Fostering Trust Through User Interface Design in Multi-Drone Search and Rescue

Ahlskog, Johanna; Bahodi, Maria-Theresa; Lugmayr, Artur; Merritt, Timothy Robert

Published in:
TAS 2024 - Proceedings of the 2nd International Symposium on Trustworthy Autonomous Systems

DOI (link to publication from Publisher):
10.1145/3686038.3686052

Publication date:
2024

Fostering Trust Through User Interface Design in Multi-Drone Search and Rescue

Johanna Ahlskog
Umeå University
Umeå, Sweden
johanna.ahlskog@outlook.com

Maria-Theresa Bahodi
Aalborg University
Aalborg, Denmark
mtoh@cs.aau.dk

Artur Lugmayr
University of Western Australia (UWA) & UXMachines Pty Ltd & Edith Cowan University (ECU)
Perth, Australia
artur.lugmayr@artur-lugmayr.com

Timothy Merritt
Aalborg University
Aalborg, Denmark
merritt@cs.aau.dk

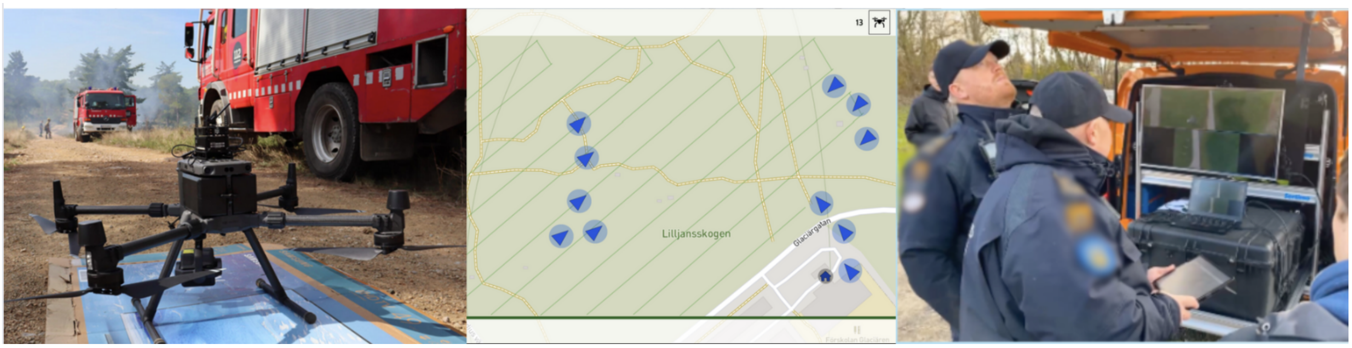

Figure 1: (Left) Single drones have become critical tools to help teams on the ground in search and rescue missions. (Center) Semi-autonomous multi-drone systems hold potential to reduce search time, yet the added complexity of such user interfaces may impact trust and task performance. (Right) Development and evaluation of functional prototypes with professionals helps to guide design and identify issues impacting trust and performance. ***(Used with permission from [12])***

ABSTRACT

Unmanned Aerial Vehicles (UAVs), or drones, are increasingly used in search and rescue (SAR) missions, with pilots transitioning from manual control of single drones to more collaborative tasks orchestrating semi-autonomous fleets. Designing user interfaces to support UAV pilots effectively is crucial to improving the success of search missions. We developed two versions of a multi-drone SAR system prototype to simulate SAR missions and evaluated them with professional UAV SAR pilots in Sweden. Both versions showed the flight paths of the UAVs, yet in one version, a heatmap was overlayed to provide information from a lost person model. We evaluated situational awareness (SA), cognitive workload, and trust in SAR scenarios. Results showed reduced cognitive load and increased trust with the heatmap-guided interface. Findings from the contextual interviews suggest three design implications for increasing trust and acceptance for future multi-drone user interfaces.

TAS '24, September 16–18, 2024, Austin, TX, USA

ACM ISBN 979-8-4007-0989-0/24/09
https://doi.org/10.1145/3686038.3686052

CCS CONCEPTS

• **Computing methodologies → Multi-agent systems**; • **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Unmanned aerial vehicle, Human-drone interaction, Cognitive workload, Trust, Search & Rescue, Drone swarm, Situational awareness, Multi-drone systems

ACM Reference Format:
Johanna Ahlskog, Maria-Theresa Bahodi, Artur Lugmayr, and Timothy Merritt. 2024. Fostering Trust Through User Interface Design in Multi-Drone Search and Rescue. In *Second International Symposium on Trustworthy Autonomous Systems (TAS '24), September 16–18, 2024, Austin, TX, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3686038.3686052

1 INTRODUCTION

Searching for missing people is a challenging task, especially with critical demands of the mission in diverse and dynamic environmental conditions and rugged terrain. *Unmanned Aerial Vehicles (UAVs)*, including drones, have become crucial in these *Search and Rescue*

(SAR) operations. While manually controlled UAVs have shown effectiveness, there's growing interest in multi-drone systems to increase coverage and reduce mission duration. Yet with additional drones, there is a need for higher levels of autonomy and advanced AI to coordinate the swarm. The development of such multi-drone systems are expected to benefit police and rescue organizations, however, transitioning from supervising single drones to multi-drone systems poses various challenges. Multi-drone systems need intuitive user interfaces to prevent confusion or misinformation, vital for mission success and trust in the system [17]. In this paper, we studied how incorporating data visualization in the interface can aid *Situational Awareness (SA)*; lower cognitive workload during stressful SAR missions; and increase the trustworthiness of the multi-drone system. To contribute to the research on the design of trustworthy autonomous systems, we provide three implications for design and discuss important next steps for the research on multi-drone systems.

