Cognitive-Aware Multi-modal Human-Swarm Interface for Optimal Collaboration

Sooyung Byeon(✉), Joonwon Choi, and Inseok Hwang

Purdue University, West Lafayette, IN 47907, USA
{sbyeon,choi774,ihwang}@purdue.edu

Abstract. This paper presents the development of an interface that enables humans to collaborate with robot swarms (RSs) to effectively accomplish missions in highly uncertain environments. Autonomous RS technology has advanced to the point where robots can perform various tasks without human intervention. However, there are still two major uncertainties in mission execution. The first uncertainty arises from the limitations of artificial intelligence (AI), such as the inaccuracy of computer vision systems. The second uncertainty comes from the decision-making aspect, where the mission requirements may not be accurately communicated to the RS. To overcome these uncertainties, it is crucial to design a system that allows humans to monitor the RS and actively participate in critical decision-making processes during missions. We design an interface to: 1) estimate human cognitive states to monitor the human partner, 2) determine the best communication methods to calibrate cognitive states, and 3) allocate functions among humans and robots to optimize team performance. We demonstrate this approach through drone search and rescue (SAR) missions.

Keywords: human-AI collaboration · robot swarm · cognitive state · multi-modal interface · function allocation

1 Introduction

Human-AI collaboration (HAC) combines the strengths of human cognitive capabilities and the accurate computational abilities of artificial intelligence (AI) to efficiently perform complex missions in dynamic and highly uncertain environments [1]. For example, in the manufacturing sector, particularly with Industry 4.0, research has focused on ways for robots and humans to share the same physical space and divide sub-tasks among themselves to accomplish a larger task [2]. In the field of vehicle control, shared control methods where humans and AI work together to perform complex maneuvers have been widely studied [3]. Additionally, human-in-the-loop research has been actively conducted to ensure that humans can participate in decision-making in a timely manner.

© The Author(s) 2026
M. Andreoni and S. Thakkar (Eds.): GENZERO 2024,
Proceedings of 1st GENZERO Workshop, pp. 155–164, 2026.
https://doi.org/10.1007/978-981-95-1050-4_19

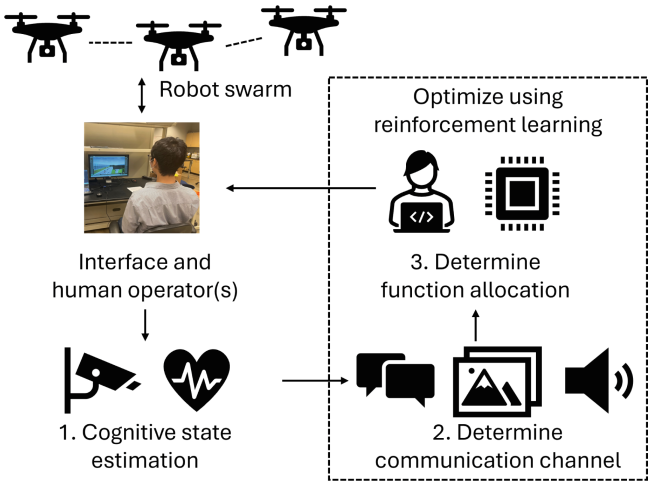


Fig. 1. The proposed human-swarm interface architecture consists of cognitive state estimation, communication channels, and function allocation.

Many studies have also focused on maintaining human cognitive states in a way that enhances close collaboration with AI [4].

For genuine collaboration, effective communication between all entities is crucial. Traditional HAC research has primarily focused on one-on-one or small group communication between humans and robots. However, with the emerging challenges of managing robot swarms (RSs) such as unmanned vehicle fleets and urban air mobility (UAM), where a few humans oversee many robots, there is a growing need to formally address the scalability of team communication. For instance, during a disaster where a RS performs search and rescue (SAR) missions, it is difficult for a human to monitor every robot. Robots must independently distribute among themselves and perform tasks, selectively relaying only essential information to humans. Humans need to set overall mission goals, understand the changing situation, make timely decisions, and issue necessary commands accordingly. Moreover, many RS operations, including SAR missions, take place in complex, uncertain, and unstructured environments. Therefore, determining how humans can efficiently collaborate with multiple robots to enhance team performance and safety is a critical issue.

