

RESEARCH ARTICLE

Contextualised generative AI in system of systems modelling: an approach for firefighting aircraft requirements

J. Lovaco¹, R. C. Munjulury^{1,2} and P. Krus¹

¹Department of Management and Engineering (IEI), Linköping University, Linköping, Sweden

²Saab Aeronautics, Linköping, Sweden

Corresponding author: J. Lovaco; Email: jorge.lovaco@liu.se

Received: 3 November 2024; **Revised:** 19 October 2025; **Accepted:** 21 October 2025

Keywords: aircraft conceptual design; agent-based modelling; firefighting; helicopter; large language models; requirements; system of systems analysis; unmanned aerial vehicle; wildfire management

Abstract

The conceptual design of mission-tailored aircraft is increasingly shifting towards system of systems (SoS) perspectives that account for system interactions using a holistic view. Agent-based modelling and simulation (ABMS) is a common approach for analysing an SoS, but the behaviour of its agents tends to be defined by rigid behaviour trees. The present work aims to evaluate the suitability of a prompt-engineered large language model (LLM) acting as the Incident Commander (IC), replacing the fixed behaviour trees that govern the agents' decisions. The research contributes by developing a prompting framework for operational guidelines, constraints, and priorities to obtain an LLM commander within a wildfire suppression, SoS capable of replicating human decisions. By enabling agents in a simulation model with decision-making capabilities closer to those expected from humans, the commander's decisions and potential emergent patterns can be translated into more defined requirements for aircraft conceptual design (ACD) (e.g., endurance, payload, sensors, communications, or turnaround requirements). Results showed that an LLM commander facilitated adaptive and context-aware decisions that can be analysed via decision logs. The results allow designers to derive aircraft requirements for their specific roles from operational outcomes rather than a priori assumptions, linking SoS mission needs and ACD parameters.

Nomenclature

<i>A</i>	rotor disk area (m ²)
<i>ABMS</i>	agent-based modelling and simulation
<i>ABS</i>	agent-based simulation
<i>ACD</i>	aircraft conceptual design
<i>AI</i>	artificial intelligence
<i>C_L</i>	lift coefficient
<i>C_T</i>	thrust coefficient
<i>C_P</i>	power coefficient
<i>ConOps</i>	concept of operations
<i>IC</i>	Incident Commander
<i>L</i>	lift force (N)
<i>LCES</i>	lookouts, communications, escape routes, safety zones
<i>LLM</i>	large language model
<i>R</i>	rotor radius (m)

V	flight velocity (m/s)
V_{stall}	stalling velocity (m/s)
S	wing area (m ²)
SoS	system of systems
T	thrust (N)
$TRDP$	technical requirements definition process
$\tanh(x)$	hyperbolic tangent function
UAV	unmanned aerial vehicle
W	aircraft weight (N)

Greek symbol

λ_h	hover inflow ratio
Ω	rotor rotational velocity (rad/s)
ρ	flow density (kg/m ³)
ρ_∞	free-stream air density (kg/m ³)
$\sigma(x)$	sigmoid activation function

1.0 Introduction

Over the last decade, wildfires have increased in number and size, earning the designation of megafires. For this reason, more resources are being allocated to preventing and fighting wildfires [1]. These resources may include increased funding for firefighting efforts and improved technology for early detection and monitoring. Additionally, collaboration between government agencies, environmental organisations, and local communities has become increasingly vital to address the growing threat of wildfires and minimise their devastating impact on both the environment and human lives [2]. To achieve this, developing an SoS framework for performing comprehensive and holistic analyses is a logical approach [3]. By considering all the constituent systems involved in a mission and assessing the resulting operational outcomes, it becomes possible to identify emergent behaviours that arise from collaboration, unmet needs, and the desired capabilities. These insights will provide a data-driven foundation for ACD.

However, despite multiple technological advancements, there is still a gap in the conceptualisation phase of aircraft design, particularly in modelling realistic operational scenarios and collaborative dynamics to fully understand the outcome of different sets of capabilities [4]. For such scenarios and dynamics, the use of ABMS is generally recognised as adequate and capable of capturing emergent behaviours of collaborative agents, which are very difficult to predict [5]. Agents represent constituent systems and are defined by sets of behaviours. These behaviours traditionally rely on decision trees such as the one shown in Fig. 1, which, while structured, constrain the fluidity and collaborative essence of real-life situations, such as those encountered in wildfire suppression operations. This limitation not only distorts the design process but also potentially impacts the operational efficacy of the designed system.

1.1 Objectives of the paper

The present study aims to conduct an agent-based simulation (ABS) whose agents’ behaviour will be driven by an LLM. The goal is to evaluate the combination of both for future SoSs analyses with decision trees removed, offering a more dynamic and realistic representation of decision-making. The feasibility and potential benefits of utilising LLMs to model human decision-making within an ABS will be evaluated. The unique challenges of ACD specifically for wildfire fighting, such as the need for rapid response, operational versatility and coordination with other systems, make the application of ABS and LLMs to the study of SoSs particularly relevant if a more accurate, holistic, and operationally relevant analysis is to be achieved.

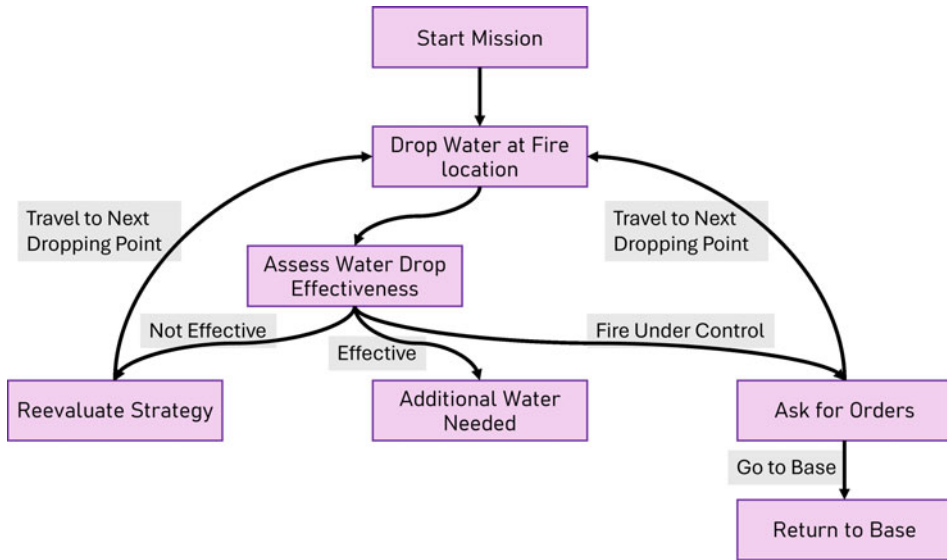


Figure 1. Example of a decision tree for a firefighting agent.

2.0 Related work

Wildfire modelling has been developed to provide support during incidents in terms of operations management, planning and even scientific analyses, with FARSITE being one such tool [6]. It was developed by the U.S. Forest Service, and the fire propagates its perimeter as expanding ellipses, with the rate of spread and intensity governed by the Rothermel model [7]. The fire simulator includes effects such as spotting and fuel moisture. Prometheus, a wildfire simulator, adopts a similar perimeter expansion methodology but is calibrated for Canadian forests [8, 9].

These tools are semi-empirical and, while robust, simplify processes such as spotting, assume continuous fuels and are highly sensitive to uncertainties in wind, fuel and moisture inputs, which can lead to incorrect predictions and biased judgements [10]. Consequently, there is increasing interest in complementing, rather than replacing, these established tools with systems based on artificial intelligence (AI). Examples of domain-specific LLMs, such as WildfireGPT, have begun to appear [11]. Although their language capabilities could be used for wildfire discussions, they would require very specialised training data and the algorithmic precision necessary for adequate wildfire behaviour predictions. However, LLMs can also serve as an additional layer to support resource allocation and communication under uncertainty, improving human decision-making [12].

The potential of LLMs extends beyond wildfire scenarios. In healthcare, advanced reasoning models have demonstrated diagnostic performance comparable to human physicians in medical scenarios [13]. Similarly, in disaster response, LLMs, such as disaster response GPT, have been able to translate brief situation reports into structured plans of action, although human validation was still required [14]. In other fields, such as logistics and supply chain management, LLMs and forms of generative AI have been used to find new ways to enhance decision making, automate operations, or reduce disruptions, but they were found to be prone to hallucination issues [15].

The trend is to use generative AI and LLMs for enhancing the speed and adaptability of human reasoning, provided their outputs are properly supervised and evaluated. Their integration offers rapid decision-support systems for dynamic environments such as wildfires, clinical emergencies, disaster response and logistics.

3.0 Theory

This section introduces the theoretical basics to understand the various concepts needed to perform the present work.

3.1 System design

System design is one of the technical processes that are part of systems engineering [16]. The system design process is a structured approach comprising four fundamental stages: developing stakeholder expectations, defining technical requirements, performing logical decompositions, and creating design solutions. A thorough understanding of stakeholder expectations, which include mission objectives, constraints, operational needs, and mission success criteria, is necessary to satisfy them. These expectations translate into requirements that guide a project throughout its life cycle. The Technical Requirements Definition Process (TRDP) transforms stakeholder expectations into problem definition first and into validated technical requirements later on, which are expressed as ‘shall’ statements [16]. Inputs to TRDP include top-level requirements, expectations and a description of how a system can be operated. The TRDP focuses on defining functional and behavioural expectations, establishing performance requirements for each function. The outputs are comprehensive technical requirements and technical measures that guide the development process. Well-crafted requirements establish a clear agreement between stakeholders and developers regarding the product’s intended functions, reduce development effort, provide a reliable basis for cost estimation, serve as a baseline for validation and verification processes, and as a foundation for future enhancements, ensuring that the system remains effective and adaptable over time.

3.2 Agent-based modelling and simulation (ABMS)

Field-testing data is sometimes scarce or even impossible to obtain. Such is the case with wildfires, where creating a real fire situation is not feasible due to the obvious costs and disastrous side effects. Therefore, a model is necessary, but modelling complex systems in a dynamic and interactive environment is a challenging task [17]. ABMS becomes particularly convenient in this context, thanks to its representation of singular entities, or ‘agents’, each with unique behaviours and decision-making capabilities, which is a very valuable aspect in systems engineering [18]. However, it is particularly relevant for wildfire management scenarios, where diverse agents interact dynamically with the environment. ABMS enhances the exploration of emergent behaviours that arise from these interactions with the environment and the agents with one another, providing additional insights into the systems’ performance both individually and collectively [19].

3.3 System of systems (SoS)

An SoS is a set of independent constituent systems working together to achieve a common goal with certain specific characteristics [20]. During the life cycle of an SoS, different levels of centralisation might be seen and be used to categorise the type of SoS [21], with the command centres, departments and assets operating independently yet coordinated towards a shared mission. Under this distributed and multi-agency configuration, emergent behaviours will be exhibited. Whether these new behaviours are improvements or conflicts, they are not found in any of the systems alone; thus, the communications, operating conditions and collaborations that affect the overall SoS need to be explored and understood.