2 RELATED WORK

2.1 Search and Rescue Operations with Drones

SAR operations face significant challenges including support technologies, effective team communication, comprehensive training, and time constraints [25]. The integration of UAVs or drones into SAR missions has revolutionized the field, offering solutions to many of these challenges. Drones provide rapid response capabilities, enhanced situational awareness, improved detection and recognition, communication support, and even lighting for night operations. However, they also face limitations in flight time, weather conditions, and aviation regulations [20]. To overcome these limitations and maximize efficiency, multi-drone operations or swarms have been developed. Drone swarms involve multiple UAVs collaborating towards specific goals, offering faster coverage of large search areas, enhanced data collection and processing, and improved resilience [24]. There are recent offerings from defense contractors offering multi-drone systems, yet the interfaces to support the operator are adapted from the single drone systems. Red Cat Holdings introduced a four-drone integrated swarm system for public safety and defense markets, allowing one pilot to oversee multiple drones simultaneously [1].

The *International Aeronautical and Maritime Search and Rescue (IAMSAR)* guidelines outline standardized procedures for SAR operations, including common search patterns such as the Expanding Square search and Parallel Sweep search. The Expanding Square search is suitable for small search areas with a defined starting point [7], while the Parallel Sweep search is effective for large search areas with uncertain target locations [26].

2.2 Human Factors in Drone-Assisted Search and Rescue

The effectiveness of drone-assisted SAR operations heavily relies on the interaction between human operators and the drone systems with safety as a primary concern [21]. Key considerations include monitoring tasks across operational levels and the use of multiple dynamic displays in ground control stations [3]. Cognitive workload, which refers to the mental effort needed for task completion, is a critical factor in SAR missions [23] and the management of stress and fatigue is crucial for maintaining optimal performance. Trust plays a vital role in human-drone interaction, influencing collaboration, decision-making, and overall system performance [4].

Situational Awareness (SA) is another crucial aspect of effective decision-making in drone operations. Endsley's model [8] defines three levels of SA: perception of elements in the environment; comprehension of the current situation; and projection of future status. Complementing SA with context awareness refers to a system's ability to understand and adapt to its surroundings, offering personalized services based on information related to people, locations, timing, and events [1, 14]. One such information source in SAR is the *Lost Person Model*, which aids in predicting the behavior of missing individuals, typically using a spatial 'ring model' or Euclidean distance model to indicate potential distances based on past incidents [6, 22]. This model, combined with effective data visualization techniques such as heatmaps, has been shown to help pinpoint search areas and optimize SAR operations [9, 28].

Current research focused on drone swarms in SAR missions explores the balance between human oversight and drone autonomy [2, 10, 11, 27]. Cloud-based applications for real-time monitoring of multiple UAVs have been developed [15], and there's an emphasis on creating mission-specific user interfaces to minimize cognitive workload [13]. Despite these advancements, several gaps remain in the field. There's a need for interfaces designed for independently operating UAVs in remote locations, particularly within drone swarms. The challenge lies in managing varied map views and information during UAV transitions, requiring rapid comprehension and adaptation. Additionally, translating theoretical concepts of situational awareness, cognitive workload, and trust into practical interface designs remains a significant hurdle. Källbäcker and Bjurling [17] emphasize the need for comprehensive system design, workload reduction strategies, and trust-building methods to enhance SAR team efficiency. Future research should focus on addressing these human factors simultaneously in user interface design for multi-drone systems, especially in demanding SAR scenarios. This holistic approach will be crucial in further improving the effectiveness and efficiency of drone-assisted search and rescue operations.

3 RESEARCH PROBLEM

While there has been exciting advancements in the technical capabilities of drones for SAR, in order to develop effective multi-drone SAR systems, it is important to understand how user interface design affects the pilot's ability to manage and control missions. Our investigation examines relevant human factors in relation to interactions with a multi-drone SAR system and guided by the following research questions:

(1) How do pilots maintain situation and mission goal awareness when low level flight information (e.g. battery status) is provided during a mission?
(2) How to design UIs and visualisations to ensure trust and achievement of mission goals in multi-drone systems on the example of the lost person model?

[1]https://www.unmannedairspace.info/latest-news-and-information/red-cat-holdings-brings-world-first-fully-operational-multi-drone-system-to-market/

(3) Which other human factors as cognitive workload, task monitoring, or mission goals affect pilots during SAR missions?

4 METHOD

To study the challenge of user interface design for multi-drone SAR systems and to approach the research questions, we developed a semi-functional prototype taking inspiration from previous related examples [2, 5, 10–12] and then evaluated with ten drone professionals.

4.1 Participants

All participants in the prototype's user study ($N = 10$) had previous experience in working with single-drones and SAR operations across various institutions and SAR organizations in Sweden, as detailed in Table 1. Among these participants, there was 1 female participant and 9 male participants. The participants' ages ranged from 25 to 65 years, with a mean age of 44 years ($SD = 12.07$). Their experience in SAR varied from 2 to 22 years, with a mean experience of 10.60 years ($SD = 7.49$).

According to Table 1, a significant majority (90%) of the participants are formally educated UAV pilots. Their experience with drones varied from 1 to 10 years, with a mean experience of 4.10 years ($SD = 3.10$). Except for UAV piloting, participants mentioned diverse roles such as police officers, police assistants, leaders/managers in SAR operations, maritime crew members, and individuals involved in SAR research and development.

The application context and experiences of drones mentioned among these participants involved traffic and public gathering monitoring, surveying, filming, crime prevention, crime scene investigations, missing person searches on land, maritime rescue, and firefighting. All participants claimed that drone technology is critical within their missions.

4.2 Multi-drone SAR Prototype

Essential components of the drone prototype include the map view of the search area, drones, flight paths, system messages, and video feeds from the drones. In one condition, a heat map provided a visualization of a lost person model overlaid on the map, in the other condition, the heatmap was not shown. The algorithm employed within the drone prototypes operates as a basic parallel sweep search algorithm for drone swarms, navigating a grid-based environment from a home base as shown in Figure 2.

The drone swarm follows the path during the mission and the participants are presented with system messages such as low battery notification and suggestions as shown in 3.

