# Framework for Assurance of Emergent Behaviour for use in Autonomous Robotic Swarms

Author1_Name, Author2_Name, ...AuthorN_Name,

***Abstract*—Abstract of the paper.**

***Index Terms*—Assurance, swarm robotics, trustworthiness, safety, transparency, ethics, autonomous systems**

## I. INTRODUCTION

The main contribution of this paper is a novel framework for assurance of emergent behaviour for use in autonomous robotic swarms based on the AMLAS, SACE* and SOCA* guidance (* under consideration). We illustrate the framework using a public cloakroom case study.

## II. BACKGROUND AND RELATED WORK

### A. Background

*1) Specification Challenges and Standards:*

*2) Robotic Swarms:*

*3) Case Study:*

***[Author Guidelines: Please use the cloakroom case study to illustrate the framework in Section III. If the examples in cloakroom case study are not sufficient, other swarm use cases listed below can be considered.]***

*a)* ***Cloakroom****:* The case study describes a public cloakroom where swarm of robots assist customers looking to deposit their jackets at an event [1]. It describes cases where customers are depositing jackets, handing a jacket to a robot for storing, and retrieval of jackets back to the customer.

*b) Other Swarm Use Cases:*

**Fault detection, diagnosis and recovery – Monitoring fires in a natural environment**. Fault detection model shall be trained to high level of accuracy. Thresholds for fault tolerance shall be set appropriately such that misclassification of a fault is a rare event. An agent experiencing minor faults shall not be immediately removed, should the fault not impact the task at hand.

**Social swarm – Brainstorming at an event**. Humans follow robots which cluster based on input. Minimise blocking paths of other humans and agents. Maintain situational awareness of humans and agents in the environment. Before the task, provide a clear explanation of the steps of the activity. Clear guidance during the task. Provide information about how the swarm/robot works.

See research paper "Mutual shaping in swarm robotics: User studies in fire and rescue, storage organization, and bridge inspection" [2].

### B. Related Work

*1) Assurance of Machine Learning in Autonomous Systems (AMLAS): Assurance of Machine Learning for use in Autonomous Systems (AMLAS)* provides guidance on how to systematically integrate safety assurance into the development of machine learning components based on offline supervised learning [3]. AMLAS provides an explicit and structured safety case that the system is safe to operate in its intended context of use. AMLAS contains six stages, and the assurance activities are performed in parallel to the development of machine learning component. The process is iterative by design and feedback is used to update previous stages.

*2) Safety Assurance of Autonomous Systems in Complex Environments (SACE):* [4]

*3) Societal Acceptability of Autonomous Systems (SOCA):* [5], [6]

## III. FRAMEWORK

### A. Overview of Framework

### B. Stage 1: EB Safety Assurance Scoping

***[Lead: WP1]***
***[Author Guidelines: 900–1800 words / 1–2 pages (maximum);***
***Format/structure: Describe adapted AMLAS activities, inputs and outputs using cloakroom case study examples. Activities: 1, 2; Inputs: A, B, C, D, F; Outputs: E, G]***
See Fig. 1

### C. Stage 2: EB Safety Requirements Assurance

***[Lead: WP1; Other: WP2, WP3]***
***[Author Guidelines: total 7 pages (maximum);***
***Format/structure: Describe adapted AMLAS activities, inputs and outputs using cloakroom case study examples.***
***WP1 = (Activities: 3, 4, 5; Inputs: E, I; Outputs: H, J, K: 2700 words / 3 pages maximum))***
***WP2 = (List of Ethical Requirements and Description: 1800 words / 2 pages maximum)***
***WP3 = (List of Socio-Technical/Regulatory Requirements and Description: 1800 words / 2 pages maximum)]***

Author1 and Author2 are with the Department of Dep1_Name (email: , )
AuthorN is with the Department of Dep2_Name (email: )

