SURFACE VEHICLE INFORMATION REPORT

J3321™	MAR2026
Issued 2026-03	

Verification and Validation of AI/ML-Based Systems in Ground Vehicles

RATIONALE

Software and hardware components realizing Artificial Intelligence (AI) and Machine Learning (ML) algorithms are becoming commonplace in next-generation ground vehicle applications and services. Many of these applications will have a bearing on the safety of vehicles. The development life cycle of AI/ML applications is atypical in the sense that it is highly data driven, performs high-level, complex human-like tasks (e.g., perception, recognition, comprehension, decision, etc.), and is opaque by design. Typical use cases involve human-centric concepts and domains (e.g., roadside objects, lanes, traffic signals, pedestrians etc.), each of which have different levels of abstractions. Robustness and certainty of the performance of AI-based systems are some of the major issues facing such systems.

The systematic verification and validation (V&V) of AI-based systems is therefore critical when these systems are used in automotive applications and features impacting the trustworthiness, including safety, of the vehicle functions. As such, the V&V of AI-enabled components and systems requires new, enhanced approaches as existing methods are often narrowly focused and relatively immature. Additionally, end users and practitioners face challenges due to insufficient experience in developing new AI-based systems, the absence of uniform industry standards, and a lack of focused and reliable commercial tools. Given the complex network of OEMs and suppliers in the automotive domain, this lack of a standardized approach can very often lead to the use of ad hoc techniques for V&V with varying degrees of component/system performance and reliability. Fortunately, recent and ongoing developments by both academic and industrial researchers in this area have led to new insights into the formal and standardized approaches for V&V of AI/ML applications.

The purpose of this information report is to:

- Review the current state-of-the-art (SOTA) in the verification and validation (V&V) of AI/ML-based systems and to provide a comprehensive guide for incorporating these methods into the systematic V&V of automotive AI-based systems.
- Identify gaps and deficiencies in the existing V&V approaches with respect to their application to AI-based SW and define requirements for their enhancements.
- Provide guidelines on the application of existing and new approaches for the V&V of AI-based components/systems.

SAE Executive Standards Committee Rules provide that: “This report is published by SAE to advance the state of technical and engineering sciences. The use of this report is entirely voluntary, and its applicability and suitability for any particular use, including any patent infringement arising therefrom, is the sole responsibility of the user.”

SAE reviews each technical report at least every five years at which time it may be revised, reaffirmed, stabilized, or cancelled. SAE invites your written comments and suggestions.

Copyright © 2026 SAE International

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, or used for text and data mining, AI training, or similar technologies, without the prior written permission of SAE.

TO PLACE A DOCUMENT ORDER: **Tel: 1-877-606-7323 (U.S. and Canada only)**
Tel: 1-724-776-4970 (outside U.S. and Canada)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: **http://www.sae.org**

For more information on this standard, visit
https://www.sae.org/standards/content/J3321_202603/

Documents Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

TABLE OF CONTENTS

1. SCOPE 4
1.1 AI/ML Systems 4
1.2 V&V of AI/ML-Based Systems 4
1.3 Quality and Dependability 5

2. REFERENCES 6
2.1 Applicable Documents 6
2.1.1 SAE Publications 6
2.1.2 ISO Publications 6
2.1.3 Other Publications 6

3. DEFINITIONS 7

4. AI/ML DEVELOPMENT LIFE CYCLE 8
4.1 Problem Definition and Requirements Description 8
4.2 Data Acquisition and Preparation 9
4.3 Model Development and Training 9
4.4 Model Evaluation and Refinement 9
4.5 Model Deployment and Monitoring 9
4.6 MLOps 10

5. VERIFICATION AND VALIDATION (V&V) OVERVIEW 10
5.1 Problems/Challenges in AI System Verification 10
5.1.1 Unpredictable Behavior 11
5.1.2 Robustness 11
5.1.3 Bias and Fairness 11
5.1.4 Explainability/Transparency 11
5.1.5 Cybersecurity 11
5.1.6 Safety 11
5.2 AI System Verification 12
5.2.1 Requirements and Properties of AI/ML systems 12
5.3 Process Verification 13