3.4 Neural networks and large language models

An artificial neural network is a computational architecture based on biological neural networks. These artificial neural networks contain computational units, known as neurons, which are connected through weights that scale their inputs. These weights scale the computed function of the neuron, and the

results are propagated through the neural network layers as outputs. Multilayer neural networks used for machine learning or artificial intelligence typically consist of:

- **Input layer:** Receives the inputs
- **Hidden layers:** Intermediate layers where the computations take place
- **Output layer:** Produces the final outputs

The chain structure of the layers forming a neural network consists of the input features x , the weights W , the bias b , the activation function g , and the output h_i :

$$h^{(1)} = g^{(1)} (W^{(1)\top} x + b^{(1)}) \quad (1)$$

$$h^{(n)} = g^{(n)} (W^{(n)\top} h^{(n-1)} + b^{(n)}) \quad (2)$$

Based on advanced forms of statistical language models, which model the likelihood of a word appearing in a given sequence [22], combined with machine learning methods, such as recurrent neural networks or reinforcement learning [23] and transformer architectures [24], LLMs are trained iteratively to learn how to provide an outcome expected by humans. This capability is relevant in many fields, such as SoS modelling, to obtain evolving models that better reflect real-world decision-making [25].

3.4.1 Universal approximation theorem

The universal approximation theorem states that a feed-forward neural network with enough hidden layers can approximate the linear output layer to a non-zero amount of error for any Borel measurable function from a finite-dimensional space to another one. A continuous function is Borel measurable if it is continuous for a closed and bounded subset of $\mathbb{R}^n$ [26, 27]. Feed-forward networks with hidden layers require activation functions to introduce non-linearity into the model for the hidden layers [28], being some common examples of them:

- **Sigmoid:** $\sigma(x) = \frac{1}{1+e^{-x}}$
- **Hyperbolic tangent:** $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

This theorem implies that, with enough training, a neural network can be trained to obtain outputs with an error very close, but never equal, to zero.

3.5 Operational concepts

Planning for emergencies involves making assumptions about the conditions that might exist when such crises occur. This type of planning includes developing strategies for deploying, employing, sustaining and redeploying assets in response. When a crisis emerges, an existing plan with a similar scenario might be used as a starting point for further planning. The Concept of Operations (ConOps) is defined as a statement that clearly and concisely expresses what the joint force commander intends to accomplish, and how it will be done using available resources [29]. It outlines assumptions, phases of the operation, prioritised missions and requirements for deployment of the assets. Through detailed and simulated planning, the requirements are determined for likely operational needs, which allows for rapid adaptation by comparing hypothetical conditions with real-time needs [30]. The joint planning process is a method used for military planning. Planners utilise it to develop ConOps, force plans, deployment plans and supporting plans that include multiple courses of action. These provide flexible options for joint forces, enabling adaptation to changing operational environments, resource identification and risk mitigation. The organisation of joint forces and the establishment of command relationships depends on the complexity and control required by the operation or ConOps. This process involves determining the types of subordinate commands and the level of authority delegated to each. Commanders also define the authority and responsibilities of functional component commanders, adjusting them as necessary during operations [30].

3.5.1 Force planning and operational design

The goal of force planning is to identify all the assets needed to execute the ConOps. This process involves determining force requirements based on operational phase, mission priority and operating area. It includes reviewing asset requirements, integration planning and refining asset lists [30].

Operational design serves as an analytical framework for planning, helping commanders and planners understand and organise the operational environment as a complex system. Commanders must comprehend the audience and political environment to provide the best advice to civilian decision-makers. Planners ensure that the language is understood, with inter-agency partners being crucial to this process. Operational design is intertwined with the planning process, filling gaps in guidance and information, and providing a framework for addressing the complexity of the operational environment. It supports mission analysis, course of action development and the creation of ConOps with a high likelihood of success [30].

3.6 Basics of rotary-wing aerodynamics

Helicopter aerodynamics studies the numerous complex interactions during the different flight regimes, such as the main rotor wake structure and its effect on the tail rotor, transonic flows over the blades, stalled regions, etc. There are, however, several results based on momentum theory that are simple enough for quick estimations [31, 32]. The thrust coefficient is one of the most relevant values and is defined as:

$$C_T = \frac{T}{\rho A \Omega^2 R^2} \quad (3)$$

Where T is the thrust, A is the rotor disk area, Ω is the rotor rotational velocity, R is the rotor radius and ρ is the flow density. In a similar manner, the power coefficient is defined as:

$$C_P = \frac{P}{\rho A \Omega^3 R^3} \quad (4)$$

Being P power, A rotor disk area, Ω , R rotor radius and ρ the flow density. Then, assuming one-dimensional flow for a hovering rotor case, the inflow ratio λ_h can be defined based on momentum theory as:

$$\lambda_h = \frac{v_h}{\Omega R} = \sqrt{C_T/2} \quad (5)$$

3.6.1 Initial rotor sizing example in conceptual helicopter design

This section includes a small illustrative example of rotor sizing for a conceptual helicopter. It is common to start by considering the hover state [33] and use Equation (3). Historical trends can be used together with weight build-up methods for an initial value of the thrust coefficient [4], although the additional weight of a water bucket should be considered. An initial consideration for the process could be to keep the thrust coefficient value similar to other helicopters [34], and modify the rotor geometry:

$$C_T = \frac{T_{ref}}{\rho A_{ref} \Omega^2 R_{ref}^2} = \frac{T_{new}}{\rho A_{new} \Omega^2 R_{new}^2} = \frac{T_{ref} + \Delta T}{\rho \pi \Omega^2 (R_{ref} + \Delta R)^4}$$

From the additional weight ΔT requirement related to the water bucket load, it is possible to use reference values and the relationship above to obtain the variation of the rotor radius ΔR for the new geometry. After some manipulation, the increment can be computed using the new relationship:

$$R_{ref} \left(\sqrt[4]{1 + \frac{\Delta T}{T_{ref}}} - 1 \right) = \Delta R$$

The new rotor radius needs to be checked against operational constraints, such as the need to land in remote wooded terrains, where finding open areas with adequate blade clearance from trees will become more challenging for larger rotor blades. The velocity at the blade tip ($V_{tip} = \Omega R$) is another constraint, as values beyond $M = 0.7$ will be in the transonic region, leading to issues such as additional noise.

It is possible to continue the sizing by using blade element analysis [33], which provides the analytical result that relates the thrust coefficient to the mean lift coefficient of the entire blade:

$$\bar{C}_L = 6 \left(\frac{\pi R}{N_b c} C_T \right) = C_{l_a} \bar{\alpha} \quad (6)$$

With C_{l_a} defined as the blade profile lift-slope, $\bar{\alpha}$ as the mean angle-of-attack of the blades, N_b as the blade count and c as the blade chord. It is convenient to introduce the definition of solidity (σ) [33] as:

$$\sigma = \frac{N_b c}{\pi R} \quad (7)$$

The different values in Equation (4) show that if the goal is to keep the thrust coefficient constant at a specific value, the variations of the geometrical values and the aerodynamic properties are going to be correlated. To continue with the example for the initial considerations for the rotor, remaining equations (Equations (4) and (5)) can be combined with Equation (3) to relate the ideal power coefficient to the thrust coefficient:

$$C_P = \frac{P}{\rho A \Omega^3 R^3} = \frac{T \cdot v_h}{\rho A \Omega^3 R^3} = \frac{C_T^{3/2}}{\sqrt{2}}$$

This relationship is useful for computing power, but it is more convenient for generating a figure of merit (FM) to relate the ideal power to hover and an estimation of a more realistic value of the non-ideal power required [33]:

$$FM = \frac{P}{\rho A \Omega^3 R^3} = \frac{T \cdot v_h}{\rho A \Omega^3 R^3} = \frac{C_T^{3/2} / \sqrt{2}}{C_{P_{real}}}$$

The power needed can be approximated without measurements by considering losses. If only induced and profile power losses are considered, it is possible to write:

$$C_P = C_{P_i} + C_{P_0} \quad (8)$$

Where the induced power losses are considered by adding a correction factor κ with a typical value of 1.15 [33], then:

$$C_{P_i} = \kappa C_T^{3/2} / \sqrt{2} \quad (9)$$

Then, for a rectangular blade with an assumed constant blade section profile drag coefficient (C_{d_0}), the profile power coefficient can be found to be [33]:

$$C_{P_0} = \frac{\sigma C_{d_0}}{8} \quad (10)$$

It is possible now to plot the figure of merit with respect to thrust coefficient variations. Helicopter typical values [34] for solidity and blade section profile drag coefficient are, respectively, $\sigma = 0.1$ and $C_{d_0} = 0.01$, resulting in Fig. 2. A good performance for hovering rotors is between 0.7 and 0.8 in the figure of merit [33, 35]. The analysis is not extended further for this example, as the process would continue with different considerations and discussions for going deeper into the new concepts' specifics [36], which is beyond the scope. Variations in the number of blades, their chord length, aerodynamic profiles, etc., should be considered then.

3.7 Basics of fixed-wing aerodynamics

For the case of fixed-wing aircraft, there are some useful closed-form solutions under the assumption of inviscid and incompressible flow, especially for finite wings [37]. An important value is the lift coefficient:

$$C_L = \frac{2L}{\rho_\infty V_\infty^2 S} \quad (11)$$

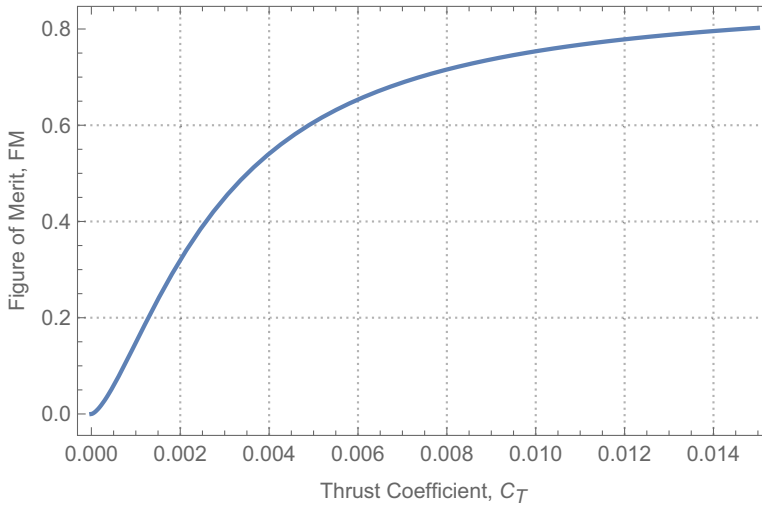


Figure 2. Figure of merit for a rotor during hover.

