

# AERoS: Assurance of Emergent Behaviour in Autonomous Robotic Swarms

Dhaminda B. Abeywickrama<sup>†1</sup>, James Wilson<sup>†1</sup>, Suet Lee<sup>1</sup>,  
Greg Chance<sup>1</sup>, Peter D. Winter<sup>1</sup>, Arianna Manzini<sup>1</sup>, Ibrahim Habli<sup>2</sup>,  
Shane Windsor<sup>1</sup>, Sabine Hauert<sup>1</sup>, and Kerstin Eder<sup>1</sup>

<sup>1</sup> University of Bristol, Bristol, UK  
`{firstname.lastname}@bristol.ac.uk`

<sup>2</sup> University of York, York, UK  
`ibrahim.habli@york.ac.uk`

**Abstract.** The behaviours of a swarm are not explicitly engineered. Instead, they are an emergent consequence of the interactions of individual agents with each other and their environment. This emergent functionality poses a challenge to *safety assurance*. The main contribution of this paper is a process for the safety assurance of emergent behaviour in autonomous robotic swarms called *AERoS*, following the guidance on the Assurance of Machine Learning for use in Autonomous Systems (AM-LAS). We explore our proposed process using a case study centred on a robot swarm operating a public cloakroom.

**Keywords:** Assurance · Safety · Emergent behaviour · Guidance · Swarms.

## 1 Introduction

Swarm robotics provides an approach to the coordination of large numbers of robots inspired by swarm behaviours in nature. The overall behaviours of a swarm are not explicitly engineered in the system [11, 1]. Instead, they are an emergent consequence of the interactions of individual agents with each other and the environment, we call this emergent behaviour (EB) [1]. EB can be difficult to model and characterize and this property of swarms poses a critical challenge to *assurance* [2]. Assurance tasks comprise conformance to standards, verification and validation (V&V), and certification. Assurance criteria for autonomous systems (AS) include both functional and non-functional requirements such as safety [2].

Existing standards and regulations of AS are either implicitly or explicitly based on the V lifecycle model [4]. However, this model is unlikely to be suitable for systems with EB; for example through interaction with other agents and the environment, as is the case with swarms. ISO standards have been developed for the service robotics sector (non-industrial) (e.g. ISO 13482, ISO 23482-1), and the industrial robotics sector (e.g. ISO 10218-1, ISO/TS 15066) [1]. However, although these standards focus on ensuring the assurance of robots at the individual level, they do not cover safety at the *swarm* level that may arise through EB.

**Related Work** There are several standards and guidance related to machine learning (ML) in aeronautics, automotive, railway and industrial domains [8], for example the guidance on the Assurance of Machine Learning for use in Autonomous Systems (AMLAS) [3], the European Union Aviation Safety Agency (EASA) concept paper, the DEpendable and Explainable Learning (DEEL) white paper, the Aerospace Vehicle System Institute (AVSI) report, the Laboratoire National de Métrologie et d’Essais (LNE) certification, and the UL 4600 standard. However, none of these approaches targets robot swarms. In this work we used AMLAS [3] as the foundation for developing an assurance process for autonomous robotic swarms. AMLAS provides guidance on how to systematically integrate safety assurance into the development of the ML components based on offline supervised learning. AMLAS contains six stages where assurance activities are performed in parallel to the development activities. AMLAS has the advantage of ensuring safety assurance for complex AS where the behaviour of the system is controlled by ML algorithms. Our study takes inspiration from AMLAS, but adapts it to focus on emergence as the driver for complexity, rather than learning.

**Main Contribution and Paper Outline** In this work, we propose a process for the safety assurance of EB in autonomous robotic swarms (AERoS), adapted from AMLAS [3]. AERoS covers six EB lifecycle stages: safety assurance scoping, safety requirements elicitation, data management, model EB, model verification, and model deployment. The AERoS process is domain independent and therefore can be applied to any swarm type (e.g. grounded, airborne). In particular, a system of components can be considered as analogous to a swarm if the system is decentralized with local interactions between components. AERoS may be applicable in such cases as well. In this paper, we explore the AERoS process using a case study centered on a robot swarm operating a public cloakroom at events with 50 to 10000 attendees [5]. In the cloakroom, a swarm of robots assist attendees to deposit, store, and deliver their belongings (e.g. jackets) [5]. As the swarm operates in a public setting, the system must prioritise public safety (e.g. avoid harm from collisions, fire hazards caused by blockages). The rest of the paper is organised as follows. In Section 2, we discuss the six stages of AERoS. Section 3 provides a brief discussion and concludes the paper.

