Modelling Hybrid Human-Artificial Intelligence Cooperation: A Call Center Customer Service Case Study

Laura H. Kahn, Onur Savas, Adamma Morrison, Kelsey A. Shaffer, Lila Zapata
Accenture Federal Services
Arlington, VA USA
https://orcid.org/0000-0001-8523-7856

***Abstract*—As autonomous systems become an essential part of augmented decision-making in the workforce, we have the opportunity to change the relationship between human and machine into a more collaborative one. The future of industry, commercial and public services point in a direction where humans and artificial intelligence (AI) increasingly work together. AI systems are increasingly extending and enriching decision support by complementing and augmenting human capabilities. To further elevate this partnership, we need to form organic human-AI teams that communicate with, adapt to, and learn from each other. We create a new human-in-the-loop hybrid spectrum, that expands existing definitions of human and machine teaming. For a given situation and a team of humans and AI systems, we are interested in testing variations on human-AI cooperation outcomes. We examine a call center use case to determine how variations in human and machine teaming affects average handle time and response quality outputs that affect customer service. We have evaluated three scenarios: 1) human-only, 2) AI-only, and 3) human + AI collaboration. Under the parameter space we studied, we found that human + AI collaboration is optimal.**

I. Introduction

A critical aspect in the effectiveness and adoption of disruptive technologies like artificial intelligence (AI) that are often overlooked include how well AI works with people [9] given a wide array of sociotechnical contexts. Many companies deploy AI technologies with a short-term view of productivity gains, rather than a longer-term human plus machine teaming strategy [17]. As the future of work and AI technologies become more ubiquitous and important to business, understanding the interactions between autonomous systems and humans will continue to be a topic of research [3] and application. Today, it is becoming glaringly obvious that there are more opportunities where AI can partner with people to accomplish more together than alone [5]. This teaming, however, requires a thoughtful approach to maximize the sum of the parts.

Although technologies have been an integral part of human society since the inception of the human race, many of the adoption discussions have been centered around human-machine teaming measured by efficiency metrics, with the implication that success is defined by improvements to task speed [10, 8]. Prior research has assumed that if an autonomous system augments human decision-making by speeding up the time to decision and/or improves the quality of the decision than what the human could do alone, then it is a successful implementation of AI [2, 11]. This narrow definition of success fails to account for and quantify the trade-offs between costs, accuracy and customer service success.

II. Background

The early days of human-computer interaction laid the foundation for the more nascent concept of a *human-in-the-loop* with AI systems. Recent emphasis has been on the importance of keeping the human-in-the-loop with the hypothesis that humans are better suited to certain tasks than machines [13, 15]. Creatively thinking about creating frameworks where humans flourish with AI in various contexts is an important AI maturity transition from the current dichotomy of human as director and AI as servant. As non-human agents such as AI-powered virtual assistants become more indistinguishable from human agents in traditional definitions of efficiency and performance, many questions arise about what these interactions should and could be like [7].

Our research accomplishes two primary goals: 1) create a new, human plus AI levels of autonomy and 2) compare trade-offs between cost, accuracy, and performance efficiency using a digital twin. Automation levels adapted from traditional industrial engineering manufacturing use cases [16] can provide a functional blueprint for translating autonomy levels to a new human plus AI levels. Our research uses simulation techniques to quantify and understand the trade-offs that occur between humans and AI as they perform tasks. Our framework provides a quantitative approach that contributes to conversations about how to realize a more symbiotic teaming between human and machine.

III. Method

A. Human and Machine Autonomy Levels

We describe the human and AI autonomy levels as a scale that delineates the continuum of human and AI's role by level number and sets general guidelines as to the amount of human and AI involvement in decision-making tasks. Specific values assigned to each level could be established and customized according to the respective use case.

