus, we integrated a copy function, the co-agent provides full-sentence recommendations, and the text box remains interactive. After instantiating the two high-fidelity prototypes, we introduced the co-agents in a second workshop again to the same group of IT support managers and FSEs and conducted a user test. Given the feedback from the experts regarding the performance, the structure of the AI responses, and the database records format, we revised our design and adjusted our prompts iteratively.

4.5 Evaluation

To evaluate our meta-designs we conducted an experiment, using the final prototypes in a controlled setting. We recruited 102 participants on Prolific and conducted a pre-experiment survey consisting of 13 questions to collect the participants' background, expertise, and AI experience. The participants were randomly assigned to the *supporter* and *collaborator* conditions. We assessed their perceived mental load after each task. Participants completed a post-survey after the experiment, rating their perceptions on a 7-point scale (1 = ”do not agree at all” and 7 = ”fully agree”) on task fulfillment, co-agent perception, AI acceptance, trust, confidence, and workload [1]. We carefully cleaned participants' records to ensure data integrity. We removed 27 participants, leaving a final number of n=75 after verifying the experiment and attention test completion (*supporter* = 42, *collaborator* = 33). Our participant group reflects a broad age range with an average age of 48 years and is almost equally split between males (40%) and females (60%). We conducted an initial correlation analysis between the demographic characteristics and the subjective and objective measurements, which revealed no anomalies that required further investigation.

In the next step, we calculate the mean of different constructs (latent variables) consisting of different items (manifest variables). The constructs are examined through descriptive statistics and frequency analysis [19]. For the final analysis of the ISDT framework, we use linear regression to gain insights into the relationship between the degree of human-AI interaction and mental workload.

Table 1. Perceived Usefulness and Perceived Ease of Use

'Supporter'

	Minimum	Maximum	Mean	Std. Deviation
Perceived Usefulness	4	7	5.506	.746
Perceived Ease of Use	3	7	5.796	.737

'Collaborator'

	Minimum	Maximum	Mean	Std. Deviation
Perceived Usefulness	4	6	5.520	.604
Perceived Ease of Use	5	7	5.868	.605

Perceived Ease of Use and Perceived Usefulness. We analyzed the perception of ease of use and usefulness (α = .903(*ease of use*), α = .789(*usefulness*)). To generate questions on ease of use, we referred to the sources [5,59], and for usefulness, we considered [29,59]. Our findings, presented in Table 1, highlight the differences between the two levels of interaction degree. We observed a slight increase in the mean value for the *collaborator* condition. This could be attributed to the additional interaction with the co-agent, which allowed participants to better understand the system's functionalities and share an additional response with the customer.

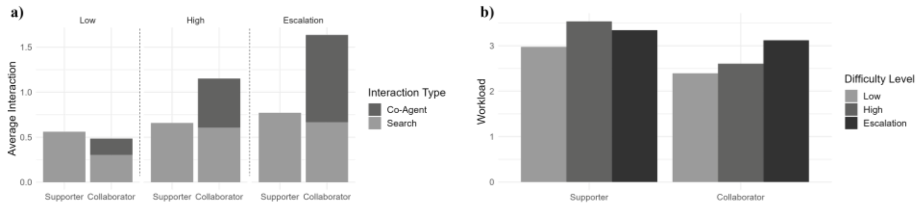

Fig.4. a) Average interaction with co-agent and search by condition and by task difficulty; b) Average workload by condition and task difficulty

Task Performance. In both conditions, the results revealed that the accuracy of solving the tasks remains comparatively low and decreases with increased task difficulty, as expected. For instance, under the *supporter* scenario, the average accuracy for resolving less complex, FAQ-based tasks stands at 92.68 %, while accuracy plummets to 31.43 % for cases necessitating escalation to an expert. Hence, accuracy was mainly influenced by the quality of the underlying data source, with distinct outcomes observed between FAQ-based and forum-based information. Specifically, when the generative co-agent utilized FAQ data, a high level of accuracy was achieved, underscoring the reliability of structured and well-curated content in supporting decision-making processes. To compare task performance, we developed a measure consisting of accuracy and time for each case by normalizing the values. That way, we found that despite showing a slightly lower accuracy, the *collaborator* enables a slightly improved overall task performance (*supporter* = 1.191, *collaborator* = 1.305). Although our regression analysis did not reveal any significant results, we can conclude that the treatment does not adversely affect performance. Instead, task performance is slightly improved.