## 2 The AERoS Process

This section discusses the six main stages of AERoS targeting autonomous robot swarms. AERoS is iterative by design, and the assurance activities are performed in parallel to EB development (see Fig. 1). For each stage, we describe its inputs and outputs, main assurance activities and their associated artefacts.

### 2.1 Stage 1: EB Safety Assurance Scoping

Stage 1 contains two activities which are performed to define the safety assurance scope for the swarm (see Fig. 2).

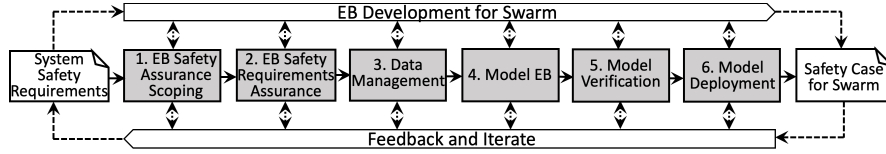


Fig. 1. The AERoS process with the six stages adapted from AMLAS.

**Activity 1. Define Assurance Scope for the EB Description and Expected Output** The goal of Activity 1 (Fig. 2) is to define the safety assurance scope for the EB and expected output. The requirements defined in this stage are independent of any EB technology, which reflects the need for the robot swarm to perform safely regardless of emergence.

*[A] System Safety Requirements:* The system safety assessment process generates the safety requirements of the swarm, and covers the identification of hazards (e.g. the blocking of critical paths in the cloakroom) and risk analysis. Figure 3 illustrates how individual robot failures propagate through the neighbourhood to swarm-level hazards: we can then derive safety requirements in the form of concrete failure conditions at the level of the whole swarm which capture, implicitly, all levels of the swarm.

*[B] Environment Description:* It is essential to consider the system environment when allocating safety requirements to the swarm. In the cloakroom, a swarm of robotic agents collects and delivers jackets, which are stored in small box-like containers. The agents are required to navigate a public space between collection and delivery points.

*[C] System Description:* In the cloakroom, we can consider three inputs: sensor availability, neighbourhood data (used because there is no access to global data in real-world deployments), and swarm parameters (see Fig. 4). The *sensors* available to agents can be cameras and laser time-of-flight sensors. The neighbourhood data of the swarm can be specified through the communication systems available to agents, in this case Bluetooth. Through the use of this short-range communication, agents can access neighbourhood data, such as approximate position or current state of local agents. As for the swarm-level parameters, we can consider options specified by a user, that is, the number of agents deployed, and the maximum speed of agents. Once defined, the three inputs are then fed to the individual agents to instruct their behaviour. This behaviour enacted by

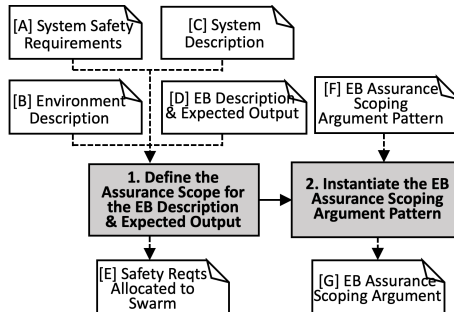


Fig. 2. Stage 1: The AERoS emergent behaviour assurance scoping process.

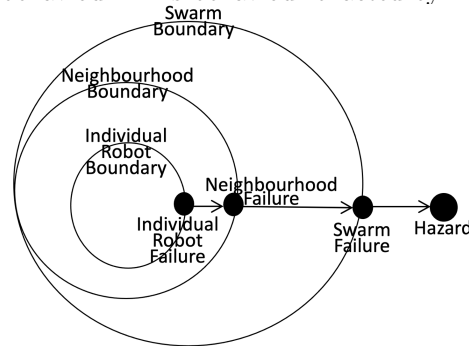
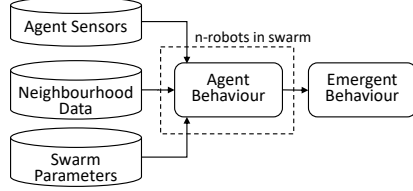
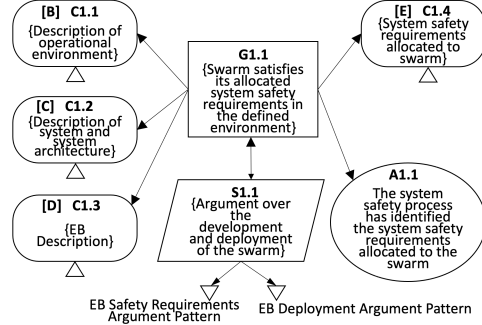


Fig. 3. Failure conditions in a swarm adapted from DO-178C and AMLAS.