6. V&V TECHNIQUES 13
6.1 Requirement/Property-Based Testing 14
6.2 Verification of KPI Metrics 15
6.3 Testing for Robustness and Resilience 15
6.4 Explainability and Interpretability Testing 16
6.5 Bias Testing 16
6.6 Interface/Structure-Based Testing/Verification 16
6.7 Latent Feature Testing 17
6.8 Architectural Analysis and Verification 18
6.9 Failure/Risk-Based Testing 18
6.10 Adversarial Perturbation 18
6.11 Distribution Shift Testing 18
6.12 Safety Counterfactuals 19
6.13 Scenario-Based Testing 19
6.14 Safety Zone-Based Testing 19
6.15 Simulation-Based Testing 19
6.16 Hardware-in-the-Loop (HIL) Testing 19
6.17 Data Augmentation and Validation 20

7. TRACEABILITY IN THE AI V&V PROCESS 20

8. GENERATIVE AI AND V&V PRACTICES 22
8.1 Brief Overview of LLM and Generative AI Technologies 22
8.2 GenAI Use Cases in Automotive Applications 23
8.3 V&V Practices 23

AVA Documents Document Sharing Hub

Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

9. CONCLUSIONS 25

10. NOTES 25
10.1 Revision Indicator 25

APPENDIX A 26

Table 1 5

Documents
Document Sharing Hub

Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

1. SCOPE

1.1 AI/ML Systems

This document intends to focus on AI/ML systems deployed on board and/or off board the vehicles, including the methods and tools used anywhere in the life cycle of automotive system development. This includes, for instance, AI/ML systems used in the identification of parameters, calibration of classical or neural network-based control algorithms, system level modeling and validation, data process automation pipeline such as automated labeling and annotation, data quality checks, etc. Furthermore, with the continued success of Generative AI (GenAI) capabilities, it is becoming a common practice to leverage these capabilities beyond just generating synthetic data. Multiple artifacts such as code, testing scenarios, and benchmarks could be developed using GenAI tools, necessitating the additional verification of these “generative” development steps. The fundamental challenge with generative methods is grounding, especially if data, scenes, and tests are used for the development of AI-enabled products and features.

AI and ML systems have been the subject of several national and international standards and guideline documents. A brief selection of these are given in Section 2. One, ISO PAS 8800:2024, focuses on safety over the entire life cycle of AI development and deployment.

This document is intended to serve the following purpose:

a. The document is an information report with no mandatory requirements.

b. The focus of the document is only on V&V of the development life cycle of AI-based components or systems.

c. The V&V methods are general in nature. The methods discussed include, but are not restricted to, safety requirements or properties.

1.2 V&V of AI/ML-Based Systems

V&V of AI/ML-based components and systems differs from that of conventional software and systems and requires special attention. Unlike conventional software, AI/ML-based systems often use very complex, high dimensional, and unstructured information sources as inputs (e.g., images, video, audio, natural language, etc.). Additionally, these systems often grapple with human-based concepts of cognition and reasoning, perform functions such as analysis or processing (e.g., detection of vehicles, pedestrians and animals, prediction of accidents and traffic jams, understanding speech and text), and perform analysis of sentiments and prediction of accidents, etc. More importantly, these systems can assist with decisions or increasingly take decisions autonomously (agentic AI). These systems often employ probabilistic algorithms or models (neural networks, graphs, Bayesian networks) and produce stochastic outputs with associated confidence levels. For such systems, changes in the model inputs that are outside of the model training/validation dataset may produce different results that are misaligned with the expectation. The degree of deviation is a function of the model capacity or expressivity and a function of distance of the distribution of the perturbed inputs relative to the training/validation data. Such out of distribution (OOD) of shifted inputs is often observed via real-world phenomenon or events that may be subjected to uncertainties and noise. The models themselves are constructed using a large but often incomplete sample representation of the real-world inputs which also can/will vary over time. The models contain parameters and rely on stochastic optimization methods to compute appropriates values for these parameters that optimize an objective function (e.g., loss function, likelihood function).

In view of the above differences relative to traditional deterministic models, there is a strong need for reviewing existing V&V methods and developing new methods and/or adapting existing methods and best practices for a robust V&V of data-driven AI models.

Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

Table 1 summarizes the scope of this document.

Table 1

Element	In-Scope Elements	Out-of-Scope Elements
Ground Vehicle Life Cycle Stages	Research, development, system integration, V&V, use case validation, deployment, operation in target ODD	Non-AI-related system level V&V, vehicle performance validation
AI/ML Development Life Cycle Stages	Analysis and validations of artifacts at different stages of the life cycle, e.g., requirement analysis, dataset validation, model verification, validation of training and implementation, and field monitoring	Design, development methods and tools, target platform selection, code generation
Class of AI solutions	Machine learning, computer vision, deep learning, generative AI, natural language processing	End-to-end learning, expert systems
Class of AI/ML Techniques	Supervisory & unsupervised learning, ANNs, CNNs, DNNs, transformer networks, and reinforcement learning	Symbolic AI, semantic web-based approaches
V&V Attributes	Functionality V&V, property-based verification, robustness, absence of bias, generalizability, explainability, uncertainty quantification, resilience, trustworthiness, safety, security	Functional safety testing for certification, security verification for certification
Types of Tools and Their Categories	Dynamic testing tools, formal verification tools, neuro-symbolic tools	Offline generative AI tools used for code generation

1.3 Quality and Dependability

The overarching theme of this report is to provide detailed information on V&V methods for ensuring quality and enabling the development of reliable AI-enabled systems. There are several quality attributes expected by the end users and included either implicitly or explicitly in the requirements of a system. The quality requirements identify several quality attributes that the system is expected to have and, depending upon the context and application, some or all of these attributes that the system under test is expected to have. In the context of AI/ML applications, some of the quality attributes that are often emphasized are robustness, generalizability, and explainability, in addition to accuracy, correctness, and performance. For the sake of keeping the report succinct, it focusses less on the details of different quality attributes and more on their verification. We further restrict the report to the V&V of only a few selected quality attributes, which include functionality, safety, and robustness. We rely on some recent standards and reports on the broader topic of quality, reliability, and dependability published by the ISO/SAE/IEC/IEEE communities. These include three ISO publications: ISO/IEC 25010:2023, ISO/IEC TS 25058:2023, and ISO/IEC 25059:2023.

Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

2. REFERENCES

2.1 Applicable Documents

The following publications form a part of this specification to the extent specified herein. Unless otherwise indicated, the latest issue of SAE publications shall apply.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (U.S. and Canada only) or +1 724-776-4970 (outside U.S. and Canada), www.sae.org.

SAE J3298 Artificial Intelligence Data for Ground Vehicle Applications

2.1.2 ISO Publications

Copies of these documents are available online at www.iso.org/store.

ISO 26262 Road vehicles - Functional safety

ISO/IEC 25010:2023 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Product quality model

ISO/IEC TS 25058:2023 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Guidance for quality evaluation of artificial intelligence (AI) systems

ISO/IEC 25059:2023 Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems

ISO/PAS 8800:2024 Road vehicles - Safety and artificial intelligence

2.1.3 Other Publications

[1] Papers on AI Safety. Available at https://aisafety.stanford.edu/.

[2] Abbas, A. et al. (2014). *User-Centered Design (UCD) for Automotive User Interfaces*.

[3] Abbas, A. et al. (2017). *Scenario Based Requirements Definition for Autonomous Vehicles*.

[4] Abdollahpouri, H. et al. (2019). *Managing Popularity Bias in Recommender Systems with Personalized Re-ranking*. Available at https://arxiv.org/abs/1901.07555.

[5] Brynjolfsson, E. et al. (2011). *Goodbye Pareto Principle, Hello Long Tail: The effect of Search Costs on the Concentration of Product Sales*. https://dspace.mit.edu/handle/1721.1/74642.

[6] Carmichael, A. et al. (2017). *Guidelines for Model-Based Systems Engineering (MBSE)*.

[7] Chen, T. et al. (2020). Machine Learning for Vehicle Electronics Control: Promise and Challenges. *IEEE Transactions on Intelligent Vehicles*.

[8] Eberle, A. et al. (2020). A Survey on Safety Engineering for Autonomous Vehicles. *ACM Computing Surveys*.

