Real-World Scenarios as Key for Trustworthy Simulations in Validation of Automated Driving Functions

Thorsten Püschl, Christopher Wiegand, Andre Skusa

Simulation Models & Scenarios
dSPACE GmbH
Rathenau Straße 26
33102 Paderborn
TPueschl@dspace.de
CWiegand@dspace.de
ASkusa@dspace.de

Abstract: Caused by the liability shift from the operator of the vehicle to the producer, one of hardest challenges in the development of automated driving functions is ensuring safety under all conditions in the operational design domain (ODD) which is required for homologation. In the case of continuously updating the autonomous vehicles' software stack, this is becoming even more critical.

Software-in-the-loop (SIL) simulations are recognized to be the means of choice to reproducibly deal with the complexity of the vehicle's environment in the validation process by performing an enormous number of tests in a reasonable amount of time.

Beside many other aspects, one key puzzle piece in a consistent safety argumentation is the ability to prove trustworthiness of the virtualization of road tests. Hence, a major task is the continuous validation of the entire SIL environment itself - including all models and interfaces - to prove a sufficiently accurate representation of the complex environment in the simulator. At the same time, the used test library has to include sufficiently representative scenarios that cover the test space within the ODD.

This paper depicts a scalable solution for the above-mentioned challenges, based on a continuous transfer of recorded real-world scenarios into simulation. The paper presents a method to prove validity of the simulation environment as well as approaches for real-world-data-based design of virtual simulation scenarios and test campaigns matching the vehicles ODD.

1 Introduction

In the last decade, the technology for automated driving (AD) on L3 and L4 has made enormous progress. Several pilot activities in a limited domain with limited local focus have proven that the developed principles work in everyday life. However, to finally take the step from these pilot applications to a broader scale – especially with a broader local scope – the type approval (homologation) is the final hurdle vehicle makers must take.

Even though there is still a lack of regulations or defined protocols to follow for type approval of ADs by authorities for L2 to L5 systems, there are high level requirements e.g., introduced by [UNE21] that demand assurance that the AD acts safely in all rationally foreseeable situations. Therefore, software-in-the-loop (SIL) simulation is proposed as a key method in the validation process which enables an enormous number of test drives to be performed in a time-efficient and cost-efficient manner by means of virtualization using simulation models and scenarios.

As a consequence, it is an essential precondition for a coherent safety argument that the meaningfulness of virtualized tests is retained through the step of replacing real world tests by simulation [IAM21], [MIE21], [SCH22]. Hence – beside other aspects – two topics must be addressed when setting up a simulation environment for virtualization of tests for validation / homologation to ensure trustworthiness of the simulation:

1. Using validated simulation models that reflect the relevant physical effects on a sufficiently accurate level
2. Using the right simulation scenarios that cover the relevant real-world scenarios

To achieve this, the integration of data from real-world test drives into the virtualization is essential to prevent a drift of reality and simulation. This becomes even more obvious when the following three considerations are taken into account:

- **Validation of simulation is a continuous task during the whole product life cycle**
  Due to performance, models used in ADAS/AD validation are typically designed to cover physical effects just up to the level of fidelity which is required to solve the intended tasks. The suitability for that is initially proven in the design and implementation phase of the models, e.g., by evaluating simulation results with comparable, selected, representative real-world situations. This already gives a valuable hint for the trustworthiness of the simulation, but it does not prove that the models generalize accurately in all possible situations. Hence, continuous evaluation of the model performance based on real-world data is key for a plausible trustworthiness statement, also because the domain of AD might change over time and new aspects of the environment not considered before might become relevant.

- **Scenarios are ODD-specific**
  The scenarios an AD experiences in operation are specific for the operational design domain (ODD). Scenarios from generic scenario databases probably cover wide parts of the ODD, but these databases do not claim to be complete for all possible ODDs. Therefore, the scenarios from databases need to be complemented by an ODD-specific collection of scenarios to achieve a high coverage of the test-space within the target ODD of the AD.

- **Scenarios are driving-function-specific or even version-specific**
  As traffic situations result from interaction between the involved traffic participants, the behavior of a traffic participant influences the behavior of others. Assuming that one participant is controlled by an automated driving function, it is the driving function's behavior that influences the behavior of others. A more "aggressive" driving function might cause other reactions of the surrounding traffic participants than "careful going". Going one step further, even new versions or parameterizations of a driving function might lead to different scenarios.

