When Is an Automated Driving System Safe Enough for Deployment on the Public Road? Quantifying Safety Risk Using Real-World Scenarios

Olaf Op den Camp[a], Erwin de Gelder[b] and Jeroen Broos
Department of Integrated Vehicle Safety, TNO, Automotive Campus 30, Helmond, The Netherlands

Keywords: Safety Assessment, Risk Quantification, Scenarios, Virtual Simulation.

Abstract: To ensure the safe and responsible deployment of vehicles equipped with Automated Driving Systems (ADSs) onto the public road, a safety assessment of such vehicles should be passed successfully. The assessment results should be unambiguous, easily understood by experts in the field, and explainable to authorities and the general public. An important metric in such a framework is the residual safety risk. The concept of risk is widely understood, and basing the safety assessment on that concept helps to come to a fair and acceptable assessment process. In this paper, we propose a method how to determine estimates for the residual safety risk, and how this safety risk estimate relates to the requirements posed by the UNECE that an activated ADS shall not cause any collisions that are reasonably foreseeable and preventable.

1 INTRODUCTION

The development of automated driving technology that supports the human driver or even completely takes over the driving task for parts of a trip, is a large challenge. However, the development of the safety assessment methods that are required to ensure that the risk associated with the deployment of such innovative and complex technologies onto the public roads is acceptable to consumers, authorities and system developers (vehicle manufacturers and suppliers), appears to be an even bigger challenge.

In this challenge, the following aspects need to be considered:

- The (residual) safety risk associated with the deployment of automated driving technologies is the result of many different factors, in- and outside of the Automated Vehicle (AV). These factors range from the technical state of the vehicle (the performance of the vehicle except for the automation), functional safety aspects, the vehicle's vulnerability to cybersecurity threats, the perception capabilities of the vehicle's sensor system, the driving skills of the Automated Driving System (ADS) in response to the sensor inputs, the interaction with the operator, and the capabilities of the human driver of the vehicle (depending on the level of automation).
- AV functions and systems become increasingly smart and complex, are more and more integrated and become increasingly dependent on machine learning technology. There is no complete overview of failure modes and the appropriate safety assessment and risk estimation methods are difficult to define appropriately.
- There is no complete overview of situations and circumstances in which the systems will be used during its lifetime, which makes it even more difficult to identify the potential failure modes and safety risks.

To put a legal framework to the deployment of automated vehicle systems onto the public road, regulations are being implemented by the UNECE, e.g., for Automated Lane Keeping Systems or ALKS (UNECE, 2021) (UNECE, 2022) and the EC, i.e., for ADS in four use cases (European Commission, 2022). UN Regulation No. 157 (UNECE, 2022) uses the following formulation for the safety requirements to ALKS:

[a] https://orcid.org/0000-0002-6355-134X
[b] https://orcid.org/0000-0003-4260-4294

Op den Camp, O., de Gelder, E. and Broos, J.
When Is an Automated Driving System Safe Enough for Deployment on the Public Road? Quantifying Safety Risk Using Real-World Scenarios.
DOI: 10.5220/0011958000003479
In *Proceedings of the 9th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2023)*, pages 297-304
ISBN: 978-989-758-652-1; ISSN: 2184-495X

- The activated system shall perform the Dynamic Driving Task (DDT), shall manage all situations including failures, and shall be free of unreasonable risks for the vehicle occupants or any other road users.
- The activated system shall not cause any collisions that are reasonably foreseeable and preventable. If a collision can be safely avoided without causing another one, it shall be avoided.

Following the formulation of the ALKS regulation, an AV should be free of reasonably foreseeable and preventable safety risks. The challenge is how to quantify what is considered to be "reasonably foreseeable" and "reasonably preventable", and to consider all the realistically possible situations. For authorities and the industry, this leads to similar but slightly different challenges:

- For authorities, the challenge is how to assess whether the AV meets the legal requirement that it is free of reasonably foreseeable and preventable safety risks.
- For AV developers, the challenge is to provide evidence that a newly developed AV is safe to be deployed on the road.