Where L is the lift force, ρ is the flow density, V is the flight velocity, and S is the wing area. From Equation (11), using the maximum lift coefficient achievable, the stalling velocity can be obtained, which is particularly relevant when designing air tankers that will drop water or retardant:

$$V_{stall} = \sqrt{\frac{2W}{\rho_{\infty} S C_{L,max}}} \quad (12)$$

3.8 System capabilities

Stakeholders define the desired capabilities for the systems to be designed. These capabilities will constrain the requirements for new system concepts, whether they are for aircraft or any other kind. Several of the capabilities that will define the system requirements can be shared between concepts. Some capabilities to satisfy operational needs are defined and listed here, alongside the proposed requirements that will be common to all systems proposed:

1. **Adaptability:** Ability to respond to dynamic and unforeseen changes in the environment. Common requirements for all systems include modular design for new technologies' integration and easier manufacturing [38]; capable of handling various weather conditions and terrains.
2. **Communication:** Effective use of communication protocols for coordination between different units. Common requirements for all systems include advanced communication systems for interoperability with ground units and other aircraft; encrypted communication.
3. **Environmental Assessment:** Ability to assess and respond to environmental factors like fire magnitude and weather conditions. Common requirements for all systems include onboard sensors and cameras for monitoring in real-time fire and weather; assessment of terrain features, vegetation types and weather patterns.
4. **Operational Efficiency:** Ability to perform its intended functions by minimising resources needed and maximising output.
5. **Safety and Emergency Response:** Ability to respond to medical emergencies and ensure safety protocols. Common requirements for all systems include fire-resistant features to avoid degradation [39].
6. **Autonomy:** Ability for autonomous operation with minimal human intervention.



Figure 3. Workflow of capability into requirements translation (adapted from (40)).

3.8.1 Capability and requirement engineering

Capability engineering begins with an understanding of a need, and an exploration of the multiple possible functionalities to achieve a capability, which fulfils the aforementioned need [40]. Available resources are related to existing functions that can contribute to obtaining the new capability. Each potential solution is evaluated to measure performance and cost-effectiveness. Analyses of the SoS are used for the translation of capabilities into requirements in a repeatable and traceable manner. A clear understanding of SoS dynamics is necessary, as it will evolve, resulting in variable performance expectations. The process of translating capabilities into requirements that will support trade-off decisions for the expected level of performance against cost. The process is shown in Fig. 3.

Requirements engineering is a systematic process that ensures stakeholder needs are accurately captured and translated into system specifications. To start this process, the necessary capabilities were previously translated into high-level requirements. These are decomposed into detailed functional and performance requirements. This decomposition process involves:

- **Decomposition** into lower-level requirements that specify what the system must do to achieve functional and performance requirements.
- **Allocation** of these requirements to specific system elements, subsystems, or components.
- **Validation** of each level of requirements against stakeholder expectations.

This iterative process continues until a comprehensive set of designs to requirements is established at all levels of the system hierarchy. Traceability is crucial and necessary throughout this process, ensuring that every requirement can be linked back to a stakeholder need or higher-level capability, and that the rationale for each requirement is documented.

Managing requirements effectively involves capturing not only the requirements statements, but also associated metadata from the outset, which is necessary for clarification and linkage among requirements. A general guideline of what each should include is:

- **Requirement Statements:** Clear expressions of what the system must do to achieve the desired capabilities.
- **Metadata:** Information such as the requirement's unique identifier, source, rationale, owner, verification method and verification level. These points ensure that the intent and context of each requirement are clear, enhancing communication among project stakeholders and facilitating changes when necessary.

The verification methods are beyond the scope of the present work; however, when they are documented early in the process, traceability is improved and the inclusion of any unforeseen requirements found during a verification stage is easier.

3.8.2 Logical decomposition and system architecture development

Logical decomposition is the process of breaking down a system's required functionalities into discrete elements [16]. The steps in the process are:

- **Functional Analysis:** Identifying and defining the functions to meet any set of desired capabilities.
- **Functional Allocation:** Assigning these functions to system elements and their respective subsystems/components.

- **Interface Definition:** Identifying the interfaces between system elements and/or the outside world.

This process is iterative and relies on its connection with the ConOps to guide the understanding of how a system will be used. Logical decomposition ensures that all required system behaviours are identified and that the role of each subsystem is clearly defined. With this decomposition, the system design concepts are refined through system architecture development. It is a creative and iterative process whose aspects include:

- **Concept Generation:** Propose multiple design concepts that meet the project's goals.
- **Stakeholder Involvement:** Engaging stakeholders to provide feedback and ensure that the architecture and needs are aligned.
- **Analysis and Trade-off Studies:** Using modelling, simulation and trade-off analyses to evaluate different design options and select the most suitable design.

The process is recursive, with feedback loops necessary to refine requirements, design solutions and provide a robust concept that meets all requirements and is feasible.

3.8.3 Technical standards considerations

Technical standards contribute to compatibility, interoperability, safety and quality while often reducing costs and development time. They provide the normalisations and guidelines that inform various aspects of system development, which, depending on the case, may include specifications for performance, safety, reliability, materials used in manufacturing, verification procedures and interfaces. Appropriate standards reduce risk and increase interoperability with other systems.

4.0 Method

The present work was performed following the design research methodology [41]. The objective is to improve product development by employing simulation as a means to refine requirement definition based on operational needs and system interactions of aircraft. To this end, an evaluative artefact must be generated to assess the outcomes of the simulated operations. This artefact should support the examination of whether the captured results are sufficient for decisions on required functionalities, and whether the identified uncertainties originate from limitations in the simulation model itself.

For a qualitative evaluation of the simulation and the decisions, a questionnaire was sent to fire suppression operations pilots. The current goal is to generate a sacrificial concept to understand the behaviour of the model and gain knowledge to point out the direction to be taken [42]. The iterative loop in Fig. 4 shows how the design research methodology helps in generating prototypes, gaining knowledge in the process and finding an agreement on the complexity needed with a holistic view of what is necessary for obtaining a satisfactory product.

4.1 Interviews with firefighting pilots

The authors provided the following questions to firefighter pilots, and this section collects the summarised answers.

1. **Mission:** *When is a wildfire incident considered under control and/or successfully concluded?*
Both interviewees converge on the view that wildfires are seldom 'extinguished' in an absolute sense; instead, operational closure is achieved through phases and gradual demobilisation. Firefighter-1 defines 'under control' as the point at which reignition is no longer expected after assets are withdrawn, beginning with higher-cost aerial assets while ground crews continue to secure. Firefighter-2 provides a formal phase taxonomy: stabilised, controlled, finished and observation, clarifying that interior hot spots may persist through the first two phases without implying

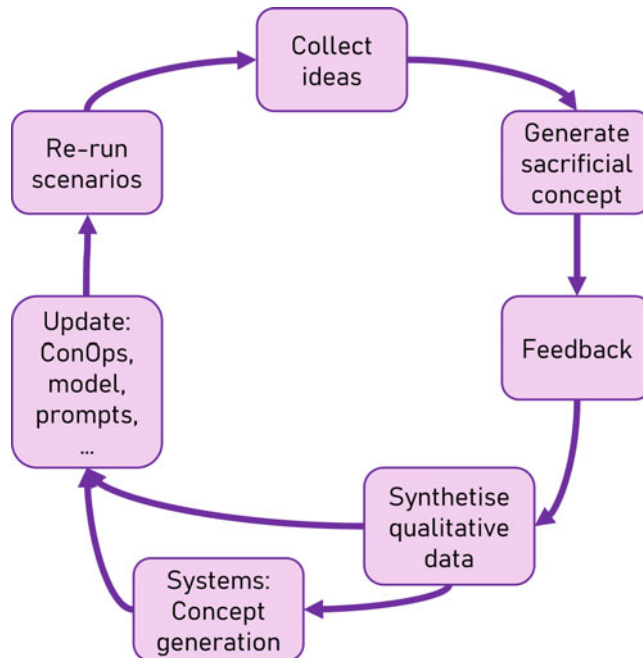


Figure 4. Creativity focused prototyping loop for sacrificial concepts.

loss of control. Together, these perspectives indicate the need for both practical simulation termination criteria and post-incident monitoring.

2. **Situational awareness:** *What information must be visible before any deployment?* Both respondents agreed on a minimum situational awareness set comprising precise fire location, active assets, and communications routing to the Incident Commander. Firefighter-1 stresses unambiguous identification via coordinates and a unique name, plus required frequencies and IC contact. Firefighter-2 adds operationally critical details: presence and frequency of an aerial coordinator; confirmation of ground crews to prevent hazardous drops; availability of nearby loading sites; and environmental/operational checks (visibility, on-scene wind verification, terrain escape planning via reconnaissance). The combination suggests a pre-launch information (coordinates, communications, assets, ground-crew status, visibility) and a mission-planning tier (loading sites, coordinator info, route/egress planning).
3. **Realism:** *Which environmental factors must the model represent to be considered reliable?* Firefighter-1 considered the question too broad and preferred not to answer. Firefighter-2, on the other hand, considered the proper evolution of the fire and the local orography particularly relevant: the factors that shape a fire's evolution are usually the wind (changes in both direction and intensity) and its ability to carry embers that start spot fires. Temperature is another critical element. So is the relative humidity in the area, both in the air and in the vegetation. Orography is very important: slopes, valleys, gullies, etc. This affects not only how the fire develops, but also often limits the capabilities of aerial resources, especially the heavier ones.
4. **Topography:** *What level of detail in maps/fuels is necessary to plan firelines, strategies, attack plans, etc.?* Both firefighters considered that this question should be addressed and answered by a forestry engineer.
5. **Communication:** *Which communication behaviours are most important to simulate?* Both emphasise the centrality of an aerial coordinator to reduce workload and deconflict traffic, and both describe multi-frequency operations. Firefighter-1 specifies a three-frequency architecture

(air-to-air/coordination, ground/incident-management-authority, and company/formation) and notes that the absence of a coordinator forces aircraft-to-aircraft self-coordination with a substantial workload increase. Firefighter-2 highlights typical medium-to-high radio saturation, and the need to model standard phraseology, radio discipline and local-language constraints; an emergency ‘radio-silence’ control is identified as essential for critical events. Jointly, these inputs motivate simulations that involve coordinator presence, explicitly meter channel occupancy, and enforce procedure-aware message flows.

6. **New systems:** *What unmanned aerial vehicle (UAV) roles and information cadence would be necessary?* Both interviewees cautioned against UAV presence within active manned airspace due to collision and distraction risks. Firefighter-1 proposes a vertical separation convention, manned aircraft if concurrent, while acknowledging high-altitude UAV sensing as a compromise. Firefighter-2 argues for excluding UAVs during manned operations entirely, citing poor UAV situational awareness (no transponder, limited communications) and recommending sensing to be performed from a coordinating aircraft; UAVs are endorsed for night or post-ops mapping. This comparison supports a conservative policy baseline (no mixed ops), with a clearly parametrised alternative allowing strictly controlled vertical separation when mixed operations are unavoidable.
7. **Operational emergencies:** *What steps should be taken when a team member is injured?* Firefighter-1 considers this question as a ground concern. Firefighter-2 answers based on events experienced and provides an anecdote: if crews are in severe danger – due to entrapment, a fire involving a pumping unit, or similar events – the best course is for the aerial operations coordinator to order radio silence on the channel and issue highly directive instructions about what is needed. As example, during an operation in Portugal, a ground crew member became trapped; radio silence was imposed, and all Canadair aircraft were directed to the crew’s coordinates to drop water, wet the area, and lower temperatures to avert a tragedy. The key in such cases, if aerial support is required, is radio silence, clearing the area of resources that will not participate, and giving very clear instructions on what is required.
8. **New events:** *How should the chain of command behave in the simulation when a second ignition appears or priorities change?* Both agree that secondary ignitions require immediate attention to prevent rapid spread, and note that aerial assets frequently provide first detection. Firefighter-1 stresses rapid redirection, highlighting the higher success probabilities when spots are small at first engagement. Firefighter-2 frames prioritisation in terms of area sensitivity, implying a risk-weighted allocation rather than a purely first-come response. Together, these inputs argue for a detection-to-tasking pipeline that minimises latency and ranks emergent targets by criticality.
9. **Uncertainties:** *What types of uncertainty should be present?* Firefighter-1 emphasises logistical and infrastructural constraints such as fuel resupply limitations, airport congestion that decouples aircraft count from drop throughput, and accommodation/transport frictions (diminishing returns from additional aircraft). Firefighter-2 focuses on environmental and obstacle uncertainties such as wind shifts, adverse weather and clouds, and evolving hazards such as new power lines or reservoir modifications, and incomplete information on loading sites.
10. **For this type of simulation:** *What would make it useful for training or for an after action review or debriefing?* Both endorse standardisation and realism as prerequisites for training value. Firefighter-1 calls for internationally standardised procedures (sectoring, drop types, priorities, formation management) and highlights the off-season as an opportunity for cross-regional training. Firefighter-2 specifies training objectives – resource management, workload handling, prioritisation and safety under changing meteorology and terrain – and identifies comprehensive recording (multi-frequency comms and operational timelines) as essential for debriefs that surface errors and lessons learned. Collectively, these elements define a training feature set that would combine standards-based scenario templates, stress-inducing variability and telemetry capture to enable procedure adherence checks and evidence-based feedback.