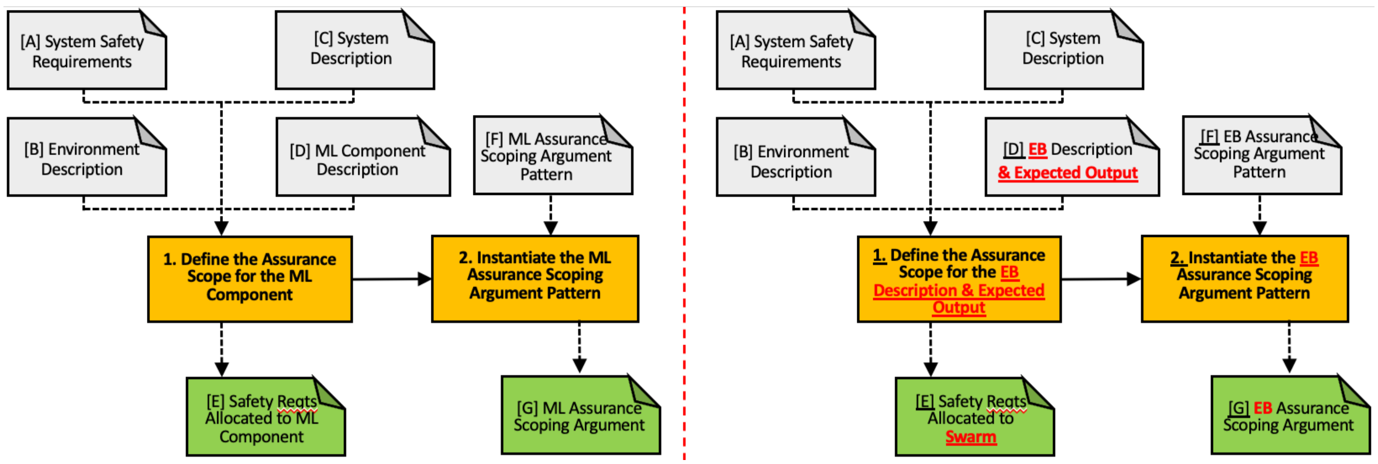


Fig. 1. Adapted AMLAS emergent behaviour assurance scoping process (right).

**Stage 2 Requirements (Input H)**
**Cloakroom: Performance Requirements:**

| | **Requirements for Faultless Operations** |
|---|---|
| RQ1.1 | The swarm *shall* experience < 1 low impact (V < 0.5m/s) collisions across 1000 seconds of faultless operation. |
| RQ1.2 | The swarm *shall* experience < 1 high impact (V > 0.5m/s) collisions across a day of faultless operation. |
| | **Requirements for Failure Modes (Graceful Degradation):** |
| RQ1.3 | The swarm *shall* experience < 10% increase in low impact collisions across 1000 seconds of operation with 10% injection of full communication fault to the swarm. |
| RQ1.4 | The swarm *shall* experience < 0.1% increase in high impact collisions across a days operation with 10% injection of full communication fault to the swarm. |
| RQ1.5 | The swarm *shall* experience < 10% increase in low impact collisions across 1000 seconds of operation with 50% injection of half-of-wheels motor faults to the swarm. |
| RQ1.6 | The swarm *shall* experience < 0.1% increase in high impact collisions across a days operation with 50% injection of half-of-wheels motor faults to the swarm. |
| | **Requirements for Worst Case:** |
| RQ1.7 | The swarm *shall* experience < 2 low impact (V < 0.5m/s) collisions across 1000 seconds of faulty operation. |
| RQ1.8 | The swarm *shall* experience < 2 high impact (V > 0.5m/s) collisions across a day of faulty operation. |