[9] Fleder, D. et al. (2009). *Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity*. Available at https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0974.

Documents
Document Sharing Hub

Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

[10] Goodrich, M. et al. (2016). *Moral Machines: Ethical Reasoning in Autonomous Vehicles*.

[11] Gupta, S.C. and Sinha, M.K. (1994). *Improving Software Testability by Observability & Controllability Measures*. 13th World Computer Congress, IFIP Congress 1994.

[12] Hoda, R. et al. (2016). Requirements Engineering for Safety Critical Systems: A Systematic Literature Review. *Information and Software Technology*.

[13] Huang, X. et al. (2020). A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defense, and interpretability. *Elsevier Computer Science Review*.

[14] International Council on Systems Engineering (INCOSE). (2010). *INCOSE Systems Engineering Handbook v4.0*.

[15] Katz, G. et al. (2019). *The Marabou Framework for Verification and Analysis of Deep Neural Networks*. Available at https://theory.stanford.edu/~barrett/pubs/KHI+19.pdf.

[16] Katz, G. et al. (2017). *Reluplex: an efficient SMT solver for verifying deep neural networks*. Majumdar, R. and Kunˇcak, V. (Eds.). CAV 2017. *LNCS, 10426*, 97-117. Springer.

[17] National Highway Traffic Safety Administration (2020). *Guidance for Manufacturers on Crash Avoidance Technology Testing and Reporting*.

[18] Niu, B. et al. (2018). *Safety Requirements Engineering for Autonomous Vehicles*.

[19] SAE J3298, Artificial Intelligence Data for Ground Vehicle Applications.

[20] nuScenes Detection Leaderboard. Available at https://www.nuscenes.org/object-detection.

[21] Topan, S. et al. *Interaction-Dynamics-Aware Perception Zones for Obstacle Detection Safety Evaluation*. Available at https://arxiv.org/abs/2206.12471.

[22] Bansal, S. et al. (2017). *Hamilton-Jacobi Reachability: A Brief Overview and Recent Advances*.

3. DEFINITIONS

AI MODEL: A construct containing logical operations, arithmetical operations, or a combination of both to generate an inference or prediction based on input data or information without being completely defined by human knowledge. [Source: ISO/PAS 8800:2024, 3.1.7]

AI SYSTEM: A system consisting of hardware/software components containing one or more AI models.

NOTE 1: An AI system may sometime be referred to as a component when it is part of larger system.

NOTE 2: An AI system would sometime be referred to as an AI/ML system.

DYNAMIC TESTING: Executing or running the software or system under consideration with specific inputs and conditions and checking the response/output of the software or system.

EDGE CASE: An input that causes the component to give a performance (output or behavior) that is close to or same as the performance limit of the component.

REACHABILITY ANALYSIS: A static verification technique that enumerates explicitly or implicitly the state space of a software or system and checks whether there exists a path from one or more initial states to a state where a specific condition holds.

Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

SCENARIOS: A set of one or more inputs and environmental conditions and the corresponding outputs of a component or system under consideration such that the inputs and environmental conditions directly or indirectly influence the performance of the system/component.

STATIC VERIFICATION OR ANALYSIS: Analyzing any of the quality attributes (e.g., functionality, performance, safety, etc.) of a software or system without running or executing the system. Typically, a (computer representation of the) model of the software/system is constructed, and an algorithmic procedure is applied on the model to compute the given quality attribute(s).

TEST CASE: A pair of inputs and outputs of the software or system which when executed with the prescribed inputs produces the expected outputs.

USE CASES: A specific set of scenarios expected of the system/component under consideration.

VALIDATION: Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled. [Source: ISO/PAS 8800:2024, 3.5.9]

VERIFICATION: Confirmation, through the provision of objective evidence, that specified requirements have been fulfilled. [Source: ISO/PAS 8800:2024, 3.5.10]

4. AI/ML DEVELOPMENT LIFE CYCLE

The successful development of AI/ML systems depends on a well-defined development life cycle with clear requirements at each stage. ISO/PAS 8800:2024 provides guidelines and recommendations on the different life cycle steps, especially in the context of safety-critical automotive systems. This section includes a brief overview of a typical AI/ML development life cycle, often referred to MLOps. The reader is referred to other ISO standards for detailed discussions on this topic.