Though ADS are complex and the assessment procedure might be complicated, the assessment results should be unambiguous, easily understood by experts in the field, and explainable to authorities and the general public. An important metric in such a framework is the residual safety risk when a vehicle is allowed onto the road (SAKURA, SIP-adus and HEADSTART projects, 2021). The residual safety risk is for example expressed as the probability of a fatality or serious injury per hour of driving. The concept of risk is widely understood, and basing the safety assessment on that concept helps to come to a fair and acceptable assessment process.

The United Nations Economic Commission for Europe (UNECE) WP.29 Working Party on Automated/Autonomous and Connected Vehicles (GRVA) has developed the New Assessment/Test Methods (NATM) Master Document (UNECE, 2021), where a multi-pillar approach is envisaged, as shown in Figure 1. The Master Document shows the process to be followed for safety assessment; it does not state how to quantify the results of safety assessment. In this paper, we propose how to determine estimates for the residual safety risk, how this safety risk estimate relates to the terms "reasonably foreseeable" and "reasonably preventable", and how this is in agreement with and a further fulfilment of the UNECE multi-pillar approach.

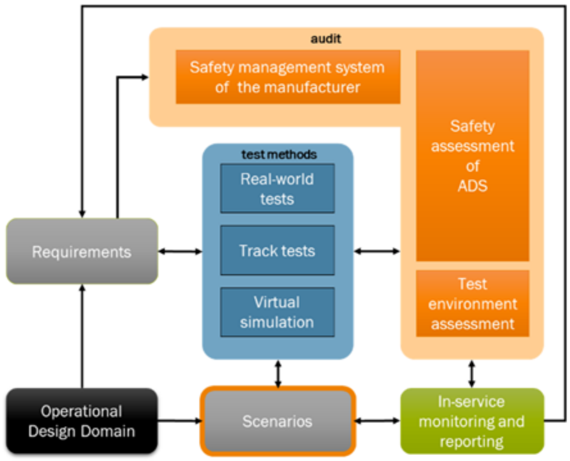

Figure 1: The multi-pillar approach. Figure is adapted from (Donà, et al., 2022).

2 AUTOMATION AND SCENARIO-BASED SAFETY ASSESSMENT

In this section, we will reference the different levels of automation that are related to the role of the driver in the vehicle. The reason is that safety assessment considers the safety of the complete vehicle which obviously includes the driver, whether this is a human driver, a robot driver, or a combination of both.

The SAE J3016 (SAE On-Road Automated Driving (ORAD) committee, 2018) is a commonly used scheme to present the change of the role and responsibility of the human driver for different levels of automation. Driver support functions (Level 0, 1, or 2) support the human driver when certain conditions are met during a drive. In these cases, the human driver must be continuously in the loop to monitor the vehicle systems and the environment, and that human driver is responsible for correct intervention when needed. For a Level 3 system, the human driver must be ready to take over control from the automation system if the Level 3 system requests to do so. For Level 4 (and Level 5) systems, the human driver will not be required to take over control, and consequently, the robot driver needs to respond appropriately in all situation the vehicle encounters, without the fallback option of a human driver.

To assess whether a Level 3 or 4 ADS can safely handle and appropriately respond to all situations and all conditions that it potentially encounters on the road during its lifetime, requires a structured approach in testing the ADS for all these situations.

Stakeholders (AV developers, type approval authorities, regulatory bodies, and automotive research institutes and academia) in Europe, Japan and Singapore, share the vision, that this approach can only be successful by using a (data-driven) scenario-based approach (SAKURA, SIP-adus and HEADSTART projects, 2021) (de Gelder, Op den Camp, & de Boer, Scenario categories for the assessment of automated vehicles, 2020). Object level sensor data from hundreds of thousands of (naturalistic) driving kilometres on public roads are used to build a scenario database that shows what scenarios occur on the road and which variations are seen in the scenarios and the conditions (e.g., weather and light). The scenarios are parameterized, and data collection campaigns are organized in such a way, that the probability density functions (PDFs) for the scenario parameters can be determined, e.g., for statistical analyses of the scenarios.

