Document ADS-20-08
20th ADS IWG Session
27 April – 1 May 2026 (Bangkok)

Testing GID

Note:

The document uses the terminology "assessor" to refer to the actor performing the safety case assessment and the confirmatory testing. Under UNR, "assessor" should be understood as the "Type-Approval Authority or the Technical Service on its behalf". Under GTR, the "assessor" will be defined by the applicable approval/certification regime.

Location:

- 5.3.2.15, 6.3.2.2.3 GTR;
- 7.3.2.15, 8.3.2.2.3 UNR;

GTR/UNR 5./7.3.2.15.: *The safety concept shall describe the manufacturer's approach to scenario selection to cover the reasonably foreseeable situations and conditions that the ADS will encounter including how the following aspects are covered:*

*(a) The selection of **sufficient** scenarios in which the ADS needs to initiate a fallback response (e.g., approaching the ODD boundaries),*

GTR: 6.3.2.2.3.: *The assessment shall verify that the set of scenarios and situations resulting from the manufacturer's scenario generation and identification process is suitable for demonstrating the ADS safety case. This includes covering reasonably foreseeable situations and conditions that the ADS will encounter during its real-world operations. In particular, the assessor shall verify that the set of scenarios and situations selected as evidence to support the ADS safety case includes:*

(a) Scenarios and situation in which the ADS needs to initiate a fall-back response (e.g., approaching the ODD limits), and…..

UNR: 8.3.2.2.3.: *The approval authority or its designated technical service shall verify that the set of scenarios and situations resulting from the manufacturer's scenario generation and identification process is suitable for demonstrating the ADS safety case. This includes covering reasonably foreseeable situations and conditions that the ADS will encounter during its real-world operations. In particular, the approval authority or its designated technical service shall verify that the set of scenarios and situations selected as evidence to support the ADS safety case includes:*

(a) Scenarios and situation in which the ADS needs to initiate a fall-back response (e.g., approaching the ODD limits), and…..

Subject: Interpretation of the concept of "sufficient" scenarios for fall-back response as part of safety case testing

Interpretation

In the context of this requirement. 'sufficient' refers to two aspects:

1) The number and type of scenarios in which a fall-back response is initiated by the ADS feature;

2) The number of users participating in assessing the effectiveness of the user interaction aspect of the ADS fall-back process.

'Sufficient' in both 1) and 2) refers to the scope of the tests that have been performed on the fall-back response. If either aspect is not 'sufficient', then the evidence that is generated may be considered as insufficient to support the claim.

The manufacturer is expected to use proper statistical techniques to justify the sufficiency of their evidence to support a reliable inference.

With respect to the number and type of scenarios, the manufacturer is expected to describe the reasoning behind their choice of scenarios, why they claim they are representative, and the justification for the overall number chosen. It is acknowledged that the number and diversity of relevant scenarios are inherently dependent on the intended use case and its operational design domain.

With respect to the number of users, statistical methods have been developed to assess whether the sample size is 'sufficient'. These methods focus on the sample size in terms of the number of participants. The manufacturer is expected to provide calculations that determine and explain its choice for the number of participants and what assumptions are used in the calculations. It is acknowledged that the choice of the number of participants also depends on what an ADS user needs to do / understand. In case of an ADSF-1 a fall-back user needs to do and understand different things than the passenger of a vehicle with an ADSF-2.

Subject: Interpretation of the representativeness of the other road user for the assessment of the safety case testing activities

Location:

- 6.3.2.4.1.4 GTR;
- 8.3.2.4.1.4 UNR.

GTR 6.3.2.4.1.4.: *"For the specific case of ADS interaction testing, the assessment shall:*

*(a) Verify that the people involved are **representative** of the expected general population of ADS users and other road users where applicable,…."*

UNR 8.3.2.4.1.4.: *"For the specific case of ADS interaction testing, the approval authority or its designated technical service shall:*

*(a) Verify that the people involved are **representative** of the expected general population of ADS users and other road users where applicable,…."*

Interpretation

The concept of representativeness aims to ensure that tests involving other road users provide credible, generalisable evidence about ADS behaviour in real traffic. The reference to "representative of the expected general population of ADS users and other road users" is intended to prevent safety claims being based on narrowly selected participants, atypical or overly compliant behaviour, or participants making choices that do not reflect the diversity of behaviours encountered in the real-world. Testing involving other road users is intended to capture the spectrum of normal and unusual behaviours that the ADS might encounter in real traffic so that the response of the ADS can be properly investigated.

In this context, representativeness does not mean a perfect demographic mirror of society. Instead, it means that the characteristics and behaviours of other road users involved in testing are diverse but sufficiently typical of those the ADS can reasonably be expected to encounter, such that conclusions drawn from the tests are valid.