4.1 Problem Definition and Requirements Description

Any ML model building exercise begins with a clear definition of the problem to be solved. The problem not only dictates the model architecture but also defines the data needs. The problem definition, in the context of safety-critical automotive applications, includes:

- Nominal function and safety concept expected to be achieved by the solution.
- Specific goals and success metrics such as Key Performance Indicators (KPI) and Measure of Effectiveness (MOE) of the solution.
- Clearly enumerated user requirements and performance expectations along with any underlying assumptions are also defined.
    - In the case of vehicle level features, the Operational Design Domain (ODD) of the feature should also be defined.
- A feasibility assessment considering technical capabilities necessary, data requirements, and availability/sourcing is also defined. Any related ethical implications are also considered.

Some more details on carrying out this step in the context of AI/ML-based components can be found in Section A.1.

Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

4.2 Data Acquisition and Preparation

Data is critical to any ML model and requires the implementation of robust data governance as part any ML development cycle.

- High-quality, representative, and diverse data aligned with the problem definition.
- Sufficient data volume and diversity for developing (training and testing) robust models and testing their performance in real-world situations.
- Data cleaning procedures to address missing values, inconsistencies, and errors.
- Data labeling (if applicable) for supervised learning tasks.
- Established data governance practices for secure storage and access control.

4.3 Model Development and Training

- Selection of appropriate AI/ML algorithms and model structure based on the problem type, run time constraints, and data characteristics.
- Feature engineering to extract meaningful insights from the data.
- Training and hyperparameter tuning to optimize model performance.
- Model validation, testing, and error analysis.
- Model interpretability methods and metrics that help to understand model behavior and decision-making.

4.4 Model Evaluation and Refinement

Typically, this step involves techniques for assessing how a model will generalize to an independent dataset.

- Rigorous evaluation methods using techniques such as hold-out or cross-validation.
- Performance metrics aligned with the defined success criteria (e.g., accuracy, precision, recall).
- Techniques to identify and address model bias and fairness issues.
- Iterative process of refining the model based on evaluation results and new data, if any.

4.5 Model Deployment and Monitoring

This step involves assessing in-use performance of a developed model when deployed on the application hardware.

- Robust infrastructure for deploying the model in a production environment.
- Scalable computing resources to handle real-time predictions.
- Continuous monitoring of model performance to detect degradation or drift in model and data.
- Automated retraining procedures to refresh models and maintain model accuracy.
- HW profiling for model performance in single and batch analysis, including TOPS, memory, throughput, and inference efficiency.

In-use field monitoring forms an important activity in the life cycle of next-generation software systems. Some details on the why's and how's of field monitoring in the context of AI/ML-based components are given in Section A.2.

Documents Document Sharing Hub

Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

4.6 MLOps

- Streamlined workflow for managing the entire AI/ML life cycle.
- Tools for version control, collaboration, and model deployment.
- Automated pipelines for data processing, model training, and monitoring.
- Continuous integration, deployment, and monitoring.

5. VERIFICATION AND VALIDATION (V&V) OVERVIEW

The development life cycle of AI/ML systems includes V&V activities.

As defined in ISO/PAS 8800:2024 and elaborating in our context, verification is the process of confirming that the AI/ML system meets its performance requirements or specifications as intended by design at all stages of development. In contrast, validation assesses whether the units, components, elements, or subsystems meet the requirements or specifications at any given hierarchical integration level. Both verification and validation are to be accompanied with supporting evidence in the form of analysis, simulations, and testing such that a rational argument, confirming that the claim of compliance with the requirements, is justified.

Verification focuses on the AI system under test and involves ensuring that the system meets its requirements. Besides checking the correctness of the system under test, verification activities include process verification that involves checking that all the process steps involved in the development of the component are carried out correctly. This includes analysis of the requirements for consistency and completeness, decomposition of requirements into subsystem requirements, etc. In the case of AI-based systems, the development involves identifying appropriate ML architecture, creating a correct dataset for training, and validating the AI model.