In one situation, the visualization of the lost person model is provided as a heatmap overlay on the map. The lost-person behavioral model estimates the *Probability of Area (POA)* for each location within the search area over time. This drone prototype visualizes these POA values using a heatmap, pinpointing the location with the highest likelihood of finding the lost person as shown in Figure 4. The visualization marks the initial planning point based on the *Last Known point (LKP)* as a red triangle and outlines four symmetrical rings centered around the LKP. The 2 outer rings are marked red, whereas the inner two rings are a part of the heatmap where darker colors denote higher probability areas and lighter colors for decreasing probabilities. These rings indicate distances of 25%, 50%, 75%, and 95%.

A screen with current video feeds as shown in Figure 5 enables the participant to view the detected objects/people and adds realism to the search missions. The video feeds were synthetic and did not show actual feeds from the search area, however they did show terrain matching the type found at the position of the respective drone.

4.3 Procedure

In alignment with GDPR, all participants were requested to provide their consent before the user study, outlining the processing of their personal information and data. This consent was essential, enabling video and camera recording during the user study. The user study was conducted online using either Zoom or Microsoft Teams. Video recordings were used within the software platforms to observe participants and visual cues during the interaction and their screen. A short training session was provided, followed by two missions lasting approximately ten minutes each, one with the heatmap and one without. The order was balanced and randomized. During the missions, the screen was randomly frozen three times during which, the participant was asked three SAGAT questions to measure situation awareness. After each mission, participants filled out a questionnaire to measure their experience, trust, and contextual interview questions.

For each mission, participants were presented with a user scenario aiming to provide context for the mission similar to a real-world search and rescue situation. The scenario was described to participants as follows. The first mission tasks did not invovle the use of heatmap as additional visualization component (Task 1);

> *"You are working as a drone operator tasked with a SAR mission. Your role involves using drone swarms via a SAR application to monitor and control them during the mission. Upon receiving an emergency call, a flight path is generated within the SAR application outlining the search area. Your responsibility is to oversee the drones' status while the swarms autonomously navigate through the defined flight path. Prepare for takeoff to execute the mission."*

The second scenario invoved the lost person model heatmap as additional visualisation element (Task 2). The following additional details have been provided to the participants:

> *"... in response to an emergency call, you now have access to a heat map displayed on the drone map view, created from the last known location of the missing person. As the drone swarms autonomously traverse the flight path, your careful attention is required, especially when flying over the heat map area. Prepare for takeoff to execute the mission."*

Participants were then instructed monitor the drone swarm until they found the missing person. They were asked to think aloud and verbally express their thoughts and observations while interacting with the drone prototype. Throughout the session, the test leader observed the participants' behavior, noted their comments, and system interactions in a note-taking application. Additionally, the

Table 1: Background of participants ($N = 10$)

Institution/SAR organization	Participants	Educated UAV pilot
The Swedish Police	4	Yes
Swedish Sea Rescue Society	3	Yes (2), No (1)
The Swedish Maritime Administration	1	Yes
Missing People Sweden	1	Yes
Swedish Emergency Services	1	Yes

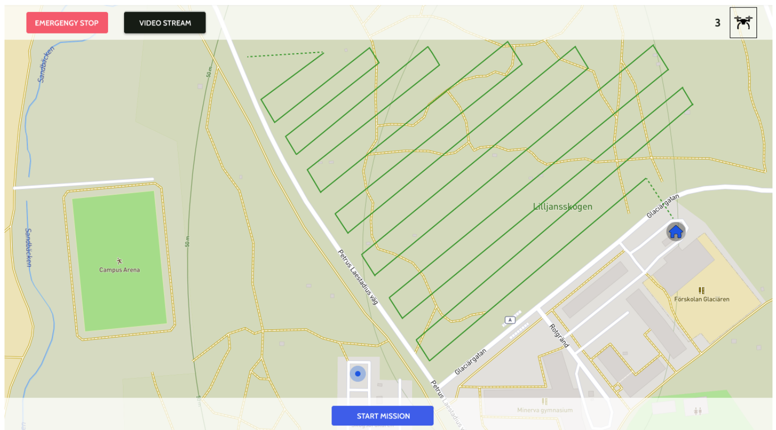

Figure 2: Start mission screen: *home* icon representing the starting point of the drone swarm and the *blue line* in the mission area indicating the parallel sweep search flight path.

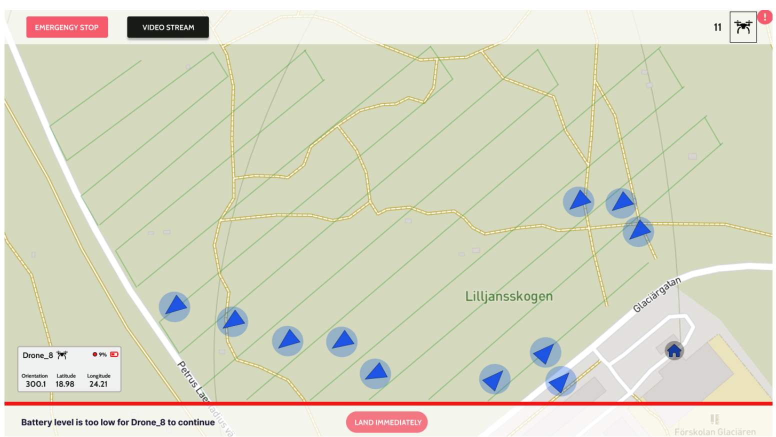

Figure 3: Message with a notification concerning aerial drone systems (low battery) to test drone pilots in situation awareness. The correct action to be taken woudl be to land the drone immediately.

test leader reviewed camera and video recordings afterwards to capture any visual cues or details from the participants.

4.4 Measures

To measure user experience, trust and task performance, we utilized a freeze-probe protocol with SAGAT questions, post mission questionnaires and semi-structures intrview questions.

