Submitted by the OPI on VT Credibility Assessment

Document ADS-03-11
3rd ADS IWG session
9-12 July 2023

Transposition of Virtual Testing Credibility Assessment Requirements[1] into provisions for the forthcoming ADS UNR/GTR

The present document provides a first proposal to transpose the provisions included in the "Guidelines and recommendations for ADS safety requirements, assessments and test methods to inform regulatory development" (GRVA-19-15r1e) and concerning the requirements to ensure the credibility of the virtual testing toolchain(s) used to perform the safety assessment of the ADS. The paragraph numbers (not to be considered relevant for the present proposal) reported in the present document as well as the reference text is taken from the aforementioned Guidelines. Amendment proposals are reported in track-changes. Terms and definitions reported in section 3 of the Guidelines considered valid also for the regulatory text.

The resulting proposed text is in line with Annex III - Part 4 of the Commission Implementing Regulation (EU) 2022/1426[2].

[…]

4. Overview of ADS safety requirements, assessment, and validation

[…]

4.21.1.1. Virtual testing uses different types of simulation toolchains to assess compliance of an ADS with safety requirements across a wide range of traffic scenarios, including some of which would be difficult (if not impossible) to reproduce in physical settings.

[…]

4.21.2.1. Having determined performance boundaries and identified situations involving ADS responses to manage conflicts and mitigate risks under the virtual testing, concrete test scenarios ~~can~~ shall be defined for track testing based on the parameters of the corresponding virtual scenarios. Comparison of performance between a virtual test and a track test when executing the same scenario enables assessment of the accuracy of the virtual testing toolchain.

[…]

5. Audit, Safety Assessment, and Manufacturer's System Documentation

5.10.4.4. The auditor ~~should~~ shall perform an assessment of the application of these analytical approaches, including:

[…]

(d) Inspection of the documentation that ~~should~~ shall demonstrate the validation/verification plans and results including appropriate acceptance criteria. It ~~should~~ shall include testing appropriate for validation, for example, Hardware in the Loop (HIL) testing, vehicle on-road operational testing, testing with real end users, or any other testing appropriate for validation/verification. The auditor/assessor ~~should~~ shall perform an assessment of the physical testing (proving ground and/or public road) environment and should assess the documentation of the virtual toolchain(s) provided by the manufacturer. The auditor/assessor ~~may decide to~~ shall carry out tests of the complete integrated tool to assess the credibility of the virtual toolchain. Results of validation and verification ~~may~~ shall be assessed by analysing coverage of the different tests and setting minimal coverage thresholds for various metrics. See Annex 5-Appendix 1 for more information on the credibility assessment.

[…]

[1] From GRVA-19-15r1e: Guidelines and recommendations for ADS safety requirements, assessments and test methods to inform regulatory development

[2] https://eur-lex.europa.eu/eli/reg_impl/2022/1426/oj

Annex 5

Virtual testing and credibility assessment

I. Types of simulation toolchain approaches

1. The simulation toolchain used for virtual testing may result in the combination of different approaches. In particular, there are many ways that tests can be performed:

(a) Entirely inside a computer (referred to as Model or Software in the Loop testing, MIL/SIL), with the model of the elements involved (e.g., a simple representation of the control logic of an ADS) interacting in a simulated environment; and/or

(b) With a sensor, a subsystem, or the whole vehicle interacting with a virtual environment (Hardware or Vehicle in the Loop testing, HIL/VIL). For VIL testing, the vehicle can either be in:

(i) A laboratory where the vehicle would be standing still or moving on a chassis dynamometer or on a powertrain test bed and is connected to the environment model by wire or by direct stimulation of its sensors; or

(ii) A proving ground where the vehicle would be connected to an environment model and would interact with virtual objects by physically moving on the test-track.

(c) With a subsystem interacting with a real driver (Driver in the Loop testing, DIL).

II. Interaction between the system and the environment

2. The interaction between the system under the test and the environment can either be an open- or closed-loop.

3. In open-loop virtual testing a data provision unit provides input stimuli to an ADS. The data provision unit can provide data that was collected from a real-world drive or from a different data source. For example, data can be generated during a test using an environment simulator. In any case, the provided data establishes an environment for the ADS. Compared to closed-loop testing there is no feedback between the data provision unit and the ADS. As a common use case is the re-computation of recorded drives, open-loop testing is sometimes referred to as re-compute, replay or re-simulation. A useful property of open-loop testing is the inherent small gap between a virtual test and a corresponding collected real-world situation, as the open-loop test can be as realistic as the used collection mechanism allowed for, with, under ideal circumstances, no additional error introduced by the open-loop approach. Potential applications of open-loop testing include:

- Regression tests for previously resolved issues as well as tests for newly introduced ADS features.
- Re-validation of previously validated features, e.g., as part of the validation of an improved ADS, especially for features that have no associated functional change.
- The testing of non-functional properties of the ADS. For example, evaluating scheduling or timing behavior of executables.