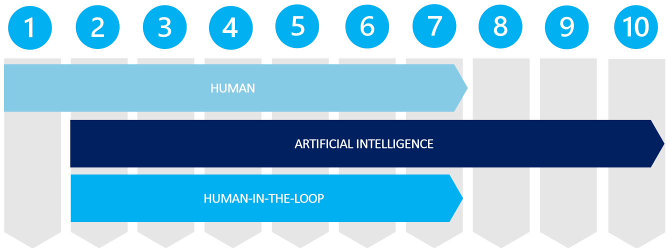

Fig. 1. The Spectrum of Human and AI Collaboration Levels

The spectrum of human and AI involvement in decision making ranges from Level 1 - Level 10 as shown in Fig. 1. For **Level 1**, there is no AI and the Human alone considers alternatives, makes and implements decisions. For **Level 2**, AI offers all alternatives, which human may ignore or accept. For **Level 3**, AI offers one or more alternatives, and human decides which one to implement. For **Level 4**, AI offers one or more alternatives and suggests one, and human decides which one to implement. For **Level 5**, AI offers one or more alternatives, suggests one alternatives, and implements if human approves. For **Level 6**, AI makes an automatic decision without human input, but the human is still nominally in-the-loop by having the option to veto the AI decision prior to final implementation. **Level 6** marks an important autonomy level philosophically and in practice, when AI makes a decision with no human input initially. For **Level 7**, AI makes and implements decision, but must inform human after the decision is made. Human cannot veto AI decision. For **Level 8**, AI makes and implements decision, and informs human only if asked. For **Level 9**, AI makes and implements decision, and informs human only if the decision meets criteria that informing is warranted. For **Level 10**, AI makes and implements decision under certain conditions, and informs human only if the decision meets criteria that informing the human is warranted.

Many factors outside the scope of this paper affect whether Levels 7-10 autonomy should even be considered for one or more industrial, commercial or government contexts from a technology readiness, user agency and end user well-being point of view [13, 14]. As our scenarios continue to evolve in time and more granular information about the interactions arise, we anticipate adding variables that proxy ethical concerns and other human-centered design factors. A thoughtful approach to designing for variations on our scenarios as well as how AI will work alongside humans is essential before deploying a real-world AI system. It should be noted that the authors are applying this methodology to start discussions about which scenarios may warrant certain levels of human plus machine collaboration levels within a call center environment.

B. Human and Machine as a Digital Twin

Human and AI systems have been modelled with a digital twin representation, wherein each of the individual components are represented virtually. Unlike digital twins described by [6, 12], our simulation accounts for the human and AI components, and also the interactions between the components. We simulate the arrival process for incoming call center questions of various complexities to human or AI agents, and compare results of three scenarios by question accuracy, *a* and average handle time, *aht*. Question accuracy and average handle time, *aht*, among other metrics are common proxies to measure customer service success within a call center [1], and therefore serve as the utility function within our digital twin.

In our call center context, we set the probability value, *p* that an incoming question, *q*, with easy, medium, or hard difficulty *d*, will arrive with time *t*, and be assigned to either a human agent, *h*, with experience level, *e1*, or *e2*, an AI agent, *ai*, or some combination therein, wherein each agent has a processing cost, *hce1*, *hce2*, or *aic*, corresponding to processing cost for human with experience level 1, human with experience level 2, or AI. In this instance, the probability value, *p*, follows a random distribution of difficulty *d*. For purposes of illustration, the human agent with experience level 1 has some amount of experience less than human agent with experience level 2. The hourly rate (another way of accounting for cost to answer a question) is approximated from nationwide average agent pay [4]. For our scenarios, an AI agent costs about 90% less than a senior human agent, and about 95% less than a junior human agent.

Using these variables, we calculate and compare the benefit, *b*, weighted cost, *wc*, and value, *V*, of agents in three scenarios as described in Equation 1, wherein the value, *V*, is described as

$$V = b_h - wc \tag{1}$$

And, wherein, the weighted cost, *wc*, is a function of agent type *he1, he2, or ai* (human experience level 1, human experience level 2, or ai), and processing cost, *ce1*, *ce2*, or *c* as described in Equations [2,3].

$$wc(h) = h(ce1, ce2) \tag{2}$$

$$wc(ai) = h(aic) \tag{3}$$

Moreover, wherein, the benefit, *b*, is the percent questions answered accurately, according to a particular call center's bank of approved answers. We arrive at our parameters based on a federal call center use case and expert elicitation. The call center's purpose is for a customer to either receive information or complete a transaction. Our use case has implemented an AI agent to replace human agents to a certain degree in answering customer questions.