Representativeness has three complementary dimensions:

1) Behavioural representativeness, which ensures behavioural patterns of other road users reflect real traffic behaviour (e.g., compliance and non-compliance with traffic rules, reaction times and decision-making variability, assertive, cautious, inattentive, or ambiguous behaviour, natural interaction dynamics (e.g., gap acceptance, yielding, hesitation). Assuming all other (mainly human) road users follow "traffic rules" is unlikely to be a realistic assumption unless the ODD is limited. Animal behaviour may also be complex, e.g., moving as a pack, herding, pack splitting up, etc.)

2) Population representativeness, which ensures that the group of other road users is representative of the expected population (e.g., age range, etc.), physical characteristics relevant to perception (e.g., height, clothes, etc.), mobility characteristics (e.g., walking speed, cycling style, etc.), driving experience and style of driving. Animals also come in many shapes and sizes and may

present different aspects to the vehicle or be covered in “materials”, e.g., hair, fur, etc. with different material properties.

3) ODD-specific relevance, which ensures that the representativeness is always relative to the declared ODD.

Subject: Interpretation of the scope of Confirmatory testing

Location:

- 6.3.3.1 GTR;
- 8.3.3.1 UNR.

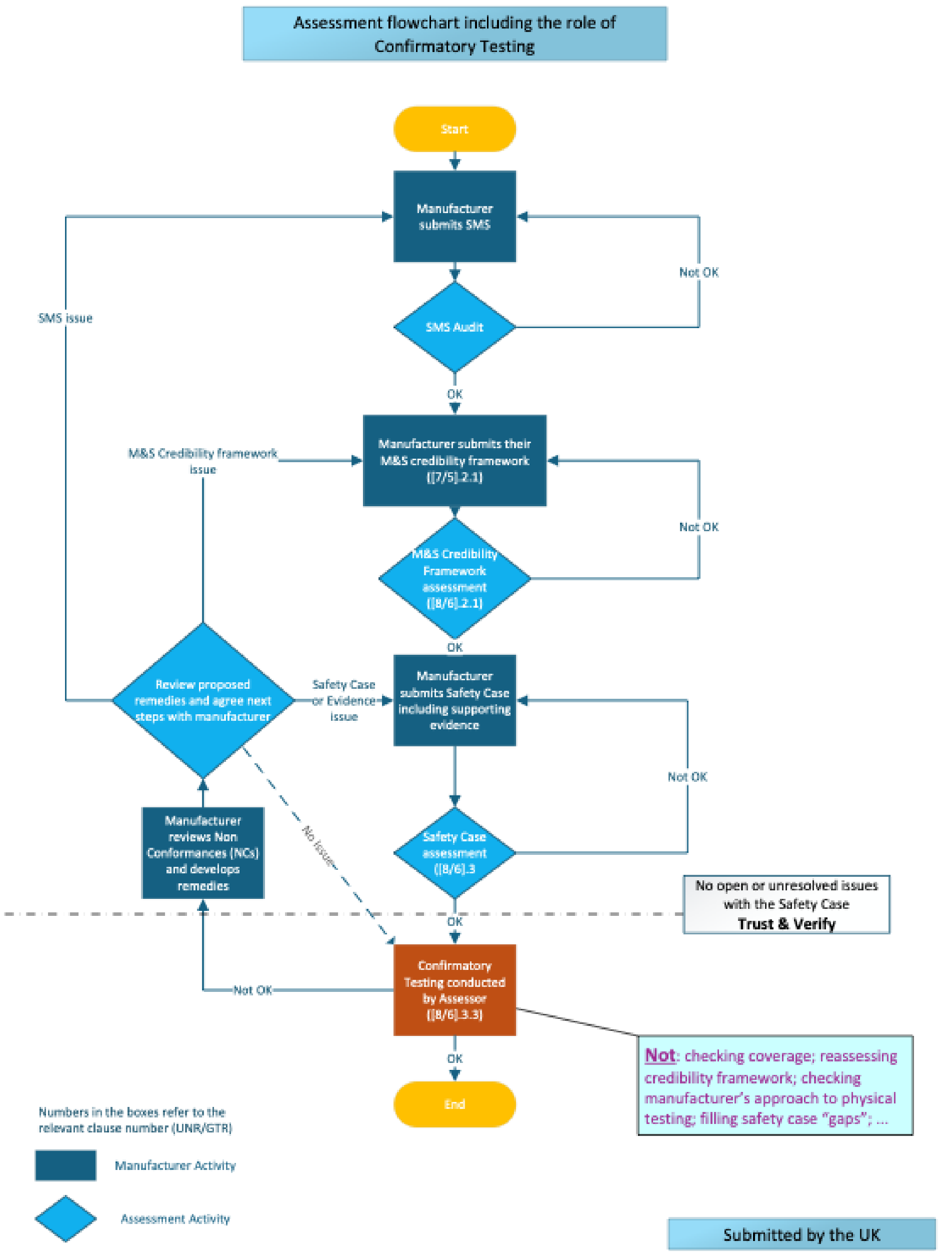
GTR 6.3.3.1.: *“At the option of the Contracting Party, confirmatory testing may be required to use one or more test methods and pre-defined and repeatable test protocols to confirm that the evidence provided by the manufacturer accurately represents the ADS performance. The confirmatory tests shall cover a range of driving conditions representative of the ODD, including at least and as appropriate:”*

