CETRA2026 27-30 May 2026, Bol, Island Brač, Croatia
9th International Conference on Road and Rail Infrastructure

LEVERAGING DATA-DRIVEN SAFETY ASSURANCE AND DIGITAL-TWIN VALIDATION FOR SAFE AND RELIABLE AUTONOMOUS PUBLIC TRANSPORT SYSTEMS

Tom Jansen, Rolf Graafland
Ricardo Nederland B.V., The Netherlands

Abstract

The deployment of Autonomous Vehicles (AVs) in public transport networks demands a new approach to safety assurance that can keep pace with fast-evolving software and communication technologies. Traditional certification frameworks are not designed for systems that continuously learn, exchange data, and update over their lifecycle. Based on lessons learned within actual commercial implementations, Ricardo has developed a data-driven safety assurance framework that combines best practices of functional safety engineering with state-of-the-art approaches such as digital twin validation and continuous verification. A digital twin, a dynamic virtual representation of a vehicle and its operating environment, enables extensive virtual testing of automated driving functions and communication interfaces, helping build confidence before real-world trials begin. By integrating real-time data from sensors, vehicle-to-infrastructure (V2X) links, and cloud analytics, the approach supports continuous comparison between simulated and actual performance to detect anomalies, predict failures, and maintain compliance with safety standards such as ISO 26262. This paper illustrates this methodology through practical experience taken from Ricardo's AV projects, including commercial autonomous shuttle deployments (MASDAR, Rivium Parkshuttle, Terhills), truck-platooning pilots, Automated Train Operation initiatives (including Europe's Rail R2DATO) and autonomous freight and terminal trucks. Taking these examples, the paper demonstrates how properly layered simulation, virtual validation, and cybersecurity assurance increase dependability, reduce testing costs, and accelerate regulatory approval.

Keywords: connected and automated vehicles, digital twin, functional safety, virtual validation, cybersecurity, data-driven certification, AI-powered infrastructure, predictive safety assurance, real-time monitoring

1 Introduction

Autonomous mobility development cycles are rapidly advancing, challenging linear verification and validation processes such as the classical V-model. The result is a widening assurance gap: by the time a waterfall phase is "closed", the system has already evolved, or new data has shifted the operational risk context. Gathering sufficient "mileage driven" based evidence in such an environment is an infeasible approach. While fragments of methods are available as part of the solution, a holistic lifecycle approach is currently lacking. While several previous works discuss individual methods, this paper provides a unified safety assurance framework that integrates these elements into a single and continuous safety assurance framework. It builds upon established functional-safety practices by incorporating more recent developments in data-driven and agile approaches, having validated their feasibility in multiple real-world commercial deployments.

This enables operators and authorities to establish a more scalable and evidence-based safety case compared to conventional mileage-based arguments.

2 Background and industry challenges

Traditional verification and validation approaches involve driving a significant number of real-world kilometers to demonstrate acceptable levels of reliability and safety. However, statistical estimations suggest that an autonomous vehicle (AV) would need to drive billions of kilometers to prove it is safer than a human driver [1]. This “distance-based” approach is economically unviable and time-prohibitive. Conventional development cycles also assume stable requirements and late-stage integration testing, making rapid software development and frequent changes difficult [3]. Furthermore, current development pipelines struggle to systematically anticipate “edge cases” - the rare, safety-critical events that often evade conventional testing approaches. Many high-profile accidents are triggered by these unforeseen situations, forming major “unknown unknowns” of autonomous developments. Additionally, reliance on deep learning architectures makes it difficult for engineers to diagnose faults and validate behavior under diverse real-world conditions [3].

3 A data-driven safety assurance framework

To address the limitations of traditional assurance frameworks, a multi-layered and continuous assurance framework is required. It should build upon digital verification approaches such as digital twins and cloud-based simulations to evolve the traditional frameworks towards data and scenario driven arguments rather than distance-driven arguments. At the same time, the framework should still maintain compliance with relevant safety standards such as ISO 26262 [4]. In this chapter such a framework is introduced, combining the contributions of modern verification and validation methods into a holistic lifecycle model based on safety assurance experience in actual AV deployment projects. An overview of this framework is presented in figure 1 below. A selection of essential aspects of this framework is explained in the upcoming chapters.

