

Agentic AI for Safety-critical Multi-drone Systems: Challenges and Opportunities

Timothy Merritt^{1*}, Alejandro Jarabo-Peñas², Juan Bravo-Arrabal², Maria-Theresa Bahodi¹ and Anders Lyhne Christensen²

¹Human-centered Computing (Aalborg University), 300 Selma Lagerlöfs Vej, Aalborg Øst, 9220, Denmark

²The Maersk Mc-Kinney Møller Institute (University of Southern Denmark), Campusvej 55, 5230, Odense, Denmark

Abstract

Multi-drone systems are increasingly positioned for safety-critical missions such as search and rescue (SAR) and critical infrastructure monitoring. Yet, real-world adoption remains constrained not only by autonomy performance, but by the difficulty of integrating agentic behavior into professional work: operators must understand, trust, and govern automation under uncertainty, time pressure, and accountability. This position paper synthesizes the ambitions and lessons from two ongoing efforts: NAMUR, which explores LLM-supported robot control in SAR and firefighting contexts, and PERSIST, which explores persistent drone operations for monitoring and security at critical infrastructure sites. We argue that agentic AI should be approached as a socio-technical design problem, where interfaces, oversight mechanisms, and evaluation practices are as critical as algorithms. We outline a human-centered, participatory, and iterative research approach aimed at uncovering stakeholder needs, shaping agent capabilities through successive prototypes, and producing transferable proof-of-concept systems and evaluation strategies for other safety-critical contexts.

Keywords

Agentic AI, Multi-drone systems, Human-centered AI, Participatory design, Safety-critical work, Critical infrastructure, Search and rescue

1. Introduction

Multi-robot and multi-drone systems promise new capabilities for safety-critical work, including emergency response [1] and security and inspection at critical infrastructure [2]. At the same time, these domains impose strict demands [3, 4, 5]: operations are uncertain, accountability is high, and workflows are governed by professional roles, protocols, and risk management. In such settings, it is rarely sufficient for autonomy to “work” in a technical sense; it must also be understandable, reliable, governable, and adoptable.

Controlling multiple robots as a *single operational system* amplifies these demands [6, 7, 8]. As fleet size grows, the traditional one-operator/one-vehicle model breaks down [9]: coordination overhead increases, maintaining shared situation awareness becomes harder, and small uncertainties can cascade into safety issues (e.g., conflicting task priorities, deconfliction problems, or ambiguous responsibility for who approved what). Operators therefore need to interact at multiple levels of granularity [10]. That might entail setting mission-level intent and constraints for the group, while still being able to inspect, redirect, or take control of a specific drone or subset when conditions change.

Recent progress in agentic AI and large language models (LLMs) has renewed interest in natural-language tasking, mixed-initiative planning, and autonomous execution in robotics [11]. However, these capabilities introduce new interaction challenges [12]: how should agent intentions be represented; how do operators supervise, approve, and override decisions; how can correct behavior be guaranteed in the messy realities of critical field operations?

Joint Proceedings of the ACM Intelligent User Interfaces (IUI) Workshops 2026, March 23-26, 2026, Paphos, Cyprus

*Corresponding author.

✉ merritt@cs.aau.dk (T. Merritt); alejp@mami.sdu.dk (A. Jarabo-Peñas); juanba@mami.sdu.dk (J. Bravo-Arrabal); mtoh@cs.aau.dk (M. Bahodi); andc@mami.sdu.dk (A. L. Christensen)

🌐 <https://www.ixd.net> (T. Merritt); <https://anderslyhnechristensen.com/> (A. L. Christensen)

🆔 0000-0002-7851-7339 (T. Merritt); 0000-0002-3312-7062 (A. Jarabo-Peñas); 0000-0002-4046-1528 (J. Bravo-Arrabal); 0000-0003-4512-400X (M. Bahodi); 0000-0002-9994-2908 (A. L. Christensen)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