UNR 8.3.3.1.: *“Confirmatory testing conducted or required by the approval authority or its designated technical service shall use one or more test methods and predefined and repeatable test protocols to confirm that the evidence provided by the manufacturer accurately represents the ADS performance. The confirmatory tests shall cover a range of driving conditions representative of the ODD, including at least and as appropriate:”*

Interpretation

The ADS assessment process under this Regulation is centred on the evaluation of a manufacturer’s safety case. The confirmatory testing is the last part of the assessment and is conducted once the safety case and the evidence provided by the manufacturer have been assessed and accepted. Up to this point, all evidence has been provided by the manufacturer, so the confirmatory testing is used to “confirm” or verify that the evidence (including considerate variations of the same), provided by the manufacturer, is trustworthy.

The flowchart below illustrates a hypothetical ADS certification/approval process with the different steps of the assessment process and clarifies the stage at which confirmatory testing takes place. The flowchart is only provided as an explanatory visual aid and is not intended to be a prescriptive procedural requirement. The approach shown does not include all the activities and does not preclude different strategies, including undertaking some steps in parallel where appropriate.



Confirmatory testing conducted by the assessor is a targeted and proportionate activity aimed at providing confidence in the credibility and internal consistency of the evidence supplied in support of the safety case. It takes place after the latter has been assessed. The confirmatory testing is intended to verify selected evidence and is not meant to replicate in its entirety or replace the manufacturer's validation and verification campaign. Confirmatory testing consists of representative samples of scenarios and operating conditions that are part of the safety case.

The assessor can accept, allow, or introduce variations to the tested scenarios or conditions, where this is considered necessary to verify that the manufacturer's evidence is representative and is able to support the safety case. Such variations are expected to be reasonable and compatible with the limitations/abilities of the toolchains[1] and allow comparison with the results provided by the manufacturer or be assessable against the expected behaviour of the system. In this context, reasonable means that the ADS is tested in scenarios/situations that are likely to occur in the real world, rather than in unrealistic situations that do not conform to the logic of the real world.

[1] It is expected that the assessment of the credibility and suitability of the simulation toolchains takes place before the confirmatory testing to ensure the assessor is aware of any potential limitations and can make an informed choice about the verification tests that can be performed.

Variations may include, for example, adjustments to the initial conditions, parameter ranges, or environmental assumptions within the bounds or at the boundaries of the declared ODD but may also consider the behaviour at ODD exit or outside the ODD. Variations may also arise inherently from the nature of the testing environment. Considering the real-world confirmatory testing, the assessor is unlikely to be able to replicate the manufacturer's test method, protocols and/or results. The real-world confirmatory testing may be conducted on test routes different from those of the manufacturer, provided that the route selection is suitable for the ADS and is representative of the ODD. The selection allows the manufacturer's results to be assessed inside, at or crossing the boundaries of and outside the ODD. Similarly, while carrying out proving ground testing, it might not be possible to control all the variables, and there might be stochastic effects in the controlled variables. The variations, planned or otherwise, that occur during the confirmatory testing help to verify the overall robustness and consistency of the manufacturer's results.

Subject: Guidance on the minimum list of scenarios to be included for confirmatory testing

Location:

- 6.3.3.1, 6.3.3.3.8, 6.3.3.3.9 GTR;
- 8.3.3.1, 8.3.3.3.8, 8.3.3.3.9 UNR.

GTR/UNR 6./8.3.3.1.: *The confirmatory tests shall cover a range of driving conditions representative of the ODD, including at least and as appropriate:*

a) Failure situations,

b) Behaviours in the presence of vulnerable road users,

c) Situations with a large number of other road users, traffic disturbance, unlikely road infrastructure, uncommon road conditions, and/or atypical environmental conditions,

d) User interactions,

e) Compliance with traffic rules,

f) Collision avoidance and mitigation,

g) ODD boundaries and fallbacks to MRC, and

h) Conditions that trigger DSSAD and ISMR functions.

GTR 6.3.3.3.8.: *The confirmatory testing shall include scenarios where the behaviour or position of other road users requires the ADS to react to their movement or presence.*

GTR 6.3.3.3.9.: *The confirmatory testing shall use track testing to also confirm that user(s)-related aspects are in line with the ADS safety case.*

UNR 8.3.3.3.8.: *The approval authority or its designated technical service shall select scenarios where the behaviour or position of other road users requires the ADS to react to their movement or presence.*

UNR 8.3.3.3.9.: *The approval authority or its designated technical service shall use track testing to also confirm that user(s)-related aspects are in line with the ADS safety case.*

Guidance

The applicability and relevance of specific scenarios selected as part of the confirmatory testing depend on the ADS feature, its ODD, and the scenarios used to demonstrate the safety case.