To address the above-mentioned topics, we propose in this paper to complement the validation processes by a continuous data-driven approach based on data from test drives. Real-world data is the basis for evaluating the model's/simulator's fidelity, and additionally it allows the user to build up a scenario database with ODD-specific and driving-function-specific scenarios. The ability to automatically create simulation scenarios representing a digital twin of a recorded real-world scenario is an essential component in the proposed approach.

1.1 A continuous data-driven approach for validation

Figure 1 shows the main steps in the proposed continuous process. A central component in the proposed process is an ODD-based scenario- and data-management system. It structures the measurement data and the scenarios according to an ODD taxonomy and is the basis for future coverage analysis.

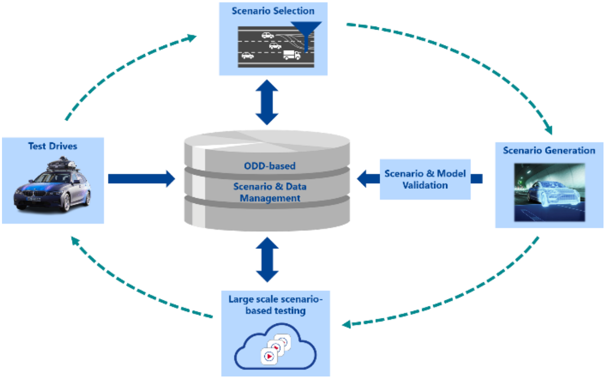

Figure 1: Continuous data-driven process for a reliable validation argument

Considering **Test Drives**, not every recorded minute of driving includes challenging, relevant scenarios. Hence, the relevant scenarios must be identified in the recordings by **Scenario Selection**. If a raw scenario is identified as a possible valuable new input to the scenario database, the **Scenario Generation** subsequently creates a digital twin of the traffic situation and static environment denoted as “simulation scenario”. This serves as a basis for **Scenario & Model Validation** by comparing the behavior of your real system with the behavior of the virtualized system (see chapter 2 Validation of the AD simulation based on real-world scenarios).

After successful validation of the model with this scenario, the significance of the tests utilizing this scenario is proven and hence it can be added to the scenario database (see chapter 3 Collection of the right scenarios to cover the ADs ODD).

Finally, the scenarios collected in the ODD-based scenario- and data-management are utilized for validation of existing and new software releases by **Large Scale Scenario-Based Testing** before deploying the software to the vehicles. With new versions of the AD software, new scenarios might occur in the test drives. These scenarios are also integrated into the validation process by the continuous approach proposed here.

This approach reliably demonstrates the trustworthiness of the simulation by ensuring the usage of validated models and scenarios. At the same time, it builds up a library of scenarios which are suitable and especially relevant for the specific operational design domain and driving function.

2 Validation of the AD simulation based on real-world scenarios

As described above, an important step in the proposed process is the validation of the AD simulation (models, scenarios, and simulator incl. interfaces) using real-world measurement data and the virtual representations of real-world scenarios. The basic concept is to correlate real-world measurement data with simulation data for evaluation of different quality indicators on several levels in data-processing pipeline.

The high-level structure of the data-processing pipeline incl. the simulator and a subset of typical data produced by the different components is shown in Figure 2. In this pipeline several quality indicators addressing different components of the simulator can be implemented as described in the following chapters.

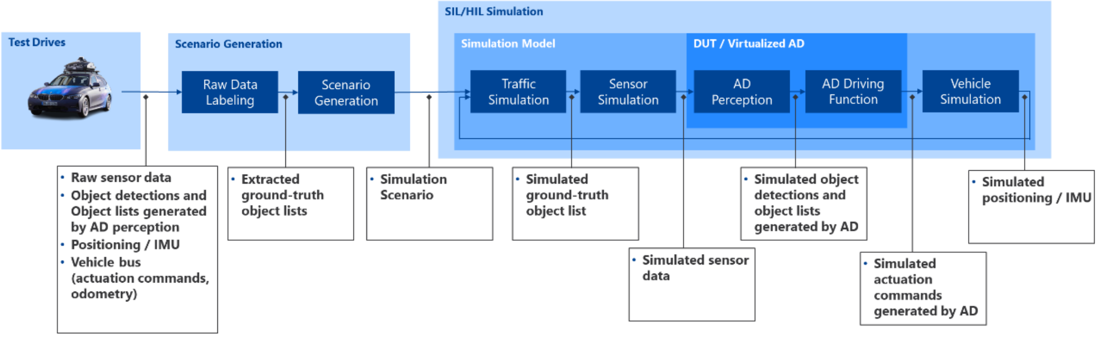

Figure 2: Data-processing pipeline and simulator including interface-data

2.1 Quality Indicator I: Validation of the generated scenario and the traffic simulation

The basis for all subsequent validation steps is the simulation scenario representing a digital twin of the real-world scenario. Hence, it needs to be ensured that the simulation of the scenario leads to same motion of ego and traffic vehicles in the simulation as in the real world.