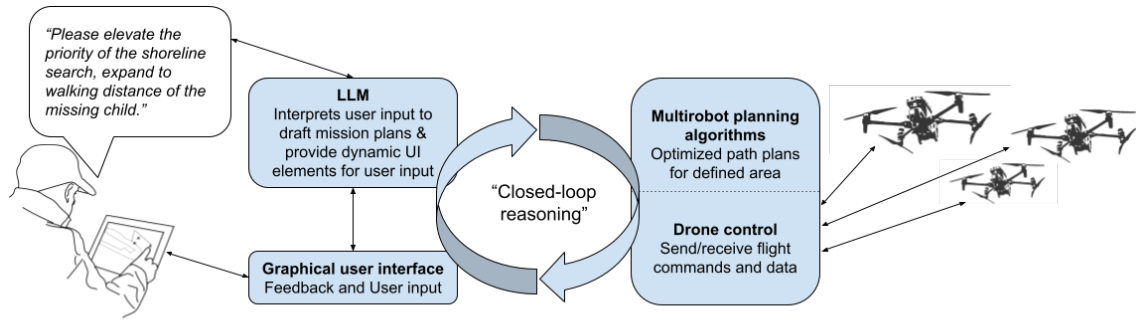


Figure 1: NAMUR concept illustration: an LLM-driven natural language interaction system with “closed-loop reasoning” to support human tasking and robot control in emergency response.

This paper is organized as follows. Section 2 introduces the two motivating project contexts—NAMUR (SAR and firefighting) and PERSIST (persistent critical-infrastructure monitoring)—and distills shared interaction requirements for multi-drone control. Section 3 synthesizes lessons from our prior field engagements to frame the key opportunities and safety-critical tensions that motivate *governable* agentic autonomy, and outlines our participatory, iterative approach for shaping such systems with stakeholders. Building on these requirements, Section 4 presents an LLM-based multi-agent system (LLM-MAS) that decomposes natural-language intent into reviewable sub-tasks, constrains execution through deterministic tools, and enforces operator preview and approval. Finally, Section 5 concludes with implications for future agentic multi-drone systems in safety-critical work.

2. Case Contexts and Project Objectives

Our position is shaped by two complementary projects that rely on agentic AI in different safety-critical conditions: (i) time-critical emergency response (NAMUR) and (ii) long-horizon, persistent operations at critical infrastructure (PERSIST). We summarize each context in terms of operational setting, stakeholders, objectives, and the agentic-AI questions that emerge.

2.1. NAMUR: Agentic support for SAR and firefighting

NAMUR investigates how LLM-supported, agentic interaction can help teams coordinate a multi-drone system during emergency response work such as search and rescue and firefighting support (see Figure 1). The operational setting is characterized by time pressure, dynamic hazards, incomplete and rapidly changing situational information, and a strong need for coordination across roles [1, 13]. Deploying multi-robot systems in the field also depends on robust communication and edge-cloud infrastructure [14], and requires interfaces that scale across heterogeneous robot platforms. Because operators with different roles and expertise tend to interact differently with each robot type, agentic AI can serve as a unifying abstraction layer—providing a more coherent, comprehensible relationship between users and diverse physical systems. In these contexts, technology is only useful when it integrates with established command structures and communication practices, and when responsibility for high-stakes decisions remains clear.

Stakeholders and work setting. The primary stakeholders include incident commanders and coordinators, field responders operating in hazardous environments, and technology operators responsible for robotic assets. The work is organized through role-specialized decision-making and structured communication, where updates and tasking must be concise, timely, and auditable [15, 16]. This setting foregrounds interaction demands that go beyond natural language control: operators must be able to supervise, confirm, and constrain what an agentic system does under uncertainty.

Project objectives. NAMUR’s objectives are to: (1) enable higher-level tasking of robotic assets (e.g., translating intent into feasible robot actions); (2) support mixed-initiative interaction where the system proposes actions, but humans authorize execution; (3) make agentic behavior interpretable enough for operators to assess risk, timing, and consequences; and (4) study these interactions in realistic, mission-oriented exercises to surface failure modes, governance needs, and usability constraints that would be invisible in purely lab-based evaluations.

Why agentic AI is compelling here. Agentic AI can help transform fragmented inputs (radio updates, map cues, observations) into actionable suggestions and structured plans [17, 18], and can reduce coordination overhead by maintaining continuity across rapid task switches. At the same time, emergency response makes the limits of agentic AI especially visible: uncertainty is unavoidable, and behavior designed for convenience can become harmful without explicit authorization workflows [19, 20], conservative defaults, and clear escalation boundaries.

