Customer Service Combining Human Operators and Virtual Agents: A Call for Multidisciplinary AI Research

Sarit Kraus, Yaniv Oshrat, Yonatan Aumann, Tal Hollander, Oleg Maksimov, Anita Ostroumov, Natali Shechtman

Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel
(sarit@cs.biu.ac.il)

Abstract

The use of virtual agents (bots) has become essential for providing online assistance to customers. However, even though a lot of effort has been dedicated to the research, development, and deployment of such virtual agents, customers are frequently frustrated with the interaction with the virtual agent and require a human instead. We suggest that a holistic approach, combining virtual agents and human operators working together, is the path to providing satisfactory service. However, implementing such a holistic customer service system will not, and cannot, be achieved using any single AI technology or branch. Rather, such a system will inevitably require the integration of multiple and diverse AI technologies, including natural language processing, multi-agent systems, machine learning, reinforcement learning, and behavioral cloning; in addition to integration with other disciplines such as psychology, business, sociology, economics, operation research, informatics, computer-human interaction, and more. As such, we believe this customer service application offers a rich domain for experimentation and application of multidisciplinary AI. In this paper, we introduce the holistic customer service application and discuss the key AI technologies and disciplines required for a successful AI solution for this setting. For each of these AI technologies, we outline the key scientific questions and research avenues stemming from this setting. We demonstrate that integrating technologies from different fields can lead to a cost-effective successful customer service center. The challenge is that there is a need for several communities, each with its own language and modeling techniques, different problem-solving methods, and different evaluation methodologies, all of which need to work together. Real cooperation will require the formation of joint methodologies and techniques that could improve the service to customers, but, more importantly, open new directions in cooperation of diverse communities toward solving joint difficult tasks.

Introduction

Virtual agents (bots) are increasingly used for providing online assistance to customers, even for complex or moderately complex tasks. However, even though a lot of effort is dedicated to the research (Luo et al. 2022), development, and deployment of these virtual agents, customers are frequently frustrated when the virtual agent is not capable of providing an appropriate service (Ashfaq et al. 2020) and would like to interact with humans (Panko 2019; Press 2019). It reached the point where some organizations advertise their services by announcing that their customers will be able to interact with humans rather than virtual agents. A hybrid customer-service system where virtual agents work along with human operators could address this difficulty.

The second difficulty is the cost and domain knowledge required when initially deploying the agent and the constant need for its update (BluePrint 2020; Carr 2020). Without constant updates of the virtual agents, they become non-relevant. A hybrid customer-service system provides many opportunities for addressing this second difficulty as well.

We propose a holistic approach of virtual agents and human operators working together to provide satisfactory service. The goal is that the virtual agents will incrementally learn from human operators and be able to handle the majority of the requests quickly and efficiently, while the human operators are employed for those requests that cannot be attended to in this way, and for cases where clients specifically seek human assistance. Implementing this configuration can provide a speedy, inexpensive service, attending to more clients with fewer human operators, and providing satisfactory service without the extensive need for explicit domain knowledge acquisition and integration.

The development of such a holistic approach yields many challenges, partially and separately addressed in different communities within Artificial Intelligence, as well as other domains of Computer Science, Operations Research and the Social Sciences. However, solving each of these challenges separately will not lead to the desired goal. Rather, true integration of methods from the different fields is needed. We will elaborate on this need throughout the paper and especially in Section .

HoCuS: A Meeting of Disciplines. There are four entities that participate in such a *Ho*listic *Cu*stomer *S*ervice system (HoCuS): the *customer*, the *virtual agent*, the *human operator*, and the *management* of the entire operation of the customer service center.

The virtual agent consists of several modules, most commonly built using NLP and ML/DL techniques but also using knowledge sciences techniques, social sciences, and Computer-Human Interactions (CHI) methodologies.

The customer interaction with the virtual agents and the human operators is studied using technologies of CHI, knowledge sciences including data mining, ML, GUI/UI, social sciences and Operation Research methodologies. The performance of the human operators and their interaction with the customers have been studied in CHI, GUI/UI, knowledge sciences, ML and OR communities. Supporting human operators in their pressing situations has been studied in MAS. Finally, customer service center management has been studied extensively in Operations Research, theoretical CS, and MAS. Problems such as handoffs between virtual agents and human operators, evaluation of the operators and the overall system have been studied in social sciences, CHI and ML.

Paper Outline. In this paper, we first discuss relevant work that was done in each of the respective research fields, and outline the key scientific questions and research avenues we identify as stemming from the HoCuS setting. Unfortunately, it is not possible to review here all the relevant approaches and related works. Indeed, the associated virtual agents' literature alone is already too large to fully review (see, for example, the (Allouch, Azaria, and Azoulay 2021) survey on conversational agents, consisting of 48 pages and citing 252 references). Therefore, the discussion will be focused on a setting of a service center that provides help to clients who are engaged in performing a complex activity that consists of a list of tasks. The tasks are sequential and may depend on each other, i.e., to be able to perform a certain task, there may be a need to perform a previous task(s) successfully. In this context, we will refer to a relatively simple hybrid customer-service system where a Q&A virtual agent collaborates with human operators.

After demonstrating the interdisciplinary nature of the HoCuS setting, we present a methodology to develop such a system, focusing on the development of the virtual agent in a manner that requires little domain expertise and adopts techniques from various research disciplines. We show that such an agent, though relatively easy to develop, can significantly improve the service (measured by both customers' waiting times and the number of human operators required to operate the service) while maintaining customer satisfaction.