**Fig. 4.** Inputs fed into individual agent behaviour producing overall swarm emergent behaviour.



**Fig. 5.** Emergent behaviour safety assurance scoping argument pattern.

multiple agents then produces a swarm-level EB as the individuals interact with one another and their environment.

*[D] EB Description and Expected Output:* By expected output, we refer to the gains that can arise from the system by deploying multiple agents. In the cloakroom, the output is a collaborative system capable of collecting, sorting, and delivering jackets in a public setting. To achieve this, the EB of the system needs to arise from the available behaviours of the individual agents, their interactions, and the constraints outlined in the system description.

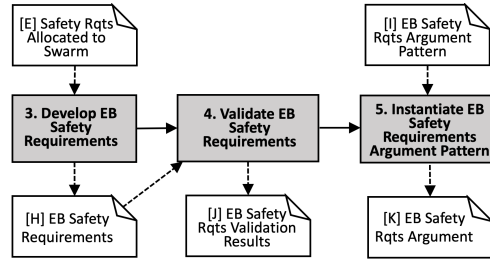
**Activity 2. Instantiate EB Assurance Scoping Argument Pattern** Each stage of the AERoS process includes an activity to instantiate a safety argument pattern based on the evidence and artefacts generated in that stage. *Argument patterns* [3], which are modelled using the Goal Structuring Notation, can be used to explain the extent to which the evidence supports the relevant EB safety claims. In Activity 2, we use the artefacts generated from Stage 1 (i.e. [A–E]) to instantiate the EB Assurance Scoping Argument Pattern ([F] – see Fig. 5). The instantiated argument [G] along with other instantiated arguments resulting from the other five stages of AERoS constitute the safety case for the swarm. The activities to instantiate argument patterns of the other stages follow a very similar pattern so are not shown due to space limitations.

## 2.2 Stage 2: EB Safety Requirements Assurance

Stage 2 contains three activities (Fig. 6), which are performed to provide assurance in EB safety requirements for the swarm.

**Activity 3. Develop EB Safety Requirements** We define EB safety requirements to specify risk controls for the swarm-level hazards by taking into account the system architecture defined and the operating environment.

In the swarm context, we consider four types of requirements: *performance*, *adaptability*, *human safety*, and *environment*. Taking inspiration from AMLAS, these requirement types are primarily safety focussed, but we go beyond the performance and robustness safety concerns of ML systems to take into account the swarm operations in a real-world environment. The environment requirements



**Fig. 6.** Stage 2: The AERoS emergent behaviour safety requirements assurance.



**Fig. 7.** 3D simulation created to validate several emergent behaviour safety requirements.

capture the need for the system to be robust to variation in the operative space. We consider several safety metrics under each requirement category: (i) performance: low-impact and high-impact collisions (the swarm should operate below a *critical number* of such collisions); (ii) adaptability: percentage of swarm stationary outside of the delivery site, number of stationary agents, time since last agent moved; (iii) human safety: velocity or average velocity of agents, swarm size, rate of humans encountered, proximity to humans; (iv) environment: sum of objects in an area (the density of objects in the environment should not block swarm operation). As the robot swarm is composed of many agents, there is potential for a large number of faults to occur at any given time [9]. This motivates three further sub-categories for each of the performance, adaptability, and human-safety requirements: *faultless operations*, *failure modes (graceful degradation)*, and *worst case*. A key output of Activity 3 is [H], which describes the EB Safety Requirements relating to: performance, adaptability, environment, and human safety (see Table 1). These example requirements have been generated following three main considerations: the hazards for each safety requirement type, the metrics [9] available to assess these hazards, and the realistic thresholds [6] given the specification of the system.

**Activity 4: Validate EB Safety Requirements** The required input to Activity 4 is the EB Safety Requirements [H]. These are validated by both review and simulation. Firstly, the requirements derived for the cloakroom have been reviewed by a safety-critical systems engineering expert to ensure that the specified EB safety requirements for the swarm will deliver its intended safe operation. Secondly, we validated all safety requirements for the cloakroom system using the Gazebo 3D simulator. This simulation is an exact replica of the 4m x 4m lab environment [6] used for hardware implementation (see Fig. 7). In **Activity 5**, the artefacts generated in this stage are used to instantiate the EB Safety Requirements Argument Pattern [I].