2.2. PERSIST: Persistent multi-drone operations for critical infrastructure

PERSIST investigates persistent, multi-drone operations for monitoring, inspection, and security at critical infrastructure sites (see Figure 2). In contrast to emergency response, the dominant challenge here is not a single high-tempo mission but sustained operations across days and shifts, where drones must repeatedly execute routine tasks, respond to anomalies, and integrate into existing organizational workflows for security and maintenance. Recent work on large language model (LLM) agents in industrial automation highlights the broader relevance of agentic approaches for orchestrating complex operational systems beyond the lab [21].

Stakeholders and work setting. Key stakeholders include site security personnel, operations and maintenance staff, and organizational decision-makers responsible for safety, compliance, and continuity of service. The setting includes constrained physical environments (restricted zones, sensitive assets), operational rhythms (shift handovers, scheduled inspections), and an expectation that systems behave predictably and recover gracefully from routine disruptions (weather, connectivity, access changes, false alarms).

Project objectives. PERSIST’s objectives are to: (1) move from ad hoc drone flights to persistent, repeatable operations that can be delegated and scheduled; (2) reduce the barrier to entry for professional drone use through interface support for planning, monitoring, and handover; (3) enable scalable supervision where one operator can manage multiple vehicles and mission types; and (4) prototype and validate the approach in realistic infrastructure settings, producing proofs of concept with transfer potential to other critical sites.

Why agentic AI is compelling here. Agentic AI can support long-horizon orchestration: selecting and parameterizing routine missions, adapting schedules as conditions change, and assisting with anomaly triage and reporting. The risk profile differs from NAMUR: failures may be less acute in the moment, but persistent operations amplify the importance of reliability, drift management, operator fatigue, and organizational trust. Here, the central interaction challenge is sustained governability: users must be able to understand what the system has been doing over time, why exceptions occurred, and what remains unresolved at shift handover.

2.3. Cross-cutting themes and contrasts

Together, NAMUR and PERSIST highlight complementary stressors for agentic multi-drone systems.

Tempo and horizon. NAMUR emphasizes rapid decision cycles and short operational windows; PERSIST emphasizes persistent activity, repeatability, and organizational integration across shifts.