We conclude the paper by discussing the challenges of adopting such a multidisciplinary approach. Currently, different research communities often use different terminologies for the same or similar concepts, methods, techniques, and evaluation methodologies. Our experience and those of others indicate that overcoming these differences will not be easy. However, we believe that the customer service application is a fitting platform and goal for mobilizing collaboration: It is an important application, cooperation is necessary, and success criteria are easy to agree upon.

Multi-discipline Relevant Research for HoCuS

In this section, we will review the four entities that participate in HoCuS, the main issues regarding each entity in the context of the HoCuS design, and the scientific domains in which they are dealt with. We also detail the main research questions that should be considered when implementing such a system. As mentioned in the introduction, the review and discussion of this paper are mostly focused on HoCuS for Q&A virtual agent collaborations with human operators to support clients who are engaged in performing a complex activity that consists of a list of tasks. In this context, we wish to demonstrate the wide range of academic and commercial domains related to HoCuS and conversational agents. Therefore, after each citation and research question, we denote in brackets the domain in which this item is investigated. The domains are:
(NLP) – natural language processing,
(MAS) – multi-agent systems and intelligent systems,
(CHI) – computer-human interaction,
(ML/DL) – machine learning / deep learning,
(RL/BC) – reinforcement learning / behavioral cloning,
(UI/GUI) – user interface / graphic user interface,
(OR) – operation research,
(social science) – psychology/marketing / sociology/economics / business, and
(knowledge science) / (informatics) / (data mining).
We also present **(review)** papers and **(commercial)** websites that shed light on relevant issues. We focus on the methods and components of the Q&A virtual agent collaborations with human operators to support clients, but we also mention other methods and components in the domain of conversational agents in order to demonstrate the multidisciplinary effort that is being made and should be made to encircle the many aspects of the domain.

Virtual Agent

The goals of the virtual agent in HoCuS are to attend to customers and to be able to answer most of their questions satisfactorily. It also has to sense the customers' content regarding its answers and identify when its performance is unsatisfactory, hence the conversation should be transferred to a human operator. The handoff process, transferring the control from the virtual agent to a human operator, is discussed in Section .

In order to provide an answer, the agent first needs to understand the client's question. A major issue is acquiring a sufficient number of examples to train the virtual agent. Several works discuss domain adaptation and transfer learning (Li 2012) **(review)** – taking a framework or a model that was built upon a specific domain and implementing it in another domain. One such example is a domain adaptation framework based on curriculum learning and domain-discriminative data selection (Ma et al. 2019) **(NLP)**. Other works were stimulated by the COVID-19 pandemic to cope with the lack of COVID-19-related data sets when there was a great need to provide information to the crowd. One such case is in (Sun and Sedoc 2020) **(NLP)**.

Another approach, implemented for the sake of the same COVID-19 distress, is matching current customers' questions to previously stored questions in order to use the latter's answers (McCreery et al. 2020) **(Knowledge Science/-Data Mining)**. This approach is often based on the ability to rapidly find similarity between sentences and/or questions, a topic that was researched extensively in recent years, and produced efficient models such as BERT (Devlin et al. 2019) **(NLP)**, RoBERTa (Liu et al. 2019) **(NLP)** and SBERT (Reimers and Gurevych 2019) **(NLP)**. After understanding the question, the virtual agent needs to generate a proper answer (Fu and Feng 2018; Wei and Zhang 2019) **(NLP)**, (Gao et al. 2021; Deng et al. 2022) **(informatics)**. In this model, which we espoused in our work, the question is taken from the DB of previous questions-and-answers. Thus, finding the right question also indicates the correct answer. Nevertheless, a single question in the DB might have more than one existing answer, and there is still a need to find the most adequate one to give to the client.

It should be noted that the traditional method of rule-based conversational agents, which was first introduced during the 1960's (Weizenbaum 1966) **(ML)**, is still very common today in academic research (Singh, Joesph, and Jabbar 2019) **(NLP)**, (Thorat and Jadhav 2020) **(review)**, commercial applications (Franceschi 2020; LandBot 2022; Zabój 2022) **(commercial)** and user guides (Kalinin 2020; Shahid 2020) **(commercial)**. Indeed, these implementations usually do not contain learning abilities and self-development of the model during its work, but in many cases they provide sufficient solutions to the practical need (Leah 2022) **(commercial)**.

In the context of our model of HoCuS, the main research questions regarding the virtual agent are:

1. At the beginning of the process of collecting pairs of question-and-answer and filling up the DB, there is insufficient data for the agent to offer any practical help to customers, and they should be routed directly to a human operator. As the DB is populated, the ability of the agent to provide answers gradually increases. When is the right point (in terms of DB readiness) at which to start routing clients to the virtual agent, knowing that its ability to provide answers is good enough to be worth the bother to the clients? **(CHI, social science)**
2. How to estimate the match between the customer's intention and questions in the DB? Is it acceptable to ask the customer whether this was their intention? If it is acceptable on the first attempt, is it still so on the following ones? **(NLP, CHI, social science)**
3. How many attempts to match the customer's question with a previous question can we make before transferring the control to a human operator? Is it possible to track signals (voice, prosody etc.) with which the customer is about to lose patience? **(social science, CHI, ML)**
4. How to choose the best answer from a set of previous answers given to a certain question? Is it necessary to choose, or can we offer more than one answer and let the client decide? **(NLP, social science)**