### 2.3 Stage 3: Data Management

When designing EB, input data for training an algorithm comes from local sensing of individual agents, both onboard the agent itself and in its local environment. The activities and outputs in this stage take into account the complexities of interactions between multiple agents.

**Table 1.** Examples of performance, adaptability, environmental, and human-safety safety requirements for the cloakroom scenario.

|                                    |  |
|------------------------------------|--|
| <b>RQ Performance Requirements</b> |  |
| 1.1                                | The swarm <i>shall</i> experience < <b>1 high-impact</b> ( <b>V &gt; 0.5m/s</b> ) collisions across <b>a day</b> of <b>faultless</b> operation.  |
| 1.2                                | The swarm <i>shall</i> experience < <b>0.1%</b> increase in <b>high-impact</b> collisions across <b>a day's</b> operation with <b>full communication faults</b> occurring in <b>10%</b> of the swarm.        |
| 1.3                                | The swarm <i>shall</i> experience < <b>0.1%</b> increase in <b>high-impact</b> collisions across <b>a day's</b> operation with <b>half-of-wheels motor faults</b> occurring in <b>50%</b> of the swarm.      |
| 1.4                                | The swarm <i>shall</i> experience < <b>2 high-impact</b> ( <b>V &gt; 0.5m/s</b> ) collisions across <b>a day</b> of <b>faulty</b> operation.   |
| 1.5                                | The swarm agents <i>shall</i> <b>weigh</b> < <b>3kg</b> and shall have <b>acceleration</b> < <b>4m/s</b> so that the <b>maximum collision force</b> in the swarm is within acceptable bounds.                |
| 1.6                                | The swarm agents <i>shall</i> only carry objects of <b>weight</b> < <b>2kg</b> .   |
| <b>Adaptability Requirements</b>   |  |
| 2.1                                | The swarm <i>shall</i> have < <b>10%</b> of its agents <b>stationary*</b> outside of the <b>delivery site</b> at a given time. *Agents are considered stationary once they have not moved for > <b>10s</b> . |
| 2.2                                | All agents of the swarm <i>shall</i> move at least every <b>100s</b> if outside of the <b>delivery site</b> .  |
| 2.3                                | The swarm <i>shall</i> experience < <b>10%</b> increase in the <b>number of stationary agents</b> at any time with <b>half-of-wheels motor faults</b> occurring in <b>50%</b> of the swarm.                  |
| 2.4                                | The swarm agents <i>shall</i> experience < <b>10%</b> increase in <b>stationary time</b> with <b>half-of-wheels motor faults</b> occurring in <b>50%</b> of the swarm.                                       |
| 2.5                                | The swarm <i>shall</i> experience < <b>10%</b> increase in <b>number of stationary agents</b> at any given time with <b>full communication faults</b> occurring in <b>10%</b> of the swarm.                  |
| 2.6                                | The swarm agents <i>shall</i> experience < <b>10%</b> increase in <b>stationary time</b> with <b>full communication faults</b> occurring in <b>10%</b> of the swarm.   |
| 2.7                                | The swarm <i>shall</i> have < <b>20%</b> of its agents <b>stationary</b> outside of the <b>delivery site</b> at a given time.  |
| <b>Environmental Requirements</b>  |  |
| 3.1                                | The swarm <i>shall</i> perform as required in environmental density levels <b>0-4 p<sub>o</sub> of objects</b> (sum of boxes and agents per m <sup>2</sup> ) in the environment.                             |
| 3.2                                | The swarm <i>shall</i> perform as required when <b>floor incline</b> is <b>0-20 degrees</b> .  |
| 3.3                                | The swarm <i>shall</i> perform as required in a <b>dry environment</b> .   |
| <b>Human-Safety Requirements</b>   |  |
| 4.1                                | The swarm agents <i>shall</i> travel at speeds of less than <b>0.5m/s</b> when within <b>2m</b> distance of a <b>trained human</b> (a worker who has received relevant training).                            |
| 4.2                                | The swarm agents <i>shall</i> travel at speeds of less than <b>0.25m/s</b> when within <b>3m</b> distance of an <b>attendee</b> .  |
| 4.3                                | The swarm agents <i>shall</i> only come within <b>2m</b> distance of a <b>human</b> < <b>10</b> times collectively across <b>1000s</b> of <b>faultless</b> operations.                                       |
| 4.4                                | The swarm <i>shall</i> experience < <b>10%</b> increase in <b>human encounters</b> across <b>1000s</b> of operation with <b>full communication faults</b> occurring in <b>10%</b> of the swarm.              |
| 4.5                                | The swarm <i>shall</i> experience < <b>10%</b> increase in <b>human encounters</b> across <b>1000s</b> of operation with <b>half-of-wheels motor faults</b> occurring in <b>50%</b> of the swarm.            |
| 4.6                                | The swarm agents <i>shall</i> only come within <b>2m</b> distance of a <b>human</b> < <b>20</b> times collectively across <b>1000s</b> of <b>faulty</b> operations.  |