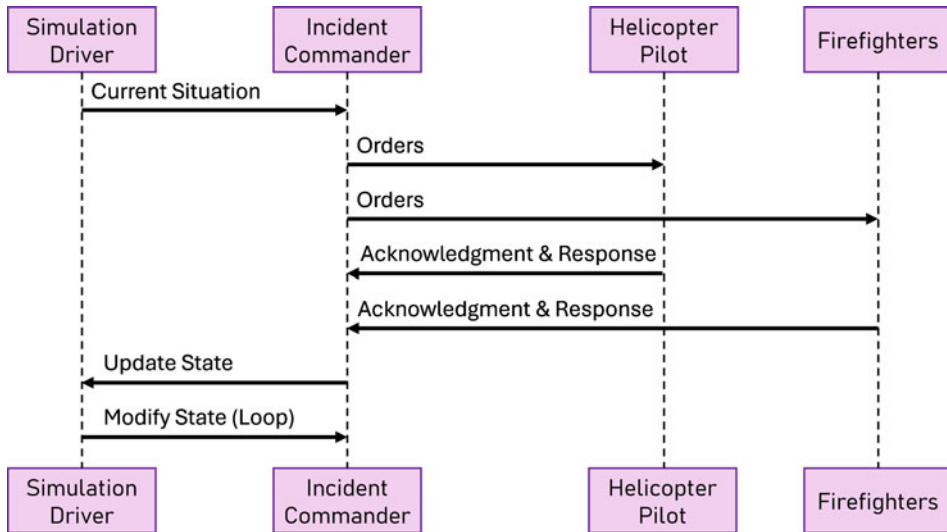


Figure 5. Sequence diagram for wildfire incident management.

4.2 Responsibilities and actions of an Incident Commander (IC)

The main points extracted from ConOps found in the reviewed literature are presented in this section [43, 44]. They will be used as prompts for the selected state-of-the-art LLM [45] and constrain its responses to the required operational concepts. Figure 5 shows a sequence diagram for a directed SoS driven by the LLM-commander. The prompts given follow a latex-based structure, which the authors have found to be easier to interpret by the LLM due to its clear structure.

The IC guidelines in terms of responsibilities and communication are listed in this section. All the following points listed here are used as a prompt for the LLM.

- **Provide complete briefings:**
 - Clearly communicate the task, purpose, and end state of the assignment to subordinates.
 - Ensure subordinates have a clear understanding of the assignment.
- **Evacuation procedures:**
 - Plan how to get injured personnel to a road or helispot.
 - Identify the required personnel and equipment for evacuation.
 - Calculate the time needed to get the injured to a hospital and determine the closest hospital.
 - Decide on air or ground transportation and consider factors that could affect the timeline.
- **Recognise common denominators of fire behaviour in tragic fires:**
 - Be aware of the conditions under which fatal or near-fatal fires often occur, such as:
 - * Small or quiet areas of large fires.
 - * Light fuels like grass and light brush.
 - * Unexpected shifts in wind direction or speed.
 - * Critical burn period between 1400 and 1700 hours.
- **Radio traffic and discipline:**
 - Maintain disciplined and concise radio communication.
 - Use clear text for all operations.
 - Follow established procedures for radio calls, identifying the resource being called and the frequency used.

- **Issuing air traffic information and advisories:**

- Provide clear air traffic information and ensure pilots acknowledge it.
- Pass on new information and ensure no missions are launched in conflict areas.
- Follow standard scripts for communication to ensure clarity and consistency.

- **Risk management principles:**

- Accept no unnecessary risk and make risk decisions at the appropriate level.
- Balance risk against benefits and integrate risk management into planning and execution at all levels.

4.3 Responsibilities and duties of helicopter personnel

The helicopter personnel guidelines in terms of responsibilities and communication are listed in this section. It is assumed that the pilot is responsible for ensuring that all of them are followed; thus, only the pilot is considered for the communication. All the points listed here were used as a prompt for the LLM.

- **General responsibilities:**

- Adhere to Federal Aviation Regulations (FAR) and agency-specific regulations.
- Coordinate with dispatchers, helicopter managers and/or helibase managers.
- Ensure aircraft and communication equipment are in good condition and operable.

- **Flight operations:**

- Do not deviate from the flight plan without notifying the appropriate dispatch office.
- Do not descend below 500 feet above ground level without prior authorisation or for high-level reconnaissance.

- **Risk management:**

- Accept no unnecessary risk.
- Make risk decisions at the appropriate level.
- Accept risk only when the benefits outweigh the costs.
- Integrate risk management into the planning and execution phases.

4.3.1 Helicopter-specific procedures

- **Landing area selection:**

- Choose a flat area free of obstructions (trees, poles, wires, etc.).
- Ensure the area is clear of stumps, brush, rocks, and any objects over 18 inches high.
- Consider wind direction for landing and takeoff.

- **Aerial delivery of retardant/water drop considerations:**

- Maintain communication with aerial resources.
- Identify and communicate flight hazards to aerial supervisors.
- Provide specific target information using clear descriptions and signal markers.
- Clear the area to avoid direct flights over ground personnel and equipment.

- **Aircraft mishap response actions:**

- Prioritise life preservation and secure the area.
- Exercise caution due to potential hazards at the wreckage site.

4.3.2 Communication protocols

- **Radio discipline and clear text usage:**

- Ensure all aircraft and ground personnel have compatible radios and frequencies.

- Maintain disciplined and concise radio traffic.
- Use clear text for all operations, keeping messages brief and to the point.
- **Standard scripts for communication:**
 - **Flight following departure script:**
 - * Identify the tail number or other designated identifier.
 - * State the frequency being used.
 - * Provide departure location, number onboard, fuel onboard, estimated time enroute (ETE), and destination.
 - * Confirm automated flight following (AFF).
 - **FTA (flight traffic advisory) calls in the blind:**
 - * Identify the receiving unit.
 - * State the tail number or designated identifier.
 - * Provide distance and direction from the incident, altitude, intent, and frequency.
- **Landing and departure communications:**
 - **When helicopter contacts helibase:**
 - * Provide helicopter number and helibase identification.
 - * Report wind speed and direction.
 - * Notify about traffic status (inbound/outbound aircraft).
 - **Before helicopter departs helibase:**
 - * Provide helicopter number and helibase identification.
 - * Report wind speed and direction.
 - * Notify about traffic status (inbound/outbound aircraft).

4.4 Responsibilities and actions of firefighters in firefighting operations

The guidelines for the firefighters responsibilities and communications are listed in this section. It is assumed that the team leader is responsible for ensuring that all of these steps are followed and performing communication when needed. All the following points listed here are used as a prompt for the LLM.

4.4.1 Fireline construction and safety

- **Establishing Firelines:**
 - Use direct attack whenever possible, constructing a fireline as close to the fire edge as conditions safely permit.
 - If an indirect attack is required, locate the fireline at an adequate distance from the main fire to allow for completion, firing and holding, considering the predicted rate of fire spread.
 - Make the fireline as short and straight as practical, using topography to your advantage.
 - Use existing natural and human-made barriers.
- **Communication Requirements and Procedures:**
 - Assign experienced and competent lookouts at good vantage points with knowledge of crew locations, escape and safety locations, and trigger points.
 - Confirm radio frequencies and establish backup procedures and check-in times.
 - Provide updates on any changes in the situation and sound alarms early.
- **Radio Traffic and Discipline:**
 - Use clear and concise radio communication.
 - Identify yourself and the resource you are calling first, followed by the message.

- Follow established procedures for radio calls and ensure disciplined radio traffic.
- Confirm receipt of critical information and acknowledge messages.

4.4.2 Reporting issues

- **Reporting Injuries:**

- Immediately report any injuries to the supervisor or IC.
- Provide details about the injury, including the nature and severity of the injury, and the number of injured personnel.
- Communicate the need for medical support and transportation.

- **Reporting Stress:**

- Report signs of fatigue or stress among crew members to the supervisor.
- Monitor crew members for symptoms of overexposure to smoke, such as headaches, visual impairment and impaired decision-making.

4.4.3 Specific communication scripts

- **Initial Contact Script:**

- Identify the resource being called and your identification.
- Provide a brief message regarding the situation.

- **Requesting Medical Support Script:**

- Identify the resource being called and your identification.
- State the nature of the injury and the need for medical support.
- Provide your location and any relevant details.

4.5 Agent-based simulation setup

The simulation is performed using NetLogo [46], a software designed for modelling and simulation of complex systems using agents. Advanced models for simulating wildfires can be built to include features such as compatibility with Geographic Information System data, dynamic changes in weather conditions, and smoke [5]. For the present work, however, a simplified version is used and shown in Fig. 6. The model includes a variety of agents, each assigned with a specific role in wildfire suppression scenarios: helicopters, UAVs and firefighters. Among these, the agents representing aircraft are of particular significance for this study, more so than those performing ground operations. The reason for this is the specific interest of the authors in generating requirements for performing ACD and systems engineering. Firefighters will remove material around the fire to stop the spread. The helicopters will assist with water drops, but they can also drop or rescue firefighters at designated points. The UAVs can fly around the map and report areas on fire or loiter over the firefighters. Both firefighters and helicopters can report random emergency calls during the simulation, such as engine problems for the helicopters or health issues for the firefighters. The map limits are $(-50, -50)$ to $(50, 50)$. Water areas (shown in blue) and wildfire starting points are generated randomly during the initialisation. It is also possible to start additional fires during the simulation to represent spotting. Time steps are taken dimensionless to simplify, and all agents move 1 position in the grid per time step when they decide to do so. A wildfire spreads every 500 steps in all directions where trees are found. Such a difference in time steps is to ensure the spread of the fire due to the lack of wind in the model, which generally drives the spread and can be approximated using a 10% rule of thumb [47]. In the simulation, everything coloured in yellow is the information the commander can request: UAV's vision and the locations of firefighters and helicopters. The only information the agents can see or report is provided if and only if an event is triggered or the LLM-commander requests it, and given through the GPT-4 web-based interface.