Customer

The customers of HoCuS are both a part of the system and its end user. Therefore, we should investigate the customers' behavior during the interaction with other parts of the system (in terms of efficiency, effectiveness etc.), and their satisfaction with the system as its clients. Note that these two aspects may contradict one another: It may be that a certain feature is effective in the performance of the system (e.g., time-saving), but at the same time annoying to the clients. In the former aspect, there is much research regarding customer services in general and virtual agent-assisted customer service in particular. Major issues related to HoCuS are the user interface (Angga et al. 2015) **(knowledge science/data-mining)**, (Liu and Dong 2019) **(CHI)**, waiting time habits and preferences (Wintersberger, Klotz, and Riener 2020) **(CHI)**, (Olivares, Musalem, and Yung 2020) **(social science)**, difference between various types of customers (Molina-Castillo, López-Nicolás, and Bouwman 2008) **(knowledge science/informatics/data-mining)**, (Fotheringham and Wiles 2022) **(social science)** and various customers' emotional states and intents (Pawlik et al. 2022) **(CHI)**, (Crolic et al. 2022) **(social science)**, (Xu et al. 2020) **(knowledge science/informatics)**, (Chen et al. 2019) **(ML)**, (Bojanić, Delić, and Karpov 2020) **(CHI, ML)**.

In the latter aspect, it is crucial to measure customers' satisfaction. Satisfaction may be examined as the general attitude towards the technology, such as evaluating the maturity of the service (Ashktorab et al. 2019), **(CHI)**, studying organizational factors affecting the successful implementation of virtual agents for customer service (Zhang, Følstad, and Bjørkli 2021) **(social science)**, or investigation of drivers of users' satisfaction and continuance intention toward agent-based customer service (Ashfaq et al. 2020) **(social science)**. Satisfaction is also investigated in the system-specific context: There is a plethora of work regarding customer satisfaction from concrete virtual agent services (experimental and commercial). To mention a few, there is a case of a women's intimate apparel manufacturing firm (Ngai et al. 2021) **(social science)**, a case of a software company (Nguyen et al. 2019) **(OR)**, an agent for citizen services (Kotsuka, Murakami, and Arima 2019) **(CHI)** and many more.

Questions regarding the client that should be investigated in the context of building better HoCuS systems include:

1. What parameters affect the willingness of clients to be served by virtual agents, and how do they affect it? **(CHI, social science, OR)**
   (a) Personality characteristics.
   (b) The urgency or immediacy of the call.
   (c) The domain of reality with which the service deals (e.g., the difference between technology-related, health-related, and recreation-related services).
2. Would it be effective to sense the mood and mental state of the client (e.g., by making them speak and collecting voice samples) and consider this parameter in the decision whether to route the call directly to a human operator or to try a virtual agent first? **(social science, OR)**
3. What is clients' tolerance to multiple attempts by the virtual agent to assist? Should we transfer the client to a human operator after a single failed attempt by the virtual agent, or is it better to try a second or even a third time? (**OR**)

4. Is a graphic user interface sufficient, or should we add other components (maybe vocal)? **(GUI/UI)**
5. What is the reaction of different types of clients to idle waiting for service? Would they prefer long waiting for a human operator over getting an immediate response from a virtual agent? Would it be helpful to ask them about their preference explicitly? **(CHI, social science, OR)**
6. How should the satisfaction of the clients of the service be measured? What other parameters should be measured? **(CHI, OR)**

Human Operator

There is a lot of research regarding human operators in call centers. One of the main reasons for this special interest is the high rates of turnover in this profession, as has been reported by many employers and market-observers for years (Holman, Batt, and Holtgrewe 2007; Reynolds 2015; MyWorkPlace 2020) **(commercial)**. There are works that try to isolate the factors that influence this phenomenon using real-world case studies from various countries (Ballard 2012; Kim and Choi 2015; Zito et al. 2018; Griffin 2021) **(social science)**, (Valle, Ruz, and Masías 2017) **(ML)**, (Reissová and Papay 2021) **(social science, informatics)**. In (Brown et al. 2022) **(social science)**, the role of professional pride is studied, and this perspective is especially interesting in the case of HoCuS because of the change in the framing of the human operator in such systems (as also discussed in Section). Other works try to predict the turnover (Valle and Ruz 2015) **(ML)**, (Rodrigo and Ratnayake 2021) **(informatics)**.

Besides the turnover issue, there is naturally a wide selection of works regarding the operators in call centers, their influence on quality of service (Chicu et al. 2019) **(social science)** and their perception of it (Ramseook Munhurrun, Naidoo, and Lukea-Bhiwajee 2010) **(social science)**, their performance (Castilla 2005) **(social science)** and how to measure and predict it (de Oliveira 2019) **(ML)**. Intelligent agents that support operators working on complex tasks were studied in several domains, including call centers. Intelligent agents that support humans in their complex activities need to be able to predict the user's behavior and decisions (Aviv et al. 2021; Azaria et al. 2016; Rosenfeld et al. 2016; Kraus 2016; Rosenfeld and Kraus 2018) **(MAS)**.

Questions regarding the human operator that should be investigated in the context of building better HoCuS systems:

1. The use of virtual agents to answer easy questions liberates the human operator from the need to answer such questions themselves and leaves only questions that the virtual agents could not answer. This situation may be either good or bad for the operators, with opposite arguments: On one hand, the issues that the operators deal with are less routine, therefore the work might be more interesting and with a lower rate of attrition. On the other hand, there are no easy questions that the operators can answer "automatically". They have to make a considerable effort in each case. The issue is which parameter has a stronger influence on the operators and whether it depends on the personality characteristics of the operators **(CHI, social science)**.
2. In the sequence of HoCuS, in which the call is primarily transferred to a virtual agent, and only if it fails it is transferred to the human operator, the clients are likely to be more irritated and less patient once they encounter the operator. Some of them were skeptical from the beginning, and others were frustrated by the failure of the agent and the need to "start all over again". This situation requires the human operator to be more aware and understanding and to use pre-fetched information instead of asking it from the client (again). This situation should be investigated from the operator's point of view, as well as its consequences on the whole system and its managers (as elaborated on in Section) **(social science, OR)**.