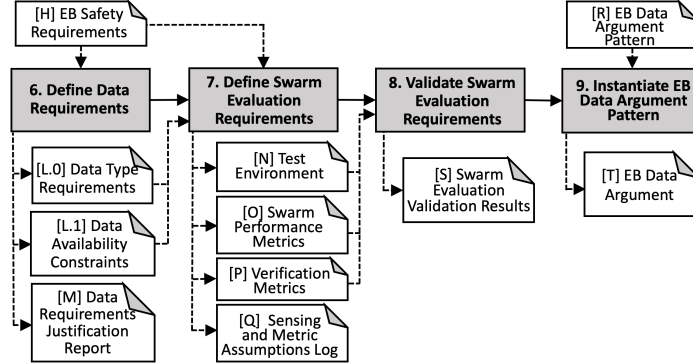


Fig. 8. Stage 3: The AERoS data management process.

**Activity 6. Define Data Requirements** We take the EB Safety Requirements [H] outlined in Stage 2 as an input (see Fig. 8) which guide the data requirements in this activity, feeding into the data specification outlined here. We split the data requirement outputs into two multi-agent focused requirements: [L.0] Data Type Requirements and [L.1] Data Availability Constraints.

*[L.0] Data Type Requirements:* This element focuses on the *relevance*, *completeness*, *accuracy*, and *balance* of the information that will be used to construct the swarm behaviour (see examples in Table 2), and subsequently, to test the EB of the system before deployment. The *relevance* of the data used in the development of the EB specifies the extent to which the test environment must match the intended operating domain of the robot swarm. The *completeness* of the data specifies the conditions under which we test the behaviour, that is, the volume of tests that will be run, the variety of tests executed, and the diversity of environments expected to be used in the testing process. The aim is to cover a representative sample of conditions for testing. *Accuracy* in this context relates to how well the data captures the parameter space defining the performance of the robot swarm. *Balance* refers to the evenly distributed trials executed in the testing process of the EB algorithm. By considering balance, we expect the number of tests conducted for failure modes or environments to be justified, ensuring that there is not an unrealistic bias in testing towards a particular scenario.

*[L.1] Data Availability Constraints:* With the introduction of multiple agents comes the issue of data availability. Distributed communication is a key feature found in emergent systems. As such, it is crucial to define how much information each agent is expected to hold, how easily data may transfer between agents, and across what range agents should be able to transfer information between one another. Feasible constraints include *storage capacity*, *available sensors*, *communication range* and *operator feedback* [6].

*[M] Data Requirements Justification Report:* This report is an assessment of the data requirements, providing analysis and explanation for how the requirements and constraints ([L.0] and [L.1]) address the EB Safety Requirements [H].

**Activity 7. Define Swarm Evaluation Requirements** Taking the outputs [L.0] and [L.1] from Activity 6, the evaluation requirements consider how the

**Table 2.** Examples of requirements for output [L.0].

| <b>RQ Relevance Requirements Examples</b> |   |
|---|---|
| 5.1                                       | All simulations <i>shall</i> include environments with ranges of incline between 0-20°.   |
| 5.2                                       | All simulations <i>shall</i> be conducted in a dry environment.   |
| <b>Completeness Requirements Examples</b> |   |
| 6.1                                       | All simulations <i>shall</i> be repeated to include occurrences of faults representative of full communication faults.  |
| 6.2                                       | All simulations <i>shall</i> be repeated a sufficient number of times to ensure results are representative of typical use.  |
| 6.3                                       | All simulations <i>shall</i> be repeated in multiple environments representative of those expected in real-world use of the system.   |
| <b>Accuracy Requirements Examples</b>     |   |
| 7.1                                       | All boxes <i>shall</i> only be considered ‘delivered’, if all four of the boxes’ feet are positioned within the delivery zone.  |
| 7.2                                       | All boxes <i>shall</i> only be considered ‘delivered’, once they are no longer in direct contact with a swarm agent.  |
| <b>Balance Requirements Examples</b>      |   |
| 8.1                                       | All simulations <i>shall</i> be repeated to obtain representative evaluations for each possible mode of failure (defined under performance, adaptability and human-safety requirements in Stage 2). |
| 8.2                                       | All simulations <i>shall</i> be repeated equally across all test environments.  |

EB of the swarm will be assessed, specifying the testing environment and the metrics to be used to assess the test results.