4.4.1 Situation Awareness Global Assessment Technique Questions. While performing the user task in the prototypes, SAGAT questions were asked in sets of three, with each set containing one question related to each of the three levels of SA. This questioning procedure was done when the prototype was paused unexpectedly at certain points in the prototype flow, using the so called the freeze technique. Moreover, the SAGAT questions of the three levels of SA would incorporate the pilot's present understanding of the various design components within the interface. Table 2 presents SAGAT questions used in the user study categorized according to various design components *C1 - C6* and the three levels of SA. The scores of SA

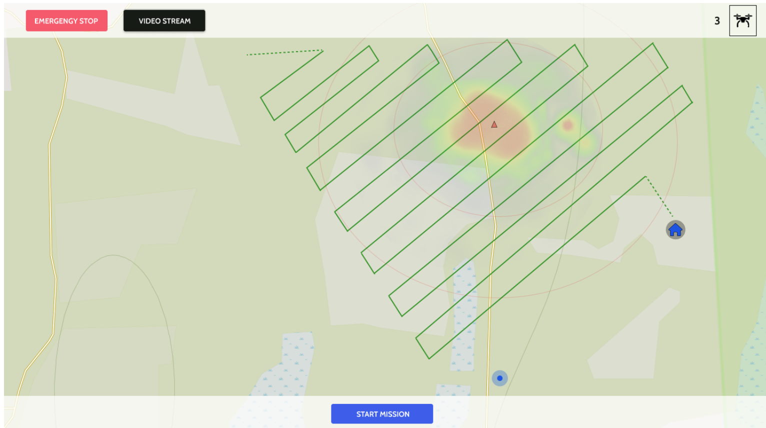

Figure 4: Map with a heatmap layer generated through the lost person model. Darker colors denote higher probability of detection of person in that area.

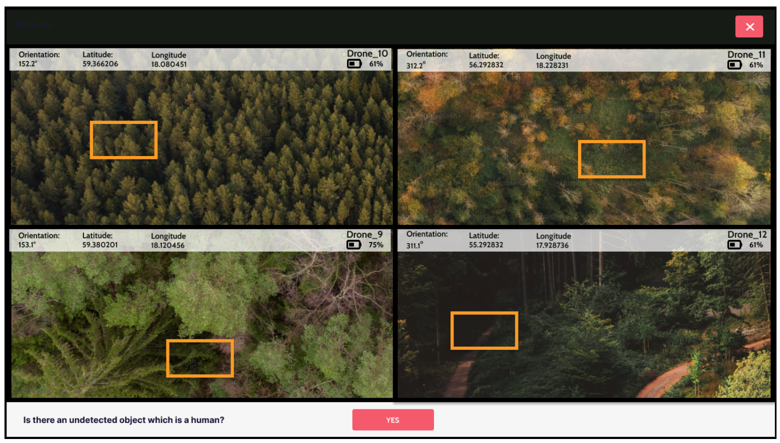

Figure 5: Video stream of multiple drone views showing telemetry data for each single UAV.

are determined based on the correctness of responses to SAGAT questions.

4.4.2 Post mission questionnaires. After each mission, participants answered questionnaires to measure cognitive load, trust, and to gather open-ended feedback. The *Nasa-TLX* questionnaire was administered to measure cognitive workload. The *Human-Computer Trust Scale* measured trust across the three components of ability, integrity, and benevolence (rated on a 1−7 Likert scale). Open-ended questions focused on suggestions for improving the prototype and any additional explanations offered by the participant for their actions during the mission.

4.4.3 Semi-structured interviews. To gather participants' feedback, semi-structured interviews were conducted following the post-mission. The open-ended questions centered on suggestions for improving the prototype and any additional explanations provided by the participants about their actions during the mission.

5 RESULTS

5.1 Situation Awareness Global Assessment Technique

The quantitative SA scores were calculated from participants' accuracy in responding to SAGAT questions for each component and SA level, which each SAGAT question was described earlier in Table 2. Unexpectedly interrupted during prototype use, each participant encountered three sets of SAGAT questions per session. Each set contained three individual questions corresponding to SA levels 1 through 3, each linked to a specific UI component.

The main results are presented in Table 3 and it reveals that the UI overall, resulted a moderate level SA, achieving an accuracy of 54.44% across all 18 SAGAT questions. Notably, SA varied based on different levels: SA level 3 (projection) achieved the highest accuracy of 63.33%, followed by SA level 1 (perception) at 53.33%, while SA level 2 (comprehension) achieved a relatively lower accuracy of 46.66%.

Table 2: SAGAT questions in the user study, covering various UI components and the three levels of situation awareness.

#	Component	Question for SA Level 1 PERCEPTION	Question for SA Level 2 COMPREHENSION	Question for SA Level 3 PROJECTION
C1	Main map view	Which drone caused the object alert?	What is the current battery level of the drone, shown as a drone card?	How many active drones are currently flying after you have clicked the button?
C2	Notification icon	Which drone caused the object alert?	What is the current battery level of the drone, caused the object alert?	What will happen next in the application?
C3	System messages	Which color was the latest system message?	What was the latest message?	How many drones will you see in the next view of the application?
C4	Drone status in drone side menu	Which drone caused the object alert?	What was the last system message to alert on the drone menu?	What will happen next in the application?
C5	Video Stream	Which drone did show a human in the camera feed?	What was the last system message in the video stream view?	What will happen next in the application?
C6	Heatmap	Which drone caused the object alert?	How many active drones are currently flying over the heat map in the map view?	What will happen next in the application?