When selecting the list of scenarios/situations for confirmatory testing, the assessor may consider the following general guiding principles to establish the relevance:

- The ADS system description and its safety concept: this guiding principle aims at ensuring that the intended use(s) of the ADS and its architecture are considered when planning confirmatory testing;

- The safety claims presented in the safety case: this aims at ensuring meaningful and robust confirmatory testing through the selection of tests relevant to safety claims in the safety case. The choice of tests with high-severity hazards, complex mitigation strategies, or critical architectural assumptions helps to provide reassurance that some of the more critical manufacturer tests have been confirmed as trustworthy.

Confirmatory testing may be undertaken using virtual testing. The testing performed is expected to be within the scope and capability of the toolchain that is being used. As with other confirmatory testing it can be within, at the boundary of or outside the ODD.

Based on the regulatory text, the following list of scenario/situation categories [may/could] be included as applicable at the confirmatory testing phase using the most suitable testing environment:

- Failure scenarios/situations relevant to the ADS safety concept (e.g., sensor failure that could lead to an unreasonable safety risk if fallback/mitigation does not perform as stated in the safety case);
- Testing with safety-relevant scenarios (e.g., scenarios with high safety-risk implications, or scenarios demanding evasive manoeuvring by the ADS to avoid collision);
- Testing with different applicable road-users (e.g., scenarios involving vulnerable road-users, or scenarios involving different reasonably foreseeable motorized vehicles);
- Testing with ODD-specific elements (e.g., scenarios with country-specific elements, or scenarios at the boundary of the ODD);
- Testing with traffic disturbances (e.g., scenarios with many traffic participants interacting with the ADS vehicle and producing perturbations and/or obstacles blocking the lane such as lost debris);
- Testing to verify the safety claims related to the robustness of the ADS (e.g., scenarios evolving outside the ODD, fallback response, or scenarios involving variations to verify ADS consistency);
- Testing for DSSAD/ISMR triggers. As part of the confirmatory testing, collisions or other triggering situations for DSSAD and/or occurrence reporting may occur. In such cases, there is the expectation that the assessor may verify the correct data recording/monitoring capabilities. Additionally, confirmatory testing may consider dedicated confirmatory testing targeting DSSAD/ISMR capabilities, such as artificial signal injections in controlled testing environments.
- Fallback behaviour (where applicable).

Subject: Guidance on the inclusion of unlikely/unusual ODD-relevant elements during real-world confirmatory testing

Location:

- 6.3.3.4.9 GTR;
- 8.3.3.4.9 UNR.

GTR 6.3.3.4.9.: *The real-world confirmatory testing shall ensure that the selection of test routes utilises appropriate strategies to enhance the probability of ADS encountering situations that involve a large number of other road users, unlikely road infrastructure, or abnormal geographic/environmental conditions, by examining when and where specific elements (e.g., high- or low-density traffic) typically occur. It is understood that it may not be possible to encounter all traffic situations during a real-world test.*

UNR 8.3.3.4.9.: *The approval authority or its designated technical service shall ensure that the selection of test routes utilizes appropriate strategies to enhance the probability of ADS encountering situations that involve a large number of other road users, unlikely road infrastructure, or abnormal geographic/environmental conditions, by examining when and where specific elements (e.g., high- or low-density traffic) typically occur. It is understood that it may not be possible to encounter all traffic situations during a real-world test.*

Guidance

During real-world confirmatory testing, the ADS may encounter unlikely or unusual elements that may or may not have been reasonably foreseeable. Such naturally occurring situations do not represent exhaustive edge-case testing, but may be considered as one element among the others being relevant to verifying safety-critical behaviours, assumptions, or fallback strategies claimed in the safety case. If the ADS does not behave in an appropriate manner, then the event is identified and recorded for further investigation and resolution by the manufacturer.

In real-world testing, it is expected that the assessor include as much variability as reasonably possible to verify the ADS performance and include as many unusual or unlikely scenarios as possible without creating a risk for the occupants or other road users.

Unlikely or unusual elements deemed relevant for confirmatory testing purposes that cannot reasonably be encountered during real-world confirmatory testing may be assessed through simulation or controlled proving ground testing.

Subject: Guidance on confirmatory test duration/coverage metrics for termination

Location:

- 6.3.3.4.14.1, 6.3.3.4.14.2 GTR;
- 8.3.3.4.14.1, 8.3.3.4.14.2 UNR.

GTR/UNR 6./8.3.3.4.14.1.: *The real-world test drive shall cover the functions required to perform the entire DDT in the ODD pursuant to the outcomes of the safety case analysis.*

GTR/UNR 6./8.3.3.4.14.2.:. *The test should be terminated only when all relevant parts of paragraph 8.3.3.4.14.1, excluding safety-critical and failure-related scenarios, have been monitored and assessed.*

Guidance

Confirmatory testing concludes once sufficient confidence has been obtained that the manufacturer's test evidence is trustworthy. The duration and extent of confirmatory testing are determined by the adequacy of the evidence collected, rather than by predefined numerical thresholds. Factors dictating the duration of the confirmatory testing may include:

- the complexity of the ODD,
- the different behavioural capabilities of the ADS, or
- the amount of claims/evidence within the safety case.