By comparing simulated trajectories (= virtual object list) and ground-truth object lists extracted from the raw data in the labeling process, a quality metric for this first quality indicator (I) can be derived as shown in Figure 3. The maximum or average Euclidean distance of vehicle trajectories in each time-step or deviations in orientation, relative distances, or velocities of the vehicles are possible criteria. With this quality indicator, the Scenario Generation process and the Traffic Simulation model are validated.

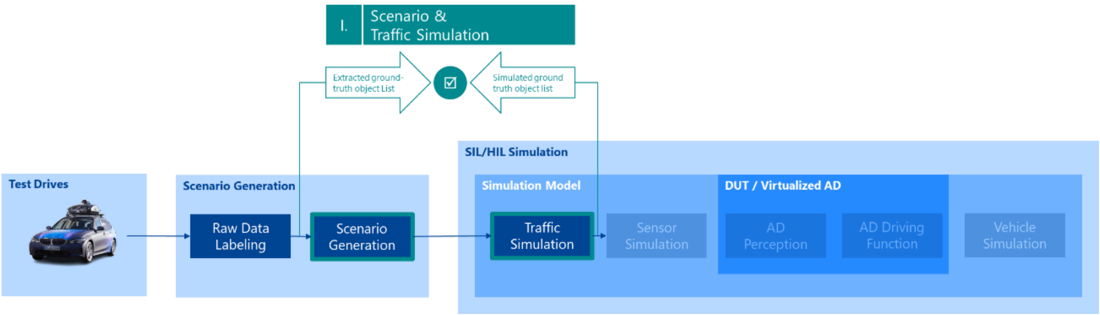

Figure 3: Validation of the generated scenario and the traffic simulation

2.2 Quality Indicator II: Validation of the sensor simulation

The second quality indicator (II) implements a comparison on the level of raw, unlabeled data (Figure 4).

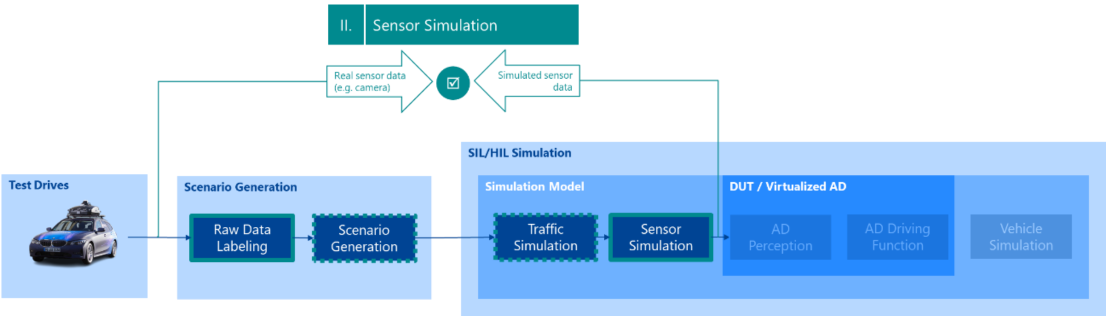

Figure 4: Validation of the sensor simulation

To evaluate the quality of synthetic camera data, similarity measures of various kinds can be an appropriate method. In the research project [KID22], such metrics on different levels (pixel level, image level, and DNN-feature level) are discussed and evaluated. It turned out that there is no easy way to set these measures into a relation to the compared images. Therefore, a further perspective is applied: It should be remembered that the main goal of this evaluation is to ensure that the synthetic images resemble the real ones in such a way that makes no difference, e.g., whether a neural network object detector "looks" at real or synthetic images. That means comparing the object detection performances on both the real and synthetic images can reveal which image distances, measured by the distance metrics, play a role at all in the differentiation of real and simulated images (from the viewpoint of a neural net).

Approaches for evaluating the similarity of Lidar data are discussed in [WAL22].