**Cloakroom: Adaptability Requirements:**

| | **Requirements for Faultless Operations** |
|---|---|
| RQ2.1 | The Swarm *shall* have < 10% of its agents stationary* outside of the delivery site at a given time. *Assumption: Agents are considered stationary once they have not moved for > 10 seconds. |
| RQ2.2 | All agents of the swarm *shall* move at least every 100 seconds if outside of the delivery site. |
| | **Requirements for Failure Modes (Graceful Degradation):** |
| RQ2.3 | The swarm *shall* experience < 10% increase in number of station agents at any given time with 50% injection of half-of-wheels motor faults to the swarm. |
| RQ2.4 | The swarm agents *shall* experience < 10% increase in stationary time with 50% injection of half-of-wheels motor faults to the swarm. |
| RQ2.5 | The swarm *shall* experience < 10% increase in number of station agents at any given time 10% injection of full communication fault to the swarm. |
| RQ2.6 | The swarm agents *shall* experience < 10% increase in stationary time 10% injection of full communication fault to the swarm. |
| | **Requirements for Worst Case:** |
| RQ2.7 | The Swarm *shall* have < 20% of its agents stationary* outside of the delivery site at a given time. *Assumption: Agents are considered stationary once they have not moved for > 10 seconds. |
| RQ2.8 | All agents of the swarm *shall* move at least every 200 seconds if outside of the delivery site. |

Metric: Swarm waiting time/time-in-area

**Cloakroom: Human Safety Requirements:**

| | **Requirements for Faultless Operations** |
|---|---|
| RQ3.1 | The agents in the swarm *shall* travel at speeds of less than 0.5m/s when within 2m distance of a Trained Human* |
| RQ3.2 | The agents in the swarm *shall* travel at speeds of less than 0.25m/s when within 3m distance of a member of the public. |
| RQ3.3 | The agents in the swarm *shall* only come within 2m distance of a human $<$ 10 times collectively across 1000 seconds of faultless operations. |
| RQ3.4 | The swarm *shall* only allow $<$ 5 agents to request intervention from a Trained Human* at a given time |
| RQ3.5 | A Trained human *shall* monitor 5-20 agents at a given time. |
| RQ3.6 | The swarm *shall* only allow 1 agent to request input from a member of the public at a given time. |
| RQ3.7 | A member of the public *shall* receive $<$ 5 agents of swarm information at a given time. |
| | **Requirements for Failure Modes:** |
| RQ3.8 | The swarm *shall* experience $<$ 10% increase in human encounters across 1000 seconds of operation with 10% injection of full communication fault to the swarm. |
| RQ3.9 | The swarm *shall* experience $<$ 10% increase in human encounters across 1000 seconds of operation with 50% injection of half-of-wheels motor faults to the swarm. |
| | **Requirements for Worst Case:** |
| RQ3.10 | The agents in the swarm *shall* only come within 2m distance of a human $<$ 20 times collectively across 1000 seconds of faulty operations. |

*Trained Human in this case refers to workers within the case study setting. We assume that this individual has received relevant training & experience in the use of the swarm system.

**Cloakroom: Environmental Specification:**

| | |
|---|---|
| RQ4.1 | The swarm *shall* perform as required in environmental density levels 0-4 $p_o$* of objects (sum of boxes and agents) in the environment. |
| RQ4.2 | The swarm *shall* perform as required when floor incline is 0-20 degrees. |
| RQ4.3 | The swarm *shall* perform as required in a dry environment. |
| RQ4.4 | The swarm *shall* perform as required in smooth-floored environments with step increases no greater than 0.5cm. |
| RQ4.5 | The swarm *shall* only operate in environments where humans have devices that identify the human's whereabouts to the swarm agents. |

*$p_o$ = sum of objects / $m^2$

See Fig. 2

### D. Stage 3: Data Management

***[Lead: WP5]***
***Author Guidelines: 900–1800 words / 1–2 pages (maximum); Format/structure: Describe adapted AMLAS activities, inputs and outputs using cloakroom case study examples. Activities: 6, 7, 8; Inputs: H; Outputs: L0, L1, M, N, O, P, Q, S***
See Fig. 3