This does not mean that metrics, such as the number of scenarios executed, parameter variations applied, or operating time accumulated, may not be used as supporting information as to when sufficient confirmatory testing has been undertaken. However, such indicators are understood as aids in making a decision and not as mandatory termination criteria.

The assessor may extend confirmatory testing beyond the initially suggested amount, if needed. As the amount is not predetermined, there may be reasons why, during the confirmatory testing, additional testing is deemed to be necessary. This may happen for several reasons. As an example, during real-world testing, the ADS may not encounter sufficient variability and therefore more time needs to be allocated.

Subject: Guidance on scenario selection and allocation

Location: N/A

Guidance

Existing legal and regulatory frameworks for ADS may be used as non-binding references to inform the identification of applicable scenarios for specific ADS use-cases, where relevant. Such references may include, for example:

- GRVA 18-50 – Annex IV Integration Document contains general principles to derive use-case agnostic scenarios.
- UN Regulation No 157-00, and later iterations, that cover ADSF-1 for motorway use-cases.
- EU Regulation 2022/1426 covers a selection of ADSF-2 use-cases, including:
    a) *Fully automated vehicles*, including dual mode vehicles, designed and constructed for the carriage of passengers or carriage of goods on a predefined area.
    b) *'Hub-to-hub'*: fully automated vehicles, including dual mode vehicles, designed and constructed for the carriage of passengers or carriage of goods on a predefined route with fixed start and end points of a journey/trip.
    c) *'Automated Valet Parking'* (AVP): dual mode vehicles with a fully automated driving mode for parking applications within predefined parking facilities.
- GB/T 41798 Intelligent and connected vehicles—Track testing methods and requirements for automated driving functions.
- DOT HS 812 623 - A Framework for Automated Driving System Testable Cases and Scenarios.

Relevant standards that can support scenario identification include:

- ISO 34502 - Road vehicles — Test scenarios for automated driving systems — Scenario based safety evaluation framework
- ISO 34504 - Road vehicles — Test scenarios for automated driving systems — Scenario categorization
- ISO 34505 - Road vehicles — Test scenarios for automated driving systems — Scenario evaluation and test case generation
- ISO 23374-1:2023 - Intelligent transport systems — Automated valet parking systems (AVPS) — Part 1: System framework, requirements for automated driving and for communications interface, Section 10: Test scenarios for automated vehicle operation.

Scenario allocation across simulation, proving ground testing, and real-world testing may be guided by the characteristics of the scenario, including its safety criticality, interaction complexity, technical feasibility in the given testing environment, and statistical frequency within the ODD.

Some concrete metrics to document the scenario allocation process may be[2]:

- Required minimum fidelity;
- Test complexity;
- Safety hazard to personnel involved;
- Real-time needs;
- Test throughput;
- Need for a controllable environment.

---

[2] https://ccam-sunrise-project.eu/wp-content/uploads/2025/08/D8.1_Final-report-to-vehicle-safety-bodies_V1.0.pdf

Subject: Coverage and representativeness including combined coverage of virtual/track/real-world testing as part of the assessment of safety case testing activities

Location:

- 5.3.2.14, 5.3.2.15, 6.3.2.1.2 GTR;
- 7.3.2.14, 7.3.2.15, 8.3.2.1.2 UNR.

GTR/UNR 5./7.3.2.14.: *The safety concept shall describe the scenario identification and generation approach and how that approach addresses the following:*

*a) **Coverage** of the appropriate nominal, critical and failure situations*

b) […]

*c) Inclusion of elements (especially dynamic elements) that are **representative** of existing traffic conditions in the expected operating conditions, and*

GTR/UNR 5./7.3.2.15.: *The safety concept shall describe the manufacturer's approach to scenario selection to **cover** the reasonably foreseeable situations and conditions that the ADS will encounter including how the following aspects are covered*

GTR 6.3.2.1.2.: *The assessment shall verify that the combined coverage of the testing results from all pillar (virtual, track, real world) is sufficient to support the ADS safety case claims.*

UNR 8.3.2.1.2.: *The approval authority or its designated technical service shall verify that the combined coverage of the testing results from all pillars (virtual, track, real world) is sufficient to support the ADS safety case claims.*

Guidance:

The regulation requires the manufacturer to generate and select scenarios to ensure the coverage of appropriate nominal, critical and failure and reasonably foreseeable situations. It is expected that the appropriate coverage is obtained through a mixture of virtual testing, proving ground testing, and real-world testing. The regulation also mandates the assessor to verify the resulting combined coverage.

This guidance is provided based on available best practices concerning the quantitative evaluation of the coverage. The target recipient of this guidance is both the manufacturer creating the safety case and the assessor verifying the safety case.