Interaction between FSE and AI. By comparing *collaborator* and *collaborator* conditions, we observed notable differences in interaction dynamics between FSE and the provided tools. Overall, the *collaborator* condition showed an increase of 65.76 % in terms of interactions on average, involving both search activities and communication with the co-agent. This uptick was particularly pronounced in more complex and unsolvable cases (Figure 5a), indicating that collaboration intensifies when challenges escalate. Our design promoted a higher degree of engagement and interaction with the generative AI-based co-agent, suggesting a strategic shift towards more collaborative problem-solving approaches. Interestingly, the search engine usage was almost as frequent in the *collaborator* scenario as it was in the *supporter* case.

Workload. Despite higher time investment and interaction count, participants in the *collaborator* condition reported a lower mental load per case while maintaining performance levels (Figure 4b), indicating that interaction with the co-agent contributes to a reduced workload. This mental load measure was collected at the end of each task. On average, the meta-design *collaborator* reduced the perceived mental load by 0.58 points. In a regression analysis assessing the impact of the *collaborator* condition on workload, the condition was associated with a reduction in reported workload (β = −0.575, *SE* = 0.220, t = −2.612, p = 0.00965). However, we could not find any moderating effects of task difficulty. Furthermore, these findings imply that AI collaboration could positively affect employee retention, potentially reducing turnover rates. Results from the post-survey confirm these insights. As Table 2 shows, the subjective measures reveal that the overall mean perceived workload (α = .555(*overall workload*)) is lower in the case of the *collaborator* condition.

Table 2. Perceived mental load per task and overall perceived workload

'Supporter'

	Minimum	Maximum	Mean	Std. Deviation
Mental load per case and user	1	6	3.310	1.137
Overall mental load	1	7	3.857	1.601

'Collaborator'

	Minimum	Maximum	Mean	Std. Deviation
Mental load per case and user	1	5	2.707	1.304
Overall mental load	1	6	3.606	1.223

5 Discussion

In light of these results, we can summarize an *interaction effort paradox* as an unexpected finding: Despite having significantly more interaction with the search engine and the co-agent overall, mental load is significantly lower in the case of the *collaborator* condition with the ability to prompt the co-agent. Thereby, we contradict prior research in human-AI collaboration that highlights the mental effort of interacting with digital technology - such as artificial agents or co-bots [7,57]. In that case, information systems not only benefit human users but cause techno-stress or, more specifically, techno-overload [7]. However, based on our results, we argue that engaging with a co-agent through prompts is more likely to alleviate FSE's task demands than to impose mental effort. The paradox can, therefore, be viewed from a perspective of cognitive offloading [53]. Similar results were found when tracking several objects [60] or in human-AI collaboration in industrial tasks [20]. According to our study, cognitive offloading should be further examined as critical in highly demanding real-time customer interactions at the service frontline. An additional explanation could be that the natural language-based recommendations with the co-agent reduce information overload when compared with the full knowledge base articles from the database [44]. In conclusion, research and practice are urged to examine the positive consequences of AI collaborators in challenging service tasks [28].