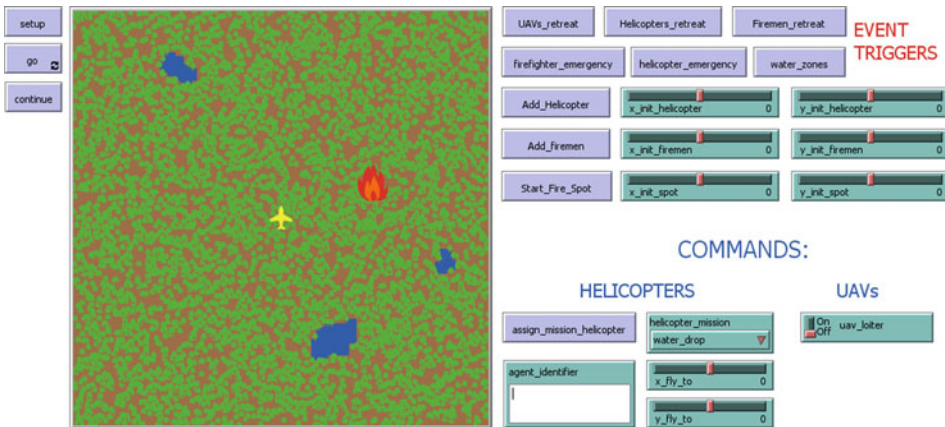


Figure 6. Agent-based model user interface.

4.6 Incorporation of large language models

The decision tree that the present work aims to replace for an LLM IC was shown in Fig. 1. For the present proof of concept, a web-based chat interface is used for the communication between LLM and the simulation. The roles of the commander and the different agents involved are taken by several instances of the LLM, whereas the information regarding the simulation situation is provided as prompts by the user.

5.0 A use case scenario implementation and evaluation

This section presents the results from the interaction with the LLM and its interactions with the wildfire simulation. Due to the length of the LLM outputs, the most relevant part of the conversation has been extracted and presented here. The ConOps from Section 4.2 are given as an initial context prompt for the commander-LLM, whereas Sections 4.3 and 4.4 are given respectively to the helicopter pilot-LLM and the firefighters-LLM. The simulation driver is the human in the loop, providing prompts to the different LLMs that respectively play the role of IC, helicopter pilot, and firefighter teams and using the output to reconfigure the simulation.

LLM-Incident Commander Reply – 0

‘Initial Briefing and Intent

Task: Our primary task is to contain and control the wildfire while ensuring the safety of all personnel involved.

Purpose: The purpose of this mission is to prevent wildfire from spreading to populated areas, protect natural resources and minimise damage to property and infrastructure.

End State: The end state of this operation is to have the wildfire fully contained with no injuries to personnel, minimal damage to property and a thorough plan for monitoring and managing hotspots.

Summary of Communication Requirements:

- Clear and complete briefings.
- Pre-operational medical emergency planning.
- Tactical pauses and reassessments.
- Establish and communicate LCES (lookouts, communications, escape routes, safety zones).
- Safe transitions between shifts.
- Disciplined radio communication.
- Standardised air traffic advisories and acknowledgements.
- Integration of risk management into all phases of operations’.

Simulation Driver Prompt – 1

‘The map of a remote area is divided into coordinates from -50 to 50 in the X axis and -50 to 50 in the Y axis. A UAV is patrolling over the area, and a wildfire is reported. It is unknown how long it has been burning and its actual size. One helicopter and two firefighter teams are available in the local agency. Please state your decision and keep consistency with the IC CONOPS’.

LLM-Incident Commander Reply – 1

‘[. . .] Summary of Actions

- UAV: Conduct initial fire location and assessment.
- Brief teams: Ensure all personnel are aware of their roles, LCES, and medical plans.
- Medical preparedness: Confirm medical personnel, equipment, and evacuation routes.
- Deploy helicopter: For reconnaissance or water drops as needed.
- Deploy firefighter teams: From strategic direction to create containment lines.
- Tactical pause: Reassess the situation around 1400 hours.
- Continuous communication: Maintain disciplined and clear radio communication’.

Simulation Driver Prompt – 2

‘The wildfire coordinates are reported to be: $(-17.88, 36.07)$. Upon checking the map of the area, the following are the coordinates of the lakes with water available for water drops: 1) Lake A: $(7, -11)$; 2) Lake B: $(-31, 47)$. The entry point to the map from the helicopter is $(-31, -50)$. Please state your decisions and how you would communicate them’.

LLM-Incident Commander Reply – 2

‘[. . .] Communication Plan:

- To Helicopter Crew: ‘Helicopter team, this is Incident Command. Enter the map at $(-31, -50)$ and proceed directly to Lake B at $(-31, 47)$ to fill the water tanks. After filling up, head to the wildfire at coordinates $(-17.88, 36.07)$ for water drops and aerial reconnaissance. Maintain communication on primary frequency Alpha and report any significant changes. Over’.
- To Firefighter Team A: ‘Team A, this is Incident Command. Proceed to coordinates $(-17.88, 50)$ and establish a line approaching the wildfire from the north. Ensure LCES is established and ready for a tactical pause at 1400 hours. Maintain communication on primary frequency Bravo and update status every 30 minutes. Over’.
- To Firefighter Team B: ‘Team B, this is Incident Command. Proceed to coordinates $(-17.88, -50)$ and establish a line approaching the wildfire from the south. Ensure LCES is established and ready for a tactical pause at 1400 hours. Maintain communication on primary frequency Bravo and update status every 30 minutes. Over’.
- To UAV Operator: ‘UAV Operator, continue monitoring the wildfire and provide real-time updates on fire behaviour, spread, and intensity. Report any critical changes immediately on primary frequency, Charlie. Over’.

[. . .]Summary

- Helicopter: Fill water at Lake B, perform water drops, and conduct aerial reconnaissance.
- Firefighter Teams A and B: Approach wildfire from the north and south, respectively, ensuring safety protocols (LCES) are in place.
- UAV: Provide continuous monitoring and updates.
- Communication: Maintain disciplined, clear, and regular communication with all teams, ensuring updates and adjustments are communicated promptly.

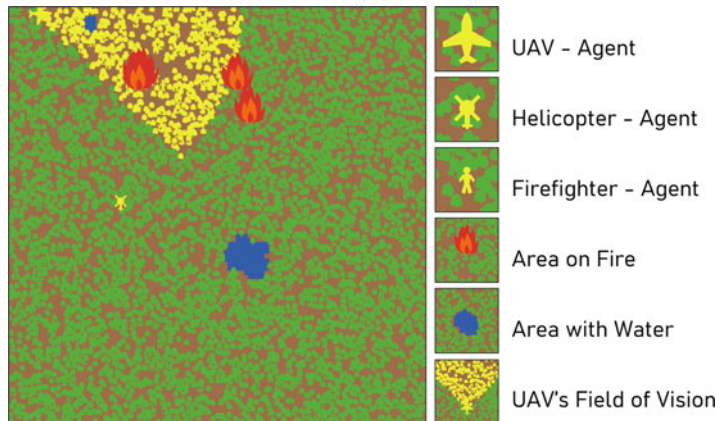


Figure 7. UAV detecting fires.

Figure 7 shows the situation after following the commander's orders. Notice the helicopter flying towards the required destination with water.

Simulation Driver Prompt – 3

The simulation continues. The helicopter pilot reports this upon reaching the wildfire: 'Alpha frequency, Helicopter six eight x-ray (68X) arrived at coordinates $(-17.88, 36.07)$ for water drops and aerial reconnaissance. Beginning operations now. Over'. At the same time, another fire spot has been reported by the UAV: $(-19, 36)$

LLM-Incident Commander Reply – 3

'Acknowledgement and Instructions to Helicopter Pilot:

- To Helicopter Pilot (68X): 'Alpha frequency, this is Incident Command. Acknowledged, Helicopter 68X. Begin water drops at coordinates $(-17.88, 36.07)$ and continue aerial reconnaissance. Be advised that a new fire spot has been reported by a UAV at coordinates $(-19, 36)$. After initial water drops, proceed to assess and report on new fire spots. Maintain visuals on both spots and update on progress. Over'.
- To Firefighter Team A: 'Bravo frequency, this is Incident Command. Be advised that a new fire spot has been reported at coordinates $(-19, 36)$ in proximity to your current target. Adjust your approach to include containment of both fire spots. Maintain LCES protocols and ensure regular status updates. Over'.
- To Firefighter Team B: 'Bravo frequency, this is Incident Command. Be advised that a new fire spot has been reported at coordinates $(-19, 36)$ near the initial fire location. Adjust your approach accordingly to assist Team A in containing both fire spots. Maintain LCES protocols and ensure regular status updates. Over'.

[...]Summary of Actions

- Helicopter 68X: Perform initial water drops, then assess and report on both fire spots.
- Firefighter Teams A and B: Coordinate to contain both fire spots with adjusted strategies.
- UAV: Provide detailed real-time updates and monitoring of fire behaviour for both spots.
- Tactical Pause: Reassess the situation at 1400 hours to ensure effectiveness and safety."

Simulation Driver Prompt – 4

Team B reports the following: 'Command, Team B, reporting a medical emergency. Firefighter experiencing heat stroke symptoms. We need immediate medical assistance at our current location. Over'.

LLM-Incident Commander Reply – 4

‘Acknowledgement and Instructions to Helicopter Pilot:

- To Helicopter Pilot (68X): ‘Alpha frequency, this is Incident Command. We have a medical emergency at Team B’s current location: a firefighter experiencing a heat stroke. Requesting immediate MEDEVAC (medical evacuation). Proceed to Team B’s coordinates. Secure a safe landing zone and prepare for transport. Over’.
- To Firefighter Team A: ‘Bravo frequency, this is Incident Command. Be advised that Team B is experiencing a medical emergency. A firefighter with heat stroke symptoms. Continue current operations, but prepare to assist with medical evacuation if required. Over’.
- To Firefighter Team B: ‘Bravo frequency, this is Incident Command. Acknowledged, Team B. Medical emergency confirmed. Immediate medical assistance will be dispatched to your current location. Ensure the firefighter is moved to a shaded area and start cooling measures immediately. Maintain hydration and monitor vital signs. Stand by for further instructions. Over’.

[. . .]Summary of Immediate Actions

- Team B: Administer first aid for heat stroke, move the firefighter to a shaded area and prepare for MEDEVAC.
- Helicopter 68X: Divert to Team B’s location for MEDEVAC, ensure safe landing and transport.
- UAV: Provide aerial monitoring and support for MEDEVAC operations.
- Team A: Continue firefighting operations, and remain on standby to assist with evacuation if needed.
- All Units: Maintain clear communication and follow updated instructions from Incident Command.

Simulation Driver Prompt – 5

The helicopter is en route to Team B for evacuation. Team A is alone without aerial support from the helicopter 68X. It is possible to ask the UAVs to loiter over the firefighters and help other agencies with an additional two helicopters and three firefighter teams.