Coverage is understood throughout the regulation as the extent to which a set of scenarios covers the relevant aspects of an ODD for the safety argumentation of the ADS. The concept of coverage is multi-dimensional. For instance, coverage can be evaluated based on[3]:

- Operational Environment: evaluates the presence of relevant ODD attributes leveraging standardized taxonomies. An example may include using ISO 34503 and demonstrating that all the applicable ODD attributes for the ADS are encompassed in the scenario suite.

- Behavioural competency: evaluates the presence of certain behaviours produced by either the ego-vehicle or other road users, leveraging standardized taxonomies. An example may include using BSI Flex 1891 and demonstrating that all applicable behavioural competencies are included within the scenario suite.

- Scenario-type/category: calculates the presence of certain a functional/logical scenario type within the scenario suite. Examples can be cut-in, cut-out or pedestrian crossing as applicable for the specific ADS.

- Rules of road compliance: checks that applicable rules of the road within the ODD are present in the scenario suite.

---

[3] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11407490

Subject: Guidance on approaches to establish the criticality of simulation toolchains within the simulation credibility framework

Location:

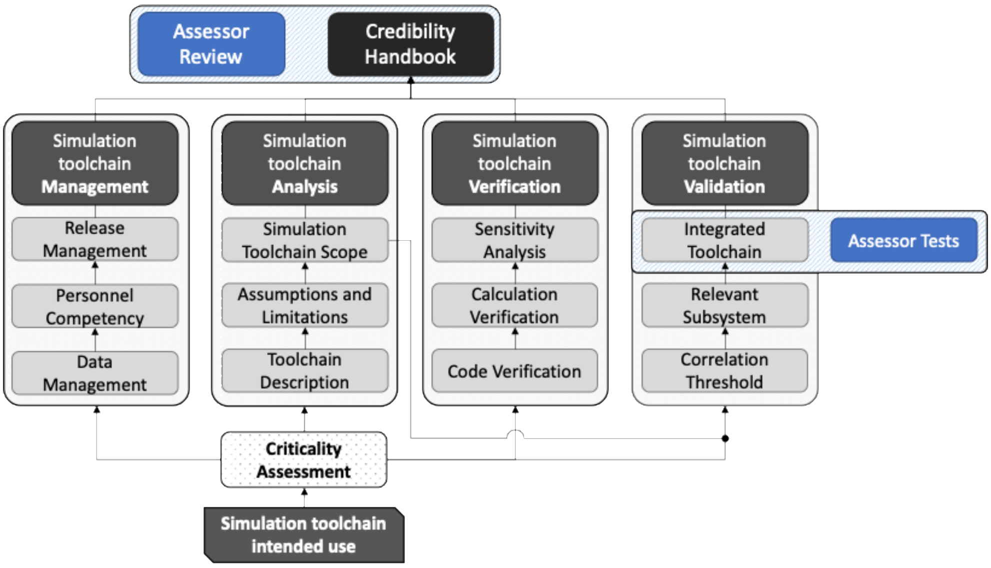
- 5.2.1.9.1 GTR;
- 7.2.1.9.1 UNR.

GTR/UNR 5./7.2.1.9.1.: *The manufacturer shall review the error estimates of the simulation toolchain(s) to assess their* ***criticality*** *and the effect these would have on the manufacturer's claims about their safety case.*

Guidance:

As part of the simulation credibility framework, the regulation requires the manufacturer to identify the *criticality* of the various simulation toolchains and their usage within the overall testing strategy.

The identification of the criticality by the manufacturer is part of the overall simulation credibility framework assessment. The assessment also includes simulation toolchain management, analysis, verification, and validation as shown in the chart below. Depending on the identified criticality level, different acceptance thresholds may be set for the different simulation toolchains and their associated usage. Based on ECE/TRANS/WP.29/2024/39, the chart below provides a graphical representation of criticality assessment and its link with the requirements of the regulation.



Criticality is the degree to which safety-relevant decisions rely on evidence generated by the simulation toolchain(s), and the potential impact of incorrect or misleading simulation results on the safety of drivers, passengers and other road users, damage to cargo, road infrastructure or to the ADS' operational capabilities.

A risk-informed approach is an appropriate method to assess criticality. To derive the criticality level of a simulation toolchain using a risk-informed method, the manufacturer may consider the following variables:

a) The consequences on human safety (e.g., severity classes in ISO 26262-8:2018) if the simulation toolchain provides incorrect, incomplete, or misleading results;

b) The degree to which the simulation toolchain(s) influence the safety case, such as:

1) exploratory use;

2) virtual evidence complementing physical testing;

3) virtual evidence partially replacing physical testing;

4) primary evidence for safety case arguments and compliance demonstration.

The table below provides an example criticality assessment matrix to demonstrate this analysis. Manufacturers may adjust this matrix to their particular use case.

Safety Case impact	Significant				
	Moderate				
	Minor				
	Negligible				
		Negligible	Minor	Moderate	Significant
		Decision consequence			

From the perspective of the criticality assessment, some possible cases for assessment may be:

1) The simulation toolchain(s) features the highest criticality level and may follow the full credibility assessment (e.g., high safety case impact and high decision consequence).