4. In shadow mode testing, an ADS that is subject to testing is connected to a data provision unit. However, the ADS tested is not controlling the vehicle itself. Indeed, it has no effect on the state or behavior of the controlling unit of the vehicle. This approach enables realistic large-scale testing with a fleet of vehicles as test platforms. Since the ADS that is subject to testing has no effect on the vehicle, using a shadow mode can be categorized as open-loop testing.

5. Closed-loop virtual tests include a feedback loop that continuously sends information from the "closed-loop" controller back to the ADS when the ADS takes

an action. Within these test systems, the digital objects in the environment could react in different ways depending on the action of the system under test.

6. Selecting an open- or closed-loop test could depend on factors such as the objectives of the virtual testing activity and the status of development of the system under test.

7. The flexibility of simulation makes it a standard test method during a vehicle's design and the development of this pillar will also make it part of the ADS validation process. For an ADS, it will be impossible to test the vehicle's behaviour in the real world for all possible situations as well as for any subsequent change in the ADS' driving logic. Virtual testing will therefore become an indispensable tool to verify the capability of the automated system to deal with a wide variety of possible scenarios. In addition, virtual testing can be beneficial in replacing real world and proving ground testing where there are concerns over safety-critical traffic scenarios. It is recommended therefore that virtual testing be used to test the ADS under safety critical scenarios that would be difficult and/or unsafe to reproduce on test tracks or public roads.

8. Virtual tests used for ADS validation can achieve different objectives depending on the overall validation strategy and the accuracy of the underlying simulation and models.

(a) Provide qualitative confidence in the safety of the full system;

(b) Contribute directly to statistical confidence in the safety of the full system (caveats apply);

(c) Provide qualitative or statistical confidence in the performance of specific subsystems or components;

(d) Discover challenging scenarios that can be tested in the real world.

9. In contrast to all its potential benefits, a limitation, of this approach, is in its intrinsic potential limited fidelity. As models provide a representation of the reality, the suitability of a model to satisfactorily replace the real world for validating the safety of an ADS has to be carefully assessed. Therefore, the validation of the simulation and models used in virtual testing is essential to determine the quality and reliability of the results compared to real-world performance.

10. It is recommended that aA certain number of virtual tests of the ADS' performance is shall be compared with its performance in the real world when executing the same scenarios. This will provide the opportunity to assess the accuracy of the virtual testing toolchain that is used. Given the high number of scenarios that virtual testing can perform compared to track testing, the validation will probably need to be performed on a smaller but still sufficiently representative subset of the relevant scenarios in order to substantiate any extrapolation beyond the scenarios used for the ADS' validation.

11. In the short-term, virtual testing might only be conducted using simulation toolchains developed and maintained by the ADS manufacturer. Since their design depends on the validation and verification strategies implemented by the manufacturer, it is recommended that simulation toolchains are not subject to regulation or standardization at this time. Rather, simulation toolchains should be explained and documented by the ADS manufacturer and its suitability assessed during the certification process. For this reason, the output of the NATM related to virtual testing ensures that documentation and data provided by the manufacturer is appropriate. Furthermore, virtual testing using modelling and simulation should be credible enough for an assessor to make sound decisions. Credibility is discussed further below.

12. It is recommended that when validating the safety of the ADS, particular attention should be placed on the interaction between virtual testing and the other test

methods. Virtual testing will have strong relationships with all the pillars of the NATM guidelines. In particular:

(a) Virtual testing supplements physical testing to account for the quantity and diversity of ADS configurations, intended uses and limitations on use. One of the strengths of virtual testing is its capacity to assess the ADS performance across multiple scenarios and across ranges of parameters within scenarios in a cost-effective manner. Virtual testing enables results of limited physical tests to be supplemented by verifiable data covering numerous instances of the test scenario, by varying parameters. Using this approach, virtual testing can demonstrate ADS coverage of safety-critical scenarios, and hence provide evidence that an ADS will perform as intended for that type of scenario in the real world. These advantages reduce the burden on physical tests (offsetting their weaknesses) and help to improve the efficiency of the overall assessment process across the pillars. Virtual testing can also be effectively used to identify and cover edge cases and other low-probability scenarios to increase confidence on the ADS' likely performances.