### E. Stage 4: Model Emergent Behaviour

***[Lead: WP5]***
***Author Guidelines: 900–1800 words / 1–2 pages (maximum); Format/structure: Describe adapted AMLAS activities, inputs and outputs using cloakroom case study examples. Activities: 10, 11; Inputs: H, N, O; Outputs: Candidate EB Algorithm, U, V, X***

See Fig. 4

### F. Stage 5: Model Verification

***[Lead: WP4]***
***Author Guidelines: 900–1800 words / 1–2 pages (maximum); Format/structure: Describe adapted AMLAS activities, inputs and outputs using cloakroom case study examples. Activities: 13; Inputs: H, P, V; Outputs: Z, AA***
See Fig. 5 and Fig. 6.

- Test bench for swarms
- Probabilistic verification ideas
- Simulation-based testing
- Verifiability?

### G. Stage 6: Model Deployment

***[Leads: WP4 & WP5; Additional: WP3]***
***Author Guidelines: 900–1800 words / 1–2 pages (maximum); Format/structure: Describe adapted AMLAS activities, inputs and outputs using cloakroom case study examples.***
***WP5 = (Activities: 15, Inputs: V, A, B, C, D, Outputs: DD),***
***WP4 = (Activities: 16, Inputs: EE, Outputs: FF),***
***WP3 = (Regulatory Considerations – 675 words / 0.75 page maximum)***
See Fig. 7

## IV. Discussion and Conclusions

## Acknowledgments

The work presented in this paper has been supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under the grant [EP/V026518/1].

## Appendix A. Supplementary Material

The supplementary material associated with this article can be found online at (https://www.).

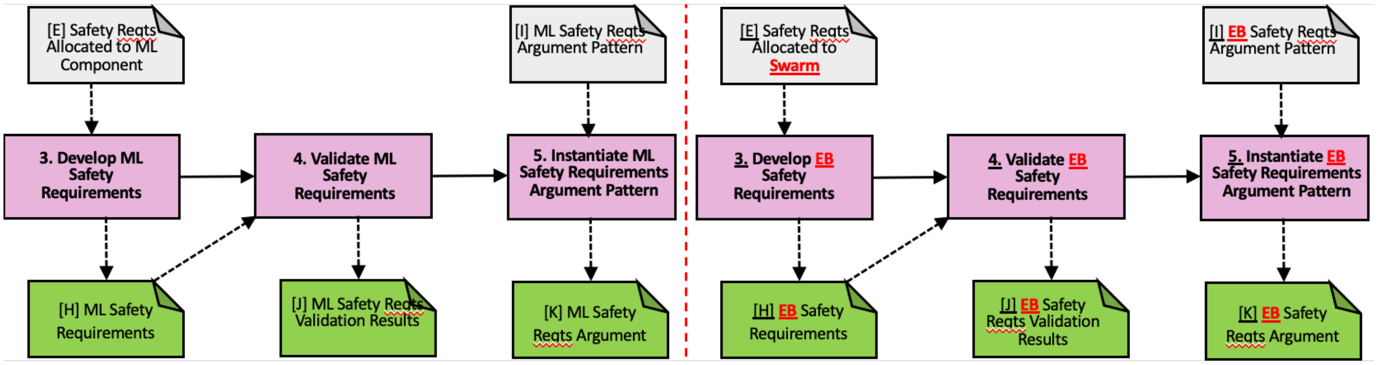


Fig. 2. Adapted AMLAS emergent behaviour safety requirements assurance process (right).