2 Problem Definition

Our goal is to develop a human-swarm interface architecture that monitors human cognitive states, optimizes multi-modal communication channels, and allocate functions to ensure efficient and safe human-swarm collaboration as shown in Fig. 1. The problem can be decomposed into three sub-problems.

The first sub-problem involves collecting, analyzing, and learning from human behavioral and physiological data to numerically estimate the human cognitive

state. Understanding the counterpart's condition is a crucial starting point for effective human-swarm team communication. If AI is unaware of the human cognitive state, effective collaboration among them is difficult to achieve [4]. Therefore, our proposed architecture monitors humans and responds interactively. To avoid restricting human behavior during collaboration, we use non-invasive sensors such as face recognition camera and heart rate sensor (Sect. 3).

The second sub-problem is to establish multi-modal communication channels that allow fast interaction between humans and robots. Advances in the large language model (LLM) technology enable robots to engage in verbal communication with humans. While maintaining the concise information delivery capabilities of auditory and visual interfaces widely used in existing HAC research, we propose to utilize the latest LLM technology to facilitate the exchange of commands and feedback between humans and robots (Sect. 4).

The final sub-problem is to optimize function allocation between humans and robots, i.e., who does what and when [5]. Function allocation can leverage human's knowledge based behaviors and robot's skill and rule based behaviors. Based on the current human cognitive state and mutual understanding through communication, we can derive an optimized function allocation policy. This function allocation must be adaptive to both humans and situations to optimize overall team performance and respond flexibly to circumstances (Sect. 5).

3 Cognitive State Estimation

People know from experience that effective communication is difficult to achieve when the interaction counterpart is busy with other tasks (i.e., overloaded). Additionally, following the decisions of someone who does not have an accurate understanding of the situation can jeopardize the team's safety. Therefore, we aim to develop methods for estimating the workload in the proposed architecture. We further plan to extend our solid foundation in cognitive state modeling approaches to address other cognitive states such as situational awareness and trust [6].

3.1 Workload

Workload represents the cognitive load a person feels while performing a task. We utilize the latest cognitive architectures to develop components that estimate workload in real-time based on the situation. There have been several works to estimate human workload based on physiological sensors. For instance, electrocardiogram (ECG) and infrared image were utilized along with features extracted using an open source software [7,8]. Motivated by this, we train models using deep learning techniques to understand how workload changes when a human collaborates with multiple robots simultaneously, communicates in real-time, and performs various functions intermittently. Our goal is to maintain the workload of humans at an optimal level, not too high and not too low (i.e., workload calibration), to ensure that humans can interact with robots in the best possible state.

3.2 Situational Awareness and Trust

As future work, we plan to extend our proposed architecture to encompass situational awareness and trust. The widely accepted theory of situational awareness categorizes human understanding of the current situation into three levels [9]. Level 1 is when a human perceives the situation, Level 2 is when a human comprehends the situation, and Level 3 is when a human can project future status based on this understanding. Quantitative models of situational awareness have been extensively studied based on this theory. Using the latest research, we plan to analyze the relationships between situational awareness, other cognitive states, and team performance. Based on this analysis, we determine the methods and depth of communication and function allocation.

The relationship between the transparency of communication (between humans and robots) and trust has been widely studied [10]. For instance, if a robot (AI) provides explanations to humans based on high transparency regarding the current situation and future predictions, it can help humans achieve Level 3 situational awareness, leading to increased trust in AI. We are planning to train and analyze a deep learning model to understand the correlations between various features such as human facial expressions, heart rate, and eye movements, and their trust levels.

3.3 Trade-Off Between Cognitive States and Performance

We note that simply increasing transparency and trust cannot be the only approach for effective mission performance between humans and RS. This is because communication transparency, workload, and situational awareness have complex trade-off relationships. If transparency is increased for all aspects of RS operations, it may lead to human overload, ultimately undermining humans' situational awareness and trust. Such situations can negatively impact function allocation and limit the human's ability to perform effectively as a team member. Therefore, we utilize computational cognitive models which we have developed [11] to numerically analyze these trade-offs and aim to validate and improve them through human user studies.