Customer Service Center Management

This section deals with the system as a whole and with the aspects of managing a hybrid service center.

A basic question that should be answered by the managers of such a service when thinking of adopting a new technology or method is what is the target function of this change. Often, clients' satisfaction is stated as the main goal, but other goals may be considered, and they might overshadow the former one. Such considerations are, for example, reducing the human resources needed to supply the service or solving the great difficulties in the constant need to recruit employees due to the high turnover rates in the business (Devi and Panchanatham 2010; Mwendwa 2017; Flood 2021) **(social science)**. The balance between these and other (sometimes clashing) goals is widely investigated in (Anderson, Fornell, and Rust 1997; Kantsperger and Kunz 2005; Rekilä et al. 2013) **(social science)** and many others.

Another issue in HoCuS that evolves from the participation of many different players in the process is the relationships and collaborations between the various entities combining the system. The trust between customers and the virtual agents, and the factors that affect this trust, are considered in an interview study in (Følstad, Nordheim, and Bjørkli 2018) **(CHI)**. The combinations of both virtual agents and human operators and the interaction/collaboration between these two entities are investigated in (Le, Sajtos, and Fernandez 2022) **(social science)**. The handoff process – transferring the control from a virtual agent to a human operator, and maybe also vice versa, is researched in (Wintersberger, Klotz, and Riener 2020) **(CHI)**, (Rozga 2018) **(NLP)**.

The issue of task allocation and staff allocation (sometimes called workforce management) is of great importance to call centers. It was approached with methods of queuing theory in (Bouchentouf, Cherfaoui, and Boualem 2021) **(OR)**, with deep neural networks and reinforcement learning in (Kumwilaisak et al. 2022) **(RL)**, and various other methods (Cohen, Reis, and Amorim 2020) **(OR)**, (Horng and Lin 2020) **(informatics, social science)**. Allocation of tasks in cases where there are operators with two levels of expertise, as our model may be described, is discussed in (Stepanov, Stepanov, and Zhurko 2019) **(OR)**.

Questions regarding the HoCuS as a whole system are:

1. In the traditional call center business, there are known and well-studied parameters of performance. Introducing

a new operation model changes the service's characteristics and requires adjustment of these parameters. Specifically, the change in the amount and the difficulty of the question routed to operators (as mentioned in Section) affects:

(a) The ratio of needed operators per customer in order to provide good service **(CHI, social science)**.

(b) The performance measurement systems and their parameters (such as calls per hour, average call times, time between calls, etc.) **(OR)**.

(c) The personal characteristics of the operators – the task is different now, and it is reasonable that new abilities and skills will be needed **(social science)**.

2. Employment models – changes in the personality characteristics of the operators and rates of attrition have considerable influence on the ways companies employ these workers. For example, the position of an"operator for hard questions" may affect the social status of this job, hence the age of employees, the length of the expected employment, and many other parameters **(social science)**.

3. The handoff process – the ways to perform the transfer of control from the virtual agent to the human operator (and/or vice versa):

(a) Should the handoff be unidirectional (from agent to human only) or bidirectional? **(CHI, Knowledge science)**

(b) How many such handoffs can be done without irritating the client? **(CHI, social science)**

(c) How to transfer information collected in early phases to later ones, to avoid the clients having to chant the same details repeatedly? **(CHI, knowledge science)**

Implementing the Holistic Approach

We decided to test the holistic approach and implement a call center service that combines virtual agents with human operators. The process and its results are described here concisely, emphasizing the interdisciplinary approach. A detailed review can be found in (Oshrat et al. 2022).

Theoretical Basis – Queuing Theory

The basis for the work was a comparison between a call service with human operators only ("pure human environment") and a call service combining humans and virtual agents ("hybrid environment").

We showed, using queuing theory-based analysis that, by a cautious choice of parameters, the hybrid system can service more clients than the purely human one, while simultaneously also reducing the expected waiting times of the clients. The analysis yielded the ranges of values for the parameters to achieve this result, but at this stage, without introducing additional techniques from other communities, it was not clear whether these ranges are realistic and can be implemented by humans and by agents. The main parameters of the process were the number of clients in the system, the rate of arrival of the questions, and the service times for a human operator and for a virtual agent (represented by average and standard deviation values).

Experimenting with Human Subjects and Building a Q&A DB

In the next phase, we performed an experiment with human subjects. The goal of the experiment was twofold: to investigate the ability to build a virtual agent based only on Q&As between clients and operators, without any other domain knowledge, and to extract actual parameter values regarding clients, questions, and communications in a call center. To this end, we used a simulation-strategy video game. The players – our clients – were asked to perform diverse tasks in order to build and fix PCs. In each session, there were up to 107 steps that the client should have performed. We recruited 41 human subjects to play the game and provided them with a textual chat channel via the Internet through which they could communicate to get help. When they encountered difficulties understanding or performing the tasks, the clients could send a question via the textual channel. The question was routed to a virtual agent that tried to match it with questions in the agent's DB, and if a matching question was found, then its answer was passed to the client. If the agent could not find a matching question, or if the answer did not satisfy the client, the question was routed to an experienced human operator who answered it. The only source of information for the virtual agent was previous questions that clients had asked and the answers that the human operators provided. We collected full data regarding the communication between the clients, the agent, and the operators.