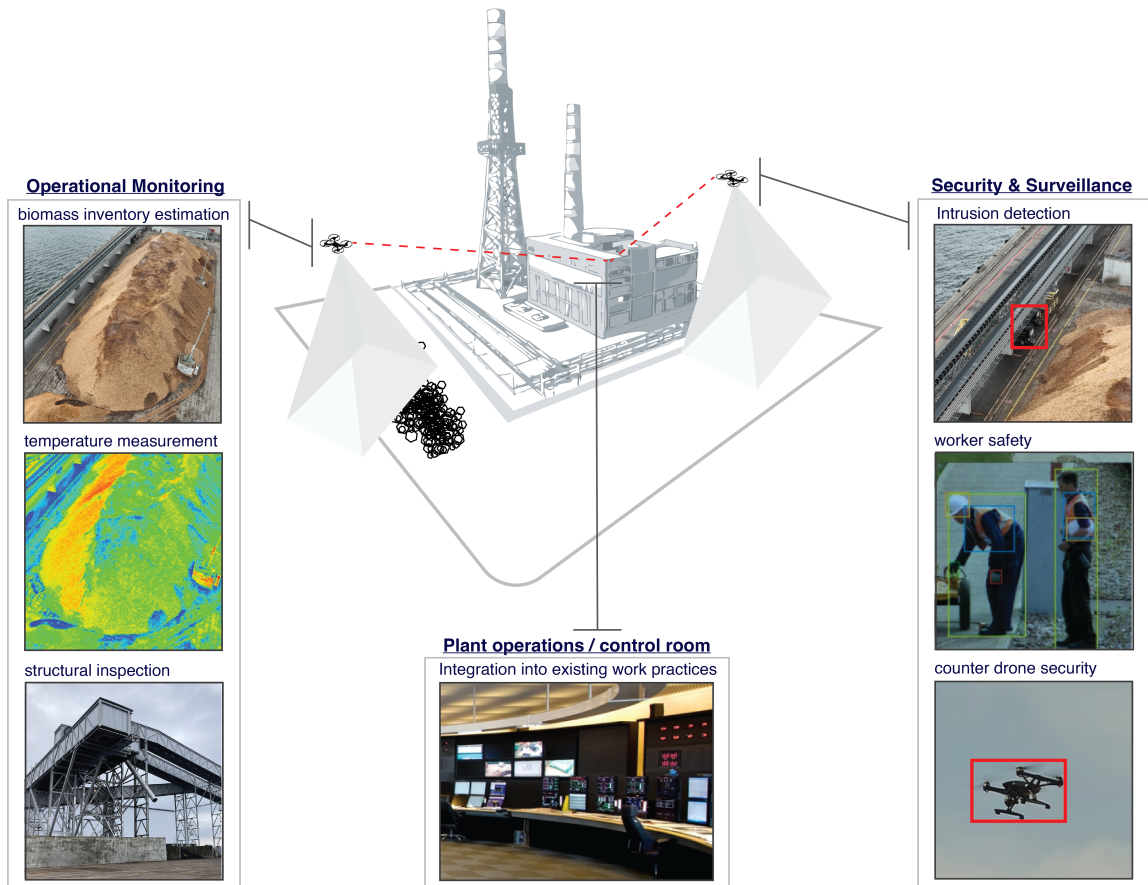


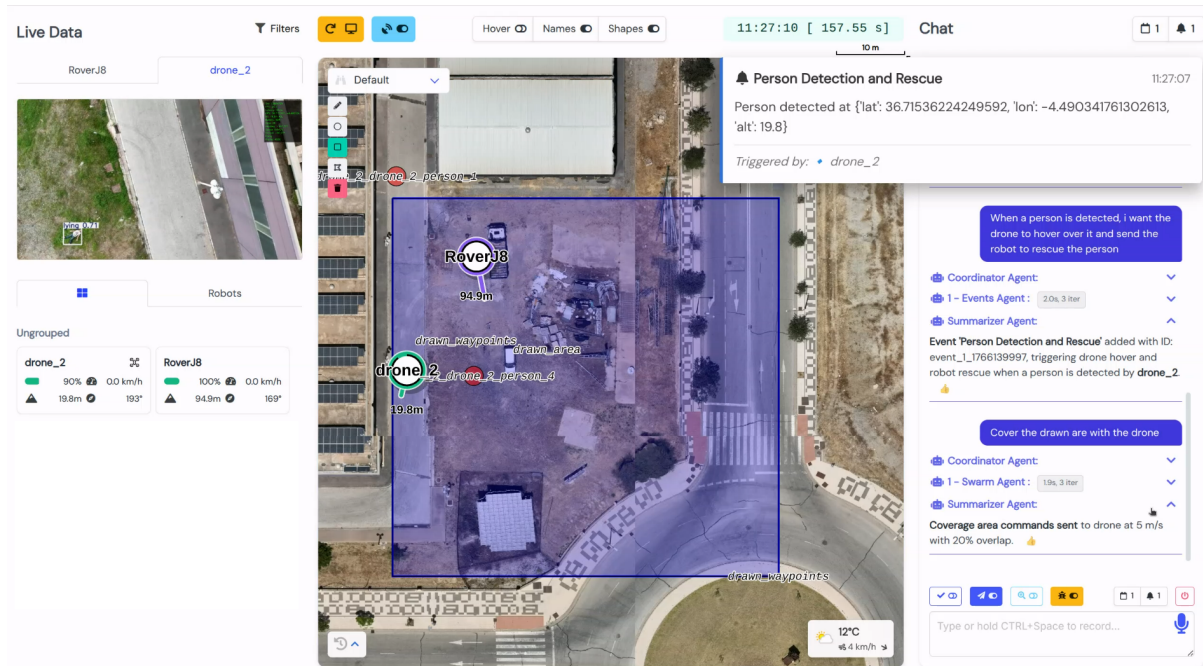
Figure 2: PERSIST overview illustration (from proposal material): examples of task types motivating persistent multi-drone operations (e.g., routine monitoring, inspection, and security workflows).

Governance and accountability. Both contexts demand clear authority and auditability, but in different forms: emergency response requires immediate confirmation and escalation pathways, while infrastructure operations require policy-aligned routines, handover summaries, and post hoc accountability.

Implication for our research approach. Across both cases, we treat agentic AI as a socio-technical design material: we study how people coordinate, decide, and verify; we prototype agentic behaviors that remain governable; and we iteratively shape the system with stakeholders rather than presuming a final autonomy solution.