4 Communication Channels

State-of-the-art LLM technology enables humans to communicate with RSs using natural language which is one of the most intuitive methods. When human commands are abstract or implicit, inferring human intention can be challenging, but translating human commands into mathematically formal methods like temporal logic can help prevent misunderstandings between humans and RS [12]. Hence, we plan to represent the state and control objectives of the RS using generate natural language to effectively communicate this to humans.

However, natural language is not the only effective communication method. For instance, in cases where a robot's path is unsafe, displaying relevant information on a graphical user interface can be a faster and more efficient communication method compared to complex natural language feedback provided by AI.

Hence, we will maintain the use of auditory and visual communication as supplementary tools while researching how communication in complex environments can be conducted through natural language.

5 Function Allocation

Function allocation involves making decisions about which functions each team member should perform to effectively achieve the team's goals [5]. Our research group has studied adaptive function allocation techniques based on a computational model which is formulated as the (partially observable) Markov decision process to optimize team performance and cognitive states [6]. For function allocation to be effective, decisions must consider the relationship between human cognitive states and transparency. While these complex relationships are challenging to model mathematically, we have built a reliable simulation platform using the computational cognition-work model (CCWM) [11]. We can determine the components of the human-swarm interface based on the CCWM model. Subsequent human user studies would allow us to precisely estimate the model parameters and verify the accuracy of the model. This model is designed to be extensible, not just for specific missions. The proposed architecture then can optimize human-swarm collaboration using a reinforcement learning technique by determining two controllable variables: the communication channel and function allocation policy as shown in Fig. 1.

6 Hardware Implementation

To validate the proposed architecture in an environment similar to real-world operations, we have developed interface software integrated with a drone swarm hardware system. For rapid development and iterative testing, we are operating multiple Crazyflie drones in an indoor environment and create a mixed reality environment to simulate complex drone operations as shown in Fig. 2. Using a motion capture system (Qualisys), we can obtain millimeter-level precision in real-time drone position data. Additionally, we have secured software development kits that allow real-time communication with various physiological sensors (e.g., Polar H-10 heart rate sensor and webcam) to create integrated software that combines graphical user interfaces, sensor data, and drone data. The hardware implementation presented in this paper is an extension of our previous work [13].

7 Preliminary Results

We conducted a pilot study to demonstrate the proposed approach, using a drone SAR scenario where a human collaborates with two drones to locate victims in a mountain area. The drones, equipped with cameras, are tasked with visiting six target points and autonomously planning their trajectories using the rapidly

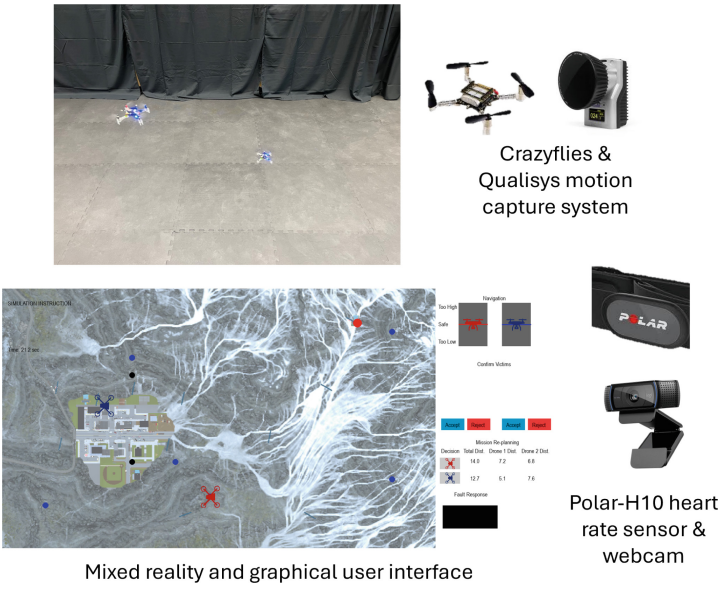


Fig. 2. The hardware elements for developing the proposed human-swarm collaboration interface.

exploring random tree (RRT). When a new target point emerges mid-mission, the drones reallocate their assigned points. They send images of potential victims, which the human confirms as true victims or misidentified objects due to the drones' limited vision.