In a second step, the ODD of the ADS is described, and tests are generated from the scenario database by sampling from those scenarios that fall within the ODD of the ADS under test.

3 RISK OF DEPLOYING AN AV ONTO THE PUBLIC ROAD

Safety assessment is about determining whether the safety risk of deploying an AV onto the public road is acceptable or not. Safety assessment procedures aim at quantifying the safety risk by determining the probability that the AV ends up in a collision and addressing the severity of the consequences of such a collision. The safety risk needs to consider each situation that the AV may encounter on the road during its lifetime.

This formulation shows that safety risk depends on a multitude of factors, such as (the list may be incomplete):

1. AV system design: is the design of the system such that it is capable to respond appropriately to all scenarios?
2. The scenarios that the AV actually encounters on the road, including the (potentially irresponsible) behaviour of other traffic participants, the layout of the infrastructure, and the (weather and lighting) conditions. The variation in each of these factors is large, and the number of variations grows further with all the possible combinations that might occur.
3. The impact on the behaviour of the AV due to a possible failure of one of its components or subsystems. This can be an internal failure, such as an overheated control unit or a malfunction of a sensor. This is a Functional Safety topic (ISO 26262, 2018). It is also possible that a component fails to work correctly (or show the intended functionality) due to causes outside of the AV. An example thereof is the failure of a camera due to, e.g., fog, glare, or a frosted lens in wintertime. This is the topic of Safety of the Intended Functionality (ISO 21448, 2021).
4. The capabilities of a possible fallback system that takes over control of the system in case the ADS feature is no longer capable to deal with the situation itself. For lower levels of automation, this is usually the human driver. For higher levels of automation, other approaches are required assuming that the human driver can no longer act as fallback option.
5. Measures taken in and on the AV to mitigate the consequences of a collision such as passive and active safety measures, such as seat belts, airbags, and AEB systems – which might be required in addition to the AV's decision and control logic.

There are two comments that need to be addressed at this stage:

- It is a common misconception that safety of an AV can be fully covered when following the main safety standards: ISO 26262 on Functional Safety and ISO 21448 on Safety of the Intended Functionality. These standards certainly need to be followed to cover the third bullet out of the list above, but conforming to the norms provided by these standards will not guarantee a safe response of the AV for all situations and under all conditions. Additional activities are needed to address the full operational safety of AVs.
- Another common misconception is that for a proper safety assessment of an AV, only those situations that are considered critical on the road have to be evaluated. This assumption is rather popular as it would drastically reduce the number of tests that are required for a proper safety assessment. Unfortunately, it is not possible to focus on critical situations only, as such situations can easily result from poor performance of the AV in any particular situation, whether such situation is initially considered a 'nominal case' or a 'non-critical case'. What appears to be a nominal case for one system, might turn out to become a critical case for another system due to differences in system design and performance.

Many criticality metrics exist, such as for instance Time To Collision (TTC). The general idea is that the smaller the TTC, the higher the risk. Westhofen et al.

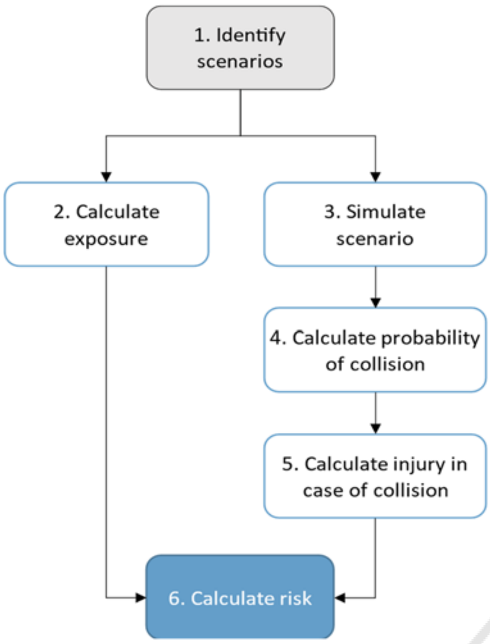

Figure 2: Overview of the risk quantification method as presented by (de Gelder, et al., 2021).