3. Agentic AI for Multi-Drone Operations in Safety-Critical Work: Lessons, Tensions, and an Agenda

Our perspective on agentic AI is grounded in prior human-centered studies of multi-robot systems in safety-critical environments, including co-design and prototype evaluations with Danish and Spanish emergency services in search-and-rescue (SAR) and firefighting contexts [1, 22], as well as co-design work with security personnel at a power plant. Across these contexts, practitioners want autonomy that is *self-sustaining* enough to be operationally viable, yet *transparent*, easily learnable, and aligned with local protocols and expertise. This motivates a cautiously optimistic stance: agentic AI can reduce the coordination burden and accelerate sensemaking, but only if paired with interaction mechanisms that keep humans able to authorize, supervise, and diagnose actions under uncertainty. In the ideal scenario, a single operator can supervise a mission involving a fleet of drones through mixed-initiative control: high autonomy for routine progress and rapid intervention at multiple levels of granularity



when risk or uncertainty increases.

3.1. Opportunities that motivate agent support

Agentic AI is compelling in these domains not because it can replace professional judgment, but because it can help teams manage *scale* (multiple vehicles, multiple data streams, multiple stakeholders) without collapsing situation awareness [23]. Across our cases, the most salient opportunities are: (i) selective attention and summarization that reduces continuous manual scanning, (ii) mixed-initiative planning and replanning that maintain coherent coverage as information changes, and (iii) long-horizon orchestration for persistent operations, including scheduling, anomaly triage, and reporting aligned to staffing constraints.

3.2. Safety-critical tensions that require governability

The same capabilities introduce tensions that cannot be addressed by autonomy alone. First, practitioners need systems that can operate without constant attention, yet resist black-box behavior that cannot be explained, constrained, or overridden. Second, strong cueing and alerting can improve short-term performance but encourage attention tunneling, reducing holistic overview during supervision [24]. Third, trust must be calibrated under imperfect perception and dynamic conditions [25]; conservative defaults, uncertainty cues, and verification workflows are essential. Finally, adoption depends on operational fit: alignment with roles, protocols, and accountability practices, including auditable authorization and decision trails.

3.3. Research approach: participatory shaping of agentic autonomy

Rather than aiming for a single, final autonomy concept, we treat agentic AI as a socio-technical design material that must be iteratively shaped through stakeholder engagement with functional prototypes [26].

We begin with needs discovery grounded in concrete scenarios and demonstrations to map work practices, decision points, and cognitive bottlenecks, then use successive prototypes (see [Figure 3](#) for an

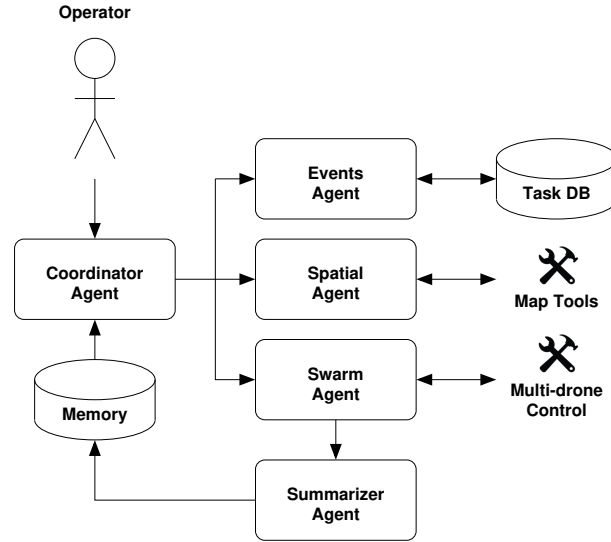


Figure 4: Architecture of the LLM-MAS use in the functional prototype.

example of our functional multi-drone control interface prototype deployed in a real-world context [22] & companion video of live control of UGVs with LLM-enabled voice commands.¹⁾ as boundary objects to negotiate what the agent should do, what must require explicit authorization, and what evidence and uncertainty cues the system must surface.

Across iterations, this process yields actionable interaction requirements for *governability*: configurable constraints that domain experts can tune quickly, clear intervention points for high-stakes actions, and observability features that let both operators and developers diagnose behavior and learn from incidents.

Bridge to architecture. Together, these cases and lessons motivate an architecture that decomposes intent into reviewable sub-tasks, constrains action through deterministic tools, supports scheduling and long-horizon memory, and enforces explicit operator preview and approval before execution.