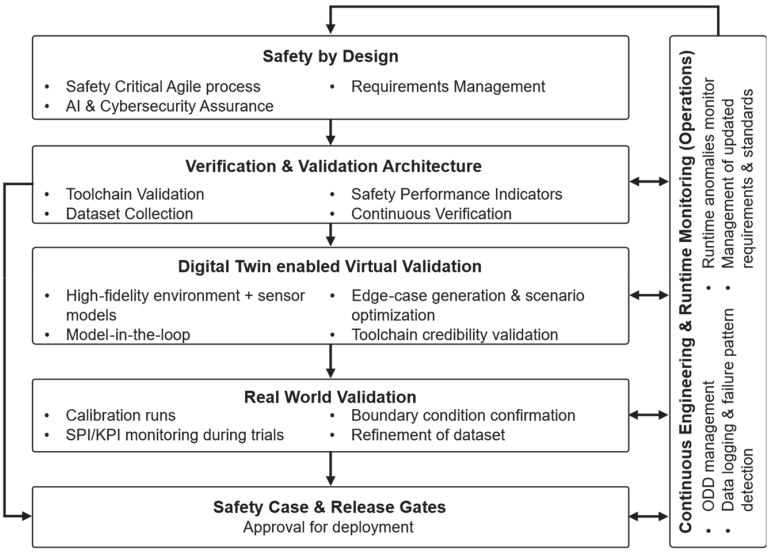

Figure 1 Overview of the data-driven safety assurance framework

3.1 Safety by design: enabling standards-based safety processes

The development and validation of automated systems are underpinned by a rigorous ecosystem of international standards across transport sectors. While such standards have been traditionally lagging behind, in recent times a largely suitable framework has become available for demonstrating the safety and compliance of autonomous public transport systems. For road vehicles, ISO 26262 now serves as the primary functional safety standard for electrical and electronic systems. Due to the fact that this standard formally does not consider autonomous vehicle developments, it is increasingly complemented by ISO 21448 (SOTIF), which focuses on identifying and mitigating hazards resulting from functional insufficiencies and performance limitations in complex environments [4]. In practice, this means that the safety analysis extends to more than just failures of the electronic systems and shall now also consider the suitability of such systems for the intended functionality.
To address further specific challenges related to the use of artificial intelligence, the emerging ISO/PAS 8800 standard provides a relevant framework for the safe integration of AI in vehicle systems. Finally, compliance with further regulations and standards such as UN-ECE 2022/1426 and ISO/SAE 21434 is required to meet broader approval, certification and cybersecurity obligations. Properly capturing such standards and regulations during requirements capture process ensures that a solid foundation is prepared against which the data-driven assurance approach can deliver the required evidence. Once captured, effectively managing requirements throughout the system lifecycle is essential. While this has traditionally been a “Waterfall” based sequential approach, modern software centric developments require shifting from rigid V-models to more Agile frameworks such SafeScrum, Scrum for Safety, or AgileAMLAS. These methodologies utilize iterative cycles to manage the complexity and uncertainty of machine learning and software-defined architectures, ensuring they fulfill all relevant requirements [5].
Requirement management fits into this model by using a hierarchical mapping where high-level objectives are broken down into implementable hierarchically structured work elements. Essential to successful requirements management in iterative approaches, is the use of high-quality requirements management tooling which enables attribution and linking of requirements throughout all lifecycles of the development. Specific Safety oriented requirements and backlogs help ensure end-to-end traceability and allow teams to manage dependencies across distributed workflows while ensuring compliance with safety standards such as ISO 26262 [6].

3.2 Multi-layered verification & validation architecture

One key aspect of any safety assurance approach is a robust verification and validation framework. While there are suitable variations of such frameworks available, they commonly consist of the following three primary layers [2]:

- V&V management layer: Defines the operational design domain (ODD), analyses the associated hazards and their safety mitigations, and derives functional and safety requirements. Based on these requirement types, a suitable scenario database can be derived to ensure the system can meet the operational needs. Additionally, key indicators such as Safety Performance Indicators (SPIs) should be defined, which enable monitoring the performance of the system and its components against the top-level safety requirements.
- Test management layer: Derives test cases and prepares them for execution while acting as a design-time monitoring layer to observe and evaluate components or systems. This layer uses the V&V management layer inputs to effectively monitor the performance of components and systems in their respective design stages and enables early detection of faulty behavior and improvements upon the established deficiencies.

- Testing platform layer: The virtual simulation environment (the digital twin) where the functionality is executed and monitored during run-time. It involves models of the developed system itself, including detailed sensor models and ego vehicle models, as well as representative world models for the environment and static and dynamic obstacles.

Essential for setting up a reliable V&V architecture is the validation of the toolchain, which must be achieved to establish credible simulation results. Calibration datasets are critical to perform in this regard, ensuring that the output of the simulation is as expected based on the calibrated dataset. Once calibrated, any updates to the toolchain will then critically need to adhere to strict change management and regression testing principles.

