CHAPTER 1 INTRODUCTION

1.1 Background

The rapid integration of artificial intelligence (AI) and machine learning (ML) technologies is transforming many critical domains, including healthcare, defence, education, manufacturing, and transportation [1]. A prominent example of this transformation is the development of autonomous vehicles (AVs), which rely heavily on AI-driven perception systems to navigate and interact safely with their environments [2]. Among these perception tasks, object recognition (OR)—the ability to detect, identify, and classify objects in real time—is essential for avoiding collisions, planning trajectories, and making informed decisions, especially in dynamic settings such as smart factories [3]. Ensuring the safety of ML-based object recognition systems remains a critical challenge due to the inherent complexity, probabilistic behaviours, and unpredictability of ML models [4]. These properties can introduce risks such as data drift, bias, and vulnerability to adversarial inputs [5], [6], [7].

Recognizing these challenges, structured safety assurance methodologies have been proposed to support the safe adoption of ML in safety-critical systems. The Assurance of Machine Learning in Autonomous Systems (AMLAS) framework offers a lifecycle-based approach for systematically identifying safety requirements, managing risks, and demonstrating compliance [8]. Existing research often adopts the Goal Structuring Notation (GSN) as the framework for representing AMLAS-based assurance cases [9], [10], [11], [12], [9], [10], [11], [12]. GSN is valued for its clear, visual representation of claims, arguments, and supporting evidence, and has been widely used in safety-critical industries as a standardized approach to safety case communication. In this thesis, the version of GSN selected for comparison is the GSN Community Standard Version 2 (v2.0), which serves as the formal, standardized reference for representing assurance cases consistently across domains.

However, while GSN is widely adopted, other assurance case frameworks also exist that provide complementary and potentially more expressive modelling

constructs. The Structured Assurance Case Metamodel (SACM) is one such framework that supports modular, traceable, and dialectical argument structures, enabling explicit modelling of claims, assumptions, context, counterclaims, and supporting evidence [13], [14], [15], [14], [15]. These features are valuable for capturing the complexity and uncertainty of ML-based systems in safety-critical applications. Despite its potential, there remains limited research exploring the adoption of SACM to support the safety assurance of AI/ML systems. A literature review conducted in this study highlights this gap, showing that while GSN is frequently applied to support AMLAS-based assurance cases, empirical studies adopting SACM in this context are scarce [13].

To address this research gap, this study proposes adopting SACM as the assurance case framework integrated with the AMLAS lifecycle to support the safety assurance of an ML-based object recognition system in an AV context. The practical scenario for this investigation is set within a smart factory environment conceptualized as a logistics-oriented industrial setting designed to emulate automated package delivery workflows typical of modern manufacturing. Drawing on the taxonomy presented by Zuehlke (2010) [16] and Lu et al. (2017) [17], the testbed mimics an intralogistics system where an autonomous ground vehicle must detect, navigate toward, and manoeuvre around packages within a constrained indoor industrial floor. This scenario reflects real-world use cases in modern smart factories that rely on cyber-physical production systems (CPPS) to enable flexible, automated material transport using mobile platforms.

To support this investigation, the Donkey Car S1 platform is employed as a low-cost, customizable simulation testbed for evaluating object recognition systems under controlled yet realistic factory-like conditions. The ML component is implemented using a YOLOv8 object detection model trained on a custom dataset representing varied factory scenarios, including different lighting conditions, occlusions, and background complexity [18]. The Donkey Car platform is equipped with upgraded hardware, including an ESP microcontroller, a smartphone-based camera system for high-quality inference, and an integrated ultrasonic sensor to support real-time collision avoidance override logic. These upgrades ensure that

object recognition outputs meaningfully inform vehicle control actions, supporting safe navigation within narrow, obstacle-rich factory pathways.

Furthermore, this study is motivated by the need for empirical evaluations of SACM's graphical notation in practical domains, as identified in future work such as Selviandro et al. [13]. By applying SACM within the AMLAS lifecycle for an SAE J3016 Level 2 [4] scenario in a simulated smart factory environment, this research contributes to addressing that gap while supporting the development of clear, maintainable, and traceable safety arguments for the deployment of ML-based perception systems in safety-critical industrial settings.