4. An LLM-MAS Architecture for Multi-drone Control

The functional prototype implements a unified agentic AI architecture that meets the requirements of the NAMUR and PERSIST projects and has been demonstrated in live control of UGVs and UAVs [22]. The architecture is implemented as a LLM-MAS composed of five specialized agents: *Coordinator*, *Event*, *Spatial*, *Swarm*, and *Summarizer*. Several agents, including the Spatial Agent and the Swarm Agent, interface with external software tools that provide deterministic capabilities such as spatial computation, multi-drone motion execution, memory storage, and task scheduling. Collectively, these components enable the decomposition of high-level natural language commands into well-defined sub-tasks [27, 28] that can be previewed, validated, and explicitly approved by an operator prior to execution (Figure 4).

The specific set of tools available to the LLM-MAS is deployment-dependent. For SAR scenarios, dedicated tools support functionalities such as compliant coverage path planning. In contrast, for persistent drone operations at biomass power plants, specialized tools are provided for generating flight paths and image acquisition plans that enable accurate biomass estimation.

In the following, we describe the responsibilities of each agent using a representative operator command as an example.

- **Coordinator Agent:** Serves as the central orchestrator of the system. It receives the operator’s command, infers the underlying intent, and decomposes it into concrete sub-queries and execution

¹Video showing live control of UGVs with LLM-enabled voice commands: <https://www.youtube.com/watch?v=Hna7AtDS5wo>

requests that are delegated to subordinate agents.

- **Events Agent:** Manages events and periodic, delayed, and future-oriented directives issued by the operator. It maintains and monitors scheduled and recurring tasks (e.g., “*Estimate biomass volume every weekday at 2 pm*” or “*Return all drones to base at 18:00*”) and ensures that they are executed or surfaced at the appropriate time and context.
- **Spatial Agent:** Provides spatial reasoning and situational awareness by accessing a GeoJSON-formatted map alongside the live positions of all active drones and tracked first responders. It identifies relevant semantic map features (e.g., the “*eastern tree line*” or the “*biomass pile next to the water*”) and accesses tools to compute spatial relationships between entities as required.
- **Swarm Agent:** Handles motion planning and task execution requests for one or more available drones. It selects suitable drone platforms based on availability and operational context and issues coordinated control commands when multiple drones are involved [29]. Given the real-world consequences of these actions, all requests generated by this agent are subject to operator preview and confirmation. Upon receiving target coordinates from the Spatial Agent, the Swarm Agent invokes an execution tool that publishes the desired actions together with the selected drone identifier(s) to a ROS topic, thereby triggering the downstream planning and flight control pipeline.
- **Summarizer Agent:** Generates a concise, human-readable summary of the actions executed and the resulting system state after each command. In addition to being presented to the operator as confirmation, these summaries are persistently stored in the system’s interaction memory. The stored summaries provide a compact, structured record of prior interactions and outcomes, which is made available as contextual input to the Coordinator Agent in subsequent dialogue turns. This persistent context enables coherent multi-turn interactions and reference resolution across commands, such as interpreting follow-up instructions that rely on previously mentioned entities or actions (e.g., “*Move drone 5 to the main building*” followed by “*Now, move it to the nearest safe landing zone*”).

By explicitly separating responsibilities across specialized agents and constraining their behavior through structured prompts and tool interfaces, the proposed LLM-MAS enables transparent, interpretable, and safe translation of natural language instructions into coordinated multi-drone mission execution. Furthermore, the integration of persistent interaction history and explicit task management supports robust multi-turn dialogue and long-horizon autonomy under continuous operator supervision.

5. Conclusion

Agentic AI holds promise for scaling multi-drone operations in SAR and critical infrastructure monitoring, but introduces challenges around oversight, trust calibration, operational fit, and evaluation. Grounded in NAMUR and PERSIST, we framed these requirements and presented an LLM-MAS architecture that operationalizes governable agentic autonomy through constrained tool use and explicit operator preview/approval. This framing supports future research in autonomy and provides interactive infrastructure required to deploy agentic systems responsibly.

Looking ahead, we see three implications for future agentic multi-drone systems in safety-critical work. First, **governability should be treated as a core system property**, not a UI add-on: architectures should make authorization points, constraints, and intervention mechanisms explicit and auditable. Second, **agentic behavior should be built from constrained, inspectable tool use** rather than unconstrained free-form action, enabling predictable failure modes, reproducible debugging, and clearer responsibility boundaries. Third, **evaluation should reflect real operational use**. This means measuring more than task success: we should assess situation awareness (and whether alerts cause attention tunneling), trust calibration under uncertainty, coordination overhead, and whether the system supports accountable after-action review.