*[N] Test Environment:* This takes into consideration the requirements specified in Activity 6, and defines the environment in which the EB will be tested. In most cases this will be multiple simulation environments featuring diverse sets of the terrain, environmental conditions, and obstacle configurations. There may also be instances in which this test environment is specified as a physical environment operating under laboratory conditions, with a hardware system acting as a test bed to observe designed behaviours.

*[O] Swarm Performance Metrics:* This output is used to quantify how well the system is performing. While there may be multiple performance metrics, these metrics should be defined with respect to the primary function of the robot swarm. Metrics that might feature in this output could include: the delivery rate in a logistics scenario, the rate of area coverage in an exploration task, or the response time in disaster scenarios.

*[P] Verification Metrics:* These metrics should be derived from the EB Safety Requirements [H] specified in Stage 2. They are intended to be used as the criteria for success within the verification process. For example, swarm density, which is used in verifying environmental safety specifications such as RQ3.1, maximum collision force experienced by agents, which could be used to verify that the swarm meets performance requirements such as RQ1.1 and RQ1.2, or the current speed of all agents, a metric relating directly to the human-safety requirements RQ4.1 and RQ4.2. Identifying [P] early, ideally during the requirements assurance stage, allows consideration of [P] during the design and development of the swarm to facilitate verification.



*[Q] Sensing and Metric Assumptions Log:* This log serves as a record of the details and decisions made in Activities 6 and 7. It should contain details of the choices made when producing the Test Environment [N], Swarm Performance Metrics [M], and the Verification Metric [P].

**Activity 8. Validate Evaluation Requirements** Taking into account outputs [N], [O], and [P] from Activity 7, this activity aims to validate these components with respect to the requirements specified in Activity 6. Should any discrepancies exist between the data requirements and the evaluation requirements, they should be fully justified and recorded in the output Swarm Evaluation Validation Results [S]. The artefacts generated in this stage are used to instantiate the EB Data Argument Pattern [R] in **Activity 9**.

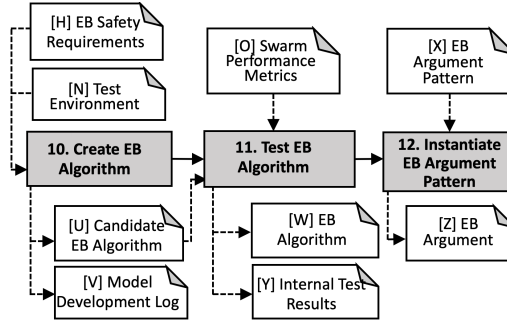
## 2.4 Stage 4: Model Emergent Behaviour

In the design of an EB algorithm, the challenge is in selecting behaviours at the individual level of the agents which give rise to the desired EB at the swarm level. We step away from the ML paradigm to allow consideration for all possible optimisation algorithms which may attain the target EB.

**Activity 10. Create EB Algorithm** This can be nature inspired, hand designed, or evolved from a relatively simple set of instructions for individual behaviour, which takes into account agent-to-agent and environmental interactions [7]. These instructions when given to a large number of agents, create a synergistic behaviour for the swarm that is more powerful than the sum of the individual agent's performance. The EB algorithm is engineered at the level of the individual agent behaviours for the Test Environment output [N] from Stage 3. The resultant EB must meet the Safety Requirements [H] defined in Stage 2 (see Fig. 9) and this will be tested in the following activity. In the cloakroom case study, the target EB for the swarm must ensure that items are stored and retrieved by individuals as quickly as possible. An example control algorithm for the individual which produces the desired collective navigation and transport behaviours is a simple random walk [10]. The swarm engineer may adjust parameters such as robot velocity to optimize for rate of task completion. However, the EB Safety Requirements [H] should eliminate parameter choices which give rise to safety hazards. For example, performance requirements RQ1.1 and RQ1.2 specify an upper bound on the low/high-impact collisions that a swarm shall experience in a given time frame. These requirements may be fulfilled by constraining the maximum velocity of individual robots or by ensuring that a robot has one or more sensory devices, such as a camera, enabling it to detect obstacles. The key output from this activity is the Candidate EB [U] for testing.