3.3 Digital twin enabled virtual validation

Digital Twin forms a virtual representation of a physical system and its operating environment that enables repeatable and controllable experimentation. In the context of AVs, digital twins allow for accessible validation at scale, at a fraction of the cost of physical testing. Virtual validation involves building detailed models of the vehicle and the environment (infrastructure, weather, traffic) [4]. For example, virtual radar sensors can be modeled using ray tracing to simulate how waves reflect off surfaces, allowing for the evaluation of object detection systems without needing physical vehicles in a testing area [7]. This reduces safety risks during the development phase and builds confidence before real-world trials commence. By implementing such Digital Twin simulations in a cloud-based environment, scenarios can be simulated with great parallelization benefits, leading to vast increases in efficiency – if the workflow is properly automated [8]. Virtual Validation tools should further employ multi-objective optimization algorithms to identify "unknown-unsafe" scenarios (i.e. edge cases), specifically targeting parameters that drive the system toward failure [9]. These scenarios are assessed using a defined set of SPIs [10] and KPIs capturing both safety-critical behavior and system performance degradation. These metrics can be defined based on the safety requirements derived during the safety engineering process, aligning with risk acceptance target values from quantitative analyses such as a Fault Tree Analysis.
Furthermore, the development process of trustworthy autonomy requires a continuous engineering loop. During testing, run-time monitors will observe system and component behavior against safety requirements. When a monitor detects a deviation, such as an object detection component failing to recognize an object in the test scenario, this data is fed back into the design process. Based on such monitoring loops, new datasets may be collected on which the model is retrained, and the improved system reapplied, closing the loop on safety assurance [2].

3.4 Real world validation

The final cornerstone of the framework is validation of the simulated results in a real-world environment. While the aim of the framework is to reduce this step as much as possible, a limited set of tests should be repeated in representative real-world scenarios to confirm the validity of the models used during simulation and the architecture of the simulation environment itself. In this step the simulated outputs are firstly compared against physical reference targets. Finally, the simulation is validated against the real world through analyzing the performances between virtual results and physical track testing [4]. In optimal cases, such tests can be mostly limited to demonstrating that scenarios which approach the established boundary values indeed still provide positive test results, thus confirming the validity of the safety concept.

Furthermore, during limited scale operational trials the established SPIs and KPIs should continue to be monitored, to confirm the continued safe operation of the AV in the actual set of real-world scenarios. In case of any misalignments occurring during the operational trials, accurate data shall be collected on such scenarios which should be used to refine the development dataset and improve the capabilities of the AV in a subsequent release.

3.5 Practical example

During one example project - a complex implementation of an autonomous shuttle system within a public transport environment in the Netherlands - frequent changes in the project requirements, combined with delays and necessary rework in both hardware and software development, created significant pressure on the deployment timeline. By applying the principles of the proposed framework, the project reduced its verification and validation cycle from the typical duration of several months to approximately 6 weeks, including a limited set of worst-case tests performed as real-world validation. This acceleration could primarily be achieved through strong management requirements and a virtual validation approach, which together provided an efficient process for determining the exact rework needed after each change. Such an approach is particularly valuable when requirements continue to evolve late in the project or when critical defects are discovered that necessitate significant rework.

4 Results and observations

The application of the approach proposed in this framework has led to several observations during its application in recent autonomous vehicle projects:

- A significant reduction in physical testing effort was often achievable, reducing real world testing time from months of duration testing to achieve edge-case coverage, to just several days of real-world validation.
- Early detection of safety critical edge cases was achieved by clear formulation of KPIs and SPIs. By basing these upon safety requirements and quantitative risk acceptance criteria, violation of the safety goals could be detected much earlier in the development process.
- Improved agility in the safety case was achieved, by combining safe agile design processes with strong requirements management and impact analysis. This meant that a change could be analyzed, and the design reworked more precisely and effectively, reducing the overall time to deployment.

These observations demonstrate that the approach proposed in this framework adds practical value to the autonomous vehicle development lifecycle, reducing project risks, costs and delays associated with traditional waterfall approaches.

5 Discussion

By combining the presented components of the data-driven safety assurance framework, significant time and effort can be spared during the development process of autonomous public transport systems. The experience gained in recent projects, combined with the available evidence from state-of-the-art research, leads to the following discussion topics:

- By utilizing the proposed framework, costly re-development cycles caused by incompliance or unforeseen safety relevant risks can be spared. In practice this means that projects tend to spend more time in the preparatory concept phases rather than rushing into the development process. However, project experience shows that this early effort greatly reduces the risk and costs of failure later in the project, eventually leading to an overall reduction in time-to-market and costs spent on redevelopment and recalls.