(Westhofen, et al., 2023) review a wide variety of criticality metrics. These are considered surrogate metrics for criticality and are particularly useful in AV development and implementation. However, no statistical analysis is possible that provides a quantification of the residual safety risk of the deployment of an AV based on surrogate criticality metrics.

4 QUANTIFYING RISK

Figure 2 (de Gelder, et al., 2021) provides a method to estimate the risk of an ADS in a quantitative manner. A data-driven approach considering real-world driving scenarios is used to rely less on subjective judgements of safety experts. The output of the method is for instance the expected number of fatal and/or severe injuries in a potential crash. The method is quantitative, the result is easily interpretable, and the result can be compared with road (crash) statistics. Hence, risk quantification is a valuable input to the safety assessment audit as part of the multi-pillar approach (Figure 1).

4.1 Scenarios

The first step of the proposed method is to identify the scenarios that the ADS encounters or may encounter in real life. The term *scenario* is used to provide a quantitative description of the relevant characteristics and activities and/or goals of the ego vehicle, the static environment, the dynamic environment, and all events that are relevant to the ego vehicle (de Gelder, et al., 2022). More informally, a scenario describes any situation on the road including the intent of the subject vehicle, the behaviour of road users, the road layout, and conditions. A drive on the road is considered a continuous sequence of scenarios – which might overlap.

In addition to the identification of scenarios and their variations that occur in the real world, this step also considers the selection of those scenarios that are part of the Operational Design Domain (ODD) of the ADS and the ego vehicle. Once deployed, the ADS needs to deal with many scenarios and the ODD in which the ADS is operating determines the variety of these scenarios. Currently, in the StreetWise scenario database governed by TNO (Op den Camp, de Gelder, Kalisvaart, & Goossens, 2023), scenario categories are used to describe most situations that occur on highways.

To describe an exemplary cut-in scenario in which a vehicle (marked T in Figure 3) changes lane into the lane and in front of an ego vehicle (marked H), several parameters can be used to describe this typical highway scenario:

v_x^H	initial longitudinal velocity of the ego vehicle [m/s]
Δv_x^T	initial relative longitudinal velocity of the target vehicle with respect to the ego [m/s]
$\overline{v_y^T}$	average lateral velocity over the duration of the lane change of the target vehicle in front of the ego vehicle [m/s]
THW_{LC}	time headway at the start of the lane change by the target vehicle [s] = $\Delta x_0/v_x^H$
Δx_0	distance between the target vehicle (rear side) and the ego vehicle (front side) when the target starts crossing the lane marking.

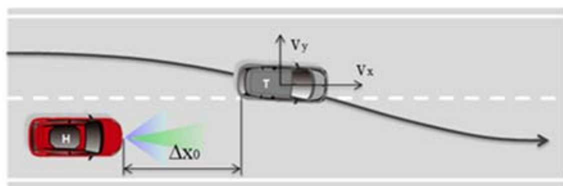

Figure 3: Schematic view of a target vehicle (T) cutting in on an ego vehicle (H).

These parameters, identified for 6.316 realizations of a cut-in in a dataset covering more than 110.000 km of highway driving in Europe, provide valuable statistical information (Paardekooper, et al., 2019) on variations of cut-in scenarios in Europe.