*[V] Model Development Log:* This should log the rationale in the design process of the EB algorithm, in particular how all Safety [H] and Data Type Requirements [L.0] have been met given the Data Availability Constraints [L.1].

**Activity 11. Test EB Algorithm** In this activity, the candidate EB algorithm will be tested against the Swarm Performance Metrics [O] produced in Stage 3. Testing ensures that the Candidate EB Algorithm [U] performs as desired with respect to the defined metrics and in the case where performance passes accepted



**Fig. 9.** Stage 4: The AERoS model learning process.

thresholds, the EB Algorithm [W] will be produced as the output of the activity. If testing fails Candidate EB Algorithm [U] can be iteratively modified and tested until it passes successfully.

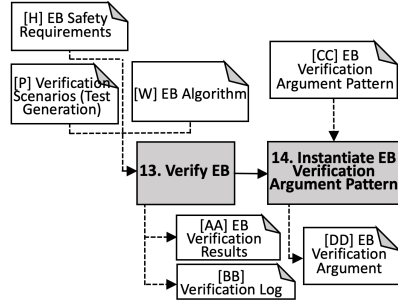
*[Y] Internal Test Results:* This output provides a degree of transparency in the testing procedure as the results may be further examined to ensure tests have run correctly. In **Activity 12**, the artefacts generated in this stage are used to instantiate the EB Argument Pattern [X].

## 2.5 Stage 5: Model Verification

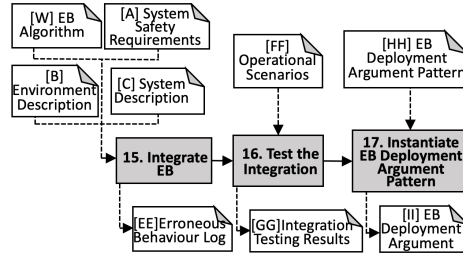
**Activity 13. Verify EB** The inputs to the verification process are the EB Safety Requirements [H], Verification Scenarios (Test Generation) [P], and the EB Algorithm [W] (see Fig. 10). The verification method and assessment process within that method will be largely determined by the specifics of the safety requirements. Some safety specifications lend themselves towards certain assessment methods due to the scenarios they prescribe. For example, to assess that the robot swarm meets the requirements for performance given a motor-fault occurrence (see RQ1.3), it may be easier to realise this in physical or simulation-based testing approaches rather than attempting to construct a formal model of robot behaviour given the complex physical dynamics of a faulty wheel. However, when considering the adaptability requirements, a formal, probabilistic verification technique of the EB Algorithm [W] is more suitable. For example, in RQ2.1, analysis using a probabilistic finite state machine of the swarm behaviour could identify the dwell period within states. Monitors could be used to observe when agents enter a stationary state, for example,  $\text{agent\_velocity}=0 \wedge \text{t\_counter} \geq 100$ , and identify if time within that state exceeds some fixed value, and ascertain a probabilistic value to this metric.

*[P] Verification Scenario (Test Generation):* In most cases there will be multiple, valid verification scenarios (test cases) applicable for each of the safety specifications. An example of a test case could include a scenario of the swarm in a hazardous environment where too many boxes create an obstacle.

Verification Results [AA] from individual assessments form entries in the Verification Log [BB]. The Verification Log identifies assessments where assurance of the EB Algorithm [W] is acceptable with respect to the Safety Requirements



**Fig. 10.** Stage 5: The AERoS verification process.



**Fig. 11.** Stage 6: The AERoS model deployment assurance process.

[H] and can be used as a set of evidence for building an assurance case. The artefacts generated in this stage are used to instantiate the EB Verification Argument Pattern [CC] in **Activity 14**.

## 2.6 Stage 6: Model Deployment

**Activity 15. Integrate EB** With the EB verified, the next step is to take [W], [A], [B] and [C] as inputs, and integrate the EB with the system to be deployed (see Fig. 11). In this activity, we use the inputs to this stage to inform the implementation of the EB and anticipate errors we might expect in the interactions between agents and the overall EB. Despite the rigorous validation and testing conducted in previous stages, there will still be a gap between the test environment and the intended, everyday-use, deployed scenario. The output, [EE] Erroneous Behaviour Log, captures these anticipated gaps between testing and reality and the differences in behaviour that may surface.