Table 3: Summary of responses from participants ($N = 10$) for the SAGAT questions covering various design components and levels of situation awareness. The summary shows correctness of participant responses (*in percentages*)

#	Component	Accuracy on SA Level 1 PERCEPTION	Accuracy on SA Level 2 COMPREHENSION *(all results in % correct)*	Accuracy on SA Level 3 PROJECTION
C1	Main map view	70%	60%	60%
C2	Notification icon	20%	60%	40%
C3	System messages	60%	20%	70%
C4	Drone status in drone side menu	50%	90%	50%
C5	Video stream	70%	30%	80%
C6	Heatmap	50%	20%	80%

Table 4: Correctness of each component based on SAGAT questions (*in percentages*).

Component	C1	C2	C3	C4	C5	C6
Accuracy	63, 33%	36, 66%	50, 00%	50, 00%	60, 00%	50, 00%

Table 4 illustrates the evaluation performed to identify the component that provided the highest and lowest SA on the user interface. SA scores for various UI components labeled as $C1, \ldots, C6$ were calculated. The findings indicate that $C5$ (Video Stream) demonstrated the highest SA accuracy, reaching 60%, while $C2$ (System messages) exhibited the lowest SA accuracy, standing at 36.66%. The SA scores for the remaining components were closely grouped, ranging between 50% and 60%.

5.2 NASA Task Load Index

After completing both *Task 1* and *Task 2* after each prototype, participants evaluated their mental workload using NASA TLX. Utilizing a 10-point scale from low to high, i.e. raw ratings, each participant assessed the six NASA TLX subscales. The figures in 6 and 7 display the raw ratings, mean, and median values for each participant and subscale, while Figure 8 illustrates a comparison between these two tasks.

The comprehensive weight for each participant, for each task, is determined by the multiplication of the raw rating and the weight rating. In Figure 9, there is a representation of the standardized overall workload for each participant, considering their individual weight ratings. On average, the total workload amounted to 51, 88, reaching a maximum of 80, 83 and a minimum of 40 for *Task 1*; and a maximum of 62, 5 and a minimum of 38, 33 for *Task 2*.

The comparison between *Task 1* and *Task 2*, illustrated in Figure 8, revealed some interesting differences. Overall, participants found *Task 2* to be slightly more mentally demanding than *Task 1*. However, *Task 2* was less frustrating, creating a more comfortable experience than *Task 1*.

In *Task 1*, participants rated the mental challenge differently, experiencing varying levels of frustration, impacting their perceived effort and performance. *Task 2*, however, was perceived as slightly less mentally taxing and less frustrating, with similar performance and perceived effort compared to *Task 1*.

In summary, both prototypes required a significant cognitive workload. Notably, *Task 2* exhibited a slightly lower cognitive workload compared to *Task 1*.

5.3 Human-Computer Trust Scales

The trust scale employed in this user study is derived from the work of Jensen et al. [16] and is designed to assess key trust dimensions of the human towards the system. The trust aspects of *ability*, *integrity*, and *benevolence* were evaluated using Likert scales, with each aspect comprising four statements. To determine whether participants' trust in the system differed between *Task 1* and *Task 2*, a paired samples t-test was conducted, specifically examining the influence of the heatmap on trust levels, as seen in Table 5. The only significant difference was found for *benevolence*, with *Task 2* including the heatmap. Average values for the 3 aspects of trust are shown in Figure 10.

5.4 Semi-structured Interviews

At the conclusion of the user study, semi-structured interviews were conducted to gather participant feedback on the drone swarm prototype. The overall impression was mixed, with participants describing it as both useful and interesting, but also stressful and demanding, largely due to the limited time available to become familiar with the system. Many acknowledged the potential of drone swarms to improve search and rescue missions by enabling more efficient searches, recognizing the critical importance of time in these scenarios. However, participants expressed the need to review how to represent important information while flying multi-drones. The participants also expressed frustration with the limitations of managing autonomous drone swarms, particularly the lack of options beyond emergency landing or mission abortion.

A common sentiment among participants was:

> *"I didn't feel like I had sufficient control of the drone when warnings appeared – For example with low battery level, it would have been good to be able to choose to send a drone home or choose to land immediately. This is needed to be able to fly in a real-life environment"* (P6)

They emphasized the need for UAV pilots to maintain authority over drone operations, rather than relying solely on autonomous systems. Monitoring multiple drones simultaneously proved challenging, with some suggesting that a more manageable swarm size would be around four to five drones.

The system messages, while helpful in focusing attention, were criticized for their overwhelming frequency. Participants suggested filtering crucial information and centralizing notifications to reduce the need for constant vigilance. They also recommended implementing a persistent log for reviewing missed alerts, as illustrated by the following comment:

> *"It would be nice if the warnings remained until I choose to remove them, maybe in a scrolling log list so that one can go back if unable to read in time."* (P3)

The heatmap feature received positive feedback, though some found it difficult to focus on during drone failures. Participants proposed integrating the heatmap with search patterns and considering the *Last Known Position (LKP)* when defining search areas. Participants also proposed improvements on how to utilize the heatmap:

> *"I like the heatmap, it could have been better if there was an option to begin the search from that specific point in the heatmap and then create the flight route from there. Personally, I would prefer to investigate that area first."* (P8)

Desired improvements included enhanced user interface features such as sound alerts, zoom functionality for video streams, and night vision capabilities. The absence of thermal cameras was noted as a significant limitation. Participants also expressed a strong desire for manual control to investigate potential sightings, rather than relying solely on autonomous swarm behavior. As one participant noted:

> *"It would have been interesting to investigate this further, if I only could manually control a drone to stop at the possible object."* (P2)

Overall, while participants saw potential in the drone swarm system for search and rescue operations, they emphasized the need for refined information presentation, increased pilot control, and additional features to enhance its effectiveness in real-world scenarios.