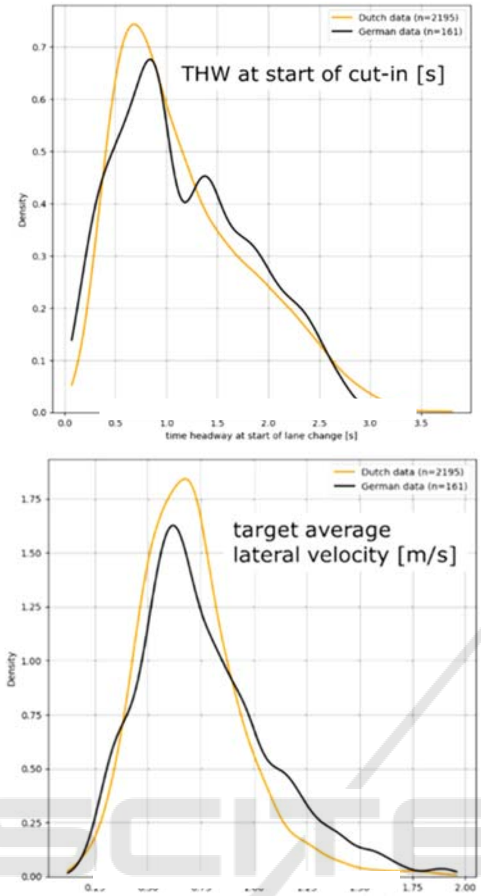

Figure 4: Probability density functions for two of the parameters describing concrete cut-in scenarios collected in approximately 35.000 km in the Netherlands (orange curves based on 2.195 identified cut-ins) and 2.500 km in Germany (black curves, 161 cut-ins).

This is illustrated in graphs of the parameter distributions, or more precisely of the probability density functions (PDF). A PDF indicates the probability that a parameter falls within a particular range, given by the area under the density function and between the upper and lower value of that range. The PDFs, combined with information about how often each specific scenario category is encountered on the road, enables us to calculate the exposure on the road for all the possible scenario variations of the scenarios. This is important for the second step in risk quantification: determining the exposure (Figure 2).

The graphs in Figure 4 indicate [top] the distance of the target vehicle with respect to the ego vehicle at the start of the lane change by the target vehicle expressed as a THW and [lower] the average lateral speed of the target vehicle while changing lane. Since the curves for Germany are based on 2.500 km of driving in which 161 cut-ins were detected, the confidence in the shape of the PDFs is limited, which makes it difficult to draw conclusions on the differences found between Germany and the Netherlands. The peak in the curves for the THW is in the Netherlands at 0.65 s and in Germany at 0.85 s. Further analysis shows that lower THW is associated with a lower average lateral speed of the cutting-in vehicle. Moreover, for a THW lower than 1.0 s, the distribution of the relative longitudinal speed of the target vehicle with respect to the ego vehicle is shifted into a more positive direction. This observation is in agreement with the expectation that for a smaller THW, the lane change is performed somewhat more carefully (slower), and the gap closing speed between both vehicles is smaller.

The example shows that a scenario database is an important tool to get insight into:

- How scenarios typically evolve on the road, what ranges of the parameters can be expected, and what the relation between parameters is;
- How frequently certain parameter values occur on the road (nominal values versus more rare occurrences) or what the probability is of the occurrence of a certain combination of parameter values (cross-correlation);
- Differences in parameter distributions for scenarios that are collected in different regions.

4.2 Virtual Simulation of a Scenario

Step 3 of risk quantification is in the virtual simulation how an ADS or even an AV behaves in each of the scenarios that is relevant for its ODD. To enable the simulation, a simulation framework is required, which is represented by five blocks (Figure 5):

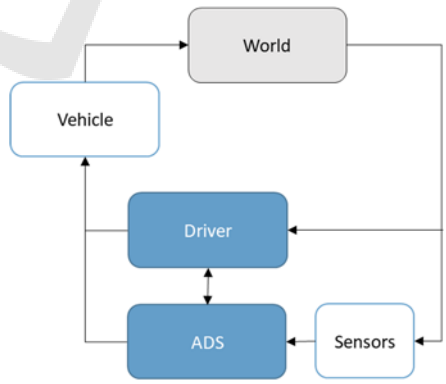

Figure 5: Scheme of the simulation framework.

- World: represents the relevant information about the environment in which the ADS operates. This includes the vehicles (and their manoeuvres) in the direct vicinity of the ego vehicle.

- Sensors: map the information that is perceived by the ego vehicle regarding the environment (both static and dynamic) to the information that can be used by the ADS.
- ADS: the control and decision logic in the vehicle that is used to perform an automated function. The ADS uses the data provided by the sensors to provide control signals to the actuators and to interact with the human driver, e.g., through an HMI.
- Driver: the actual human driver that is behind the steering wheel or that operates the vehicle remotely.
- Vehicle: the system with ADS, actuators, and driver interaction that translates the inputs into vehicle motion, manoeuvring through the world.