This quality indicator validates the generated scenario, the traffic simulation, and also the sensor simulation.

2.3 Quality Indicator III: Validation of the sensor simulation and the virtualization of the perception

Evaluating the similarity of the recorded onboard-perception output (e.g., object list or camera 2-D bounding boxes) and the output generated by the perception in the virtualized AD, additionally allows for proving the trustworthiness of the sensor simulation. This test confirms that the fidelity of the sensor simulation reaches at least a sufficient level for the intended purpose of the test. Due to the trade-off between performance and simulation accuracy, a higher level of sensor simulation fidelity is even not desired.

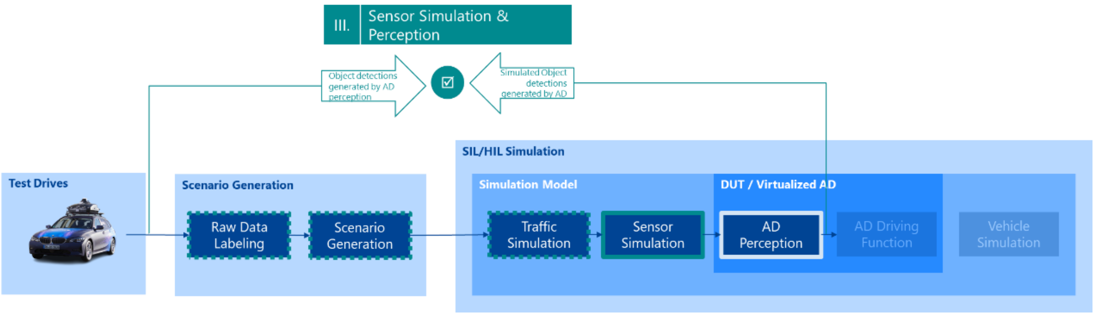

Figure 5: Validation of the sensor simulation and the virtualization of the perception

This third quality indicator (III) was implemented as a proof-of-concept for the here proposed approach using the dSPACE End-to-End Simulation and Validation Tool Suite.

For this demonstration, we implemented a prototype device-under-test (DUT) based on a pre-trained YOLOv4-Network [BOC20] for object detection in RTMaps. The same RTMaps-based DUT implementation was integrated into a SIL simulator using dSPACE ASM for Traffic and Vehicle Dynamics simulation and dSPACE AURELION for physics-based sensor simulation.

Test drive recordings of a busy Korean inner-city scenario, recorded by an AD prototype vehicle from the Korea Intelligent Automotive Parts Promotion Institute (KIAPI) serves as a basis for the process. The data was transformed into a simulation scenario using the dSPACE Scenario Generation service.

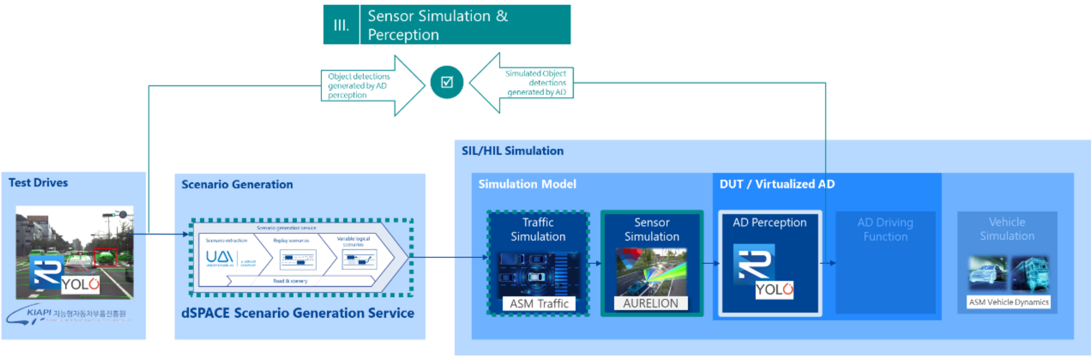

Figure 6: Demonstration setup for quality indicator III

The metric used in this example is based on the 2-D bounding boxes of the detected vehicles. An Intersection-over-union(IoU)-based approach is applied to the 2-D bounding boxes generated on real and synthetic camera data to quantify the similarity in each frame/simulation time step. By using confusion matrix and comparing the results of the object hypothesis for the simulation and the measured data an assessment of the quality of the 3D environment can be performed.