2) The simulation toolchain(s) features an intermediate criticality level and may follow a less stringent credibility evaluation with justified deviations (e.g., moderate safety case impact and minor consequence).

3) The simulation toolchain(s) features a low criticality level and may be exempted from the full application of the credibility framework (e.g., negligible safety case impact and minor/negligible decision consequence).

Alternative solutions for the criticality evaluation of the simulation toolchain(s) are also possible. Relevant best practices may include:

- ISO 26262 Tool Confidence Level (TCL) → structured tool confidence classification based on error detection and usage
- NASA-STD 7009B → NASA's credibility assessment of modeling and simulation (M&S);
- Quality Management System (QMS) derived approaches such as IATF 16949 → process assurance and traceability mechanisms.

Guidance on approaches to establish the competencies of the personnel in simulation toolchains within the simulation credibility framework

Location:

- 5.2.1.4.1 GTR;
- 7.2.1.4.1 UNR.

GTR/UNR 5./7.2.1.4.1. *The manufacturer shall document and provide the rationale for their confidence in the competency of:*

a) The personnel who developed the simulation toolchain(s) and its components,

b) The personnel who assessed the simulation toolchain(s) and its components, and

c) The personnel who used the simulation toolchain(s) to perform the testing with the purpose of validating the system.

Guidance:

The simulation credibility framework recognises that modelling and simulation (M&S) activities rely not only on technical tools and processes but also on the competence of the personnel who develop, assess, and use the simulation toolchain(s). Inadequate competence may introduce risks such as inappropriate modelling assumptions, incorrect parameterisation, insufficient validation, or misinterpretation of simulation outputs. These risks may compromise the credibility of evidence supporting the safety case.

The regulation acknowledges that different levels of competency may be required depending on role-specific considerations.

Personnel developing the simulation toolchain

Relevant competences for developing simulation models may include:

- Knowledge of modelling theory relevant to the simulated phenomena (e.g., vehicle dynamics, traffic behaviour, sensor modelling);
- Understanding of numerical methods and software implementation practices;
- Experience in verification methods and model validation techniques;
- Awareness of uncertainty sources and their propagation within the simulation environment;
- Familiarity with ADS functionalities and safety-relevant performance requirements.

Personnel assessing the simulation toolchain

Assessment activities require the ability to critically review modelling assumptions, validation strategies, and statistical understanding of the results. Relevant competences may include:

- Capability to evaluate model adequacy with respect to its intended use;
- Experience in reviewing verification and validation activities;
- Understanding of limitations and potential misuse of simulation results;
- Ability to challenge assumptions and identify methodological weaknesses.

Personnel using the simulation toolchain(s) for validation testing

Users responsible for generating safety-case evidence may demonstrate:

- Understanding of the intended use and limitations of the toolchain;
- Ability to configure scenarios and parameters appropriately;
- Competence in interpreting outputs and associated uncertainties;
- Knowledge of statistical methods used to derive conclusions from simulation campaigns.

Documentation

Depending on the role, the competency of the personnel may be supported by documenting:

- Educational background and specialized training relevant to modelling and simulation;
- Relevant personnel accreditation (e.g., ISO 9001, IATF 16949);
- Documented experience in comparable M&S or safety-critical applications;
- Participation in verification, validation, or scenario-based assessment activities;
- Records of initial and recurrent training;
- Demonstrated involvement in peer review or internal assessment processes.

The depth and formality of competence demonstration may be proportionate to the criticality of the simulation toolchain(s), as determined by the criticality assessment within the simulation credibility framework.

Subject: Guidance on approaches for the verification, and validation, including criteria to evaluate the correlation between test results and the manufacturer's data in simulation toolchains within the simulation credibility framework

Location:

- 5.2.1.14.2, 6.3.3.4.13 GTR;
- 7.2.1.4.1, 8.3.3.4.13 UNR.

GTR/UNR 5./7.2.1.14.2.: *The manufacturer shall provide evidence that each simulation toolchain's results are consistent and correlate with the results of physical tests.*

GTR 6.3.3.4.13.: *In case of track testing according to 6.3.3.3., the assessment shall compare the information generated during real-world testing with the information from track testing to ensure there is an appropriate level of correlation of the results including the performance of the ADS.*

UNR 8.3.3.4.13.: *In case of track testing according to paragraph 8.3.3.3., the approval authority or its designated technical service shall compare the information generated during real-world testing with the information from track testing to ensure there is an appropriate level of* ***correlation*** *of the results, including the performance of the ADS.*

Guidance:

The Regulation requires that simulation toolchains used to support the ADS safety case undergo verification and validation (V&V). Verification and validation address different but complementary questions:

- Verification: Was the model implemented correctly?
- Validation: Does the model adequately represent the real-world system (RWS) of its intended use?

Both activities may be carried out in a proportionate way depending on the criticality of the simulation toolchain and its use the safety case.

