Human Factors in Robots, Drones and Unmanned Systems, Vol. 218, 2026, 106–113
https://doi.org/10.54941/ahfe1007683

Human-Friendly Control of Drones and Drone Swarms Using Natural Language and AI-Based Task Decomposition

Geert De Cubber and Daniela Doroftei

Royal Military Academy, 1000 Brussels, Belgium

ABSTRACT

The deployment of drones and swarms of small drones for operational purposes is rapidly increasing the load on human operators especially if interaction relies on low-level control interfaces. This paper reports work carried out as part of a project on human-centric drone control by natural language interaction and AI-augmented task understanding. With the proposed method, operators are able to give high-level commands to single or small groups of drones and task-representations are structured from these commands and executed with predefined mission primitives. The focus of the system is to promote transparency and operator supervision through explicit feedback about the interpretation of tasks and task execution.

Keywords: Human–drone interaction, Natural language interfaces, Drone swarms, Human-centred autonomy

INTRODUCTION

The use of unmanned aerial systems (UAS), also known as drones, is becoming more prevalent in reconnaissance, inspection, surveillance, and protection roles, both in civilian and military contexts (AL-Dosari, 2023). This increasing requirement for UAS operations is creating a requirement for human operators to monitor not only a single unmanned aerial system, but potentially several unmanned aerial systems simultaneously, either in parallel or in coordinated formations (Lopez, 2017). This is creating significant cognitive demands (Schmidt, 2022), particularly in relation to low-level control interfaces, such as waypoint specification, numerous parameters in a graphical interface, or joystick-based control.

One of the important challenges, therefore, is to increase the level of abstraction at which human interaction with autonomous aerial systems takes place. Rather than controlling the trajectories of platforms manually, it should be possible to communicate high-level mission intent and monitor its execution (Casado Fauli, 2024). Natural language interaction is one area that holds promise in this regard, as it enables operators to articulate tasks in familiar operational language, thus minimizing the need for constant human input. Nevertheless, its application in safety-critical and mission-critical domains raises issues of ambiguity, expectation management, and transparency (Javaid, 2024).

Received February 5, 2026; Revised April 3, 2026; Accepted April 22, 2026; Available online July 20, 2026

In this paper, a methodology for human-friendly control of drones and small swarms using natural language tasking and structured intent representation and task decomposition is proposed. The paper is conceptual, and we outline the interaction model, system architecture, and design principles that underlie the process of translating operator intent into autonomous behaviour.

BACKGROUND AND HUMAN FACTORS CONSIDERATIONS

The human-drone interface has undergone significant changes in the past decade, with the primary aim being to make drones more accessible, such that they can be effective in operational real-world scenarios (De Cubber, 2017). Initially, most of the research efforts were concentrated on enhancing stability and low-level control support for drones, enabling non-experts to control drones more comfortably (Alexis, 2016). Nevertheless, even with enhanced flight controllers, traditional control interfaces still require manual input and are still demanding in terms of training and concentration.

With the increasing complexity of drone operations and the involvement of more drones in operations, the focus of research has shifted to higher-level modalities. Gesture-based interaction (Obaid, 2016), multimodal interfaces using vision and sound (Fernandez, 2016), and augmented reality interfaces (Mourtzis, 2024) have shown promise in improving usability. However, these interfaces often depend on the explicit specification of commands and require controlled environments for use.

Natural language interfaces have also been explored as a more natural interface modality, especially for ground robots and simple aerial operations (Nwankwo, 2024). Some existing systems have already demonstrated the capability to translate natural language commands into specified actions or parameterized operations (Liu, 2022). While these systems reduce the complexity of the interaction process, they often require the use of a fixed grammar and/or a limited context of operation, thereby limiting the flexibility of the system. On the other hand, open-ended systems leave the door open to undocumented or erratic responses from the UAS, which can lead to incidents (Doroftei, 2020). Moreover, the extension of natural language interfaces to the control of multiple drones or swarm robotics has not yet been adequately explored (Heidari, 2023), especially in the context of the increasing importance of such systems.

From the human factors' perspective, there are several critical issues that are important in the context of designing the interface for the effective operation of autonomous aerial vehicles. These include issues of cognitive workload, trust calibration, system transparency, and the operator's ability to maintain situational awareness (Doroftei, 2022). Mismatches in the interpretation of the intent of the operator and the operator's expectations can have a negative impact on the performance of the system and the pilot (Doroftei, 2024), especially in time-critical and/or safety-critical operations. Moreover, the ambiguity of natural language commands complicates the interaction process and necessitates the use of constrained interpretation mechanisms, thereby providing a high level of feedback and task decomposition (Yadav, 2021).