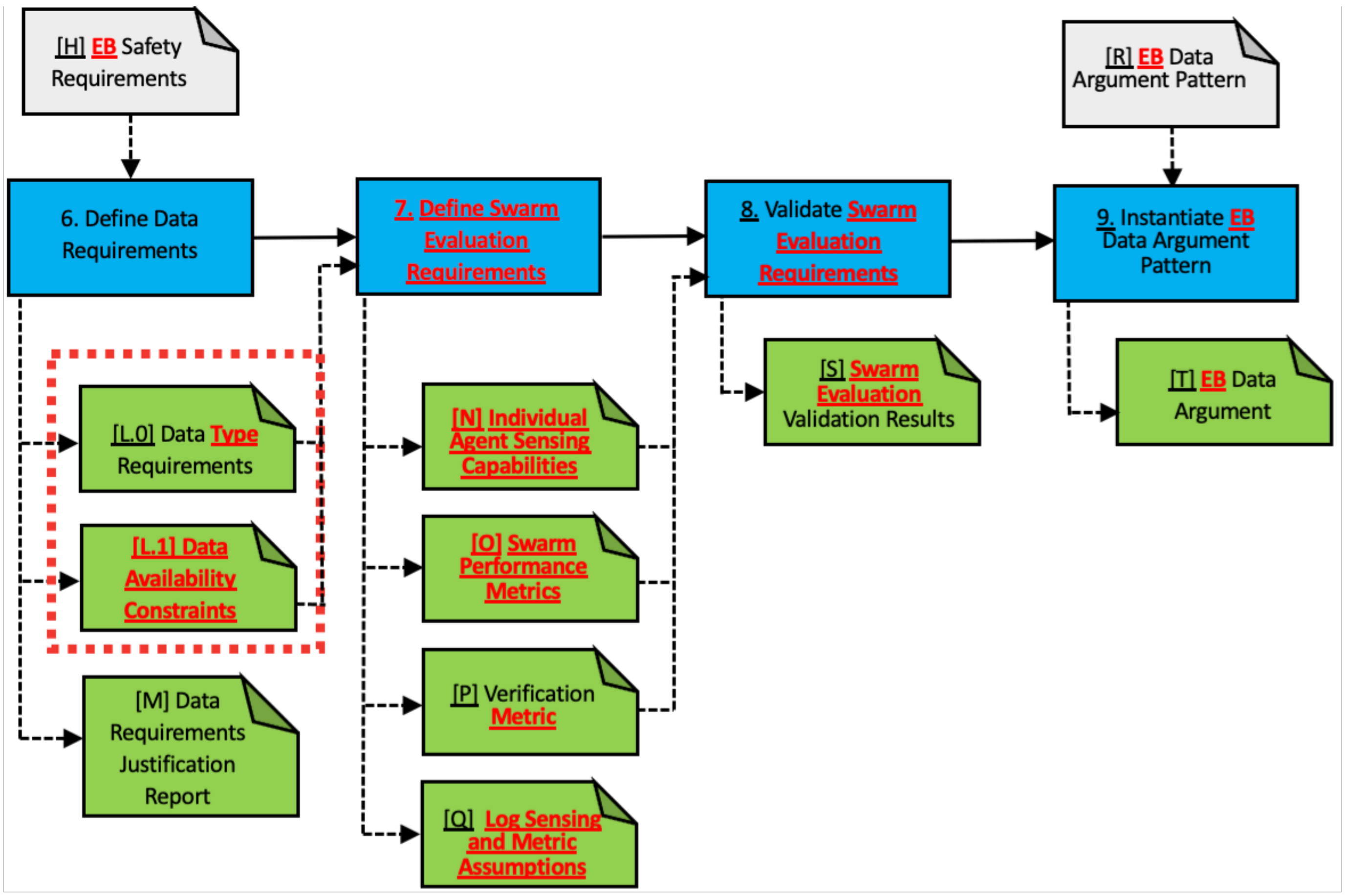


Fig. 3. Adapted AMLAS data management process.

## References


[1] S. Jones, E. Milner, M. Sooriyabandara, and S. Hauert, "Distributed situational awareness in robot swarms," *Advanced Intelligent Systems*, vol. 2, no. 11, p. 2000110, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202000110

[2] D. Carrillo-Zapata, E. Milner, J. Hird, G. Tzoumas, P. J. Vardanega, M. Sooriyabandara, M. Giuliani, A. F. T. Winfield, and S. Hauert, "Mutual shaping in swarm robotics: User studies in fire and rescue, storage organization, and bridge inspection," *Frontiers Robotics AI*, vol. 7, p. 53, 2020.