The human monitors the drones' navigation, confirms victims, and responds to environmental changes (e.g., wind gusts) via a graphical interface. They also monitor drone positions, altitude, and status and can override the drones to reassign target points or manage harsh wind conditions.

The interface consists of a graphical display, sensors, and an LLM-based feedback module. A Polar H10 heart rate sensor and webcam track the human's physiological data, while a cognitive model estimates their workload in real-time. The LLM-based module provides adaptive mission information by generating natural language feedback through the interface or an auditory text-to-speech (TTS) tool.

Three different fixed function allocation (FA) policies were tested. The functions allocated to the human for each FA are as follows:

- FA 1 (simple): Monitoring the drones' states and confirming identified victims.
- FA 2 (intermediate): All functions in FA 1, plus reallocating drones when a new target point appears.
- FA 3 (complex): All functions in FA 2, plus responding to environmental changes, if any.

In our pilot study, we gathered NASA-TLX results for each FA case, and that reveals that FA 1, FA 2, and FA 3 involve (relatively) low, medium, and high workload.

We have trained a workload model capable of estimating low and high workload levels. The long short-term memory (LSTM) network used for workload estimation includes two LSTM layers with 50 units each, followed by a fully connected layer. The input features include ECG measurements from a heart rate sensor, along with facial landmarks and gaze data from OpenFace [7].

The procedure of the user study is as follows. First, the human is asked to remain in front of the interface to calibrate the sensors. Then, the human performs three sessions of the SAR scenario. In the first session, FA 1 is applied, and the workload is estimated. Once the workload is gathered, the function allocation is adjusted according to the following rule: if the human's workload is high, simplified feedback is provided to reduce the workload, and a less complex function allocation is assigned. Conversely, if the workload is low, the feedback module offers more detailed information to improve transparency, and a more complex function allocation is assigned. Two examples of LLM-generated feedback are shown in Fig. 3.

```
Current workload is low. Overall mission status remains unchanged.

Please perform the following human tasks:
 1. Monitor the drone state.
 2. Set and update the searching area.
 3. Find the victims.

Meanwhile the following drone tasks will be performed by UAV:
 1. Monitor the drone faults.
 2. Monitor the environmental condition.
```

```
Workload status: High
warning! Answer will be short and concise to reduce human burden.

 1. Monitor the drone state.
 2. Set and update the searching area.
 3. Find the victims.
```

Fig. 3. The example LLM-generated feedback when the human's workload is low (top) and high (bottom), respectively. A text-to-speech (TTS) tool can generate auditory feedback based on the generated script.

Figure 4 presents the estimated workload of the human. The average workload aligns with the NASA-TLX results for each FA case (FA1, FA2, and FA3). An interesting observation is that high workload was recorded in the latter part of the second session, which was performed with FA2. During this period, target reallocation and victim confirmation occurred simultaneously. This indicates that real-time workload estimation can be used for adaptive function allocation during each session.

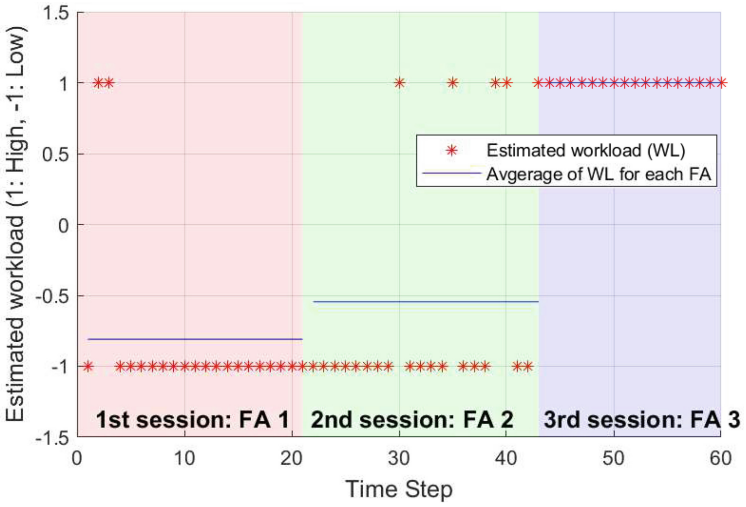


Fig. 4. The estimated workload (WL) for each session and the corresponding allocated function.