In further detail, the results regarding the perceived utility of the co-agent demonstrate great satisfaction with the co-agent. As illustrated by the high average values for perceived usefulness and ease of use illustrate that, the co-agent is likely to be accepted at the workplace of novice FSE. However, the difference between *supporter* and *collaborator* is only marginal. Thus, there are unanswered questions on improving the prompt interface's utility in the second condition. For instance, prompt examples and templates could be provided to stimulate the usage of the co-agent as shown in related work [63]. It remains unclear how novice FSE can effectively prompt RAG-based GenAI tools. A comparison of the provided search engine and the prompt interface is required. Regarding the impact of the *collaborator* on task performance, we found no significant evidence for increased productivity regardless of the difficulty of customer issues. The non-detection of an influence on performance could be due to the study design and task selection, mainly the ease and unsolvability of the low-difficulty task and escalation task. However, we observed no significant results despite assuming a potential influence on the task with a high difficulty level. Our findings indicate that accuracy primarily depends on the task difficulty and the quality of the underlying data source used in the solution material. Utilizing FAQ data resulted in high accuracy, highlighting the reliability of well-structured content for decision-making and the efficiency of GenAI-based conversational agents. Given our design of experimental conditions, the *collaborator* resulted in an intensified overall interaction of the FSE with the given co-agent and search engine. Hence, the co-agent is perceived as a helpful collaborator. Interestingly, the search engine was nearly used as often as in the case of the *supporter*. Thus, further analysis of the usage patterns, the quality of search engine results, and the coagent response are required. Despite the higher overall interaction, we observed a lower perceived workload within the *collaborator* condition. As a key challenge for companies in the customer support sector refers to retaining their employees in the long term to minimize onboarding efforts and reduce costs [48], reducing the workload is the first step towards increasing attractiveness and thus reducing the churn rate. From that perspective, our results show the potential of GenAI for help desks to improve workplace conditions.

This study contributes to both service science and DSR literature by extending frontline service research through the implementation of AI augmentation [26,36] and studying the different meta-designs of human-GenAI interaction [40]. Our empirical insights shed light on how AI can positively impact novices' perceived workload during initial customer interaction, achieved through a GPT4-based co-agent. Furthermore, we showcased the utilization of emerging GPT-based bots as simulated customers in service triads [45,11]. Through effective and iterative prompt engineering, we paved the way for promising DSR research on AI augmentation in frontline service triads. Adapting the perspective of outer and inner models according to [42], our approach illustrates the integration of experimental and design-oriented research. To this end, our experiment instantiation including the corresponding data analysis pipeline could be generalized into a method and architecture which other researchers can draw on to conduct experiments in other related domains. This study's practical contributions lie in guiding the design of AI-based co-pilots and co-agents for frontline services and offering design decisions to improve employees' workplaces. Workplaces could be further improved by also introducing and studying the effects of co-pilots in other functions within a company such as the sales department.

Given the functionality of the GenAI-based co-agent and the simulated customer bot, our empirical design study has certain limitations and leaves room for further research. First, due to the generative nature of LLMs, the co-agent does not generate invariable recommendations despite prompt engineering, configuring the temperature, and connecting the co-agent to the database. Thus, further research should evaluate the co-agent's responses to allow for the analysis of issues such as overreliance. However, by testing the RAG system we ensured that the co-agent provides accurate responses. Similarly, the customer bot implies variations. Thus, our results are restricted by the uncontrollable nature of AI. Furthermore, broadening the scope of the study to include a wider range of tasks would provide a more thorough perspective on the co-agent's performance in various contexts. Investigating the efficiency of the co-agent in comparison to the search engine, as well as the prompts and queries formulated by the users, is essential for interpreting the quality of interaction. Lastly, contrasting situations involving the co-agent with those without it will show the influence and benefit of incorporating conversational AI into FSE routines. Thus, the paper allows for numerous avenues of future research. DSR researchers could examine various designs to reach more beneficial degrees of human-AI interaction - for example, by designing mechanisms to stimulate prompting or providing example prompts and templates. After enlarging the sample size, subsequent research could prioritize text-mining techniques for a more comprehensive analysis of the transcripts of customer interactions.