After performing 41 sessions and accumulating data, we applied several pre-trained LM models to explore their ability to identify the cluster of questions in the DB for a given question. Using the paraphrase-MiniLML6-v2 model (Reimers and Gurevych 2019) with the raw clusters we achieved 70% accuracy, and using the clusters after short manual processing (e.g., adding short natural language questions) we reached 76% accuracy. The experiment's results show that it is feasible to build an effective virtual agent in a short time using real-world existing processes (such as an existing pure human service center).

We also achieved the second goal of acquiring actual values for the various parameters of our theoretical model. These values were very helpful in the next phase.

Simulation

In order to test the model and the relations between parameters, we built a simulator that implements the system as described in the theoretical model. The input parameters are:

- n – the number of clients in the system
- α – the fraction of questions that the virtual agent knows how to answer (out of the total number of questions)
- s_α, s_β – the distributions of avg. service time by humans or by virtual agent, respectively, for hard questions
- $\bar{\lambda}$ – the distributions of arrival rate of questions per client
- ϵ – the distributions of the time it takes for a virtual agent to notice that it cannot answer a question

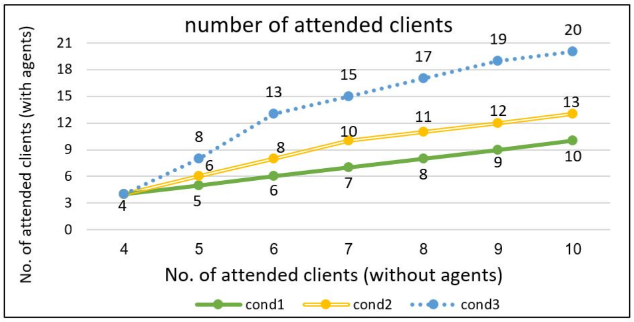

Figure 1: The number of clients a human operator can attend to in a hybrid environment, relative to the number of clients that a human operator attends to in a pure human environment without increasing the waiting time. In all three conditions, $s_\alpha \sim \mathcal{N}(1, 0.2)$, $\epsilon \sim \mathcal{N}(0.5, 0.1)$, $\bar{\lambda} \sim \text{Pois}(0.1)$. Cond1: $\alpha = 0.2$, $s_\beta \sim \mathcal{N}(3.5, 1)$. Cond2: $\alpha = 0.5$, $s_\beta \sim \mathcal{N}(3.15, 1)$. Cond3: $\alpha = 0.7$, $s_\beta \sim \mathcal{N}(2.8, 1)$.

We studied the behavior of the system with different values of parameters and tried to understand the possible optimizations that can be made. There were many interesting ideas that we lack the space here to detail, but we will demonstrate one finding: In Section we claimed that it is possible to adjust the hybrid environment's parameters to provide a win-win situation, where a single operator can serve more clients, and at the same time, the clients will experience shorter waiting times. Figure 1 presents simulations of 3 conditions and compares the number of clients that can be attended in a pure human environment to the number of clients that can be attended in a hybrid environment. For every value of no. of clients in a pure human environment on the x-axis, we matched the number of clients that can be attended to in a hybrid environment with a lower expected waiting time per client. The results seem to support our claim.

Second Experiment: Combining Human Subjects and Agent-based Simulation

The first experiment's results (Section) show that it is feasible to build an effective virtual agent in a short amount of time using real-world existing processes. Nevertheless, the quality of service, and specifically the clients' experience of the hybrid configuration, could not be measured in this experiment design. To this end, we designed a second experiment, which aimed to compare clients' satisfaction in two conditions: (1) Pure human service, in which a single human operator attends to 11 clients; and (2) Hybrid service, in which a single human operator, assisted by virtual agents, attends to 18 clients. In both conditions, one of the clients was played by a human recruit and the other clients were virtual, i.e., they were placeholders that waited in the queue and took the human operator's attention in order to make the experience of the real human client authentic. The time each virtual client needed to be "served" was generated using the real-world data we collected in the first experiment and the simulations of Section . Similarly, the values of 11/18 clients

	Num. of clients	Avg. waiting time (Sec.)	Avg. time to answer (Sec.)	Satisfaction grades		
				waiting time	service	information
with agents	18	61.8	51.2	4.7	4.7	4.8
withou agents	11	63.3	90.7	3.5	4.2	4.1
			T test	0.005	0.105	0.035

Figure 2: Results of the user satisfaction study.

for the different conditions were chosen according to our experience in the first experiment, in order to produce similar waiting times in the queue for the human operator: Obviously, in the second condition, some of the clients' questions are answered by the virtual agents, thus easing the load on the human operator. We wanted the clients in the experiment to experience equal waiting times for the human operator in both conditions, in order to measure the clients' satisfaction not as a result of shorter waiting times in Condition 1, but in relatively equal situations. 20 recruits took part in the experiment as clients. At the end of each session, as commonly done in CHI projects, the human clients were asked to grade their satisfaction with the service they received in three categories: The waiting times, the quality of service in general, and the quality of information they received from the help desk.

The results of the second experiment are presented in Figure 2. It can be seen that although the number of simultaneous clients was higher in Condition 1 (18 "with agent" vs. 11 "without agent"), the average waiting time for a human operator was similar in both conditions (61.8 vs 63.3), and the quality of service was higher in Condition 1: Clients received answers to their questions significantly faster (51.2 vs 90.7), finished the game-steps sooner and reported higher satisfaction grades.