Our methods quantitatively experiment with these levels and the associated outcomes in real world scenarios. Our research aims to model human and machine teaming with an integrated systems perspective. We achieve this by mathematically representing the interaction between the two entities in a call center context and examine the trade-offs between performance and cost.

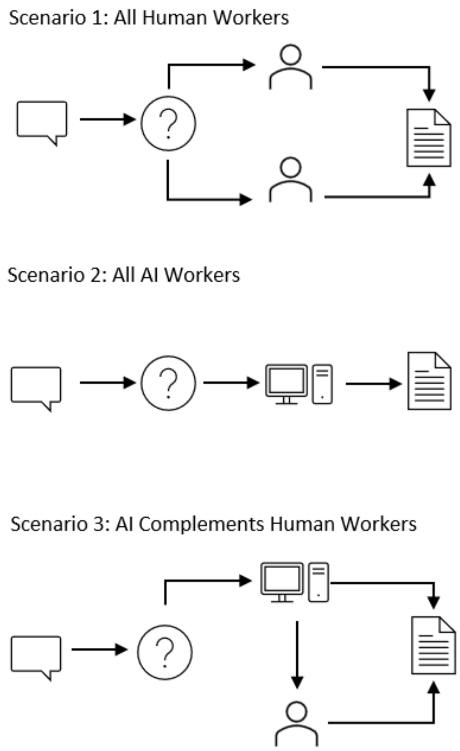

Fig. 2. Human-AI Collaboration Scenarios

IV. Discussion

Our goal is to discuss trade-offs for a utility function of accuracy, efficiency and cost given a set of parameters in at least one real-world system using simulation methods. We achieve this goal by 1) modelling humans and AI as a holistic system, 2) defining collaboration parameters between humans and AI, 3) creating three scenarios by adjusting one or more experimental factors. We discuss our results and the associated implications for human and machine collaboration within a call center context, though we believe our results could be generalized to other contexts.

Our research applies mathematically rigorous simulation techniques traditionally used to improve linear processes by optimizing for efficiency. Our approach is a new way of applying simulation to a virtual human and AI collaboration system, and quantitatively examine its impact on metrics that affect customer service. We test three scenarios in a call center environment with the following scenarios shown in Fig. 2. There are three question types (easy, medium, hard) that could be defined and varied based on the primary goal of the call center type (for a customer to provide information, complete a transaction, or some hybrid therein), or other organizational goals. We believe our approach could generalize to other instances of human-in-the-loop systems outside of a call center environment to measure these trade-offs.

Having the human-in-the-loop when handling ambiguous questions may have an affect on answer accuracy but will most likely cost more in time and money than a purely AI solution. Measuring the impact of this human intervention at various autonomy levels and in terms of performance metrics is an important step in quantifying the trade-offs in these human-machine collaboration scenarios.

V. Challenges

Representing humans, AI systems, and collaboration parameters using virtual simulation techniques presented two primary challenges: 1) variables are directional and 2) other types of human preferences and organizational structures were not included. Directional variables mean that our mathematical representation and corresponding simulation is an approximation rather than concrete values. However, more work with call center values and metrics could address this challenge. Our technique is primarily designed to test what-if scenarios for improved decision making before a change is implemented. Outputs of optimal parameters, though not absolute, could serve as tools to intentionally implement strategic human and AI teaming solutions.

We present three scenarios of human-machine autonomy teaming, where humans and AI are abstracted as shown in Fig. 2 and simulated using custom Python open-source code. Scenario 1 corresponds to level 1 (all human), Scenario 2 corresponds to levels 7-10 (all AI), and Scenario 3 corresponds to levels 2-7 (hybrid human and AI) on the spectrum shown in figure 2. These scenarios were chosen as boundary points elicited from consultation from call center experts. In these scenarios, the human icon could represent one or more humans-in-the loop answering a question in a call center environment. The computer icon could represent one or more AI systems (such as conversational AI or the like) interacting with the customer to answer a question. Depending on a variety of organizational structures and AI solution requirements, more than one human and AI collaboration level could be desirable.