Acknowledgments

This work was supported by the Innovation Fund Denmark for DIREC project U07 (the PERSIST project), and the Independent Research Fund Denmark under grant 10.46540/4264-00105B (the NAMUR project).

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly and GPT 5.2 in order to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] M.-T. O. Hoang, K. A. R. Grøntved, N. van Berkel, M. B. Skov, A. L. Christensen, T. Merritt, Drone Swarms to Support Search and Rescue Operations: Opportunities and Challenges, Springer International Publishing, Cham, 2023, pp. 163–176. doi:10.1007/978-3-031-28138-9_11.
- [2] R. H. Jacobsen, L. Matlekovic, L. Shi, N. Malle, N. Ayoub, K. Hageman, S. Hansen, F. F. Nyboe, E. Ebeid, Design of an autonomous cooperative drone swarm for inspections of safety critical infrastructure, *Applied Sciences* 13 (2023) 1256.
- [3] R. R. Murphy, S. Tadokoro, D. Nardi, A. Jacoff, P. Fiorini, H. Choset, A. M. Erkmén, Search and Rescue Robotics, in: B. Siciliano, O. Khatib (Eds.), *Springer Handbook of Robotics*, Springer, 2008, pp. 1151–1173. doi:10.1007/978-3-540-30301-5_51.
- [4] R. Ventura, P. U. Lima, Search and rescue robots: The civil protection teams of the future, in: 2012 Third International Conference on Emerging Security Technologies, IEEE, 2012, pp. 12–19.
- [5] D. S. Drew, Multi-agent systems for search and rescue applications, *Current Robotics Reports* 2 (2021) 189–200. doi:10.1007/s43154-021-00048-3.
- [6] A. L. Christensen, K. A. R. Grøntved, M.-T. O. Hoang, N. van Berkel, M. Skov, A. Scovill, G. Edwards, K. R. Geipel, L. Dalgaard, U. P. S. Lundquist, et al., 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*, 2022, pp. 1–4.
- [7] M.-T. Bahodi, N. Lau, N. van Berkel, K. A. R. Grøntved, M. B. Skov, T. Merritt, Before it falls: Supporting drone fleet management through battery visualizations, in: *IFIP Conference on Human-Computer Interaction*, Springer, 2025, pp. 347–370.
- [8] S.-J. Chung, A. A. Paranjape, P. Dames, S. Shen, V. Kumar, A survey on aerial swarm robotics, *IEEE Transactions on Robotics* 34 (2018) 837–855. doi:10.1109/TRO.2018.2857475.
- [9] L. J. Planke, Y. Lim, A. Gardi, R. Sabatini, T. Kistan, N. Ezer, A cyber-physical-human system for one-to-many UAS operations: Cognitive load analysis, *Sensors* 20 (2020) 5467.
- [10] L. H. Kim, D. S. Drew, V. Domova, S. Follmer, User-defined swarm robot control, in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–13. doi:10.1145/3313831.3376814.
- [11] J. Wang, E. Shi, H. Hu, C. Ma, Y. Liu, X. Wang, Y. Yao, X. Liu, B. Ge, S. Zhang, Large language models for robotics: Opportunities, challenges, and perspectives, *J. Autom. and Intell.* 4 (2025) 52–64.
- [12] R. Sapkota, K. I. Roumeliotis, M. Karkee, UAVs meet agentic AI: A multidomain survey of autonomous aerial intelligence and agentic UAVs, *arXiv preprint arXiv:2506.08045* (2025).
- [13] J. A. Delmerico, S. Mintchev, A. Giusti, B. Gromov, K. Melo, T. Horvat, C. Cadena, M. Hutter, A. J. Ijspeert, D. Floreano, L. M. Gambardella, R. Siegwart, D. Scaramuzza, The current state and future outlook of rescue robotics, *J. Field Robotics* 36 (2019) 1171–1191. doi:10.1002/ROB.21887.
- [14] J. Bravo-Arrabal, M. Toscano-Moreno, J. J. F. Lozano, A. Mandow, J. A. Gómez-Ruiz, A. García-Cerezo, The Internet of Cooperative Agents Architecture (X-IoCA) for Robots, *Hybrid Sensor*