6 DISCUSSION

6.1 Interpreation of Results

The study yielded several significant findings regarding the use of multi-drone prototypes in search and rescue operations. Participants generally exhibited low initial stress levels and reported average or calm experiences with the prototypes. *Situational Awareness (SA)* scores varied across different components and levels, while cognitive workload, though notable, showed a slight decrease when using the prototype with the heatmap. Task perception differed marginally, with the second task being viewed as slightly less demanding. Importantly, feedback from the second task demonstrated

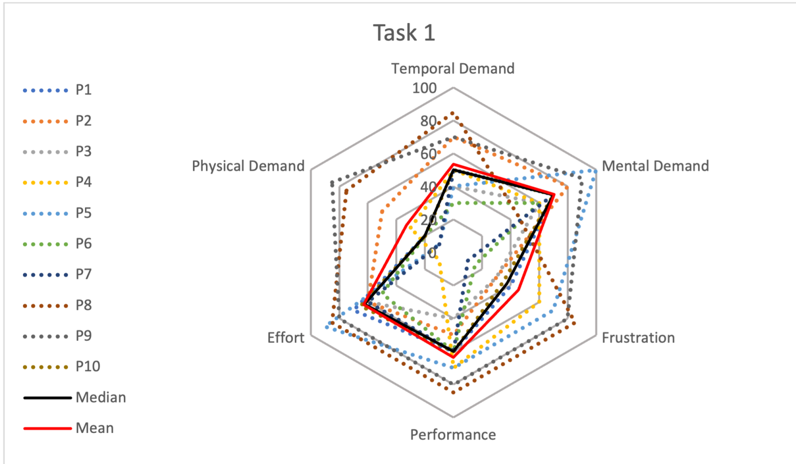

Figure 6: Results of NASA-TLX ratings from task 1

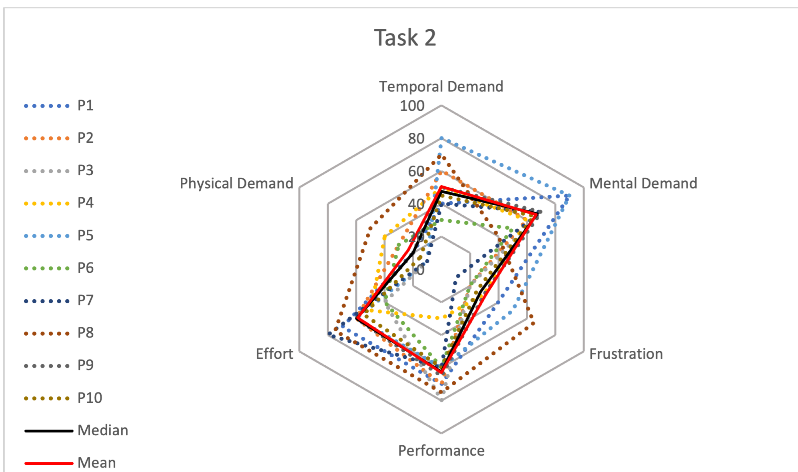

Figure 7: Results of NASA-TLX ratings from task 2

Table 5: Paired t-test raw data for comparing prototype 1 and prototype 2, related to the human-computer trust scale.

Trust Aspects	Statistical test results	
	T-statistic	P-value
Ability	[0, −0.318, 0.318, 0.152]	[1, 0.758, 0.758, 0.882]
Integrity	[1.627, 1.246, −0.802, 1]	[0.138, 0.244, 0.443, 0.343]
Benevolence	[−1.464, 1.309, 1.711, 3]	[0.177, 0.223, 0.121, **0.015**]

increased trust in the aspect of benevolence, although no significant changes were observed in ability or integrity. Qualitative data revealed participants' preferences for specific features, such as monitoring fewer drones, integrating audio warnings, and having individual UAV control. The heatmap feature was well-received, but participants expressed a need for improved system message review. These findings align with previous research by Källbacker and Bjurling [17] on UI improvements for multi-drone systems, as well as Lochner et al.'s [18] work emphasizing the importance of actionable recommendations in collaborative decision-making environments. However, the limited number of participants necessitates caution in drawing definitive conclusions from the quantitative data alone.

6.2 Trust and Heatmap Visualization

The human-computer trust scale showed an increase in the benevolence aspect of trust (see Figure 10). The heatmap received positive

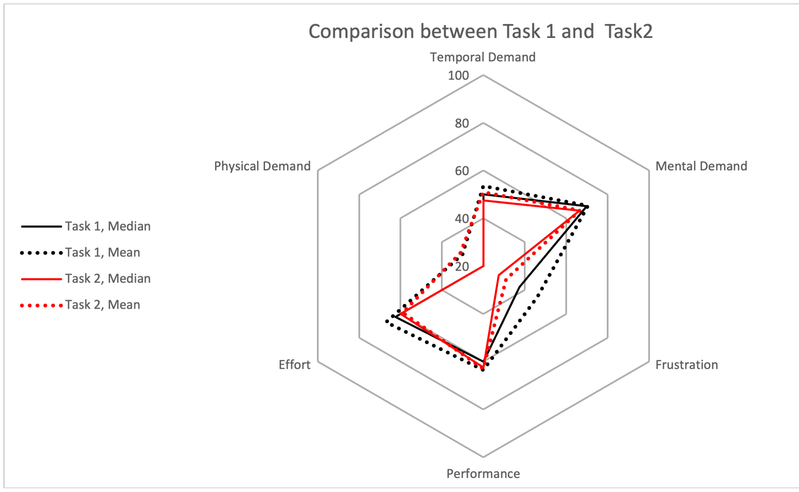

Figure 8: Comparison between *Task 1* and *Task 2* - NASA-TLX mean and median values