* In Scenario 1, a junior-level human agent, answers 100% of the questions that come into the call center, limited to questions that are of type easy or medium. A senior-level human agent answers 100% of the questions that are of type difficult.

* In Scenario 2, an AI agent answers 100% of all questions, regardless of their type.

* In Scenario 3, an AI agent answers easy or medium questions, with hard questions escalated by the AI to a senior-level human agent (including suggested answers to decrease the required answer time). Scenario 3 corresponds to any of the levels 2-7 on the autonomy scale shown in Figure 1. Scenario 3 can be thought of as one instance of a human-in-the-loop model as many more variations exist.

VI. Results

Average results are shown from each of the three Scenarios: Scenario 1 - All Human Workers, Scenario 2 - All AI Workers, and Scenario 3 - AI Complements Human Workers. Benefit, b, Weighted cost, *wc*, and Value, *V* ($) are calculated as described in Equations 1-3 and shown in Table 1.

TABLE I
AGENT SCENARIOS

Scenario	*b* (%)	*wc* ($)	*V* ($)
1	93	132	40
2	75	0.08	75
3	80	0.42	80

For each scenario, 1 million simulations were run to calculate the benefit, *b*, weighted cost, *wc*, and value, *V*, of the strategy of choosing either a blend of junior-level and senior-level human agents, an AI agent, or some combination therein, to answer easy, medium, or difficult questions. We want to minimize costs, and maximize the number of calls answered accurately since call accuracy, among other factors, has the most impact on customer service.

In Scenario 1, as questions increased in complexity to be more difficult, the benefit, *b*, decreased from 97% to 88% for the senior-level human agent. In Scenario 2, the benefit decreased from 89% to 61% for the AI agent, and from 99% to 61% for the hybrid human-AI agent in Scenario 3.

For Scenario 1, 2, and 3, the average weighted cost, *wc*, was $132, $0.08, and $0.42, respectively. For Scenario 1, 2, and 3, the average benefit, *b*, was 93%, 75%, and 80%, respectively. For Scenario 1, 2, and 3, the value *V* was 40, 75 and 80, respectively, so the highest value with these parameters was with the hybrid human-AI agent. There was a trade-off for the highest accuracy with the senior-level human agent. This trade-off was a higher weighted cost, which is over 100 times higher than both the weighted cost of the AI agent and the hybrid human-AI agents. In the case of hybrid human - AI agents in Scenario 3, the benefit is higher than Scenario 2 (AI only), but still more expensive than AI only in Scenario 2. Considering these trade-offs with a test and learn mentality improves our understanding of how the human and machine cooperation teaming blends quantitatively affects outcomes. Depending on an organization's customer service level agreements, one or more of these scenarios might be appropriate to meet customer needs.

Our goal is to discuss trade-offs for a utility function of accuracy, efficiency and cost given a set of parameters in at least one real-world system using simulation methods. If we are only solving the problem for the lowest cost to answer questions and given these constraints and assumptions, our recommendation would be to replace all human agents with AI. But that recommendation would be far from conclusive or the only course of action. Future work would include capturing other unknown nuances in more complex scenarios. Since a call center has no control of the types of questions that arrive, there is currently a need to have more experienced human agents to answer more complex and perhaps ambiguous questions. Also, given the current state of AI solutions within call center contexts, technology appropriateness should be a consideration for decision-makers when discussing which blend of human and AI is appropriate for their organization. Furthermore, there are other more qualitative issues such as accounting for a user's experiences with the service in the call center that are not within the scenarios modelled in this research.

Although our recommendations are directional rather than concrete, and have assumptions and limitations, we believe our adaptation of the human and AI cooperation spectrum is an important first step in understanding the nuances of human plus machine teaming. We use the spectrum and apply simulation techniques to quantify trade-offs between performance, accuracy and impact on customer service in a call center setting. The application of simulation for imagining the effects of human plus machine is an important contribution to improving metrics that affect customer service and have the potential to alter how we think about the future of work. More work needs to be done to more accurately model these real-world call centers and incorporate other factors such as technology appropriateness and human-centered design into the model. By making intentional steps in measuring and evaluating the trade-offs in human and machine collaborations, mutually beneficial goals are more quickly to be realized.