8 Conclusions

A computational approach for effective human-swarm collaboration was presented. The key idea of the proposed approach is to explicitly account for the human cognitive state, which can be monitored in real-time using behavioral and physiological data. Our ongoing work can be categorized into three parts. First, we are expanding our cognitive model to incorporate additional cognitive states such as situational awareness and trust. Second, a series of human user studies is being conducted to gather more behavioral and physiological data. Finally, an adaptive function allocation policy (i.e., time-varying policy) and advanced interactive feedback will be developed using the deep Q-learning approach.

Furthermore, our research group has the capability to quickly scale up our initial results using the PURT (Purdue UAS Research and Test facility) with significant expertise in the field of HAC through our work on projects[1] that encompass computational cognitive models [6] and human-autonomy teaming [11]. Based on this, we plan to expand the proposed architecture to more complex and larger scale of platform which several drones and human can interact simultaneously.

Acknowledgments. This work was partially supported by NSF CNS-1836952.

[1] Cognitive Autonomy Project (NSF CNS-1836952) and Safe Autonomy Project (NASA 80NSSC20M0160).

References

1. National academies of sciences, engineering, and Medicine: human-AI teaming: state-of-the-art and research needs. The National Academies Press, Washington, DC (2022). https://doi.org/10.17226/26355
2. Tadesse, A.A., Wang, K.J., Lin, C.J.: Task allocation problem between human–robot collaboration team. In: Intelligent and Transformative Production in Pandemic Times, pp. 473–482. Springer (2023)
3. Byeon, S., Choi, J., Zhang, Y., Hwang, I.: Stochastic-skill-level-based shared control for human training in urban air mobility scenario. ACM Trans. Hum.-Robot Interact. **13**(3) (2024). https://doi.org/10.1145/3603194
4. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J. Cogn. Eng. Decis. Making **2**(2), 140–160 (2008). https://doi.org/10.1518/155534308X284417
5. Roth, E.M., Sushereba, C., Militello, L.G., Diiulio, J., Ernst, K.: Function allocation considerations in the era of human autonomy teaming. J. Cogn. Eng. Decis. Making **13**(4), 199–220 (2019). https://doi.org/10.1177/1555343419878038
6. Byeon, S., Choi, J., Hwang, I.: A computational framework for optimal adaptive function allocation in a human-autonomy teaming scenario. IEEE Open J. Control Syst. **3**, 32–44 (2024). https://doi.org/10.1109/OJCSYS.2023.3340034
7. Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.P.: Openface 2.0: facial behavior analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 59–66 (2018). https://doi.org/10.1109/FG.2018.00019
8. Cardone, D., et al.: Classification of drivers' mental workload levels: comparison of machine learning methods based on ECG and infrared thermal signals. Sensors **22**(19), 7300 (2022)
9. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–64 (1995). https://doi.org/10.1518/001872095779049543
10. Chen, J.Y.C., Lakhmani, S.G., Stowers, K., Selkowitz, A.R., Wright, J.L., Barnes, M.: Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theor. Issues Ergon. Sci. **19**(3), 259–282 (2018). https://doi.org/10.1080/1463922X.2017.1315750
11. Byeon, S., Tian, D., Ayoub, J., Song, M., Moradi Pari, E., Hwang, I.: Optimal function and attention allocation for human-AI collaboration using computational cognition-work model. In: Proceedings of the 2024 IEEE 63rd Conference on Decision and Control (CDC), pp. 2269–2274 (2024). https://doi.org/10.1109/CDC56724.2024.10886315
12. Cosler, M., Hahn, C., Mendoza, D., Schmitt, F., Trippel, C.: nl2spec: interactively translating unstructured natural language to temporal logics with large language models. In: Computer Aided Verification, pp. 383–396. Springer (2023)
13. Byeon, S., Yuh, M., Choi, J., Jain, N., Hwang, I.: Workload classification for function allocations in human-autonomy teaming using noninvasive measurements. In: AIAA SciTech 2025 Forum, AIAA (2025). https://doi.org/10.2514/6.2025-2253

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