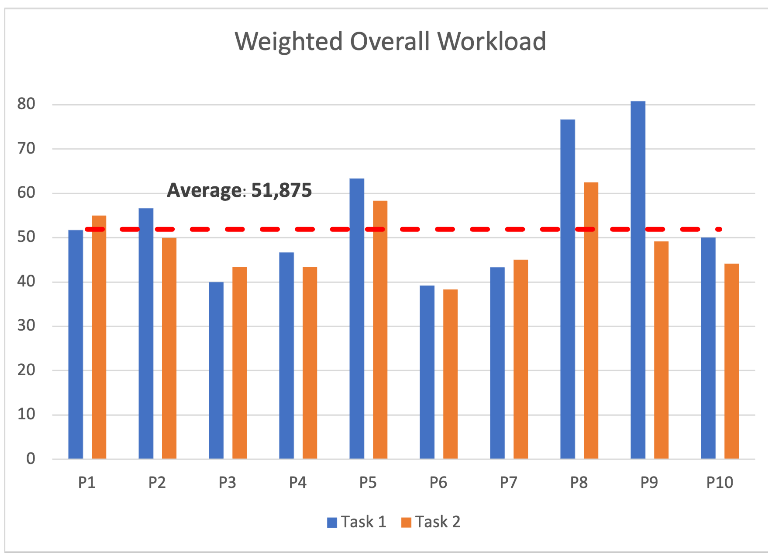

Figure 9: Graph illustrating the comprehensive weight outcomes

feedback (mean score 4.08 on a 7-point Likert scale), though challenges in ensuring accurate use of the heat map information persist. Cognitive workload averaged 51.88, with participants describing tasks as demanding and stressful.

6.3 Limitations

Time constraints limited prototype development to a single iteration and a simplified map view visualization, resulting in the absence of certain animations like live video feeds. Earlier testing with a small group could have provided timely feedback on UI aspects. Despite these limitations, the prototypes were sufficient to address the research questions.

6.4 Implications for Design

The insights from the design and evaluation process offer valuable implications for designing multi-drone interfaces tailored to experienced UAV pilots. This intermediate-level understanding is strengthened by presenting this research-backed findings alongside relevant theories and practical examples of human-multi-drone interaction [19].

Pilot adaptability and visual data representation. While some pilots may require additional training to handle the stress of managing drone swarms, others may readily explore system capabilities. User interfaces must adhere to key principles of data visualization, offering comprehensive representations without overwhelming users.

Optimizing task management in multi-drone systems. Through a limited prioritized 4-view interface, facilitating quick adjustment

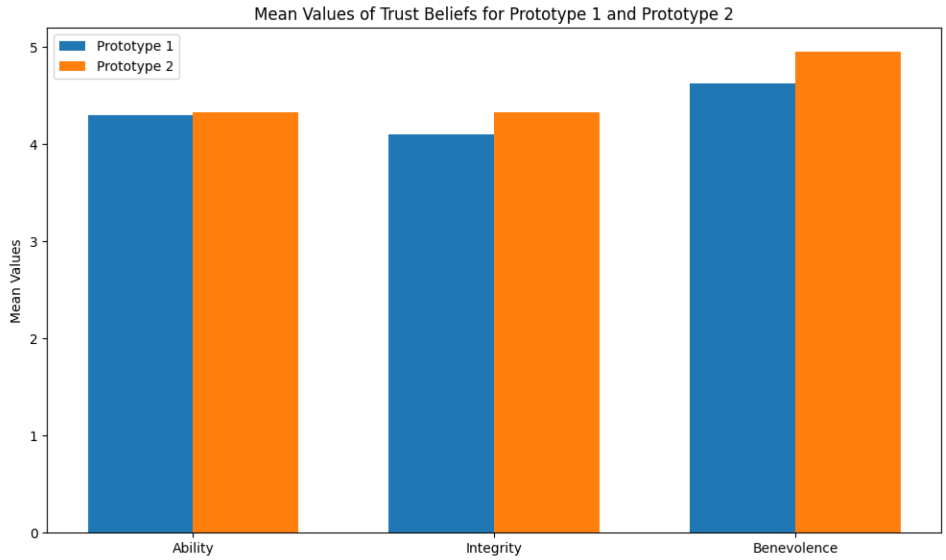

Figure 10: Comparison of the mean ratings in *Task 1* and *Task 2* for each of the three trust aspects. (7 point-likert scale, where 1 no trust, and 7 highest trust.)

for pilots with minimal training. Despite efforts to automate tasks and provide useful information, tracking all drones remains challenging for pilots, necessitating integrated UI components for better comprehension.

Enabling on-demand interaction for users. While UAV pilots could monitor autonomous drone swarms, receive sufficient information about each drone, and watch the drone swarms automatically act in emergencies, it's crucial to consider incorporating more one-to-one interaction during critical situations. However, striking a balance between mission functionality, drone autonomy, and human control poses a complex design challenge, emphasizing the ongoing need for effective solutions that prioritize pilot supervision while leveraging drone autonomy.

7 CONCLUSION

This study investigates critical human factors affecting experienced UAV professionals using multi-drone systems. Two prototypes were developed and tested in a usability test simulating a challenging SAR mission, revealing the importance of manual control and decision-making by UAV pilots. While participants managed tasks, there's a need to optimize the user interface to mitigate information overload. Heatmaps show potential in SAR missions to increase levels of trust, however, it requires thoughtful design and future work. Most requested features include thermal cameras and centralized warnings. Challenges remain in reducing cognitive workload and improving situational awareness. Insights from this study can guide ongoing research on user interfaces for real-world SAR missions with drone swarms and research involving the design of trustworthy autonomous systems.

ACKNOWLEDGMENTS

We thank the participants who are professionals in Sweden for taking time out of their schedules and for providing extensive feedback including the Police, Sea Rescue Society, Maritime Administration, Missing People, and the emergency services. This work is supported in part by the Innovation Fund Denmark for the project DIREC (9142-00001B).

REFERENCES

[1] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. 1999. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1*. Springer, 304–307.

[2] Ankit Agrawal, Sophia J Abraham, Benjamin Burger, Chichi Christine, Luke Fraser, John M Hoeksema, Sarah Hwang, Elizabeth Travnik, Shreya Kumar, Walter Scheirer, et al. 2020. The next generation of human-drone partnerships: Co-designing an emergency response system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[3] Jonathan Cacace, Alberto Finzi, Vincenzo Lippiello, Michele Furci, Nicola Mimmo, and Lorenzo Marconi. 2016. A control architecture for multiple drones operated via multimodal interaction in search & rescue mission. In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 233–239.