(b) Virtual testing can play an important role in the development of traffic scenarios.

(c) Virtual testing enables assessment of ADS performance boundaries, enabling precise definition of the boundaries between collision avoidance and crash mitigation. Through methods of randomization and scenario compositions, virtual testing enables the developer or the assessor to challenge the ADS and increase confidence in its performance when challenged with low probability events.

(d) Virtual testing will be a key element in the audit assessment. Results of virtual testing carried out both during vehicle development and in the verification and validation phase will provide valuable evidence supporting the safety audit. The manufacturers will need to provide evidence and documentation about how the virtual testing is carried out and how the underlying simulation toolchain has been validated.

(e) Results from real-world tests can improve the accuracy of simulation and models.

(f) Virtual testing can play an important role in responding to concerns identified through in-use monitoring of ADS performance. Virtual testing provides a quick and flexible approach to analyse ADS performance based on real-world events. It allows manufacturers to understand and verify the ADS behaviour and to understand why an issue may have occurred. It may identify an untested scenario, or a set of untried parameters. It may also identify the "scale" of any issue. If the virtual testing does identify unsafe behaviour it can then also help to assess the efficacy of modifications to the ADS and ultimately to improve the overall ADS performance. Where appropriate, the information and scenario descriptions can be shared and integrated into scenarios and testing regimes worldwide.

13. It is recognised that specific regulatory functional safety requirements are still under development. Virtual testing however, using a validated simulation toolchain, shows promise for assessing the following general safety requirements that are currently under consideration:

(a) The ADS should drive safely and manage safety critical situations. These are the requirements where virtual testing can play a prominent role. MIL/SIL, HIL and VIL virtual testing can all be used to assess these requirements at different stages of vehicle verification and validation.

(b) The ADS should interact safely with the user. DIL virtual testing can be helpful to support the assessment of this category of safety requirement by analysing the interaction between the driver and the ADS in a safe and controlled environment.

(c) The ADS should safely manage failure modes and ADS should ensure a safe operational state. The use of virtual testing in these two categories is also very promising but would probably require further research work. SIL virtual testing could include simulated failures and maintenance requests. HIL and VIL virtual testing could be used to assess how the system would react to the occurrence of a malfunctioning induced into the real system.

Annex 5 - Appendix 1

Credibility assessment for using virtual toolchain in ADS validation

I. Introduction, motivation, and scope

1. The use of Modelling and Simulation (M&S) is becoming widespread thanks to the increasing computational capabilities, accuracy, usability, and availability of M&S software packages. M&Svirtual testing can be beneficial for ADS safety validation because it provides an opportunity to overcome some of the limitations of real testing and to increase the number of testing scenarios. Nonetheless, M&Svirtual testing can also lead to erroneous/seemingly correct results, especially in relation to complex simulations not adequately supported by robust practices addressing all M&Smodelling and simulations aspects beyond pure validation. Therefore, higher confidence in M&Sthe credibility of the virtual toolchain(s) used for ADS assessment is needed so that virtual testing can be used instead of and in conjunction with the other pillars. In other words, M&S a virtual toolchain can be used for virtual testing if an assessor is able to consider the simulation results credible enough to make sound decisions taking into account the potential uncertainties of the same toolchainM&S.

2. If M&S is to be credible it needs to be validated. Validating the models and the simulation tools and process that make up M&S toolchain is difficult and there are limitations, which include the limited scope of the validation tests and the difficulty in gathering data to support the validation procedures. The use of M&S requires attention to all the factors influencing the quality and validity of M&S toolchain and all its separate components. The aim is to:

(a) Identify a common framework to determine, justify, assess and report the overall credibility of the M&S toolchain.

(b) Identify a way to indicate the levels of confidence in the results when a validation assessment takes place and also to determine the associated domains of applicability for the toolchain.

3. This framework should be general enough to be used for different M&S types and applications. Unfortunately, the goal is further complicated by the range and differences of ADS features and the variety of simulation tools and toolchains that are used. These considerations lead to the decision to use an (risk-based/informed) credibility assessment framework that can be applied to all M&S applications.

4. The proposed credibility assessment framework provides a general description of the main aspects needed for assessing the credibility of an M&S solutiontoolchain together with guidelines of the role played by the relevant assessor in the validation process with respect to credibility. The assessor should investigate the documentation and evidence supporting credibility during the audit phase. It is understood that the actual validation tests will take place once there is sufficient evidence that a simulation tool or toolchain produces credible results.

5. The outcome of the current credibility assessment will define the envelope in which the virtual toolchaincan be used to support the ADS assessment.