- Utilization of a cloud based digital twin validation approach enables much more efficient testing compared to real-world or manual testing. It enables the development team to identify and test edge cases early in the development cycle, rather than waiting for the perfect storm to occur during real-world testing. Additionally, cloud based high performance computing now allows digital twin simulation scenarios to be performed in parallel, significantly compressing the timeline for test campaigns [8].
- By extending the monitoring of SPIs towards the real-world validation and even towards commercial deployment, the safety performance of AVs can continue to be monitored. This will be especially valuable for systems containing on-board AI enabled components, for which the persistence of the input-output relationship will be increasingly hard to argue.

The current state-of-the-art for the employed methods still poses challenges and limitations which need to be considered:

- To overcome the potential gap with the real-world, it is essential to build the Digital Twin environment with highly accurate sensor models as well as with real-world scenario data and accurate point cloud data. Domain modeling relies on gathering high-accuracy infrastructure data through LiDAR-based 3D point clouds to create semantic maps and Digital Twins. This process formalizes the ODD, defining the environment where the system operates safely [7].
- Software and hardware must be decoupled to allow independent verification and continuous evolution post-production. Dedicated roles, such as Functional Safety Owners and RAMS engineers, work within the Agile teams to ensure independent verification and compliance with regulatory standards throughout the lifecycle [5]. This integrated approach allows for rapid response to change while maintaining the rigorous validation essential for safety-critical systems.
- One critical resource needed for representative Virtual Validation is a complete set of scenarios to be simulated. Such a set of scenarios has been complex and time consuming to compile, with the completeness being especially hard to establish. Several ongoing development projects for scenarios databases and their methodology are helpful in this regard, such as VVMethods, Hi-Drive Driving Scenario Database (DSDB) and the Safety Pool Scenario Database. At the time of writing however, no single source of scenarios can be considered complete by itself.

6 Conclusion

This paper contributes to the autonomous vehicle development lifecycle by integrating methods such as digital twin validation, scenario based testing and agile safety engineering principles into a single coherent framework. By shifting from a mileage driven argument towards evidence and coverage-based arguments based on data driven approaches, this framework provides operators and authorities with a more scalable approach to certifying complex autonomous systems. And by leveraging this framework, developers can: a) identify rare and critical edge case failures before real world deployment, b) achieve significant improvements in testing speed and cost efficiency, c) ensure continuous compliance with evolving safety and cybersecurity standards.

These benefits are essential to enable fast moving technological innovations such as autonomous public transport for large-scale deployments and to meet transportation demand. Future research will continue to bridge the gap between simulation and reality, and utilizing AI enabled infrastructure to support the large-scale rollout of autonomous public transport systems.

References

[1] Kalra, N., Paddock, S.M.: Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability, 2016.

[2] Esen, H., Liao, B.H.C.: Simulation-Based Safety Assurance for an AVP System Incorporating Learning-Enabled Components, arXiv preprint, 28.09.2023.

[3] Liu, Z., Zhang, W., Zhao, F.: Impact, Challenges and Prospect of Software-Defined Vehicles, Automotive Innovation, 5 (2022), pp. 180–194

[4] Forrai, A., Alirezaei, M., Singh, T., Gali, A., Ploeg, J.: Virtual Verification and Validation of Autonomous Vehicles: Toolchain and Workflow, IntechOpen eBooks, 2025.

[5] Hodge, J. H., Osborne, M.: Agile Development for Safety Assurance of Machine Learning in Autonomous Systems (AgileAMLAS), Array, 27 (2025)

[6] Martin, O., Fraser, L.A.S.: Continuous Requirement Refinement and Validation in Agile Automotive Development Pipelines, 2024.

[7] Haider, A., Jacobs, A., Ziegler, M., Wang, Y., Augustine, M. E., Schilling, R., Pfeiffer, C., Zeh, T.: Virtual Radar for the Railway of Tomorrow – Testing Automated Trains in a Virtual Environment, Signal + Draht, 10 (2025), pp. 35–39

[8] Samak, T., Samak, C., Martino, G., Nair, P., Krovi, V: Digital Twins in the cloud: a modular, scalable and interoperable framework for accelerating verification and validation of autonomous driving solutions, 19 May 2025.

[9] Zhang, X., Khastgir, S., Asgari, H., Jennings, P.: Test Framework for Automatic Test Case Generation and Execution Aimed at Developing Trustworthy AVs from Both Verifiability and Certifiability Aspects, IEEE International Intelligent Transportation Systems Conference, pp. 312–319, Indianapolis, IN, United States, 19-22 September 2021.

[10] Koopman, P.: Safety Performance Indicators (SPIs) for Autonomous Vehicles, Carnegie Mellon University, 2022.