Discussion

We have demonstrated that the HoCuS setting offers an application domain necessitating a rich multidisciplinary approach. However, cooperation and collaboration amongst different (academic) disciplines can also be challenging. The lack of acquaintance and information exchange between researchers is the easy part. Deeper difficulties stem from the inherent gaps in terminology, concepts, methods, and standards. Indeed, even basic scientific terms such as "correlations" and "significance" have different meanings in different faculties of academia, let alone agreement about what constitutes a "proof" or "evidence", and what is "reasonable".

That said, we suggest that the HoCuS application is a fitting platform and goal for mobilizing collaboration:

- it is an important domain with wide practical applications,
- cooperation is necessary (as we have demonstrated), and
- clear success criteria can be defined.

We hope that work in this setting will result in methodologies and techniques that could improve customer service, but more importantly, open new directions for cooperation amongst diverse AI communities (and beyond) toward solving joint difficult tasks.

Acknowledgments

This research was funded in part by JPMorgan Chase & Co and by the IBM Global University Program Academic Award. Any views or opinions expressed herein are solely those of the authors listed, and may differ from the views and opinions expressed by JPMorgan Chase & Co. or its affiliates. This material is not a product of the Research Department of J.P. Morgan Securities LLC. This material should not be construed as an individual recommendation for any particular client and is not intended as a recommendation of particular securities, financial instruments or strategies for a particular client. This material does not constitute a solicitation or offer in any jurisdiction.

References

Allouch, M.; Azaria, A.; and Azoulay, R. 2021. Conversational Agents: Goals, Technologies, Vision and Challenges. *Sensors*, 21(24): 8448.

Anderson, E. W.; Fornell, C.; and Rust, R. T. 1997. Customer satisfaction, productivity, and profitability: Differences between goods and services. *Marketing science*, 16(2): 129–145.

Angga, P. A.; Fachri, W. E.; Elevanita, A.; Agushinta, R. D.; et al. 2015. Design of chatbot with 3D avatar, voice interface, and facial expression. In *2015 international conference on science in information technology (ICSITech)*, 326–330. IEEE.

Ashfaq, M.; Yun, J.; Yu, S.; and Loureiro, S. M. C. 2020. Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, 54: 101473.

Ashktorab, Z.; Jain, M.; Liao, Q. V.; and Weisz, J. D. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *CHI-19*.

Aviv, A.; Oshrat, Y.; Assefa, S. A.; Mustapha, T.; Borrajo, D.; Veloso, M.; and Kraus, S. 2021. Advising Agent for Service-Providing Live-Chat Operators. *arXiv preprint arXiv:2105.03986*.

Azaria, A.; Gal, Y.; Kraus, S.; and Goldman, C. V. 2016. Strategic advice provision in repeated human-agent interactions. *Autonomous Agents and Multi-Agent Systems*, 30(1): 4–29.

Ballard, J. K. 2012. *Call center turnover: A study of the relationships between leadership style, burnout, engagement and intention to quit*. Ph.D. thesis, Capella University.

BluePrint. 2020. How to Reduce the Rising Costs of RPA Maintenance and Support. https://www.blueprintsys.com/blog/rpa/reduce-rising-costs-rpa-maintenance-and-support. Online, Accessed: 31.8.2022.

Bojanić, M.; Delić, V.; and Karpov, A. 2020. Call redistribution for a call center based on speech emotion recognition. *Applied Sciences*, 10(13): 4653.

Bouchentouf, A. A.; Cherfaoui, M.; and Boualem, M. 2021. Analysis and performance evaluation of Markovian feedback multi-server queueing model with vacation and impatience. *American Journal of Mathematical and Management Sciences*, 40(3): 261–282.

Brown, E. S.; Palka, J.; Helm, S. V.; and Kulikova, A. 2022. The relative importance of reputation and pride as predictors of employee turnover in an academic medical center. *Health Care Management Review*, 47(1): 66–77.

Carr, D. F. 2020. When Robotic Process Automation (RPA) bots break: 3 things to know. https://enterprisersproject.com/article/2020/2/when-robotic-process-automation-rpa-bots-break-3-tips. Online, Accessed: 2/1/2022.

Castilla, E. J. 2005. Social networks and employee performance in a call center. *American journal of sociology*, 110(5): 1243–1283.

Chen, C.; Zhang, X.; Ju, S.; Fu, C.; Tang, C.; Zhou, J.; and Li, X. 2019. AntProphet: an Intention Mining System behind Alipay's Intelligent Customer Service Bot. In *IJCAI*, volume 8, 6497–6499.

Chicu, D.; del Mar Pàmies, M.; Ryan, G.; and Cross, C. 2019. Exploring the influence of the human factor on customer satisfaction in call centres. *BRQ Business Research Quarterly*, 22(2): 83–95.

Cohen, Y.; Reis, J.; and Amorim, M. 2020. Empirical Analysis of Call Center Load & Service Level for Shift Planning. In *International Conference on Exploring Services Science*, 79–91. Springer.

Crolic, C.; Thomaz, F.; Hadi, R.; and Stephen, A. T. 2022. Blame the bot: anthropomorphism and anger in customer–chatbot interactions. *Journal of Marketing*, 86(1): 132–148.

de Oliveira, E. L. 2019. Machine learning techniques applied to predict the performance of contact centers operators. In *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–4. IEEE.