II. Components of the credibility assessment framework

6. ~~It is recommended that t~~A~~he~~ ~~M&S~~ virtual toolchain could be used for virtual testing if its credibility is established by evaluating its fitness for the intended purpose. It is recommended that credibility is achieved by investigating and assessing five ~~M&S~~ properties of the virtual toolchain:

(a) Capability – what the ~~M&S~~ virtual toolchain can do, and what are the associated risks;

(b) Accuracy – how well ~~M&S~~the virtual toolchain does reproduce the target data;

(c) Correctness – how sound & robust is the ~~M&S~~ data and the algorithms in the tools;

(d) Usability – what training and experience is needed and what is the quality of the process that manage its use;

(e) Fit for Purpose – how suitable is the ~~M&S~~ virtual toolchain for the assessment of the ADS within its ODD.

3. This framework should be general enough to be used for different ~~M&S~~ virtual toolchains types and applications. ~~Unfortunately, t~~The goal is ~~further~~ complicated by the range and differences of ADS features and the variety of simulation tools and toolchains that are used. ~~These considerations lead to the decision to use an~~ A ~~(~~risk-based/informed~~)~~ credibility assessment framework that can be applied to all ~~M&S~~ virtual testing applications is therefore defined here.

4. The credibility assessment framework provides a general description of the main aspects considered for assessing the credibility of a virtual toolchain together with principles on the role of third parties assessors in the validation process with respect to credibility. The assessor shall investigate the documentation and evidence supporting credibility during the audit phase. It is understood that the actual validation tests will take place once there is sufficient evidence that a simulation tool or toolchain produces credible results.

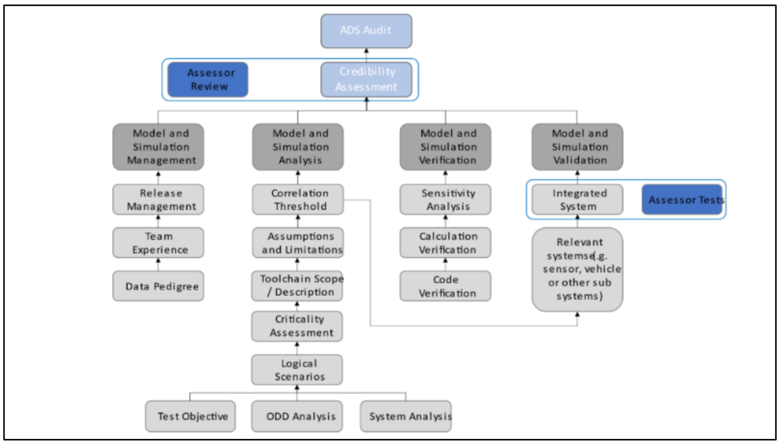
5. The outcome of the credibility assessment defines the envelope in which the virtual toolchain(s) can be used to support the ADS assessment.

II. Components of the credibility assessment framework and related documentation requirements

7. ~~Therefore, credibility requires a unified method to investigate these properties and get confidence in the M&s results.~~ The c~~C~~redibility ~~Assessment~~ assessment framework introduces a way to assess and report the credibility of ~~M&S~~ a virtual toolchain based on quality assurance criteria that allow an ~~indication~~ estimation of the levels of confidence in results. In other words, the credibility is established by evaluating the ~~key~~ influencing factors that are considered the main contributors to the performance of the virtual toolchain ~~behaviour of the models and simulation tools~~ and therefore affect ~~the~~ its overall ~~M&S toolchain~~ credibility: ~~The following all have an influence on the overall M&S credibility; organizational management of the M&S activity, team's experience and expertise, the analysis and description of the chosen M&S toolset, the pedigree of the data and inputs, verification, validation, uncertainty characterization~~. How well each of these factors is addressed indicates the level of quality achieved by ~~M&S~~ the toolchain, and the comparison between the obtained levels and the required levels provides a qualitative measure of the ~~M&s~~toolchain credibility and fitness for its use in virtual testing. A graphical representation of the relationship among the components of the credibility assessment framework is reported in the following figure.

Graphical representation of the relationships between the components of the credibility assessment framework

Commented [CB(1]: When referring to the assessor, the appropriateness for type-approval and self-certification shall be checked



A. Models and Simulation Management

8. The ~~M&s~~toolchain lifecycle is a dynamic process with frequent releases that shall ~~should~~ be monitored and documented. ~~As a result, it is recommended that m~~Management activities ~~should~~ shall be established to support the ~~M&S~~virtual toolchain(s) through typical product management processes. Relevant information on the following aspects ~~should be included in this section~~shall be provided.