V&V is a widely supported topic in industry standards and existing regulations. Some useful references that may guide the V&V activities are:

- ASME V&V 10 Standard for Verification and Validation in Computational Solid Mechanics, EASA CM-S-014 → best practices for V&V in mechanical systems simulation.
- NASA STD7009B → criticality informed framework for V&V.
- ISO 19364, 19365 → explicitly target vehicle dynamics model validation.
- EU Reg 1426/2022 Interpretation Document Appendix V → general guidance concerning simulation credibility in ADS virtual testing.

The following text provides practical methodologies derived from the cited standards to fulfil the V&V requirements set out by the Regulation.

Verification

Verification concerns the correctness of the implementation of the conceptual and mathematical models forming the simulation toolchain. Its objective is to ensure that numerical, logical, or software errors do not undermine the credibility of simulation outputs.

Verification may include the following complementary activities.

<u>Code verification</u>

Code verification aims to demonstrate that the implemented software correctly represents the intended conceptual and mathematical models.

Methodologies may include, where appropriate:

- **convergence testing**: execution of tests to demonstrate the convergence to a stationary value while iterating spatial/temporal discretization;

- **order of accuracy**: execution of tests aiming at assessing whether the solution/discretization error converges with the expected rate;
- **comparison with a known analytical solution**: whenever a known (analytical) solution is known it should be compared to the corresponding simulation model code realization;
- **unit testing**: execution of a series of low-level tests and comparison of the implemented (coded) model with the conceptual/mathematical models;
- **model (code) coverage**: execution of virtual tests to determine that all logical branches within the model are executed;
- **static testing**: checking of compilation warnings and errors, consistency analysis in the usage of the computer language;
- **dynamic testing**: code execution to investigate memory leaks.

Calculation verification

Calculation verification deals with the estimation of numerical errors affecting the toolchain.

Practical methodologies supporting calculation verification may include:

- **float operation**: evaluate the impact of the uncertainty of underflow/overflow and rounding errors;
- **solver tolerances**: evaluate the impact of the uncertainty of different solver tolerances. To achieve the maximum credibility, the applicant shall perturb the solution until convergence is achieved;
- **sampling intervals**: evaluate the impact on uncertainty/accuracy of sampling interval. To achieve the maximum credibility, the applicant shall perturb the solution until convergence is achieved.

Sensitivity analysis

Sensitivity analysis aims to quantify how input data and parameters affect output values and identify which have the greatest impact. The analysis also provides information that is useful in assessing whether the toolchain and its components can continue to satisfy the acceptance tests and criteria when subjected to small variations of the inputs and parameters.

Practical methodologies supporting sensitivity analysis may include:

- Qualitative sensitivity screening;
- Local (one-at-a-time) quantitative sensitivity analysis;
- Global sensitivity analysis (e.g., variance-based methods);
- Structured peer review of modelling assumptions.

Validation

Validation within the overall simulation toolchain credibility framework aims at verifying that the degree of discrepancy between the simulation-generated data and the corresponding real-world system (RWS) is acceptable for the intended use.

The validation analysis may be carried out using correlation methodologies. A collection of examples is provided in the table below.

Correlation approach	**Description**	**Example**
Graphical	Qualitative evaluation of the signals	• Plotting simulation output vs. RWS output
Scalar data	Quantitative evaluation of scalar quantities deriving from either native scalar outputs or from time-series data following aggregation	• Comparison of minimum distance to an object at the end of a test (native scalar data) • Aggregation of time-series using mean/median and other operators

Time-series	Quantitative evaluation of the discrepancies between two time-series using distance operators	• L_2 norm, Normalized Root Mean Square Error • Sprague and Geers and Dynamic Time Warping • Frequency domain approaches
Statistical testing	Quantitative verification of whether the null hypothesis "the model is an accurate representation of the real-world phenomena" cannot be rejected using statistical testing tools	• T-test or KS-test

---

Subject: Guidance for assessing the manufacturer's testing facilities / environment / capabilities

Location:

- 6.3.3.1.1 GTR;
- 8.3.3.1.1 UNR.

GTR 6.3.3.1.1. *The assessment shall ensure that the physical testing (proving ground and/or public road) facilities and environment and the virtual testing environment as applicable are suitable to conduct the testing and confirm the evidence provided by the manufacturer to support the safety case in accordance with the requirements under paragraph 5.2. of this Regulation.*

UNR 8.3.3.1.1. *The approval authority or its designated technical service shall ensure that the physical testing (proving ground and/or public road) facilities and environment and the virtual testing environment as applicable are suitable to conduct the testing and confirm the evidence provided by the manufacturer to support the safety case in accordance with the requirements under paragraph 7.2.*

Guidance:

Relevant standards supporting the assessment of the laboratories may include:

- ISO/IEC 17025 Testing and calibration laboratories → sets the requirements for competence, impartiality, and consistent operation of testing and calibration laboratories. Applied to ADS testing, it ensures that measurements (are traceable, validated, and reproducible—so results are technically reliable and defensible for approval/certification purposes.

- DOT HS 813 083 - Advanced Test Tools for ADAS and ADS.