Deng, Y.; Li, Y.; Zhang, W.; Ding, B.; and Lam, W. 2022. Toward Personalized Answer Generation in E-Commerce via Multi-perspective Preference Modeling. *ACM Transactions on Information Systems (TOIS)*, 40(4): 1–28.

Devi, G.; and Panchanatham, N. 2010. A Cram on Employee Attrition in Chennai Based Call Centers: An Exploratory Study. *International Journal of Marketing & Human Resource Management*, 1(1): 18–27.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.

Flood, T. E. 2021. *Strategies for Call Center Employee Retention*. Ph.D. thesis, Walden University.

Følstad, A.; Nordheim, C. B.; and Bjørkli, C. A. 2018. What makes users trust a chatbot for customer service? An exploratory interview study. In *International conference on internet science*, 194–208. Springer.

Fotheringham, D.; and Wiles, M. A. 2022. The effect of implementing chatbot customer service on stock returns: an event study analysis. *Journal of the Academy of Marketing Science*, 1–21.

Franceschi, C. 2020. How to make a chatbot in less than 30 minutes. https://joonbot.com/blog/guides/how-to-make-a-chatbot/. Online, Accessed: 30.8.2022.

Fu, Y.; and Feng, Y. 2018. Natural answer generation with heterogeneous memory. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 185–195.

Gao, S.; Chen, X.; Ren, Z.; Zhao, D.; and Yan, R. 2021. Meaningful answer generation of e-commerce question-answering. *ACM Transactions on Information Systems (TOIS)*, 39(2): 1–26.

Griffin, G. 2021. *Job burnout and turnover intentions among telecommuting call center agents*. Ph.D. thesis, Walden University.

Holman, D.; Batt, R.; and Holtgrewe, U. 2007. The global call center report: International perspectives on management and employment. https://hdl.handle.net/1813/74325. Accessed: 2022-12-12.

Horng, S.-C.; and Lin, S.-S. 2020. Staffing Optimization of One-Period Multi-Skill Call Center. In *Proceedings of the 2020 the 3rd International Conference on Computers in Management and Business*, 89–93.

Kalinin, K. 2020. How to Make a Chatbot from Scratch and Grow Your Business with AI. https://topflightapps.com/ideas/how-to-build-a-chatbot/#3. Online, Accessed: 30.8.2022.

Kantsperger, R.; and Kunz, W. H. 2005. Managing overall service quality in customer care centers: Empirical findings of a multi-perspective approach. *International Journal of Service Industry Management*.

Kim, J.-I.; and Choi, B.-R. 2015. Convergence study on emotional labor, stress response and turnover intention of call-center worker. *Journal of the Korea Convergence Society*, 6(6): 139–146.

Kotsuka, Y.; Murakami, R.; and Arima, S. 2019. Development and Evaluation of Chat-bot Systems: The applications of citizen service procedures. In *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 290–295. IEEE.

Kraus, S. 2016. Human-agent decision-making: Combining theory and practice. *arXiv preprint arXiv:1606.07514*.

Kumwilaisak, W.; Phikulngoen, S.; Piriyataravet, J.; Thatphithakkul, N.; and Hansakunbuntheung, C. 2022. Adaptive Call Center Workforce Management With Deep Neural Network and Reinforcement Learning. *IEEE Access*, 10: 35712–35724.

LandBot. 2022. The Most Powerful No-Code Chatbot Builder. https://landbot.io/. Online, Accessed: 30.8.2022.

Le, K. B. Q.; Sajtos, L.; and Fernandez, K. V. 2022. Employee-(ro) bot collaboration in service: an interdependence perspective. *Journal of Service Management*, (ahead-of-print).

Leah. 2022. How Rule-Based Chatbots Beat AI Chatbots in Business. https://www.userlike.com/en/blog/rule-based-chatbots. Online, Accessed: 30.8.2022.

Li, Q. 2012. Literature survey: domain adaptation algorithms for natural language processing. *Department of Computer Science The Graduate Center, The City University of New York*, 8–10.

Liu, R.; and Dong, Z. 2019. A study of user experience in knowledge-based QA chatbot design. In *International Conference on Intelligent Human Systems Integration*, 589–593. Springer.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Luo, B.; Lau, R. Y.; Li, C.; and Si, Y.-W. 2022. A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1): e1434.

Ma, X.; Xu, P.; Wang, Z.; Nallapati, R.; and Xiang, B. 2019. Domain adaptation with BERT-based domain classification and data selection. In *DeepLo*.

McCreery, C. H.; Katariya, N.; Kannan, A.; Chablani, M.; and Amatriain, X. 2020. Effective transfer learning for identifying similar questions: matching user questions to COVID-19 FAQs. In *KDD*.

Molina-Castillo, F. J.; López-Nicolás, C.; and Bouwman, H. 2008. Explaining mobile commerce services adoption by different type of customers. *Journal of Systemics, Cybernetics and Informatics*, 6(6): 73–79.

Mwendwa, L. 2017. *Factors influencing call center agent attrition: A case of Kenya Power call center*. Ph.D. thesis, University of Nairobi.

MyWorkPlace. 2020. What's Causing All That Turnover in Your Call Center? https://myworkchoice.com/blog/whats-causing-turnover-in-your-call-center/?utm_source=google. Online, Accessed: 29.8.2022.

Ngai, E. W.; Lee, M. C.; Luo, M.; Chan, P. S.; and Liang, T. 2021. An intelligent knowledge-based chatbot for customer service. *Electronic Commerce Research and Applications*, 50: 101098.