9. ~~It is recommended that this part should~~The models and simulation management shall:

(a) Describe the modifications within the ~~M&S~~ toolchain(s) releases;

(b) Designate the corresponding software (e.g., specific software product and version) and hardware arrangement (e.g., XiL configuration);

(c) Record the internal review processes that accepted the new releases;

(d) Be supported throughout the full duration of the virtual testing utilization.

1. Releases management

10. ~~It is recommended that a~~Any virtual toolchain's version used to release data for certification purposes ~~should~~ shall be stored. The virtual models constituting the testing toolchain(s) ~~should~~ shall be documented in terms of the corresponding validation methods and acceptance thresholds to support the overall credibility of the toolchain. The developer ~~should~~ shall establish and enforce a method to trace generated data to the corresponding toolchain version.

11. Quality check of virtual data. Data completeness, accuracy, and consistency ~~are~~ shall be ensured throughout the releases and lifetime of a tool or toolchain to support the verification and validation procedures.

2. Team's Experience and Expertise.

12. Even though Experience and Expertise (E&E) are already covered in a general sense within ~~an~~ the organization, it is important to establish the basis for confidence on the specific experience and expertise for ~~M&S~~modelling and simulation activities.

13. ~~In fact, t~~The credibility of the ~~M&s~~toolchain depends not only on the quality of the simulation models but also on the E&E of the personnel involved in ~~the~~ its validation and usage ~~of the M&S~~. For instance, a proper understanding of the limitations and

validation domain will prevent possible misuse of the ~~M&S~~virtual toolchain or misinterpretation of its results.

14. It is therefore ~~important~~ necessary to establish the basis for the ADS manufacturer's confidence in the experience and expertise of:

(a) The teams that will internally assess and validate the ~~M&S~~ toolchain(s) and,

(b) The teams that will use the validated ~~simulation~~ toolchain for the execution of virtual testing with the purpose of validating the ADS.

15. ~~Thus, if~~ A proper management of the~~a~~ team's E&E ~~is good it~~ increases the level of confidence ~~and hence~~ on the credibility of the ~~M&S~~virtual toolchain(s) and its ~~results~~ outcomes by ensuring that the human elements underpinning the ~~M&S~~modelling and simulation ~~activity~~ activities are taken into consideration and any possible human component risk~~s from the human aspect of the activity can be~~is controlled~~, through its Management System~~.

16. If the ADS manufacturer's toolchain incorporates or relies upon inputs from organizations or products outside of the manufacturer's own team, ~~it is recommended that the~~ the ADS manufacturer shall provide ~~includes~~ an explanation of measures it has taken to manage and develop confidence in the quality and integrity of those inputs.

17. The team's Experience and Expertise ~~include~~ consists of two ~~aspects~~level:

(a) Organizational level

The credibility is established by setting up processes and procedures to identify and maintain the skills, knowledge, and experience to perform ~~M&S~~modelling and simulation activities. The following processes ~~should~~ shall be established, maintained and documented:

(i) Process to identify and evaluate the individual's competence and skills;

(ii) Process for training personnel to be competent to perform ~~M&S~~modelling and simulation -related duties.

(b) Team level

Once a toolchain has been finalized, its credibility is mainly dictated by the skills and knowledge of the teams that will ~~first~~ validate ~~the M&S~~it and then use it for the validation of ADS. The credibility is established by documenting that these teams have received adequate training to fulfil their duties.

18. The ADS manufacturer ~~should~~shall:

(a) ~~Provide~~ provide the basis for the ADS manufacturer's confidence in the Experience and Expertise of the individual/team that validates the ~~M&S~~ toolchain;

(b) ~~P~~provide the basis for the ADS manufacturer's confidence in the Experience and Expertise of the individual/team that uses the ~~simulation~~ toolchain(s) to execute virtual testing ~~with the purpose of validating~~to validate the ADS.

19. The ADS manufacturer ~~should~~ shall demonstrate ~~of~~ how it applies the principles of its Management Systems, e.g. ISO 9001 or a similar best practice or standard, with regard to the competence of its ~~M&S~~ organization and the individuals in that organization and the basis for this determination. ~~It is recommended that t~~The independent assessor shall not substitute its judgment for that of the ADS manufacturer regarding the experience and expertise of the organization or its members.

3. Data/input pedigree

20. The data/input pedigree shall contain a record of and traceability of the data and inputs used by the manufacturer in the validation of the M&stoolchain is important. The manufacturer should have a record of these that allows the assessor to verify their quality and appropriateness.