[3] R. Hawkins, C. Paterson, C. Picardi, Y. Jia, R. Calinescu, and I. Habli, "Guidance on the assurance of machine learning in autonomous systems (amlas)," University of York, Guidance Version 1.1, Mar. 2021.

[4] R. Hawkins, M. Osborne, M. Parsons, M. Nicholson, J. McDermidand, and I. Habli, "Guidance on the safety assurance of autonomous systems in complexenvironments (sace)," University of York, Guidance Version 1, Jul. 2022.

[5] Z. Porter, I. Habli, and J. A. McDermid, "A principle-based ethical assurance argument for ai and autonomous systems," Mar. 2022.

[6] J. McDermid, Y. Jia, Z. Porter, and I. Habli, "Artificial intelligence explainability: the technical and ethical dimensions," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2207, Aug. 2021.

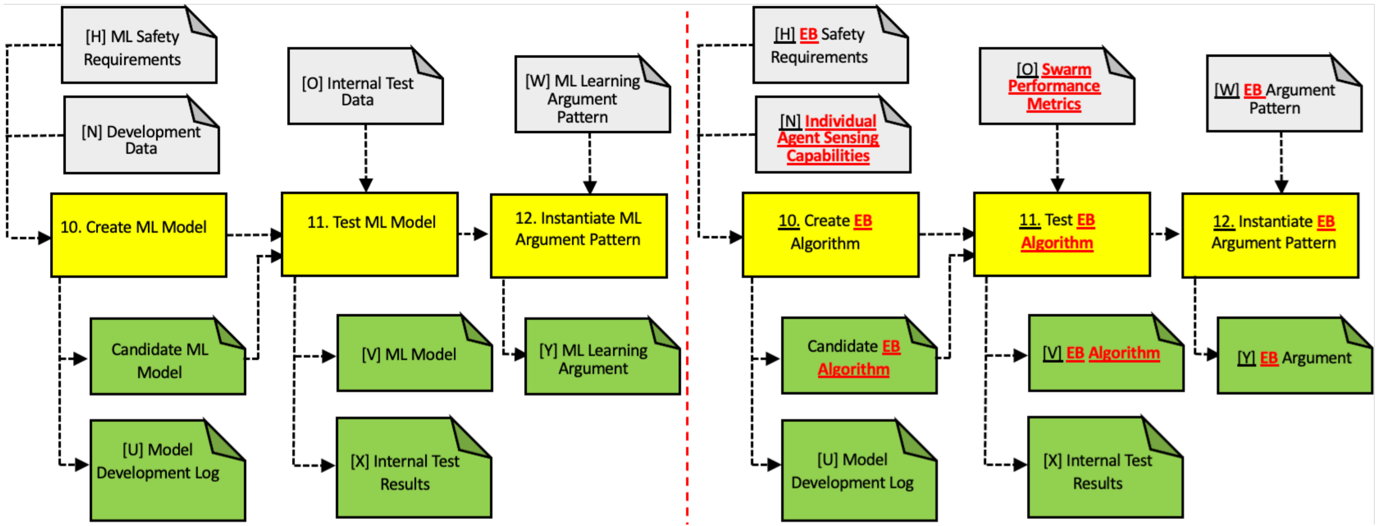


Fig. 4. Adapted AMLAS model learning process (right).