6 Conclusion

Our results show that the design of GenAI-based and employee-facing co-agents and co-pilots in customer service represents a pressing and challenging problem. Based on a practice and theory-driven DSR approach, we developed a ”whispering” co-agent based on generative AI. Apart from minor design nuances discussed with IT support FSEs and managers, the paper emphasizes a broader dimension of designing human-GenAI interaction by realizing hybrid intelligence meta-designs. Through an experimental online setup with two distinct conditions, we evaluated two meta-designs for GenAI-driven employee augmentation: a supportive co-agent offering unidirectional decision support and a collaborative co-agent facilitating interaction via. The analysis of more than 300 customer interactions reveals that increased interaction with GenAI-based co-agents can improve FSE’s workload during real-time customer service interactions.

Acknowledgements

This research is partly funded by the German Federal Ministry of Education and Research (BMBF) and supervised by PTKA (Project HISS - 02K18D060). Additionally, this work has been developed in the project Hyko and is partly funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy enabled by the European Regional Development Fund (ERDF) under grant number 1.2-StMWK-F.4-UFR-002. The authors are responsible for the content of this publication.

References

1. Herm, L.V.: Impact of explainable ai on cognitive load: Insights from an empirical study. In: ECIS 2023. Research Papers. 269 (2023)
2. Adam, M., Wessel, M., Benlian, A.: AI-based chatbots in customer service and their effects on user compliance. Electronic Markets **31**(2), 427–445 (2021)
3. Attri, R.K.: Accelerated Proficiency for Accelerated Times: A Review of Key Concepts and Methods to Speed Up Performance. Speed To Proficiency Research (2020)
4. Bao, Y., Cheng, X., de Vreede, T., de Vreede, G.J.: Investigating the relationship between AI and trust in human-AI collaboration. In: HICSS. pp. 607–616 (2021)
5. Benbasat, I., Wang, W.: Trust in and adoption of online recommendation agents. Journal of the association for information systems **6**(3), 4 (2005)
6. Bittner, E.A.C., Oeste-Reiß, S., Leimeister, J.M.: Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. In: HICSS (2019)
7. Carissoli, C., Negri, L., Bassi, M., Storm, F.A., Delle Fave, A.: Mental workload and human-robot interaction in collaborative tasks: A scoping review. International Journal of Human-Computer Interaction pp. 1–20 (2023)
8. Chen, Q., Gong, Y., Lu, Y., Tang, J.: Classifying and measuring the service quality of AI chatbot in frontline service. Journal of Business Research **145**, 552–568 (2022)

9. Das, A.: Knowledge and productivity in technical support work. Management Science **49**(4), 416–431 (2003)
10. Davenport, T., Guha, A., Grewal, D., Bressgott, T.: How artificial intelligence will change the future of marketing. Journal of the Academy of Marketing Science **48**, 24–42 (2020)
11. De Keyser, A., Ko¨cher, S., Alkire (n´ee Nasr), L., Verbeeck, C., Kandampully, J.: Frontline Service Technology infusion: conceptual archetypes and future research directions. Journal of Service Management **30**(1), 156–183 (2019)
12. Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., Ebel, P.: The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In: HICSS (2019)
13. Dellermann, D., Ebel, P., So¨llner, M., Leimeister, J.M.: Hybrid intelligence. Business & Information Systems Engineering **61**, 637–643 (2019)
14. Dellermann, D., Lipusch, N., Ebel, P., Leimeister, J.M.: Design principles for a hybrid intelligence decision support system for business model validation. Electronic Markets **29**, 423–441 (2019)
15. Ebel, P., Söllner, M., Leimeister, J.M., Crowston, K., de Vreede, G.J.: Hybrid intelligence in business networks. Electronic Markets **31**, 313–318 (2021)
16. Es, S., James, J., Espinosa-Anke, L., Schockaert, S.: Ragas: Automated evaluation of retrieval augmented generation. arXiv preprint arXiv:2309.15217 (2023)
17. Fabri, L., Haäkel, B., Oberla¨nder, A.M., Rieg, M., Stohr, A.: Disentangling HumanAI Hybrids. Business & Information Systems Engineering pp. 1–19 (2023)
18. Fink, L.: Why and how online experiments can benefit information systems research. Journal of the Association for Information Systems **23**(6), 1333–1346 (2022)
19. Fisher, M.J., Marshall, A.P.: Understanding descriptive statistics. Australian Critical Care **22**(2), 93–97 (2009)
20. Fournier, E., Kilgus, D., Landry, A., Hmedan, B., Pellier, D., Fiorino, H., Jeoffrion,´ C.: The impacts of human-cobot collaboration on perceived cognitive load and usability during an industrial task: an exploratory experiment. IISE Transactions on Occupational Ergonomics and Human Factors **10**(2), 83–90 (2022)
21. Freire, S.K., Panicker, S.S., Ruiz-Arenas, S., Rusa´k, Z., Niforatos, E.: A cognitive assistant for operators: Ai-powered knowledge sharing on complex systems. IEEE Pervasive Computing **22**(1), 50–58 (2022)
22. Graßmann, C., Schermuly, C.C.: Coaching with artificial intelligence: Concepts and capabilities. Human Resource Development Review **20**(1), 106–126 (2021)
23. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS quarterly pp. 337–355 (2013)
24. Gregor, S., Jones, D., et al.: The anatomy of a design theory. Association for Information Systems (2007)
25. Hemmer, P., Schemmer, M., Vo¨ssing, M., Kühl, N.: Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. In: PACIS (2021)
26. Henkel, A.P., Bromuri, S., Iren, D., Urovi, V.: Half human, half machine – augmenting service employees with AI for interpersonal emotion regulation. Journal of Service Management **31**(2), 247–265 (2020)
27. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. Management Information Systems Quarterly **28**(1), 6 (2008)