(a) Description of the data used for the M&Stoolchain validation

(i) The ADS manufacturer should shall document the data used to validate the models included in the tool or toolchain and note important quality characteristics;

(ii) The ADS manufacturer should shall provide documentation showing that the data used to validate the models covers the intended functionalities that the toolchain aims at virtualizing;

(iii) The ADS manufacturer should shall document the calibration procedures employed to fit the virtual models' parameters to the collected input data.

(b) The manufacturer shall quantify the Eeffect of the data quality (e.g. data coverage, signal to noise ratio, and sensors' uncertainty/bias/sampling rate) on model parameters uncertainty. This will be an input to the

The quality of the data used to develop the model will have an impact on model parameters' estimation and calibration. Uncertainty in model parameters will be another important aspect in the final uncertainty analysis of the virtual toolchain.

Formatted: Indent: Hanging: 0.79"

4. Data/output pedigree

19. The data/output pedigree shall contain of the output data is important. The manufacturer should keep a record of the toolchain outputs of the M&S toolchain used for its validation and and ensure that it is traceable to the inputs and the M&S toolchain that produced it. This will form part of the evidence trail for the ADS validation.

(a) Description of the data generated by the M&stoolchain

(a) The ADS manufacturer should shall provide information on any data and scenarios used for virtual testing toolchain validation;

(b) The ADS manufacturer should shall document the exported data and note important quality characteristics e.g. using the correlation methodologies;

(c) The ADS manufacturer should shall trace M&Sthe toolchain(s) outputs to the corresponding M&S setup:

(i) Effect of the data quality on M&Stoolchain credibility:

(c) The M&Stoolchain output data should shall be sufficient to ensure the correct execution of the validation exercise. The data should shall sufficiently reflect the ODD relevant to the virtual assessment of the ADS.

(d) The output data should shall allow consistency/sanity check of the virtual models, possibly by exploiting redundant information.

(ii) Managing stochastic models

(e) Stochastic models should shall be characterized in terms of their variance;

(f) The use of a stochastic models should shall not prohibit the possibility of deterministic re-execution.

B. ~~M&S~~Toolchain(s) Analysis and Description

20. The ~~M&s~~toolchain(s) analysis and description aim to define the whole toolchain and identify the parameter space that can be assessed via virtual testing. It defines the scope and limitations of the models and simulation tools and the uncertainty sources that can affect its results.

1. General description

21. The ADS manufacturer ~~should~~ shall provide a description of the complete toolchain along with how the ~~M&S~~ data will be used to support the ADS validation strategy.

The ADS manufacturer ~~should~~ shall provide a clear description of the test objective.

2. Assumptions, known limitations, and uncertainty sources

22. The ADS manufacturer ~~should~~ shall ~~motivate~~ describe the modelling assumptions which guided the design of the ~~M&S~~ toolchain. The ADS manufacturer ~~should~~ shall provide evidence on:

(a) How the manufacturer-defined assumptions play a role in defining the limitations of the toolchain;

(b) The level of fidelity required for the simulation models.

22. The ADS manufacturer ~~should~~ shall provide justification that the tolerance for ~~M&S~~toolchain versus real-world correlation is acceptable for the test objectives.

23. Finally, this section ~~should~~ shall include information about the sources of uncertainty in the model. This will represent an important input to final uncertainty analysis, which will define how the ~~M&S~~ toolchain outputs can be affected by the different sources of uncertainty of the ~~M&S~~ toolchain used.

3. Scope ~~(what is the model for?). It~~(it defines how the ~~M&S~~toolchain is used in the ADS validation)~~.~~

24. The credibility of virtual tool ~~should~~ shall be enforced by a clearly defined scope for the utilization of the developed ~~M&S~~ toolchains.

25. The ~~mature M&S~~toolchain(s) ~~should~~ shall allow a virtualization of the physical phenomena to a degree of accuracy which matches the fidelity level required for certification~~. Thus, the M&S environment will act as a "virtual proving ground" for ADS testing~~.

26. ~~M&S t~~The toolchains need dedicated scenarios and metrics for validation. The scenario selection used for validation ~~should~~ shall be sufficient ~~such that there~~ to achieve ~~is~~ confidence that the toolchain will perform in the same manner in scenarios outside ~~that were not included in~~ of the validation scope.

27. ADS manufacturers ~~should~~ shall provide a list of validation scenarios together with the corresponding parameter description limitations.

28. The ODD analysis is a crucial input to derive requirements, scope and the effects that the ~~M&S~~ toolchain(s) must consider to support~~ing~~ ADS validation.

29. Parameters generated for the scenarios will define extrinsic and intrinsic data for the toolchain and the simulation models.