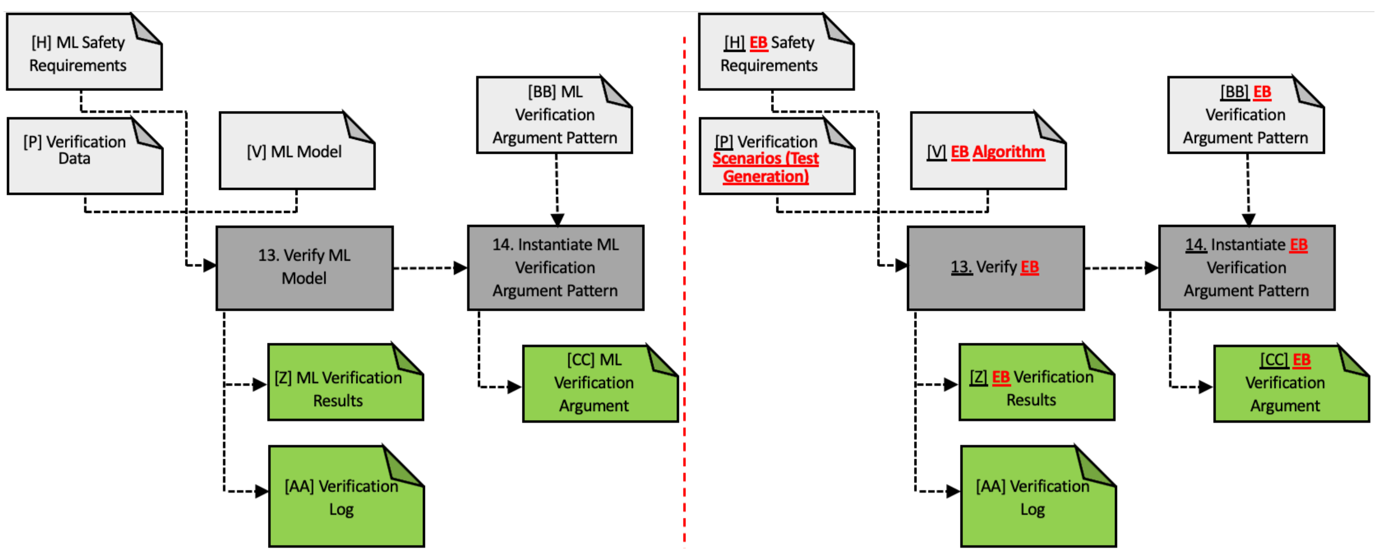


Fig. 5. Adapted AMLAS verification assurance process (right).