The methodology that is proposed in this research is based on the insights that have been gathered on the importance of structured intent representations, constrained task instantiation, and providing high levels of feedback in the execution of the tasks. Unlike the existing natural language interfaces that are often considered as a free-form interface modality, the proposed methodology considers natural language interfaces as a highly ergonomic interface modality for the supervision of the operation of autonomous systems, thereby providing the human operators the flexibility to specify the intent of the operation while providing the operators the level of oversight and control that is critical in the context of semi-autonomous systems.

METHODOLOGICAL FRAMEWORK

The proposed methodology is designed to enable operators to issue concise, mission-level commands to single drones or small swarms through natural language input, while maintaining operator oversight and ensuring predictable system behaviour. Rather than treating natural language as a free-form conversational interface, the methodology structures interaction as a sequence of intent specification, constrained interpretation, task decomposition, and transparent execution feedback. This section presents the conceptual framework underlying this process.

Interaction Model

The interaction model follows a command--response paradigm tailored to operational environments where instructions must be issued rapidly and interpreted reliably. Operators formulate mission-level intent using short spoken or textual commands, such as "observe the area north of the building" or "send two drones to patrol the perimeter." The system does not assume continuous dialogue or open-ended conversation; instead, it treats each command as a discrete input to be interpreted, instantiated, and executed under human supervision.

To support ergonomic use, the model emphasises:

- Low cognitive overhead: Operators express tasks in familiar operational terminology.
- Predictability: Only task types covered by predefined primitives can be instantiated.
- Modality flexibility: Commands may be spoken or typed, depending on operational constraints.

A conceptual overview of this interaction pipeline is shown in Figure 1. Natural language input is first processed by a speech-to-text front-end (when speech is used) and then passed to a Natural Language Understanding (NLU) component (Yadav, 2021), which produces a structured intent representation.

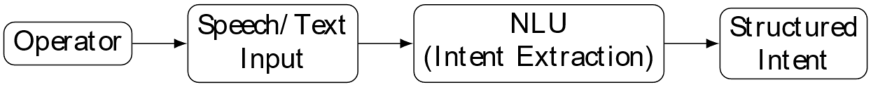

Figure 1: Conceptual interaction flow from operator input to structured intent representation.

Intent Interpretation Framework

Once a command is parsed, the system converts it into a structured intent representation. This representation forms a bridge between human expression and executable drone behaviour. It captures the task type (e.g., observation, patrol, area exploration), spatial or temporal parameters, and any explicitly stated constraints.

The interpretation process is intentionally constrained:

- Only supported task types may be instantiated, reducing ambiguity.
- Parameters are normalised into mission-relevant forms (e.g., areas, waypoints, regions).
- Implicit assumptions are avoided unless explicitly specified by the operator.

This design ensures that natural language remains an ergonomic input channel, not a source of unpredictable behaviour. Figure 2 illustrates the relationship between natural language commands, the extracted intent, and the predefined task primitives that guide subsequent decomposition.

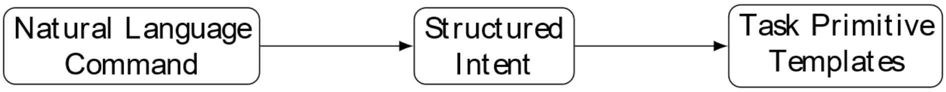

Figure 2: Intent interpretation as a mapping from linguistic expression to structured task representations anchored in predefined task primitives.

Task Decomposition Logic

Task decomposition translates the structured intent into executable plans. Unlike fully autonomous planners that reason over large action spaces, the proposed methodology uses a constrained instantiation approach based on predefined mission primitives and execution templates. This ensures predictability, supports transparency, and keeps operator oversight central to the process.

For single-drone missions, decomposition typically produces a sequence such as:

1. transit to region of interest,
2. perform an observation, coverage, or patrol pattern,
3. maintain position or return to base.

For swarm-level tasks, the system distributes subtasks across available drones using simple, interpretable allocation rules—for example, dividing a search area into subregions or assigning observation points to specific platforms. Coordination is explicit rather than emergent to ensure transparency.

Safety gates are integrated into this process:

- Collision avoidance and spatial separation are enforced before execution.
- Platform capability checks ensure tasks are feasible for selected drones.
- No-fly constraints prevent instantiation of unsafe plans.

The task decomposition process is illustrated conceptually in Figure 3.

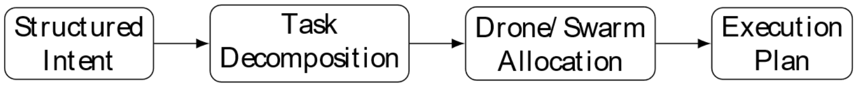

Figure 3: Conceptual task decomposition pipeline producing executable, safety-checked plans.