4. Criticality assessment

30. The simulation models and the simulation tools used in the overall toolchain ~~should~~ shall be investigated in terms of their impact in case of a safety error in the final product. The proposed approach for criticality analysis is derived from ISO 26262, which requires qualification for some of the tools used in the development process. In order to derive how critical the simulated data ~~is~~are, the criticality assessment shall consider~~s~~ the following parameters:

(a) The consequences on human safety e.g. severity classes in ISO 26262;

(b) The degree ~~in~~ to which the ~~M&S~~ toolchain(s) results influence~~'s~~ the ADS.

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

Table 4.
Criticality assessment matrix

Influence on ADS	Significant	N/A			
	Moderate				
	Minor				
	Negligible			N/A	
		Negligible	Minor	Moderate	Significant
		Decision consequence			

32. From the perspective of the criticality assessment, the three possible cases for assessment are:

(i) Those models or tools that are clear candidates for following a full credibility assessment.

(ii) Those models or tools that may or may not be candidates for following the full credibility assessment at the discretion of the assessor.

(iii) Those models or tools that are not required to follow the credibility assessment.

C. Verification

33. The toolchain(s) verification ~~of M&S~~ deals with the analysis of the correct implementation of the conceptual/mathematical models that create and build up the overall toolchain(s). Verification contributes to the ~~M&s~~toolchain's credibility via providing assurance that the individual tools will not exhibit unrealistic behaviour for a set of inputs which cannot be tested. The procedure is ~~grounded~~ based in a multi-step approach described below, which includes code verification, calculation verification and sensitivity analysis.

1. Code verification

34. Code verification ~~is concerned~~ concerns ~~with~~ the execution of ~~testing~~ tests ~~that~~ to demonstrate~~s~~ that no numerical/logical flaws affect the virtual models.

35. The ADS manufacturer ~~should~~ shall document the execution of proper code verification techniques, e.g. static/dynamic code verification, convergence analysis and comparison with exact solutions if applicable.

36. The ADS manufacturer ~~should~~ shall provide documentation showing that the exploration in the domain of the input parameters was sufficiently wide to identify parameter combinations for which the ~~M&S~~toolchain(s) ~~tools~~ show unstable or unrealistic behaviour. Coverage metrics of parameters combinations ~~may~~ shall be used to demonstrate the required exploration of the models'~~s~~ ~~behaviours~~performances.

37. The ADS manufacturer should shall adopt sanity/consistency checking procedures whenever data allows

2. Calculation verification

38. Calculation verification deals with the estimation of numerical errors affecting the M&stoolchain(s). The ADS manufacturer should shall document numerical error estimates (e.g. discretization error, rounding error, iterative procedures convergence). The numerical errors should shall be kept sufficiently bounded to not affect validation.

3. Sensitivity analysis

39. Sensitivity analysis aims at to quantifying how model output values are affected by changes in the model input values and thus identifying the parameters having the greatest impact on the simulation modeltoolchain resultsoutputs. The sensitivity study also provides the opportunity to determine the extent to which the simulation model satisfies the validation thresholds when it is subjected to small variations of the parameters, thus it plays a fundamental role to support the credibility of the simulation results.

40. The ADS manufacturer should shall provide supporting documentation demonstrating that the most critical parameters influencing the simulation toolchain(s) outputs have been identified by means of global sensitivity analysis techniques such as by perturbing the model's parameters.;

41. The ADS manufacturer should shall demonstrate that robust calibration procedures have been adopted and that this has identified and calibratedfor assigning an appropriate value to the most critical parameters, leading to an increase in the credibility of the developed toolchain.

42. Ultimately, the sensitivity analysis results will also help to define the inputs and parameters whose uncertainty characterization needs particular attention in order to characterize the uncertainty of the simulation results.

4. Validation

43. The quantitative process of determining the degree to which a model or a simulation is an accurate representation of the real world from the perspective of the intended uses of the M&stoolchain. It is recommended that tThe following items elements shall be considered when validating assessing the validity of a model or simulationthe toolchain:

(a) Measures of Performance (metrics)

The Measures of Performance are metrics that are used to compare the ADS's performance within a virtual test with its performance in the real world. The Measures of Performance are defined during the M&stoolchain analysis. Metrics for validation may include:

(i) Discrete value analysis e.g. detection rate, firing rate;

(ii) Time evolution e.g. positions, speeds, acceleration;

(iii) Analysis of state changes e.g. distance/speed calculations, TTC calculation, brake initiation.

(b) Goodness of Fit measures

The analytical frameworks used to compare real world and simulation simulated metrics are generally derived as Key Performance Indicators (KPIs) indicating the statistical comparability between two sets of data. The validation should shall show that these KPIs are met.