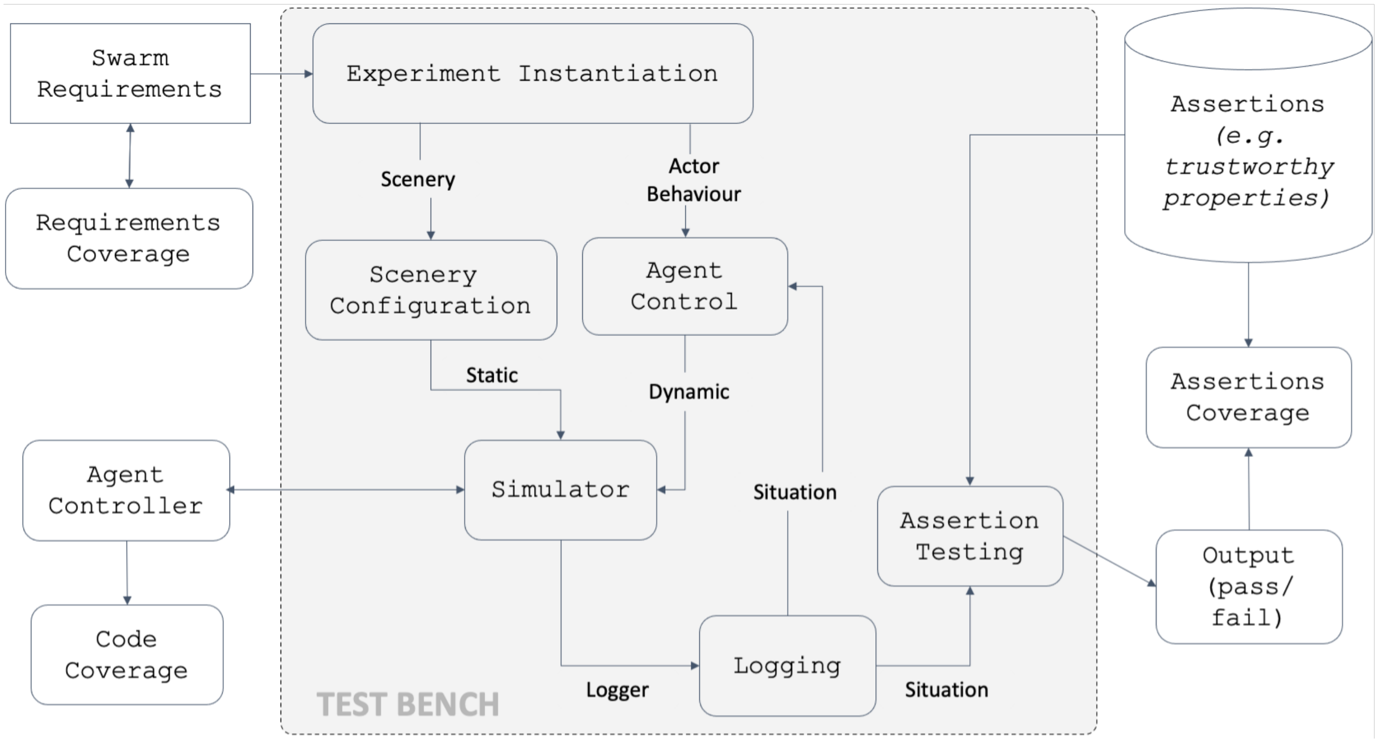


Fig. 6. Adapted AMLAS verification assurance process: test bench.

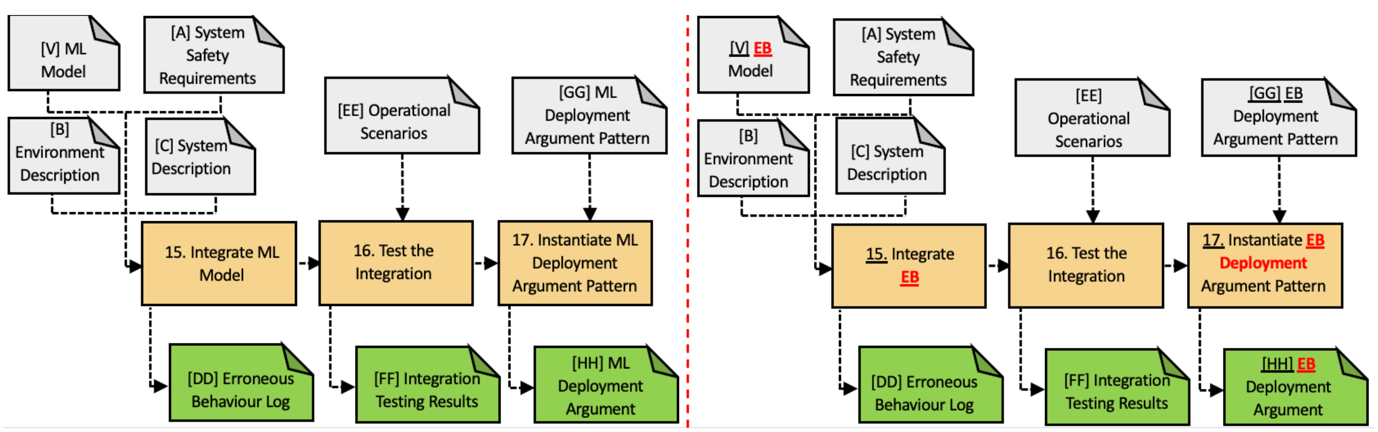


Fig. 7. Adapted AMLAS model deployment assurance process (right).