1.2 Problem Statement

The assurance of ML-based object recognition systems remains underdeveloped in safety-critical domains [4]. Existing frameworks such as Goal Structuring Notation (GSN) and Claims-Argument-Evidence (CAE) provide structured reasoning but fall short in addressing the uncertainties and dynamic behaviours inherent in ML models [19]. Although advances in ML safety methodologies have been made, there is still a need for approaches that integrate formal structured reasoning with empirical validation of model performance.

This research explores the adoption of the Structured Assurance Case Metamodel (SACM) to support the safety assurance of ML-based object recognition systems within an autonomous vehicle (AV) simulation. By investigating SACM’s application, the study aims to examine its suitability for modelling safety arguments in ML contexts and to identify the benefits and challenges associated with its use. Accordingly, the study is guided by the following research questions (RQ):

1. **RQ1:** How can SACM be adopted to support assuring the safety of an ML-based object recognition system in an AV simulation environment?
2. **RQ2:** What are the benefits and challenges in adopting SACM in the context of supporting the safety assurance of an ML-based object recognition system in an AV?

1.3 Objectives

The objectives (OB) of this study are as follows:

1. **OB1:** To adopt and implement the Structured Assurance Case Metamodel (SACM) for structuring the safety assurance of an ML-based object recognition system within a smart factory simulation environment.
2. **OB2:** To investigate the potential benefits and challenges arising from the integration of SACM (Structured Assurance Case Metamodel) into the AMLAS (Assurance of Machine Learning in Autonomous Systems) process for ML safety assurance.

Through these objectives, the study aims to contribute practical insights into applying SACM for ML-based system assurance, addressing current gaps in both theory and empirical practice.

1.4 Justification for Research

Aligned with the background of this study, existing research commonly adopts GSN as the assurance case framework in the implementation of AMLAS [8]. However, SACM offers potential benefits that support more expressive and rigorous safety assurance analysis. Therefore, it is important to investigate these benefits for adoption in studies aiming to assure the safety implementation of ML systems.

Additionally, this research contributes to the broader field of ML safety assurance by providing a structured methodology that explicitly integrates ML performance metrics with assurance case development, ensuring traceable linkage between empirical evidence and safety claims. By applying SACM in a smart factory simulation, this study offers empirical validation of its applicability and effectiveness in supporting safety assurance for ML-based object recognition systems.

Unlike prior work that often remains at the conceptual or modelling level, this research demonstrates SACM-based assurance through an integrated, real-world-inspired simulation using the Donkey Car S1 platform. The complete design, hardware-software integration, and empirical testing are detailed in Chapter 3 (particularly Sections 3.4 and 3.5), while the evaluation results and their role in

supporting the assurance argument are elaborated in Chapter 4, especially Section 4.2. This approach strengthens the empirical value and novelty of the research by ensuring the assurance case is grounded in realistic, replicable evidence.

1.5 Scope and Limitations

The definition of safety in this research is scoped specifically to obstacle detection and crash avoidance capabilities enabled by the object recognition system. The study does not extend to high-level planning, path planning, motion control, or other aspects of full autonomous navigation. Instead, it focuses on ensuring that the ML-based perception component can reliably and accurately detect critical objects in time to support safe human-supervised navigation decisions.

In terms of functional safety classification, the system aligns with SAE J3016 Level 2 (Partial Automation), where the vehicle can perform some automated functions—such as perception and limited navigation assistance—within structured environments, but still requires human oversight or pre-defined paths for overall control [18]. This scope ensures that the safety assurance case remains precisely targeted to perception-level risks, without making claims about end-to-end or fully autonomous driving capabilities.

This research focuses on developing and assuring the safety of an object recognition system using SACM. The system is trained and evaluated with the following specifications in Table 1.5.1:

Table 1.5.1 Summary of Methodology and Tools

Aspect	Tool/Approach	Purpose
Object Recognition	YOLOv8	Detect and classify objects
Dataset Preparation	Roboflow	Annotate and split datasets
Training Platform	Google Colab	Train the YOLOv8 model
Simulation Environment	Donkey Car S1	Simulate smart factory object recognition
Safety Assurance Framework	SACM (Structured Assurance Case Metamodel)	Provide structured safety assurance and dialectical reasoning

The platform utilized in this research is the Donkey Car S1 simulation, designed to emulate smart factory conditions in a controlled environment. The machine learning model employed is YOLOv8, trained using datasets prepared through Roboflow and processed on Google Colab. A custom dataset comprising 1,775 images was developed, with splits of 71% for training, 20% for validation, and 9% for testing, as detailed in Table 1.5.2.