Figure 7: Object detections by YOLOv4 on real-world data (left) compared to simulation data (right) for timeframe 00:03.238

Figure 7 shows the real camera data including the bounding boxes detected by the YOLOv4 detector on real and simulated camera data for timeframe 00:03.238. You can clearly see that all relevant vehicles are detected similarly well in both images.

Figure 8: Object detections by YOLOv4 on real-world data (left) compared to simulation data (right) for timeframe 00:10.521

In timeframe 00:10.521 (Figure 8), major differences become obvious. Due to slight inaccuracies in the generated dynamic scenario and due to different 3-D geometries used for the yellow bus, the pedestrian on the real image (left) is still occluded by the bus, whereas it is already visible in the simulation data (right). Additionally, the bus is still detected in the real data. Because of the two completely distinct detections, the IoU-based metric drops in this frame.

This difference in detections is noted by the defined metrics established to judge the similarity of simulation and reality in this quality indicator (III). This dropping metric is a trigger for further investigation of the simulation with this scenario. In a review, an engineer can check if this difference in simulation and reality is a serious issue for the meaningfulness of the test or if the simulation still sufficiently covers all required effects for the intended tests.

2.4 Quality Indicator IV: Validation of the virtualization of the driving function & vehicle simulation

Finally, the last quality indicator (IV) proves that the whole virtualized DUT behaves similarly enough in the simulator compared to the real test drive. The complete signal chain is involved in this test and additionally the control-loop is closed in the simulator.

Similarity metrics in this quality indicator (IV) can be derived by comparing the ego vehicle's motion profile and/or the actuator commands of the AD on a temporal basis, or by comparing typical KPIs for scenario criticality like time-to-collision (TTC), post-encroachment-time (PET), or gap time (GT) in the recorded real object-list and the synthetic.

A good result in this final quality indicator indicates validity of the overall simulation.

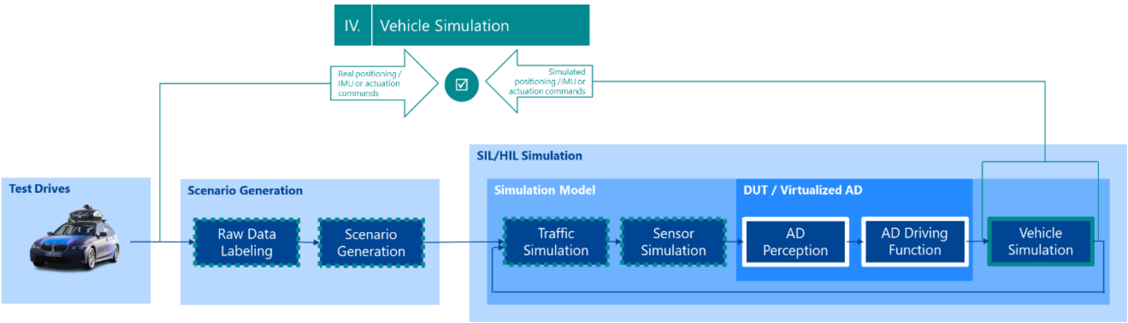

Figure 9: Validation of the virtualization of the driving function & vehicle simulation

2.5 Summary for scenario, model, and simulator validation

Each of these validation steps / quality indicators increases the confidence level of the simulator, models and scenarios used in the validation process. A good result in quality indicator IV already gives a mature hint of the accuracy and validity of the whole simulation, but also the proposed upstream quality indicators I - III help to find deviations and to understand the fidelity of the simulation. Depending on the algorithm to be developed dedicated key performance indicators (KPIs) must be defined for a quality assessment.

Each scenario that has successfully run through this process is a mature starting point for further simulation-based validation activities such as analysis of critical situations or other issues (e.g., disengagements), regression tests or large-scale validation applying scenario-based testing.

3 Collection of the right scenarios to cover the ADs ODD

To address the second key aspect for validation of ADs ("2. Using the right simulation scenarios that cover the relevant real-world scenarios") mentioned in the introduction, each relevant scenario validated as described in the previous chapter is a valuable contribution to the trustworthiness argument of the complete simulation-based validation process of the AD. These scenarios are the basis for setting up a collection of relevant scenarios fitting to the ODD.