VII. ACKNOWLEDGEMENTS

The authors would like to thank Domenico Amodeo, for his knowledge of simulation-related topics.

REFERENCES

[1] *12 Call Center KPIs to Track for Customer Success*. 2020. URL: https://www.talkdesk.com/blog/12-call-center-kpis-to-track-for-success/.

[2] H.A. Abbass. "Social Integration of Artificial Intelligence: Functions, Automation Allocation Logic and Human-Autonomy Trust". In: *Cogn Comput* 11 (2019), pp. 159–171. DOI: https://doi.org/10.1007/s12559-018-9619-0.

[3] Saleema Amershi et al. "Power to the People: The Role of Humans in Interactive Machine Learning". In: 35 (2014), pp. 105–120. DOI: 10.1609/aimag.v35i4.2513. URL: https://www.aaai.org/ojs/index.php/aimagazine/article/view/2513.

[4] "Average Call Center Agent Hourly Pay". In: (2020). URL: https://www.payscale.com/research/US/Job=Call_Center_Agent/Hourly_Rate.

[5] Gagan Bansal et al. "Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff". In: *In Thirty-Third Conference on Artificial Intelligence* (2019). DOI: https://doi.org/10.1609/aaai.v33i01.33012429.

[6] Eric Bonabeau. "Agent-based modeling: Methods and techniques for simulating human systems". In: *Proceedings of the National Academy of Sciences* 99.suppl 3 (2002), pp. 7280–7287. ISSN: 0027-8424. DOI: 10.1073/pnas.082080899. eprint: https://www.pnas.org/content/99/suppl_3/7280.full.pdf. URL: https://www.pnas.org/content/99/suppl_3/7280.

[7] Jesse David Dinneen. "Planning for Ethical Agent-Agent Interaction". In: Association for Computing Machinery, 2019.

[8] Robert Hoffman, Patrick Hayes, and Kenneth Ford. "Human-centered computing: thinking in and out of the box". In: *IEEE Intelligent Systems* 16.5 (2001), pp. 76–78.

[9] Matthew Johnson and Alonso H. Vera. "No AI Is an Island: The Case for Teaming Intelligence". In: *AI Magazine* 40.1 (2019), pp. 16–28. DOI: 10.1609/aimag.v40i1.2842. URL: https://www.aaai.org/ojs/index.php/aimagazine/article/view/2842.

[10] Walter Lasecki et al. "Real-Time Captioning by Groups of Non-Experts". In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST '12. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 23–34. ISBN: 9781450315807. DOI: 10.1145/2380116.2380122. URL: https://doi.org/10.1145/2380116.2380122.

[11] Yixiang Lim et al. "Cognitive Human-Machine Interfaces and Interactions for Multi-UAV Operations". In: Feb. 2019.

[12] Raja Parasuraman, Thomas Sheridan, and Christopher Wickens. "A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 30(3), 286-297". In: *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans : a publication of the IEEE Systems, Man, and Cybernetics Society* 30 (June 2000), pp. 286–97. DOI: 10.1109/3468.844354.

[13] Bogdana Rakova and Rumman Chowdhury. *Human self-determination within algorithmic sociotechnical systems*. 2019. arXiv: 1909.06713 [cs.CY].

[14] Bogdana Rakova and Laura Kahn. *Dynamic Algorithmic Service Agreements Perspective*. 2020. arXiv: 1912.04947 [cs.CY].

[15] Mark O. Riedl. *Human-Centered Artificial Intelligence and Machine Learning*. 2019. arXiv: 1901.11184 [cs.AI].

[16] Thomas B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA, USA: MIT Press, 1992. ISBN: 0262193167.

[17] H. James Wilson and Paul Daugherty. "Collaborative Intelligence: Humans and AI are Joining Forces". In: *Harvard Business Review* (2018). URL: https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.