Table 1.5.2 Overview of Dataset Characteristics

Aspect	Details
Total Images	1,775
Training Split	71% (1255 images)
Validation Split	20% (354 images)
Test Split	9% (166 images)
Preprocessing	No preprocessing applied.
Augmentations	No augmentations applied.

Notably, no preprocessing or augmentation techniques were applied to the dataset. This choice was intentional to preserve the raw, unaltered characteristics of the factory-like environment, including variations in lighting, occlusion, and background clutter. The goal was to ensure that model performance metrics would realistically reflect deployment conditions without overfitting to artificially enhanced data distributions (as elaborated in Chapter 4 for further discussion of this rationale).

The limitations of this study include the following: the scope is restricted to object recognition tasks and does not encompass higher-level decision-making, path planning, or navigation autonomy. The research is confined to a simulation environment and does not extend to real-world deployment or field testing. Additionally, the performance metrics and safety assurance arguments are tailored specifically to the YOLOv8 model, which may limit the generalizability of findings to alternative models or detection architectures.

1.6 Significance of the Study

The study contributes to the field of autonomous systems and safety assurance in the following ways:

1. It demonstrates the practical adoption of SACM in a simplified case study involving object recognition in smart factory simulations.
2. It highlights the integration of ML performance metrics (e.g., mAP, Precision, and Recall) with safety assurance frameworks, showcasing a structured approach to addressing safety concerns in AI-based systems.
3. It provides insights into the advantages of SACM over CAE and GSN in presenting dialectical arguments for safety assurance cases.

Furthermore, this study demonstrates characteristics of both maintainability and scalability within its assurance framework. Maintainability refers to the system's ability to be updated or extended with minimal effort or risk, achieved here through the modular nature of SACM modelling. Each SACM element (claim, context, evidence) is traceable and can be individually revised without disrupting the overall assurance structure [20]. This approach enables future safety updates—such as model retraining, integrating new sensors, or deployment in novel environments—to be incorporated without overhauling the entire safety case.

By adopting SACM in a smart factory simulation, this research also aligns with best practices for structured, adaptable assurance arguments [21], ensuring the framework remains robust even as system capabilities evolve.

Scalability, in turn, is demonstrated by the methodology's capacity to accommodate a diverse range of ML models or application scenarios beyond YOLOv8. Owing to its structured alignment with AMLAS stages and its generic mapping to SACM constructs (as elaborated in Section 3.3), the assurance framework offers a reusable and adaptable foundation. This structure enables its application to alternative use cases—such as pedestrian detection, path planning, or other perception tasks—thereby enhancing its potential utility in larger, multi-component autonomous vehicle systems.

1.7 Structure of the Thesis

This thesis is structured to systematically address the research problem and to answer the stated research questions while achieving the defined objectives.

The flow of the thesis begins with an exploration of the existing challenges in assuring ML-based object recognition systems, as articulated in the Problem

Statement (Section 1.2). Based on these challenges, two Research Questions (RQ1 and RQ2) and corresponding Objectives (OB1 and OB2) were formulated to guide this study (Sections 1.3 and 1.2).

To address these questions and objectives, a methodology combining the AMLAS lifecycle with SACM-based assurance modelling was designed and is detailed in Chapter 3. The practical implementation of this methodology and the results of the study are presented in Chapter 4, where:

1. RQ1 / OB1 is addressed through the integration of SACM within each AMLAS assurance stage (Sections 4.4.1 to 4.4.8) and the development of the structured assurance case.
2. RQ2 / OB2 is addressed through the analysis of the benefits and challenges of SACM adoption, discussed in Section 4.6 and further summarized in Section 4.5.

Finally, Chapter 5 provides a synthesis of the findings, outlines the study's limitations, and offers recommendations for future work. This structured flow ensures a clear alignment between the research problem, objectives, methodology, and outcomes, thereby providing a coherent narrative for the thesis.