3.1 Reducing the unknown risks: Scenario-based testing using ODD- and driving-function-specific real-world scenarios

A major challenge in compiling a set of scenarios and test cases for validation of a certain AD is to prove test coverage of the ODD. Scenario-based testing allows to explore the test space around each known scenario and especially to create critical situations by intelligent parameter variations. However, the covered region of the ODD is limited to the scenarios provided as a starting point for scenario-based testing. Using generic scenario databases is required to establish a good base coverage but suffers from the lack of ODD-specific scenarios and especially driving-function-specific scenarios due to the diversity of the ODD and, for example, the ODDs regional aspects. Hence, even though some regions of the required test space might be covered well, others might not be covered because the required logical scenarios are unknown as Figure 10 illustrates.

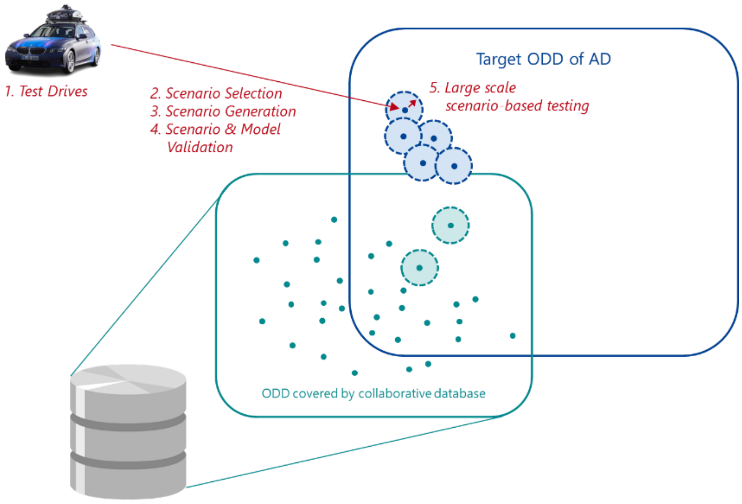

Figure 10: Schematic illustration of ODD coverage and test-space exploration by integration real-world scenarios and scenario-based testing

The development and validation process of an AD is supported by an enormous number of hours of test drives. In these test drives, the AD is exposed to realistic traffic situations that also need to be part of the simulation-based validation process. With each scenario experienced in a test drive, a new starting point for scenario-based testing can be derived by generating a simulation scenario from the recorded measurement data and adding it to the scenario library. Comparing the identified scenarios with scenarios already present in the scenario database enables targeted assembly of the scenario database. A suitable scenario- and data-management-system using an ODD-description as a structuring mechanism is a pre-condition of this approach.

With this approach, it is possible to continuously complement the scenario library by ODD- and driving-function-specific scenarios, fill up uncovered regions in the ODD and hence finally reduce the number of unknown risks for the AD.

4 Conclusion

Chapter 2 describes how we can ensure that the simulation and each used scenario accurately cover reality for the desired testing purpose. On top of that, chapter 3 describes how to ensure that the scenarios used in the simulation-based validation are relevant for the ODD and cover the required test-space. In both cases, the usage of digital twins of real-world scenarios is an important building block.

Combining both methods in a continuous process applied during pre-development, series-development, and fleet-operation ensures a comprehensive argument for trustworthiness of the simulation and hence the validity of the simulation-based homologation process. With its comprehensive data-driven toolchain for simulation and validation, dSPACE supports the proposed approach in all stages.

5 Bibliography

[IAM21] International Alliance for Mobility Testing and Standardization™ Best Practice, A Comprehensive Approach for the Validation of Virtual Testing Toolchains, 2021

[SCH22] Barbara Schütt, A Taxonomy for Quality in Simulation-Based Development and Testing of Automated Driving Systems, IEEE 1109/ACCESS.2022.3149542, 2022

[MIE21] Christoph Miethaner, TüV Süd; Jann-Eve Stavesand, dSPACE GmbH; Virtual homologation of an ALKS according to UNECE R157, White paper, 2021

[KID22] BMWi research project “KI-Data Tooling”, https://www.ki-datatooling.de/de/

[WAL22] Albert Wallace, Validating Simulation Environments for Automated Driving Systems Using 3D Object Comparison Metric, IEEE 10.1109/IV51971.2022.9827354, 2022

[UNE21] UN Regulation No. 157, Concerning the Adoption of Harmonized Technical United Nations Regulations for Wheeled Vehicles, Equipment and Parts which can be Fitted and/or be Used on Wheeled Vehicles and the Conditions for Reciprocal Recognition of Approvals Granted on the Basis of these United Nations Regulations, Addendum 156, 2021

[BOC20] Alexey Bochkovskiy, YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv:2004.10934, 2020