Transparency and Feedback Mechanisms

A key human factors requirement is that operators should understand how the system has interpreted their intent and how this interpretation translates into drone behaviour. To support predictable and trustworthy operation, the methodology incorporates transparency mechanisms:

- Intermediate task representations may be displayed to the operator.
- Planned actions are summarised before execution.
- Mission progress is reported using high-level, mission-relevant abstractions rather than low-level telemetry.

These mechanisms ensure that operators retain supervisory authority and can intervene, re-task, or abort missions when necessary. The overall framework thus balances autonomy with human oversight, enabling ergonomic interaction with single drones and small swarms while constraining system behaviour to remain safe, predictable, and explainable.

Consolidated Prototype Architecture

Figure 4 presents a consolidated view of the prototype, aligning the interaction model, intent interpretation, task decomposition, and execution/feedback into a single, inspectable pipeline. The design is modular to support transparency and supervisory control:

1. An Interaction Front-End acquires short operational commands (speech or text);
2. NLU & Intent Structuring maps language to a constrained, mission-relevant representation;

3. Task Decomposition & Safety Gates instantiate intent via predefined mission primitives, apply feasibility checks, and allocate work to single drones or small swarms;
4. An Execution & Feedback layer dispatches plans and returns high-level status abstractions to the operator. Exposing the intermediate representations (dashed callouts) supports human oversight and predictable behaviour.

The pipeline constrains interpretation to supported task types and exposes intermediate artifacts (structured intent, planned actions, allocation/ constraints) to the operator. This balances expressiveness with predictability: operators can issue high-level tasks, yet retain authority to inspect, confirm, or re-task before and during execution. Safety gates ensure feasibility and compliance (e.g., capability checks, separation, no-fly constraints), while the feedback loop reports mission-relevant progress instead of low-level telemetry, supporting reduced cognitive load in single- and small-swarm scenarios.

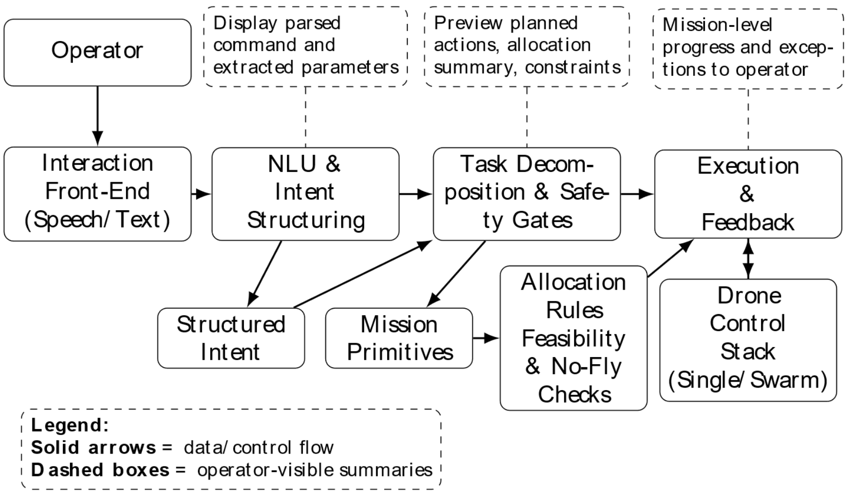

Figure 4: Consolidated architecture: ergonomic input (left) is interpreted into a structured intent, which is instantiated via predefined mission primitives under explicit safety gates and simple, interpretable allocation rules. Execution abstracts status back to the operator, exposing intermediate representations to maintain transparency and oversight.

DISCUSSION AND CONCLUSIONS

The methodology presented in this study illustrates the potential of natural language as an ergonomically appropriate, mission-level input medium for the control of individual drones and small swarms, particularly under interpretation constraints and transparency design. The structured approach to operator intent, task instantiation through mission primitives, and provision of intermediate representations achieve a balance between the benefits of autonomy and the need for continuous operator engagement. The focus on transparency is consistent with established human factors principles, including the need for predictability, cognitive manageability, and trust calibration in human-agent interaction.

The implementation of the methodology gives rise to a fundamental trade-off: increasing expressiveness in natural language interfaces requires more sophisticated interpretation and planning, but this may reduce predictability and increase operator workload. The methodology therefore prioritizes a scoped, constrained interaction space, consistent with the objective of enabling safe and interpretable supervisory control rather than fully replacing traditional interfaces.