Validation, on the other hand, focuses on assessing the alignment of the model's performance to the intended purpose. Validation takes an external view of component under test and hence focuses on testing the end system or application integrating the component. It considers the requirements and constraints of the end application and employs application-level tests that are traceable to the requirements. Failure of the application-level tests are expected to be traceable to the potential discrepancies and gaps, if any, in the assumptions and/or commitment of the AI system.

The methods and tools used for testing will depend upon the end application. For instance, if the end application is an ADS feature, then testing exercises the behavior of the feature under the different ODD conditions and evaluates the performance of the overall system. Testing should also identify potential issues in the AI/ML component that impacts the system performance. To provide feedback to the AI system developer, test cases sufficient in terms of quality and quantity are developed that have sufficient coverage of the interface constraints. Any behavioral-level discrepancies at the applications level are then traced to the issues at the component level. While it is desirable to run the tests on the end application, some early validation may be carried out using proxy models of the end system or feature which would provide continuous feedback to the AI system developer.

5.1 Problems/Challenges in AI System Verification

V&V of AI systems are critical processes to ensure that AI models are reliable, accurate, and trustworthy, especially in safety-critical industries such as healthcare, aerospace, and automotive. There are general challenges in applying traditional verification methods to AI systems, which include:

- Black box nature of the AI/ML model: While the dataset is known, far less is known of the trained model, especially the features that are being learned of the dataset. This can be one of the greatest challenges in AI/ML V&V.
- Complexity of formal verification: Due to the high-dimensional nature of the models, performing rule-based formal verification methods are difficult.
- Stochastic behavior: Due to stochastic learning process, two similar trained models can exhibit different outputs on a test dataset. While this could be adjacent to unpredictable behavior, it is quite different.

AVA Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

Specifically, from an automotive perspective, we have the following problems or challenges.

5.1.1 Unpredictable Behavior

AI systems can exhibit unpredictable or unexpected behavior influenced by domain and/or model drifts. Drifts that cause the model behaviors to differ significantly from base design can become problematic, making it difficult to ensure consistent and reliable operation, especially in safety-critical applications like self-driving cars and ADAS applications.

5.1.2 Robustness

Robustness in AI involves the system's ability to handle unexpected or extreme situations. For example, an autonomous vehicle might encounter a sudden road closure due to an accident and needs to quickly reroute the vehicle without human intervention.

In addition, AI robustness also refers to the system's sensitivity to small changes in input. For instance, in the context of adversarial robustness, even a tiny perturbation in the input - imperceptible to humans - can lead to a drastically different output from the model. Similarly, a change in hardware, such as switching to a different camera sensor, may slightly alter the input image. A robust AI system should maintain consistent performance despite such variations, ensuring reliability across different environments and conditions.

5.1.3 Bias and Fairness

Fairness ensures AI systems demonstrate equitable behavior. For example, an AI system in a vehicle avoids biased driving decisions that precludes certain classes of road objects or scenarios from certain neighborhoods.

5.1.4 Explainability/Transparency

Many AI models, especially deep learning models, are mostly opaque, making it difficult to understand how they arrive at specific decisions. Explainability/transparency in AI systems refers to the ability to trace and understand decision-making processes. An example is providing clear documentation and explanations for how a self-driving car's algorithms make decisions in real-time traffic situations.

5.1.5 Cybersecurity

In autonomous vehicles, cybersecurity is crucial to prevent unauthorized access from malicious actors. For instance, a car's communication system could be hacked to send false data to the vehicle's navigation system, potentially causing it to make unsafe decisions. Other issues include data poisoning, adversarial change of weights and biases, etc.

5.1.6 Safety

AI-based systems in vehicle features, e.g., automated driving, may perform crucial functions like pedestrian or object detection. Any misdetection arising out of the failure of the AI system may pose a threat to the safety of the vehicle and/or other road occupants. Ensuring vehicle safety requires the need for comprehensive testing to ensure that autonomous vehicles (AVs) are safe despite cyber-threats, operate transparently, behave robustly in various conditions, and function without bias. There are also ethical considerations associated with the application of AI, including automotive systems. These include issues like fairness and transparency/explainability as defined above. The ethical concerns related to AI algorithms can perpetuate biases if not properly managed, and the importance of designing AI systems that are fair and transparent to all users can hardly be underemphasized. For instance, ensuring that AI-driven vehicles do not discriminate against certain groups of people and that their decision-making processes are understandable to humans are important requirements of AI-based components used in ground vehicle applications.

Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles

5.2 AI System Verification

AI system verification involves checking that all its stated requirements are satisfied. One standard approach to verification is dynamic testing that involves developing one or more test cases for each requirement and exercising these test cases on the system under test. But this approach has several challenges when applied to AI systems. The inputs to AI-based applications typically involve complex human-based artifacts (e.g., natural language statements, images, video, etc.) and high-level semantic concepts such as pedestrians, lanes, etc. These are often multidimensional and multimodal and are subject to variations and noise. Generating test cases to capture these variations is an enormous task.

The safety requirements of AI systems include some unique aspects such as functional insufficiencies and uncertainties in the output computations. The overall system using the AI system should be able to meet its requirements with the specified uncertainties and insufficiencies in the AI/ML subsystems. The requirements need to include specification of the various trigger conditions in the environment (e.g., fog on the roadway) that causes the insufficiency.

Another distinguishing aspect of AI/ML systems that requires newer approaches is the verification of datasets used in the training and evaluation of the system. The requirements at the AI system level are appropriately translated into necessary properties of the training and validation datasets, which needs to be verified before the datasets are used in the development.

Verification is expected to provide objective evidence that the AI/ML system meets the requirements. The evidence is typically in terms of analysis results and a set of test cases passed by the system. To strengthen the evidence, the verification process includes examination and evaluation of the internal architecture of the component, design steps, and processes involved in the development of the component under question. The requirements at the overall system level are decomposed into requirements at the individual process steps which, in turn, are verified with respect to their requirements.

The AI system often interfaces with the rest of the application, and the verification is carried out at the interface boundaries. Either the full end application with their interfaces, their proxies, or models would be required for verification. Traceability and coverage of component level requirements are important considerations in verification. The nature of the requirements and the extent of verification are unique to AI systems. For instance, the requirements of an AI system may include the various properties peculiar to AI-based systems.

5.2.1 Requirements and Properties of AI/ML systems

V&V involves checking whether an AI/ML system satisfies a set of requirements which includes functional, safety, and security requirements. These requirements state a variety of properties; the chief among them is the functionality or performance properties. Several other general properties are expected to be satisfied by AI/ML components, and these include robustness, freedom from bias, generalizability, explainability, resilience, and trustworthiness.

Functional requirements often involve some KPI or metrics, and the component is expected to meet these metrics. The metrics depend upon the task being performed by the ML component. Accuracy, precision, and recall in classification tasks, intersection over union for segmentation, BLEU score, and human evaluation metrics for Large Language Models (LLM) are some of the examples. A distinguishing aspect of requirements of ML components is the uncertainty specification. Unlike the traditional component, an ML component cannot ascertain with total confidence its output. There will always be an uncertainty associated with the output. For instance, a classification network used in a pedestrian detection task is expected to provide probabilities of detection or non-detection rather than total assurance. Any meaningful perception system making use of this network is expected to look at these probability values and take appropriate actions. The uncertainty may arise out of many reasons that includes the nature of the semantic tasks performed (epistemic uncertainty, i.e., limitations of our knowledge) and the way in which these tasks are developed (aleatoric uncertainty, i.e., incompleteness of the methods and tools). Verification of such complex requirements need to provide sufficient evidence to the expected probabilities as well.

Robustness properties are used to assess how well an AI/ML system performs when the input is altered or degraded. This is one of the important practical requirements of many AI/ML-based components in automotive systems. For example, the components that process images of road scenes are expected to be robust under varying distortions, noise, weather, and lighting.

While robustness achieves variations of smaller scale, bias freedom ensures that the system under test can comprehend the variability found in high-level features of the inputs. For example, a pedestrian detection system needs to detect pedestrians irrespective of their size, shape, skin color, attire, pose, etc.

Documents
Document Sharing Hub
Full document details: SAE J3321-2026 Verification and Validation of AI/ML-Based Systems in Ground Vehicles