[4] Jessica R Cauchard, Jane L E, Kevin Y Zhai, and James A Landay. 2015. Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 361–365.

[5] Anders Lyhne Christensen, Kasper Andreas Rømer Grøntved, Maria-Theresa Oanh Hoang, Niels van Berkel, Mikael Skov, Alea Scovill, Gareth Edwards, Kenneth Richard Geipel, Lars Dalgaard, Ulrik Pagh Schultz Lundquist, Ioanna Constantiou, Christiane Lehrer, and Timothy Merritt. 2022. The HERD project: Human-multi-robot interaction in search & rescue and in farming. In *Adjunct Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1–4.

[6] Krystal Dacey, Rachel Whitsed, and Prue Gonzalez. 2022. Using an agent-based model to identify high probability search areas for search and rescue. *TheAustralian Journal of Emergency Management* 37, 4 (2022), 88–94.

[7] Yuriy Dachev and Valentin Zlatev. 2018. Search for an Object in Distress by Improved Scheme of the IAMSAR Expanding Square Methodology. *Journal Environment and Development, WSEAS* 14 (2018), 446.

[8] Mica Endsley. 1988. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting*, Vol. 32. Sage Publications Sage CA: Los Angeles, CA, 97–101.
[9] Zuguang Gu. 2022. Complex heatmap visualization. *Imeta* 1, 3 (2022), e43.
[10] Maria-Theresa Oanh Hoang, Kasper Andreas Rømer Grøntved, Niels van Berkel, Mikael B Skov, Anders Lyhne Christensen, and Timothy Merritt. 2023. Drone Swarms to Support Search and Rescue Operations: Opportunities and Challenges. *Cultural Robotics: Social Robots and Their Emergent Cultural Ecologies* (2023), 163–176.
[11] Maria-Theresa Oanh Hoang, Niels van Berkel, Mikael B Skov, and Timothy Merritt. 2022. Challenges Arising in a Multi-Drone System for Search and Rescue. In *Adjunct Proceedings of the 12th Nordic Conference on Human-Computer Interaction*. 1–5.
[12] Maria-Theresa Oanh Hoang, Niels van Berkel, Mikael B. Skov, and Timothy R. Merritt. 2023. Challenges and Requirements in Multi-Drone Interfaces. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 65, 9 pages. https://doi.org/10.1145/3544549.3585673
[13] Amy Hocraffer and Chang S Nam. 2017. A meta-analysis of human-system interfaces in unmanned aerial vehicle (UAV) swarm management. *Applied ergonomics* 58 (2017), 66–80.
[14] Jong-yi Hong, Eui-ho Suh, and Sung-Jin Kim. 2009. Context-aware systems: A literature review and classification. *Expert Systems with applications* 36, 4 (2009), 8509–8522.
[15] Mason Itkin, Mihui Kim, and Younghee Park. 2016. Development of cloud-based UAV monitoring and management system. *Sensors* 16, 11 (2016), 1913.
[16] Theodore Jensen, Yusuf Albayram, Mohammad Maifi Hasan Khan, Ross Buck, Emil Coman, and Md Abdullah Al Fahim. 2018. Initial trustworthiness perceptions of a drone system based on performance and process information. In *Proceedings of the 6th International Conference on Human-agent Interaction*. 229–237.
[17] Jonathan Källbäcker and Oscar Bjurling. 2023. Human-Swarm Interaction in Semi-voluntary Search and Rescue Operations: Opportunities and Challenges. In *Proceedings of the European Conference on Cognitive Ergonomics 2023*. 1–7.
[18] Martin Lochner, Andreas Duenser, and Shouvojit Sarker. 2019. Trust and cognitive load in semi-automated UAV operation. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. 437–441.
[19] Jonas Löwgren. 2013. Annotated portfolios and other forms of intermediate-level knowledge. *interactions* 20, 1 (2013), 30–34.
[20] Mingyang Lyu, Yibo Zhao, Chao Huang, and Hailong Huang. 2023. Unmanned Aerial Vehicles for Search and Rescue: A Survey. *Remote Sensing* 15, 13 (2023), 3266.
[21] Balita Heriniaina Rakotonarivo, Nicolas Drougard, Stéphane Conversy, and Jérémie Garcia. 2023. Cleared for Safe Take-off? Improving the Usability of Mission Preparation to Mitigate the Safety Risks of Drone Operations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
[22] Elena Sava, Charles Twardy, Robert Koester, and Mukul Sonwalkar. 2016. Evaluating lost person behavior models. *Transactions in GIS* 20, 1 (2016), 38–53.
[23] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.
[24] Anam Tahir, Jari Böling, Mohammad-Hashem Haghbayan, Hannu T Toivonen, and Juha Plosila. 2019. Swarms of unmanned aerial vehicles—A survey. *Journal of Industrial Information Integration* 16 (2019), 100106.
[25] Christopher Van Tilburg. 2017. First report of using portable unmanned aircraft systems (drones) for search and rescue. *Wilderness & environmental medicine* 28, 2 (2017), 116–118.
[26] Bruce H Wardlow and Operational Test Command. 2006. Evolving a Model for Probability of Detection within a Search Box for a Search and Rescue Mission Parallel Sweep Search. *The International Journal of Applied Management and Technology* 4, 2 (2006), 76–89.
[27] R Williams. 2020. Collaborative multi-robot multi-human teams in search and rescue.. In *Proceedings of the International ISCRAM Conference*, Vol. 17.
[28] Michał Wysokiński, Robert Marcjan, and Jacek Dajda. 2014. Decision support software for search & rescue operations. *Procedia Computer Science* 35 (2014), 776–785.