LLM-Incident Commander Reply – 5

Summary of Updated Plan of Action

- UAV Support: Redirect UAV to provide continuous aerial support and monitoring for Team A.
- Request Additional Resources: Contact other agencies for two additional helicopters and three firefighter teams.
- Team A Operations: Continue firefighting efforts with UAV support, maintain LCES and report status every 15 minutes.
- Helicopter 68X: Complete MEDEVAC, then refill water at Lake B and resume firefighting support.
- Continuous Communication: Ensure all units maintain clear and disciplined communication, updating on status and any changes.

Simulation Driver Prompt – 6

Before this prompt, during the conversation, the LLM is informed that it can request help from other agencies. After doing so, it is informed of the following: You receive the following over the radio: ‘Command, Team A, reporting signs of severe exhaustion among crew members. Requesting immediate relief and rotation of resources. Over’. ‘Command, Team C, entering the area from point (–1,50). Over’. ‘Command, Team D, we are entering from point (–1,50). Over’. ‘Command, Team E, now entering from

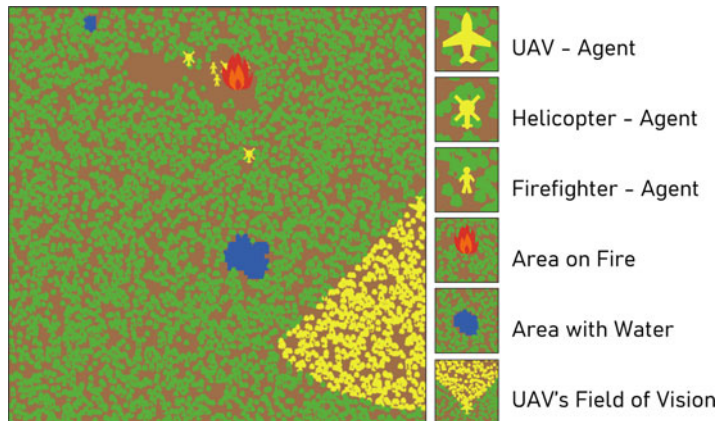


Figure 8. Wildfire under control.

point (−1,50). Over’. ‘Incident Command, this is Helicopter Five Four Zulu (54Z). Entering from point (50, 8). Ready for further instructions. Over’. ‘Incident Command, this is Helicopter two one tango (21T). Entering from point (50, 8). Ready for further instructions. Over’.

LLM-Incident Commander Reply – 6

[. . .]To all units: ‘All units, this is Incident Command. Be advised that Team A is experiencing severe exhaustion. Teams C, D and E are now entering the area for relief and rotation. Helicopters 54Z and 21T are refilling at lakes A and B, respectively, and will provide aerial support and water drops. Maintain LCES protocols and update status regularly. Over’.

Monitor and Adjust Operations as Necessary

- Team A: Provide immediate relief and ensure safe rotation with incoming teams.
- Teams C, D and E: Establish LCES, focus on their assigned fire spots and ensure a smooth transition.
- Helicopters 54Z and 21T: Conduct water drops and provide aerial reconnaissance.
- UAV: Monitor the situation and provide real-time updates.

Summary of Updated Plan of Action

- Team A: Immediate relief by Teams C, D and E. Ensure safe rotation and rest.
- Teams C, D and E: Take over firefighting operations with clear focus areas and support.
- Helicopter 54Z: Refill at Lake A, conduct water drops and provide reconnaissance.
- Helicopter 21T: Refill at Lake B, conduct water drops and provide reconnaissance.
- UAV Support: Monitor relief operations and fire behaviour, providing real-time updates.
- Continuous Communication: Maintain clear and regular communication with all units, updating on the status and any changes.

Figure 8 show the final stages of the operation. The wildfire is being reduced with the different agents around it, continuing until the end of the simulation when the wildfire is completely extinguished. Notice the area where the vegetation has been removed by either fire or the work of the agents.

6.0 Discussion

In this section, the results are discussed in terms of the performance of the LLM as IC, and in terms of operational capabilities for the generation of system requirements.

6.1 Performance of the LLM-IC

The simulation results highlight several key aspects of the LLM's behaviour and decision-making in wildfire scenarios. There are, however, issues to be addressed using the information from Section 4.1. The LLM effectively tracked the agents' actions and whereabouts, and it produced coherent and consistent intent, and disciplined radio formats, indicating that an LLM could replace, at least partially, some decision trees while preserving a correct operational structure for an SoS. For the UAV assets, coexistence rules in manned airspace were left implicit, as one of the goals of SoS analysis is to study the impact and conflict of new systems towards product development. However, as the interviewees stated, they populate a reserved airspace. Thus, an initial draft of both possible conflicts and ConOps should be created and included in the initial prompts for the UAV coexistence with manned aircraft. A conservative baseline is to have no mixed operations unless unavoidable, with strict vertical separation in such cases. The current use of the airspace without explicit rules is dangerous and should be corrected via new policies and explicit rules in the prompts. An example in the results was the IC decision of assigning a UAV from surveillance to loitering over Team A for aerial support. It also demonstrated adaptability, but there was no consideration of the potential risks.

The situational awareness was not sufficient: current wind conditions and visibility, or the presence of ground units before drops, were neglected, or even the fire's magnitude. Even though the LLM adhered to scheduled weather checks, confirmation of the availability of water loading sites or terrain escape routes was neglected. These checks should be explicit preconditions in the LLM context prompts and checks for the SoS state.

During the emergency evacuation, radio silence and clearing the airspace were omitted. Interviews emphasised the figure of a central aerial coordinator. Aerial communications should be organised around this figure for the aerial units, differentiating clearly from the IC. This figure should have multi-frequency capability, standardised phraseology, and the ability to impose radio silence and clear airspace. An unexpected behaviour was the LLM's use of the NATO phonetic alphabet ('Tango', etc.) for naming, demonstrating the value of diverse training for enhancing interoperability and communication across contexts. However, LLM directly utilised all resources available, indicating a focus on maximising effectiveness, possibly due to a lack of cost information. Although it is considered acceptable, since literature suggests that undermanned fire operations are more costly than overmanned ones [48], some reasoning was expected.

Finally, additional realism is necessary in the model. Beyond spotting, the interviewees identified as important: wind shifts and their intensity, temperature variations, the relative humidity of the air and fuel and orography (steep slopes or valleys) that both shape fire behaviour and constrain heavier aircraft manoeuvres. These should be treated as stochastic sources that constrain aircraft operational modes (e.g. approach or hover with external loads).

6.2 Hallucination issues

An emergent by-product of using LLMs is the apparition of hallucinations in the generated replies. An accepted definition of hallucination is a generated content that deviates from the real facts, resulting in unfaithful outputs, with some categories proposed to classify these hallucinations [49]:

- **Numeric Nuisance:** numeric values are generated that are inconsistent with facts.
- **Acronym Ambiguity:** incorrect expansion of acronyms.
- **Generated Golem:** an imaginary personality is fabricated.
- **Geographic Erratum:** generated locations do not correspond to the associated event.
- **Time Wrap:** a generated text combining two different timelines.

The obtained results presented hallucinations that should be addressed:

- Considered a Numeric Nuisance was the tactical pause scheduled at 1400 hours. The period between 1400 and 1700 hours is considered critical, and a re-evaluation of the situation is indeed

Table 1. Helicopter Concept 1: Firefighting helicopter

Capability	Requirement
Adaptability	Modular design for new technologies' integration and easier manufacturing. Capable of handling various weather conditions and terrains.
Communication	Advanced communication systems for interoperability with ground units and other aircraft. Encrypted communication.
Environmental assessment	Onboard sensors and cameras for real-time fire and weather monitoring. Assessment of terrain features, vegetation types and weather patterns. Integration of data for strategic planning.
Operational efficiency	Extended operational periods. High hover stability and precision in water drop targeting. Advanced targeting and navigation systems. Large water carrying capacity (min. 2000 litres) with quick refill capability.
Safety and emergency response	Fire-resistant features to avoid degradation. Automatic fire suppression systems. Real-time status monitoring of critical systems.
Autonomy	Semi-autonomous hovering capabilities with autopilot systems for precision operations. Real-time obstacle detection and avoidance systems. Flight path optimisation for quicker turnarounds.

a good idea, but no tactical pause was ever requested, nor was there any prompt whatsoever. This is considered a numeric detail that can lead to wrong planning of the SoS tasks or life cycle.

- Frequencies used (Alpha/Bravo/Charlie) are assigned arbitrarily, which drifts from the protocols given, being close to Acronym Ambiguity. This may result into misleading or confusing communications.
- The helicopter assigned for medical evacuation is considered a Generated Golem, as the capability was fabricated since the status of the crew or equipment was not clearly declared nor requested.
- The instant availability of the assets added (two helicopters and three ground teams) is considered a Time Wrap hallucination, as they are integrated into the SoS without proper inclusion into the operation planning.

6.3 Operational capabilities

The simulation results must also be discussed in terms of ACD. From the operational capabilities used to fulfil the mission, the top-level requirements are to be extracted for generating new system concepts and adding them to those that could have been given by stakeholders. Four different concepts of aircraft could be proposed, as shown in Tables 1–4, using the results of the simulation. The vehicles need to be able to handle different weather conditions and terrains, since the location of a wildfire is going to be unknown a priori. In terms of collaboration, the various constituent systems must include communication subsystems, which are crucial from an SoS point of view. The communication capabilities must safely ensure interoperability, whether for requesting evacuation or informing about new issues, as observed in the simulation when one team was in distress or the new fire spot started. Another requirement identified is the inclusion of different types of sensors and cameras to provide enhanced situational awareness throughout the mission, which was the main task of the UAV when patrolling the perimeter of the map and supporting the firefighters.

Specific requirements extracted for helicopters, shown in Tables 1 and 2. The amount of water to be carried will depend on the system being specific for water drops or for a multi-role one. The requirements

Table 2. *Helicopter Concept 2: Multi-role support helicopter*

Capability	Requirement
Adaptability	Modular design for new technologies' integration and easier manufacturing. Easy firefighting and search and rescue reconfiguration. Capable of handling various weather conditions and terrains.
Communication	Advanced communication systems for interoperability with ground units and other aircraft. Encrypted communication.
Environmental assessment	Onboard sensors and cameras for real-time fire and weather monitoring. Assessment of terrain features, vegetation types and weather patterns. Integration of data for strategic planning.
Operational efficiency	Extended operational periods. Quick deployment and retrieval of rescue teams and equipment. Advanced navigation and targeting systems. Moderate water carrying capacity (max. 1000 litres) and additional cargo space for equipment.
Safety and emergency response	Advanced medical support systems, and easy access for emergency evacuations with a stretcher. Fire-resistant features to avoid degradation. Automatic fire suppression systems. Real-time status monitoring of critical systems.
Autonomy	Enhanced navigation systems (obstacle detection and avoidance). Automated flight paths and water drops for firefighting. Real-time assisted decision-making based on data from sensors and cameras. Capability to switch between manual and autonomous modes as needed.