28. Hollender, N., Hofmann, C., Deneke, M., Schmitz, B.: Integrating cognitive load theory and concepts of human–computer interaction. Computers in human behavior **26**(6), 1278–1288 (2010)
29. Hone, K.S., Graham, R.: Towards a tool for the subjective assessment of speech system interfaces (SASSI). Natural Language Engineering **6**(3-4), 287–303 (2000)
30. Huang, M.H., Rust, R.T.: A Framework for Collaborative Artificial Intelligence in Marketing. Journal of Retailing **98**(2), 209–223 (2022)
31. Khan, S., Iqbal, M.: AI-Powered Customer Service: Does it Optimize Customer Experience? In: ICRITO. pp. 590–594. IEEE (2020)
32. Krosnick, J.A.: Questionnaire design. The Palgrave handbook of survey research pp. 439–455 (2018)
33. Kuechler, B., Vaishnavi, V.: On theory development in design science research: Anatomy of a research project. EJIS **17**(5), 489–504 (2008)
34. Larivi`ere, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wu¨nderlich, N.V., De Keyser, A.: "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. Journal of Business Research **79**, 238–246 (2017)
35. Li, M.M., Reinhard, P., Peters, C., Oeste-Reiss, S., Leimeister, J.M.: A value cocreation perspective on data labeling in hybrid intelligence systems: A design study. Information Systems **120**, 102311 (2024)
36. Luo, Qin, F..Q.: Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions. Journal of Marketing **85**(2), 14–32 (2021)
37. Maedche, A., Gregor, S., Morana, S., Feine, J.: Conceptualization of the problem space in design science research. In: DESRIST. pp. 18–31 (2019)
38. Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken. Beltz (2022)
39. McFarlane, D.C., Latorella, K.A.: The scope and importance of human interruption in human-computer interaction design. Human-Computer Interaction **17**(1), 1–61 (2002)
40. Murray, A., Rhymer, J., Sirmon, D.G.: Humans and technology: Forms of conjoined agency in organizations. Academy of Management Review **46**(3), 552–571 (2021)
41. Nicolescu, L., Tudorache, M.T.: Human-computer interaction in customer service: The experience with AI chatbots—a systematic literature review. Electronics **11**(10) (2022)
42. Niehaves, B., Ortbach, K.: The inner and the outer model in explanatory design theory: the case of designing electronic feedback systems. EJIS **25**, 303–316 (2016)
43. Noy, S., Zhang, W.: Experimental evidence on the productivity effects of generative artificial intelligence. Science **381**, 187–192 (2023)
44. Oc´on Palma, M.d.C., Seeger, A.M., Heinzl, A.: Mitigating information overload in e-commerce interactions with conversational agents pp. 221–228 (2020)
45. Odekerken-Schro¨der, G., Mennens, K., Steins, M., Mahr, D.: The service triad: an empirical study of service robots, customers and frontline employees. Journal of Service Management **33**(2), 246–292 (2021)
46. OpenAI: GPT-4 Technical Report. arXiv preprint arXiv:2303.08774 (2023)
47. Pentland, B.T.: Organizing moves in software support hot lines. Administrative Science Quarterly pp. 527–548 (1992)
48. Pierre, X., Tremblay, D.G.: Levels of Involvement and Retention of Agents in Call Centres: Improving Well-Being of Employees for Better Socioeconomic Performance. Journal of Management Policy and Practice **12**(5), 53–71 (2011)