(c) Validation methodology

The ADS manufacturer ~~should~~ shall define the logical scenarios used for virtual testing toolchain validation. They ~~should~~ shall be able to cover, to the maximum possible extent, the ODD of virtual testing for ADS validation. The exact methodology depends on the structure and purpose of the toolchain. The validation may consist of one or more of the following:

(i) Validate subsystem models e.g. environment model (road network, weather conditions, road user interaction), sensor models (Radio Detection And Ranging (RADAR), Light Detection And Ranging (LiDARs), Camera), vehicle model (steering, braking, powertrain).

(ii) Validate vehicle system (vehicle dynamics model together with the environment model).

(iii) Validate sensor system (sensor model together with the environment model).

(iv) Validate integrated system (sensor model together with the~~+~~ environment model with influences form vehicle model).

(d) Accuracy requirement

The ~~R~~requirements for the correlation threshold ~~is~~ are defined during the ~~M&S~~toolchain analysis. The validation ~~should~~ shall show that ~~these KPIs are~~the necessary accuracy is met.

(e) Validation scope (~~what~~ the part of the toolchain to be validated)

A toolchain consists of multiple tools, and each tool will use several models. The validation scope includes all tools and their relevant models that require validation.

(f) Internal validation results

The documentation ~~should~~ shall not only provide evidence of the ~~M&s~~toolchain validation but also ~~should~~ shall provide sufficient information related to the processes and products that demonstrate the overall credibility of the toolchain used. Documentation/results may be carried over from previous credibility assessments.

(g) Independent Validation of Results

The assessor ~~should~~ shall audit the documentation provided by the manufacturer and ~~may~~ shall carry out tests of the complete integrated tool. If the output of the virtual tests does not sufficiently replicate the output of physical tests, the assessor ~~may~~ shall request that the virtual and/or physical tests to be repeated. The outcome of the tests will be reviewed and any deviation in the results ~~should~~ shall be reviewed with the manufacturer. Sufficient explanation is required to justify why the test configuration caused deviation in results.

(h) Uncertainty characterisation

This section is concerned with characterizing the expected variability of the virtual toolchain results. The assessment ~~should~~ shall be made up of two phases. In a first phase the information collected from the "~~M&s~~toolchain ~~Analysis~~ analysis and ~~Description~~description" section and the "~~Data~~data/~~Input~~ input ~~Pedigree~~pedigree" are used to characterise the uncertainty in the input data, in the model parameters and in the modelling structure. Then, by propagating all of the uncertainties through the virtual toolchain, the uncertainty in the toolchain output of the model results is quantified. Depending on the uncertainty of the model resultsoutputs, proper safety margins will need toshall be introduced by the ADS manufacturer in the use of virtual testing as part of the ADS validation.

(i) Characterization of the uncertainty in the input data

The ADS manufacturer should shall demonstrate they have estimated the model's critical inputs by means of robust techniques such as providing multiple repetitions for their assessment.

(ii) Characterization of the uncertainty in the model parameters (following calibration).

The ADS manufacturer should shall demonstrate that when the a model's critical parameters of a model cannot be fully determined they are characterized by means of a distribution and/or confidence intervals.

(iii) Characterization of the uncertainty in the M&stoolchain structure

The ADS manufacturer should shall provide evidence that a proper the modelling assumptions are given a quantitative characterization by assessing the generated uncertainty generated by the modelling assumptions has been performed (e.g. comparing the output of different modelling approaches whenever possible).);

(iv) Characterization of aleatory vs. epistemic uncertainty

The ADS manufacturer should shall aim to distinguish between the aleatory component of the uncertainty (which can only be estimated but not reduced) and the epistemic uncertainty (which instead cannot be reduced) deriving from the lack of knowledge in the virtualization of the physical process.

Annex 5 - Appendix 2

Documentation structure

1. This section sets out will define how the aforementioned above information will be collected and organized in the documentation provided by the ADS manufacturer to the relevant authority.

2. The ADS manufacturer should shall produce a document (a "simulation handbook") structured using this outline to provide evidence for the topics presented.

3. The documentation should shall be delivered together with the corresponding release of the toolchain and appropriate supporting data.

4. The ADS manufacturer should shall provide clear reference that allows tracing the documentation to the corresponding parts of the toolchain and the data.

5. The documentation ~~should~~ shall be maintained throughout the whole lifecycle of the toolchain utilization. The assessor ~~may~~ shall audit the ADS manufacturer through assessment of their documentation and~~/or~~ by conducting physical tests.