Table 3. *UAV Concept 1: Fixed-wing configuration*

Capability	Requirement
Adaptability	Modular design for new technologies' integration and easier manufacturing. Easy firefighting and search and rescue reconfiguration. Capable of operating in various weather conditions and in diverse terrains.
Communication	Advanced communication systems for interoperability with ground units and other aircraft. Encrypted communication.
Environmental assessment	Onboard sensors and cameras for real-time fire and weather monitoring. Assessment of terrain features, vegetation types and weather patterns.
Operational efficiency	Extended operational periods (minimum 10 hours) for both firefighting and rescue missions. Rapid deployment. Minimisation of maintenance downtime. Advanced navigation and targeting systems for precise operations. Flight path optimisation.
Safety and emergency response	Fire-resistant features to avoid degradation. Emergency protocols for automated return or safe landing. Real-time status monitoring of critical systems.
Autonomy	Fully autonomous operation with advanced mission planning and execution capabilities. Autonomous navigation and obstacle detection for search and rescue missions. Ability to switch between autonomous and manual control as needed.

Table 4. UAV Concept 2: Tilt-rotor configuration

Capability	Requirement
Adaptability	Modular design for new technologies' integration and easier manufacturing. Easy firefighting and search and rescue reconfiguration. Capable of operating in various weather conditions and in diverse terrains. Capable of vertical take-off and landing (VTOL) and transitioning to fixed-wing flight.
Communication	Advanced communication systems for interoperability with ground units and other aircraft. Encrypted communication.
Environmental assessment	Onboard sensors and cameras for real-time fire and weather monitoring. Assessment of terrain features, vegetation types and weather patterns.
Operational efficiency	High manoeuvrability with the ability to hover and perform precise drops. Rapid deployment. Minimisation of maintenance downtime. Advanced navigation and targeting systems for precise operations. Flight path optimisation.
Safety and emergency response	Fire-resistant features to avoid degradation. Emergency protocols for automated return or safe landing. Real-time health monitoring of critical systems.
Autonomy	Fully autonomous operation with advanced mission planning and execution capabilities. Autonomous navigation and obstacle detection for search and rescue missions. Ability to switch between autonomous and manual control as needed.

for a multi-role will reduce the size of the water bucket in exchange for being able to land in forestry areas to transport or evacuate people, as happened during the simulation. Equation (3) can be used for initial predictions related to the helicopter rotor thrust needed for these different operational situations. For example, a smaller rotor radius will be convenient to be able to land in forestry areas and avoid collisions, but the performance will differ from a helicopter with a bigger radius and specifically for water drops.

For the case of UAVs, Equations (11) and (12) can be used together with common weight estimation methods during the phase of generating the concept of an aircraft [50]. These predictions are relevant because, if the UAV is requested to change its mode from surveillance to support, it will fly at a slower velocity and, depending on the needs, being able to either loiter or hover could make a difference. Hence, considering fixed-wing or tilt-rotor configurations is relevant for proposing new concepts, as well as the prediction from Equation (12). The reason is that the fixed-wing concept can be designed with endurance as the main goal, but when in loitering mode, its capabilities to support firefighters could be insufficient. Whereas a tilt-rotor concept will trade an enhanced loitering or hovering capability for reduced endurance. The process of requirements decomposition involves breaking down high-level stakeholder expectations into detailed requirements across the system hierarchy. In this context, the LLM's performance as an IC in the simulation adds valuable insights for the decomposition process. Its inclusion in a simulation allowed for the dynamic generation of operational scenarios that reflect real-world complexities better than predefined decision trees. By observing the LLM's decisions, it is possible to derive functional requirements that may not have been explicitly identified by stakeholders. Some of the LLM's decisions, such as diverting a helicopter for evacuation or requesting reinforcements during the operation, came from certain capabilities that are directly related to the operational requirements of the different aircraft. These operational requirements are satisfied by functionalities that emerge at the aircraft level and are accounted for when generating the concepts [51]. Since the procedure for including an LLM in a simulation is still under development, a proposed workflow for the analysis could be:

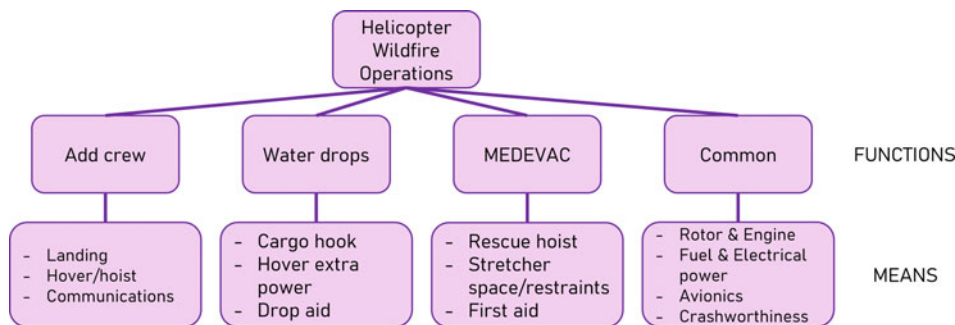


Figure 9. Example of function/means tree for a firefighting helicopter.

1. **Capture the episode** such as a concise statement of the operational context, agents involved and possible constraints; for example, emergency evacuation of ground units over a wooded area.
2. **Express as an emergent function** using a function/means structure (e.g., *MEDEVAC*).
3. **Formulate requirements** in capability terms, such as adaptability, communication, environmental assessment, operational efficiency, safety and emergency response, autonomy.

It is important to note that, for avoiding conflictive functionalities, the different modes or states must be analysed at the operational level for both the SoS and single systems. For the SoS, the operation changed from a single into a multi-aircraft operation, whereas for the single system, the firefighting mode of a helicopter switched into evacuation. This leads to conflict analysis, hover margins, etc. In addition, the simulated conditions should change to be able to consider multiple scenarios, such as having many fire spots, different elevations, no water in the vicinity, etc.

6.4 Functional analysis

The results obtained provided interactions that are also useful for logical decomposition and functional analysis, which can be done with the use of function/means trees [52] and ontologies [53]. The process involves translating actions into system functions and sub-functions and the means for achieving them, as shown in Fig. 9. Among the different interactions, as an example of the decomposition process, it can be taken the high-level functionality: helicopter wildfire operations. This can be broken into:

- **Sub-Function 1:** Facilitate emergency evacuation when required.
- **Means:** Rescue hoist, space for a stretcher, first aid resources.

Each sub-function can be allocated to specific subsystems, and their interfaces and interactions can be defined. This hierarchical breakdown provides an understanding of all system functions and their relationships, necessary for developing a coherent system architecture. At each level of decomposition, the derived requirements must be validated against higher-level requirements and stakeholder expectations. For example, the LLM's oversight in assessing the fire's magnitude indicates a gap in situational awareness capabilities, which addresses the need for real-time fire monitoring and data integration.

7.0 Conclusion

The present work has shown that behaviours modelled with long decision trees can be replaced, at least partially, with an LLM in an ABS. This provides more degrees of freedom for the simulation and extends the analysis of SoS. The ABS was used to evaluate the LLM's decision-making and its capability to follow operational guidelines as an IC, with a satisfactory outcome. Moreover, from the decisions taken, it was possible to evaluate what kind of capabilities would be required from the constituent systems that are part of the SoS. The operational capabilities found have been related to requirements that can be used for ACD and generate different aircraft concepts, such as helicopters or UAVs.

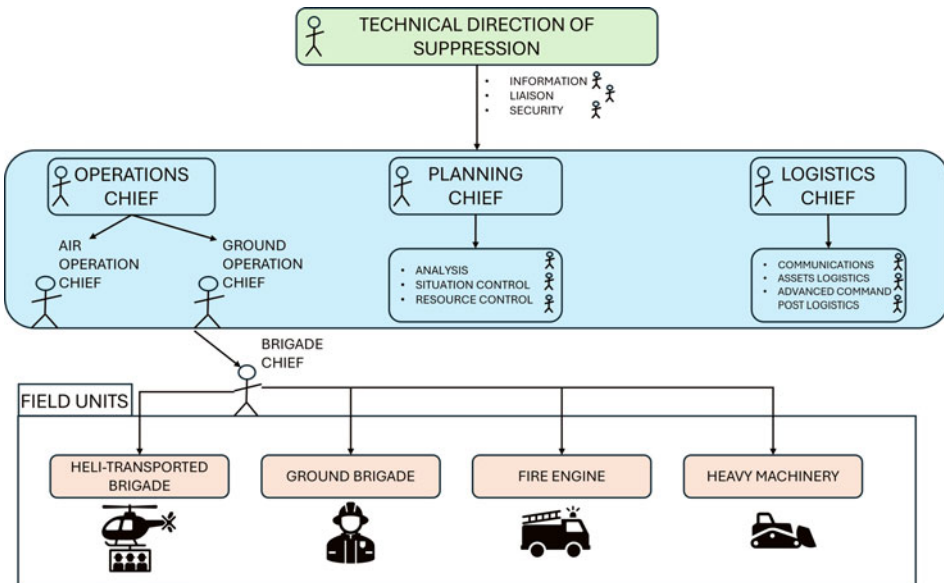


Figure 10. Incident Command structure for wildfire suppression. Adapted from (56).

Future work will be divided into two different paths for different goals:

- First the issues related to the inclusion of an LLM. Hallucination was already discussed, but other gaps need to be researched. Literature has proposed methods for benchmarking [49], or mitigating hallucinations [54], being especially interesting is the addition of another LLMs as a means for mitigation. Since the figure of an aerial commander was mentioned in the interviews, it seems particularly relevant as a means of achieving both an improved simulation in terms of realism and in terms of reduced issues related to hallucinations. A particularly intriguing configuration would be in such a way that both have access to the same information, similar to radio communication, but with the added benefit of hallucination checking. For this setup in the simulation, it is distinctly interesting to use lightweight LLMs, as they were reported to be less prone to hallucination issues [49], they require less computational power and are easier to fine-tune [55]. As another long-term goal, it is important to study several configurations of the LLMs and their temperature parameter (which adjusts the level of determinism in the answers). Being able to reproduce results is important, especially when designing experiments, and certain parameters must be kept constant. However, the authors consider highly relevant the variability in the answers for the analysis of an SoS, since the resilience of it to errors or misunderstandings is considered very important for finding an optimal configuration. Human mistakes can always happen during an operation, and they provide valuable information for understanding an SoS, designing the ConOps, or finding the configuration with the best performance.
- The initial concept proposed for the case study has brought important insights, but it was clearly insufficient. The future work for the model needs to focus on improving the model fidelity and the overall architecture of the SoS. The feedback from the interviewees emphasised the importance of capturing not only fire dynamics but also wind conditions, terrain and smoke. This low fidelity affects the operation and does not allow for proper analysis, especially in terms of commander decisions. No escape routes for ground units could be decided upon, nor could dropping paths be planned for avoiding smoke and mountains. It was also found important the role of an air operation chief [56], as shown in the structure in Fig. 10, particularly relevant when designing the SoS architecture. Once a proper model is achieved, an evaluation can be done by using performance indicators. However, before considering any indicators for the SoS or the systems in it,

experts should grade the decisions and behaviour of the agents. In doing so, it would be important for the evaluation of the performance to consider decisions related to tactics, fire behaviour evaluation, water drop planning, resource distribution, ground-aircraft collaboration and radio communication.