To conclude, natural language is a promising medium for human-centred control of UAS, particularly under the structured approach presented in this study, where intent, task, and transparency are key considerations. The architecture presented here provides a basis for incorporating these considerations in a cohesive, predictable interaction process. Future extensions should explore the potential for better handling of ambiguity, task templates, and their application in larger, more time-sensitive scenarios. However, advancing natural language interfaces for drone operations requires finding a delicate balance between autonomy, transparency, and human cognitive capabilities.

ACKNOWLEDGMENT

The research presented in this paper has received research funding from Belgian Defence and the Belgian Royal Higher Institute for Defence in the framework of the HADRON project (DAP22/06).

REFERENCES

Al-Dosari, K., Hunaiti, Z. & Balachandran, W. (2023). Systematic review on civilian drones in safety and security applications. Drones, 7(3), p. 210.

Alexis, K., Papachristos, C., Siegwart, R., & Tzes, A. (2016). Robust model predictive flight control of unmanned rotorcrafts. Journal of Intelligent & Robotic Systems, 81(3), 443–469.

Casado Fauli, A., Malizia, M., Hasselmann, K., Le Flécher, E., De Cubber, G. & Lauwens, B. (2024) HADRON: Human-friendly Control and Artificial Intelligence for Military Drone Operations. In Proceedings 33rd IEEE International Conference on Robot and Human Interactive Communication, IEEE RO-MAN.

De Cubber, G., Doroftei, D., Balta, H., Matos, A., Silva, E., Serrano, D., Govindaraj, S. Roda, R., Lobo, V., and Wagemans, R. "Operational Validation of Search and Rescue Robots," in Search and Rescue Robotics – From Theory to Practice, De Cubber, G. and Doroftei, D. Eds., InTech, 2017.

Doroftei, D., De Cubber, G. & De Smet, (2020) H. Reducing drone incidents by incorporating human factors in the drone and drone pilot accreditation process. 11th International Conference On Applied Human Factors And Ergonomics (AHFE 2020).

Doroftei, D., De Cubber, G. & De Smet, H. (2022) Assessing Human Factors for Drone Operations in a Simulation Environment. Human Factors In Robots, Drones And Unmanned Systems – AHFE (2022) International Conference.

Doroftei, D., De Cubber, G., Lo Bue, S. & De Smet, H. (2024) Quantitative assessment of drone pilot performance. Drones. {8}, 482.

Fernandez, R. A. S., Sanchez-Lopez, J. L., Sampedro, C., Bavle, H., Molina, M., & Campoy, P. (2016). Natural user interfaces for human-drone multi-modal interaction. In 2016 International Conference on Unmanned Aircraft Systems (ICUAS) pp. 1013–1022. IEEE.

Heidari, A., Jafari Navimipour, N., Unal, M., & Zhang, G. (2023). Machine learning applications in internet-of-drones: Systematic review, recent deployments, and open issues. ACM Computing Surveys, 55(12), 1–45.

Javaid, S., Fahim, H., He, B., & Saeed, N. (2024). Large language models for uavs: Current state and pathways to the future. IEEE Open Journal of Vehicular Technology.

Liu, J. X., Yang, Z., Schornstein, B., Liang, S., Idrees, I., Tellex, S., & Shah, A. (2022, December). Lang2ltl: Translating natural language commands to temporal specification with large language models. In Workshop on Language and Robotics at CoRL 2022.

López, D., Moreno, G., Cordero, J., Sanchez, J., Govindaraj, S., Marques, M., Lobo, V., Fioravanti, S., Grati, A., Rudin, K., Tosa, M., Matos, A., Dias, A., Martins, A., Bedkowski, J., Balta, H., De Cubber, G. (2017). Interoperability in a heterogeneous team of search and rescue robots. Search And Rescue Robotics-From Theory To Practice. pp. 93–125

Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2024). Unmanned Aerial Vehicle (UAV) path planning and control assisted by Augmented Reality (AR): The case of indoor drones. International Journal of Production Research, 62(9), 3361–3382.

Nwankwo, L., & Rueckert, E. (2024, March). The conversation is the command: Interacting with real-world autonomous robots through natural language. In Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction pp. 808–812.

Obaid, M., Kistler, F., Kasparavičiūtė, G., Yantaç, A. E., & Fjeld, M. (2016). How would you gesture navigate a drone? a user-centered approach to control a drone. In Proceedings of the 20th International Academic Mindtrek Conference (pp. 113–121).

Schmidt, R., Schadow, J., Eißfeldt, H., & Pecena, Y. (2022). Insights on remote pilot competences and training needs of civil drone pilots. Transportation research procedia, 66, 1–7.

Yadav, A., Patel, A., & Shah, M. (2021). A comprehensive review on resolving ambiguities in natural language processing. AI Open, 2, 85–92.