Nguyen, T.; et al. 2019. *Potential effects of chatbot technology on customer support: A case study*. Master's thesis, Aalto University School of Science.

Olivares, M.; Musalem, A.; and Yung, D. 2020. Balancing agent retention and waiting time in service platforms. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 295–313.

Oshrat, Y.; Aumann, Y.; Hollander, T.; Maksimov, O.; Anita Ostroumov, N.; and Kraus, S. 2022. Efficient Customer Service Combining Human Operators and Virtual Agents. https://biu-ai.com/agent_human_cooperation.pdf. Online, Accessed: 1.9.2022.

Panko, R. 2019. How Businesses Should Set Up Their IVR Menu. https://clutch.co/bpo/resources/virtual-assistants/how-businesses-set-up-ivr-menu/. Accessed: 2022-12-12.

Pawlik, Ł.; Płaza, M.; Deniziak, S.; and Boksa, E. 2022. A Method for Improving Bot Effectiveness by Recognising

Implicit Customer Intent in Contact Centre Conversations. *Speech Communication*.

Press, G. 2019. AI Stats News: 86% Of Consumers Prefer Humans To Chatbots. https://www.forbes.com/sites/gilpress/2019/10/02/ai-stats-news-86-of-consumers-prefer-to-interact-with-a-human-agent-rather-than-a-chatbot/?sh=220aa7252d3b. Accessed: 2022-12-12.

Ramseook Munhurrun, P.; Naidoo, P.; and Lukea-Bhiwajee, S. D. 2010. Measuring service quality: Perceptions of employees. *Global journal of business research*, 4(1): 47–58.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Empirical Methods in Natural Language Processing*.

Reissová, A.; and Papay, M. 2021. Relationship between employee engagement, job satisfaction and potential turnover. *TEM Journal*, 10(2): 847–852.

Rekilä, T.; et al. 2013. *A study of the factors influencing customer satisfaction and efficiency in contact centers: the combined effect*. Master's thesis, Aalto University School of Business.

Reynolds, P. 2015. Exploring call center turnover numbers. *Nashville: QATC*. Online, Accessed: 29.8.2022.

Rodrigo, D. S.; and Ratnayake, G. S. 2021. Employee Turnover Prediction System: With Special Reference to Apparel Industry in Sri Lanka. In *2021 6th International Conference for Convergence in Technology (I2CT)*, 1–9. IEEE.

Rosenfeld, A.; Keshet, J.; Goldman, C. V.; and Kraus, S. 2016. Online prediction of exponential decay time series with human-agent application. In *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, 595–603. IOS Press.

Rosenfeld, A.; and Kraus, S. 2018. Predicting human decision-making: From prediction to action. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 12(1): 1–150.

Rozga, S. 2018. Human Handoff. In *Practical Bot Development*, 547–570. Springer.

Shahid, U. 2020. Building a Rule-based Chatbot in Python. https://online.datasciencedojo.com/blogs/rule-based-chatbot-in-python/. Online, Accessed: 30.8.2022.

Singh, J.; Joesph, M. H.; and Jabbar, K. B. A. 2019. Rule-based chabot for student enquiries. In *Journal of Physics: Conference Series*, volume 1228, 012060. IOP Publishing.

Stepanov, S. N.; Stepanov, M. S.; and Zhurko, H. M. 2019. The modeling of call center functioning in case of overload. In *International Conference on Distributed Computer and Communication Networks*, 391–406. Springer.

Sun, S.; and Sedoc, J. 2020. An analysis of bert faq retrieval models for covid-19 infobot. https://openreview.net/pdf?id=dGOeF3y_Weh. Accessed: 2022-12-13.

Thorat, S. A.; and Jadhav, V. 2020. A review on implementation issues of rule-based chatbot systems. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

Valle, M. A.; and Ruz, G. A. 2015. Turnover prediction in a call center: Behavioral evidence of loss aversion using random forest and naïve bayes algorithms. *Applied Artificial Intelligence*, 29(9): 923–942.

Valle, M. A.; Ruz, G. A.; and Masías, V. H. 2017. Using self-organizing maps to model turnover of sales agents in a call center. *Applied Soft Computing*, 60: 763–774.

Wei, M.; and Zhang, Y. 2019. Natural answer generation with attention over instances. *IEEE Access*, 7: 61008–61017.

Weizenbaum, J. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1): 36–45.

Wintersberger, P.; Klotz, T.; and Riener, A. 2020. Tell Me More: Transparency and Time-Fillers to Optimize Chatbots' Waiting Time Experience. In *Proceedings of the 1 Human-Computer Interaction: Shaping Experiences, Shaping Society*.

Xu, K.; Fu, C.; Zhang, X.; Chen, C.; Zhang, Y.-L.; Rong, W.; Wen, Z.; Zhou, J.; Li, X.; and Qiao, Y. 2020. admscn: A novel perspective for user intent prediction in customer service bots. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2853–2860.

Zabój, D. 2022. How to build your first chatbot. https://www.chatbot.com/help/build-your-chatbot/how-to-build-your-chatbot/. Online, Accessed: 30.8.2022.

Zhang, J. J.; Følstad, A.; and Bjørkli, C. A. 2021. Organizational factors affecting successful implementation of chatbots for customer service. *Journal of internet commerce*, 1–35.

Zito, M.; Emanuel, F.; Molino, M.; Cortese, C. G.; Ghislieri, C.; and Colombo, L. 2018. Turnover intentions in a call center: The role of emotional dissonance, job resources, and job satisfaction. *PloS one*, 13(2): e0192126.