- Networks, and MEC Centers in Complex Environments: A Search and Rescue Case Study, *Sensors* 21 (2021) 7843. doi:10.3390/S21237843.
- [15] J. Jensen, S. Thompson, The incident command system: a literature review, *Disasters* 40 (2016) 158–182.
 - [16] J. Wolbers, K. Boersma, The common operational picture as collective sensemaking, *Journal of Contingencies and Crisis management* 21 (2013) 186–199.
 - [17] V. G. Goecks, N. R. Waytowich, DisasterResponseGPT: Large Language Models for Accelerated Plan of Action Development in Disaster Response Scenarios, in: *Workshop on Challenges in Deployable Generative AI at International Conference on Machine Learning (ICML)*, 2023. doi:10.48550/arXiv.2306.17271. arXiv:2306.17271.
 - [18] A. R. Sadik, M. Ashfaq, N. Mäkitalo, T. Mikkonen, Human-LLM Synergy in Context-Aware Adaptive Architecture for Scalable Drone Swarm Operation, *CoRR* abs/2509.05355 (2025). doi:10.48550/arXiv.2509.05355. arXiv:2509.05355.
 - [19] A. Robey, Z. Ravichandran, V. Kumar, H. Hassani, G. J. Pappas, Jailbreaking LLM-controlled Robots, *CoRR* abs/2410.13691 (2024). doi:10.48550/ARXIV.2410.13691. arXiv:2410.13691.
 - [20] J. Cleland-Huang, P. A. A. Granadeno, A. M. R. Bernal, D. Hernandez, M. Murphy, M. Petterson, W. Scheirer, Cognitive Guardrails for Open-World Decision Making in Autonomous Drone Swarms, *CoRR* abs/2505.23576 (2025). doi:10.48550/arXiv.2505.23576. arXiv:2505.23576.
 - [21] Y. Xia, M. Shenoy, N. Jazdi, M. Weyrich, Towards autonomous system: flexible modular production system enhanced with large language model agents, in: *2023 IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, 2023. doi:10.1109/ETFA54631.2023.10275362.
 - [22] A. Jarabo-Peñas, J. Bravo-Arrabal, D.-h. Lin-Yang, F. Pastor, R. Ladig, J. J. Fernández-Lozano, A. Christensen, A. García-Cerezo, Real-World Deployment of an LLM-Enabled Voice-Commanded UGV for Logistics in SAR Missions, in: *2025 IEEE International Symposium on Safety Security Rescue Robotics (SSRR)*, IEEE, 2025.
 - [23] M. R. Endsley, Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors* 37 (1995) 32–64. URL: <https://doi.org/10.1518/001872095779049543>. doi:10.1518/001872095779049543.
 - [24] M.-T. Bahodi, N. van Berkel, M. Skov, T. Merritt, Show Me What’s Wrong: Impact of Explicit Alerts on Novice Supervisors of a Multi-Robot Monitoring System, in: *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems, TAS ’24*, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1–17. URL: <https://dl.acm.org/doi/10.1145/3686038.3686069>. doi:10.1145/3686038.3686069.
 - [25] J. Ahlskog, M.-T. Bahodi, A. Lugmayr, T. Merritt, Fostering trust through user interface design in multi-drone search and rescue, in: *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems, TAS ’24*, Association for Computing Machinery, New York, NY, USA, 2024. URL: <https://doi.org/10.1145/3686038.3686052>. doi:10.1145/3686038.3686052.
 - [26] S. Bødker, M. Kyng, Participatory design that matters—facing the big issues, *ACM Trans. Comput.-Hum. Interact.* 25 (2018). URL: <https://doi.org/10.1145/3152421>. doi:10.1145/3152421.
 - [27] I. Singh, V. Blukis, A. Mousavian, A. Goyal, D. Xu, J. Tremblay, D. Fox, J. Thomason, A. Garg, Progprompt: Generating situated robot task plans using large language models, in: *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 11523–11530. doi:10.1109/ICRA48891.2023.10161317.
 - [28] Y. Tsushima, S. Yamamoto, A. A. Ravankar, J. V. S. Luces, Y. Hirata, Task Planning for a Factory Robot Using Large Language Model, *IEEE Robotics Autom. Lett.* 10 (2025) 2383–2390. doi:10.1109/LRA.2025.3531153.
 - [29] K. A. Grøntved, U. P. Lundquist, A. L. Christensen, Decentralized multi-UAV trajectory task allocation in search and rescue applications, in: *2023 21st International Conference on Advanced Robotics (ICAR)*, IEEE, 2023, pp. 35–41.