49. Poser, M., Wiethof, C., Banerjee, D., Shankar Subramanian, V., Paucar, R., Bittner, E.A.: Let's Team Up with AI! Toward a Hybrid Intelligence System for Online Customer Service. In: DESRIST. pp. 142–153 (2022)
50. Raisamo, R., Rakkolainen, I., Majaranta, P., Salminen, K., Rantala, J., Farooq, A.: Human augmentation: Past, present and future. International Journal of HumanComputer Studies **131**, 131–143 (2019)
51. Reinhard, P., Li, M., Peters, C., Leimeister, J.M.: Generative AI in Customer Support Services: A Framework for Augmenting the Routines of Frontline Service Employees. In: HICSS (2024)
52. Reinhard, P., Wischer, D., Verlande, L., Neis, N., Li, M.: Towards designing an ai-based conversational agent for on-the-job training of customer support novices. In: DESRIST (2023)
53. Risko, E.F., Gilbert, S.J.: Cognitive offloading. Trends in cognitive sciences **20**(9), 676–688 (2016)
54. Ritz, E., Fabio, D., Elshan, E., Rietsche, R.: Artificial Socialization? How Artificial Intelligence Applications Can Shape A New Era of Employee Onboarding Practices. In: HICSS (2023)
55. Robinson, S., Orsingher, C., Alkire, L., De Keyser, A., Giebelhausen, M., Papamichail, K.N., Shams, P.: Frontline encounters of the AI kind: An evolved service encounter framework. Journal of Business Research **116**, 366–376 (2020)
56. Schmidt, S.L., Li, M., Peter, C.: Requirements for an IT Support System based on Hybrid Intelligence. In: HICSS (2022)
57. Sun, Y., Li, S., Yu, L.: The dark sides of AI personal assistant: effects of service failure on user continuance intention. Electronic Markets **32**(1), 17–39 (2022)
58. van Turnhout, K., Terken, J., Bakx, I., Eggen, B.: Identifying the intended addressee in mixed human-human and human-computer interaction from non-verbal features. In: ICMI. pp. 175–182 (2005)
59. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS quarterly pp. 425–478 (2003)
60. Wahn, B., Schmitz, L., Gerster, F.N., Weiss, M.: Offloading under cognitive load: Humans are willing to offload parts of an attentionally demanding task to an algorithm. PLoS ONE **18**(5) (2023)
61. Wiethof, C., Bittner, E.A.: Toward a hybrid intelligence system in customer service: collaborative learning of human and AI. In: ECIS (2022)
62. Wiethof, C., Poser, M., Bittner, E.A.C.: Design and Evaluation of an EmployeeFacing Conversational Agent in Online Customer Service. In: PACIS (2022)
63. Zamfirescu-Pereira, J., Wong, R.Y., Hartmann, B.: Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In: CHI. pp. 1–21 (2023)
64. Zhang, G., Raina, A., Cagan, J., McComb, C.: A cautionary tale about the impact of AI on human design teams. Design Studies **72**, 100990 (2021)
65. Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.Y., Wen, J.R.: A Survey of Large Language Models. arXiv preprint arXiv:2303.18223 (2023)