Acknowledgements. The research presented in this paper has been performed in the framework of the COLOSSUS project (Collaborative System of Systems Exploration of Aviation Products, Services and Business Models) and has received funding from the European Union Horizon Europe programme under grant agreement No. 101097120. The Swiss participation in the COLOSSUS project is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00609. The authors would also like to acknowledge the Swedish Innovation Agency (VINNOVA) and The Swedish Defence Materiel Administration (FMV) for financial support through the grant 2019-05371. The authors also wish to acknowledge the help of the interviewed special operation pilots J. M. García and D. Moreno Gonzalez. ChatGPT is a language model developed by OpenAI in San Francisco, CA, USA, and is used to partly generate the conversation between the Incident Commander and the agents in the simulation.

References

- [1] Wintle, B.A., Legge, S. and Woinarski, J.C.Z. After the megafires: What next for australian wildlife? *Trends Ecol. Evol.*, 2020, **35**, (9), pp 753–757.
- [2] Gallagher, R.V., Allen, S., Mackenzie, B.D.E., Yates, C.J., Gosper, C.R., Keith, D.A., Merow, C., White, M.D., Wenk, E., Maitner, B.S., et al. High fire frequency and the impact of the 2019–2020 megafires on australian plant diversity, *Diversity Distrib.*, 2021, **27**, (7), pp 1166–1179.
- [3] Staack, I., Amadori, K. and Jouannet, C. A holistic engineering approach to aeronautical product development, *Aeronaut. J.*, 2019, **123**, (1268), pp 1545–1560.
- [4] Raymer, D. *Aircraft Design: A Conceptual Approach*. American Institute of Aeronautics and Astronautics, Inc., 2012.
- [5] Lovaco, J., Staack, I. and Krus, P. Environmental agent-based modelling for a firefighting system of systems, *33rd Congress of the International Council of the Aeronautical Sciences (ICAS)*, Stockholm, Sweden, ICAS, 2022, pp 1–12.
- [6] Finney, M.A. *FARSITE, Fire Area Simulator–Model Development and Evaluation*, vol. 4, The Station, 1998.
- [7] Rothermel, R.C. *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*, vol. 115, Intermountain Forest & Range Experiment Station, Forest Service, 1972, US.
- [8] Forestry Canada Fire Danger Group. Development and structure of the canadian forest fire behavior prediction system. *Information Report ST-X-3*, 1992.
- [9] Tymstra, C., Bryce, R.W., Wotton, B.M., Taylor, S.W., Armitage, O.B., et al. Development and structure of prometheus: the Canadian wildland fire growth simulation model, *Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Information Report NOR-X-417*. (Edmonton, AB), 2010.
- [10] Alexander, M.E. and Cruz, M.G. Are the applications of wildland fire behaviour models getting ahead of their evaluation again? *Environ. Model. Software*, 2013, **41**, pp 65–71.
- [11] Ramesh, M., Sun, Z., Li, Y., Zhang, L., Annam, S.K., Fang, H. and Tong, D. Assessing wildfiregpt: a comparative analysis of Ai models for quantitative wildfire spread prediction, *Nat. Hazards*, 2025, pp 1–14.
- [12] Chen, M., Tao, Z., Tang, W., Qin, T., Yang, R. and Zhu, C. Enhancing emergency decision-making with knowledge graphs and large language models, *Int. J. Disaster Risk Reduct.*, 2024, **113**, p 104804.
- [13] Vrdoljak, J., Boban, Z., Males, I., Skrabic, R., Kumric, M., Ottosen, A., Clemencau, A., Bozic, J. and Völker, S. Evaluating large language and large reasoning models as decision support tools in emergency internal medicine, *Comput. Biol. Med.*, 2025, **192**, p 110351.
- [14] Odubola, O., Adeyemi, T.S., Olajuwon, O.O., Iduwet, N.P., Aniekan, A.I. and Odubola, T. Ai in social good: Llm powered interventions in crisis management and disaster response, *J. Artif. Intell. Mach. Learn. Data Sci.*, 2025, **3**, (1), pp 3353–3360.
- [15] Boone, T., Fahimnia, B., Ganeshan, R., Herold, D.M. and Sanders, N.R. Generative AI: Opportunities, challenges, and research directions for supply chain resilience, *Transp. Res. Part E: Logist. Transp. Rev.*, 2025, **199**, p 104135.
- [16] Hirshorn, S.R., Voss, L.D. and Bromley, L.K. *Nasa systems engineering handbook*. Technical report, National Aeronautics and Space Administration, 2017.
- [17] Lane, J.A. and Epstein, D. What is a system of systems and why should i care? *University of Southern California*, 2013.
- [18] Acheson, P., Dagli, C. and Kilicay-Ergin, N. Model based systems engineering for system of systems using agent-based modeling, *Procedia Comput. Sci.*, 2013, **16**, pp 11–19.
- [19] Collins, A.J. and Etemadidavan, S. Interactive agent-based simulation for experimentation: a case study with cooperative game theory, *Modelling*, 2021, **2**, (4), pp 425–447.
- [20] Maier, M.W. Architecting principles for systems-of-systems, *INCOSE International Symposium*, 1996, **6**, (1), pp 565–573.
- [21] Franzén, L.K. *An Ontological and Reasoning Approach to System of Systems*. PhD thesis, Linköping Studies in Science and Technology. PhD Thesis, 2021.
- [22] Shannon, C.E. A mathematical theory of communication, *Bell Syst. Techn. J.*, 1948, **27**, (3), pp 379–423.
- [23] De Mulder, W., Bethard, S. and Moens, M.-F. A survey on the application of recurrent neural networks to statistical language modeling, *Comput. Speech Lang.*, 2015, **30**, (1), pp 61–98.

- [24] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners, *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp 1877–1901.
- [25] Luzeaux, D. and Ruault, J.-R. *Systems of Systems*, John Wiley & Sons, 2013.
- [26] Hornik, K., Stinchcombe, M. and White, H. Multilayer feedforward networks are universal approximators, *Neural Netw.*, 1989, **2**, (5), pp 359–366.
- [27] Hornik, K., Stinchcombe, M. and White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Netw.*, 1990, **3**, (5), pp 551–560.
- [28] Goodfellow, I., Bengio, Y. and Courville, A. *Deep Learning*, MIT Press, 2016.
- [29] Department of the Army, U.S.A. *Joint Publication 2-0: Joint Planning*. Department of the Army, 10 2013.
- [30] Department of the Army, U.S.A. *Joint Publication 5-0: Joint Planning*. Department of the Army, 12 2020.
- [31] Glauert, H. *Airplane Propellers*, Springer Berlin Heidelberg, 1935, Berlin, Heidelberg, pp 169–360. ISBN 978-3-642-91487-4.
- [32] Johnson, W. *Helicopter Theory*, Courier Corporation, 2012.
- [33] Leishman, G.J. *Principles of Helicopter Aerodynamics with CD Extra*, Cambridge University Press, 2006.
- [34] Kee, S.G. Guide for conceptual helicopter design. Technical memorandum, Naval Postgraduate School, CA 93940, USA, June 1983.
- [35] Johnson, W., Withrow-Maser, S., Young, L., Malpica, C., Koning, W.J.F., Kuang, W., Fehler, M., Tuano, A., Chan, A., Datta, A., et al. Mars science helicopter conceptual design. Technical Memorandum NASA/TM-2020-220485, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA 94035-1000, USA, March 2020.
- [36] Johnson, W. and Sinsay, J.D. Rotorcraft conceptual design environment, *The 3rd International Basic Research Conference on Rotorcraft Technology*, 2009.
- [37] Anderson, J.D. *Fundamentals of Aerodynamics*, McGraw Hill, 2011.
- [38] Desai, A. and Mital, A. *Sustainable Product Design and Development*, CRC Press, 2020.
- [39] Ellis, M., Bojdo, N., Filippone, A. and Clarkson, R. Monte Carlo predictions of aero-engine performance degradation due to particle ingestion, *Aerospace*, 2021, **8**, (6), p 146.
- [40] Lane, J.A. System of systems capability to requirements engineering, 2014 *9th International Conference on System of Systems Engineering (SOSE)*, IEEE, 2014, pp 91–96.
- [41] Blessing, L.T.M. and Chakrabarti, A. *DRM, a Design Research Methodology*, Springer, 2009.
- [42] McGowan, A.-M.R., Bakula, C. and Castner, R.S. Lessons learned from applying design thinking in a NASA rapid design study in aeronautics, *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2017, p 0976.
- [43] National Wildfire Coordinating Group. *Incident Management Field Guide*, 2014.
- [44] Department of the Interior, U.S.A. *Basic Aviation Safety*. Office of Aviation Services, 5 2013.
- [45] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. Gpt-4 technical report. arXiv preprint arXiv:2303.08774, 2023.
- [46] Tisue, S. and Wilensky, U. Netlogo: A simple environment for modeling complexity, *International Conference on Complex Systems*, vol. 21, Citeseer, 2004, pp. 16–21.
- [47] Cruz, M.G. and Alexander, M.E. The 10% wind speed rule of thumb for estimating a wildfire’s forward rate of spread in forests and shrublands, *Ann. Forest Sci.*, 2019, **76**, (2), p 44.
- [48] Parks, G.M. Development and application of a model for suppression of forest fires, *Manag. Sci.*, 1964, **10**, (4), pp 760–766.
- [49] Rawte, V., Chakraborty, S., Pathak, A., Sarkar, A., Islam, T., Tonmoy, M.S., Chadha, A., Sheth, A. and Das, A. The troubling emergence of hallucination in large language models-an extensive definition, quantification, and prescriptive remediations. Association for Computational Linguistics, 2023.
- [50] Horvath, B.L. and Wells, D.P. Comparison of aircraft conceptual design weight estimation methods to the flight optimization system, *2018 AIAA Aerospace Sciences Meeting*, 2018, p 2032.
- [51] Drego, A. and Steinkellner, S. A system integrator’s perspective on vehicle system evaluation at the aircraft concept stage, *33rd Congress of the International Council of the Aeronautical Sciences*, 2022.
- [52] Vigolo, V., Muñoz, K. and De Negri, V.J. Conceptual design of mechatronic systems based on function/means tree, *2021 9th International Conference on Traffic and Logistic Engineering (ICTLE)*, IEEE, 2021, pp 108–113.
- [53] Franzén, L.K., Staack, I., Krus, P., Jouannet, C. and Amadori, K. Ontology-represented design space processing, *AIAA AVIATION 2021 FORUM*, 2021, p 2426.
- [54] Verspoor, K. ‘fighting fire with fire’—using llms to combat llm hallucinations, 2024.
- [55] Zhalgasbayev, A., Khauazkhan, A. and Sarsenova, Z. Fine-tuning the Gemma model for Kaggle assistant, *2024 IEEE AITU: Digital Generation*, IEEE, 2024, pp 104–109.
- [56] GRUPO DE TRABAJO DE SISTEMA DE MANDO DE INCIDENTES Y ACREDITACIÓN DE LA FORMACIÓN. *ESTRUCTURA ORGANIZATIVA DEL SISTEMA DE GESTIÓN DE EMERGENCIAS EN INCENDIOS FORESTALES*. Ministerio para la Transición Ecológica y el Reto Demográfico, 5 2024.