**Activity 16. Test the Integration** Once the initial integration is complete, the physical implementation should undergo additional testing in which the system will be observed in multiple operational scenarios, as specified in [FF].

*[FF] Operational Scenarios:* These scenarios should reflect the environment descriptions specified in [B], offering real-world situations to examine the behaviour of the integrated system. The testing of the integrated system in these true-to-operation environments should be conducted in a safe manner, ensuring that the entire multi-agent system can be shut down in an emergency. In the cloakroom, an example of [FF] may take the form of a deployment of agents in a controlled storage area that will not interfere with emergency services.

*[GG] Integration Testing Results:* Results from the integration testing will be reported here, detailing how the system performs against the EB Safety Requirements [H] specified in Stage 2. The artefacts generated in this stage are used to instantiate the EB Deployment Argument Pattern [HH] in **Activity 17**.

## 3 Conclusion

Using AMLAS [3] as a foundation, we proposed the six-stage development process AERoS. This process acts as guidance for those looking to construct swarm

robot systems, particularly those that exhibit emergent behaviour through environmental and agent-to-agent interaction. The stages of AERoS break down the design of these systems to ensure that fundamental safety requirements are adhered to, even in instances of system degradation and compounded failures that should be expected, and managed, in robot swarms. We achieve this with an approach that allows for iteration of and feedback to the previous stages as issues of safety are encountered and investigated. We combine this iteration with repeated specification at each stage, observing the issue of safety through the lens of: data, modelling/behaviour design, verification, and deployment. While the iterative nature of AERoS is a key advantage, the scope of this work has been limited to investigating inherent swarm qualities and the emergent properties that arise from these. However, one can expand on this, and consider adaptation of individual robots through techniques such as ML (e.g. by applying AMLAS). Also, we can broaden the evaluation by considering additional swarm use cases (e.g. monitoring fires in a natural environment, and also a social swarm), and by providing a worked example of the entire AERoS process. While the focus of AERoS is to ensure the safety assurance of EB in swarms, the trustworthiness of an AS can be dependent on many factors other than safety, such as ethics and regulation, which we will investigate as part of future work.

## References

1. Abeywickrama, D.B., Bennaceur, A., Chance, G., and Others: On specifying for trustworthiness (2022), <http://arxiv.org/abs/2206.11421>
2. Cheng, B.H.C., Eder, K.I., Gogolla, M., and Others: Using models at runtime to address assurance for self-adaptive systems, pp. 101–136. Springer International Publishing, Cham (2014)
3. Hawkins, R., Paterson, C., Picardi, C., Jia, Y., Calinescu, R., Habli, I.: Guidance on the assurance of machine learning in autonomous systems (AMLAS). Guidance Version 1.1, University of York (Mar 2021)
4. Jia, Y., McDermid, J., Lawton, T., Habli, I.: The role of explainability in assuring safety of machine learning in healthcare. *IEEE Transactions on Emerging Topics in Computing* **10**(4), 1746–1760 (2022)
5. Jones, S., Milner, E., Sooriyabandara, M., Hauert, S.: Distributed situational awareness in robot swarms. *Advanced Intelligent Systems* **2**(11), 2000110 (2020)
6. Jones, S., Milner, E., Sooriyabandara, M., Hauert, S.: DOTS: An open testbed for industrial swarm robotic solutions (2022)
7. Jones, S., Studley, M., Hauert, S., Winfield, A.: Evolving behaviour trees for swarm robotics, pp. 487–501. Springer International Publishing, Cham (2018)
8. Kaakai, F., Dmitriev, K., Adibhatla, S., and Others: Toward a machine learning development lifecycle for product certification and approval in aviation. *SAE Int. J. Aerosp.* **15**(2) (May 2022). <https://doi.org/10.4271/01-15-02-0009>
9. Lee, S., Milner, E., Hauert, S.: A data-driven method for metric extraction to detect faults in robot swarms. *IEEE Robot. Autom. Lett.* **7**(4), 10746–10753 (2022)
10. Milner, E., Sooriyabandara, M., Hauert, S.: Stochastic behaviours for retrieval of storage items using simulated robot swarms. *Artificial Life and Robotics* **27**(2), 264–271 (2022)
11. Winfield, A.F.T., Nembrini, J.: Safety in numbers: Fault-tolerance in robot swarms. *International Journal on Modelling Identification and Control* **1**(1), 30–37 (2006)