The outcome of a simulation is used to determine whether a scenario ends up in a crash or not. In case of a crash, the simulation can also be used to determine the extent of harm that is caused by the crash. This is strongly related to steps 4 and 5 in Figure 5.

4.3 Calculate Risk

With the identification and selection of relevant scenarios in the ODD, the calculation of the exposure to each of these scenarios and the calculation of the possible harm caused by the scenarios, all information is available to calculate the total risk posed by the system onto the road (Figure 5). As explained in (de Gelder, et al., 2021), the risk associated with a scenario category C_i is the combination of the probability of occurrence of that scenario and the expected extent of harm given the scenario. De Gelder (de Gelder, et al., 2021) also shows how to determine an upper bound of the total risk, based on the combination of the risks for the individual scenario categories that potentially are encountered:

$$\text{Risk}\left(\sum_i C_i\right) \leq \sum_i \text{Risk}\,(C_i) \qquad (1)$$

where the equality applies if none of the scenario categories C_i show an overlap. Note that the collection of scenario categories C_i should cover the complete ODD of the ADS. Those situations in which the ADS unintentionally leaves its ODD need also to be considered in risk quantification.

In (de Gelder & Op den Camp, How certain are we that our automated driving system is safe?, 2023), it is shown how to additionally quantify the uncertainties that are associated with safety risk estimation. These uncertainties result from the fact that the database is not complete leading to uncertainty of the exposure and uncertainty in the scenario parameter distributions, and from the fact that only a limited number of simulations can be performed. Using a probabilistic framework, all results are combined to estimate the residual risk as well as the uncertainty of this estimation.

5 DISCUSSION

A structured approach in testing an AV for all possible situations that the AV may encounter on the road during its lifetime, is provided by scenario-based safety assessment. Application of the scenario-based approach is a pre-requisite to calculate the safety risk associated with the deployment of a particular AV onto the public road. The concept of safety risk (the probability that the AV ends up in a collision, considering the severity of the injury resulting from the collision) is widely understood, and basing the safety assessment on that concept helps to come to a fair and acceptable assessment process that is fully in agreement with the multi-pillar approach as proposed by the UNECE (Donà, et al., 2022).

Scenarios and scenario statistics are essential in safety assessment of AVs. Scenarios are typically stored in a database. The use of scenarios for development and testing puts requirements to scenario databases that need to be established. A scenario database should provide a (complete) view on scenarios (and their variations, also depending on region, traffic rules, and driving culture) that a vehicle can encounter on the road during its lifetime. This includes how scenarios evolve over time with the changes in the mobility system. Scenarios should cover nominal everyday driving and more rare and extreme (challenging) cases. Most important is the possibility to determine scenario statistics, with metrics such as:

- Exposure: what is the probability of encountering a scenario within certain parameter ranges or given characteristics, e.g., expressed in the number per 100.000 km of driving;
- Completeness: a quantitative metric that determines how well the scenarios (and their variations) included in the scenario database cover the occurrence of scenarios in the real world. De Gelder (de Gelder, Paardekooper, Op den Camp, & De Schutter, 2019) shows how to determine completeness using the PDFs of the scenario parameters.

Currently available scenario databases, similar to TNO's StreetWise, are far from complete. In other words, the known scenarios do not cover all possible scenarios in the real world. Not only is it difficult to provide a reliable description of the ODD of a function when the scenario database is not sufficiently complete, also the relevance of the selection of test cases is limited in that case. In other words, the function might encounter a scenario on the road for which the function has not been tested, if tests are based on an incomplete scenario database only.

It is for this reason that industrial stakeholders, such as OEMs, TIER1 suppliers and AV developers are building collaborations in scenario mining on public roads (SUNRISE, 2023). It needs to be collaborative efforts, as it is (almost) infeasible for a single party to collect information from all public roads at which a vehicle is potentially deployed.

Also, authorities have an interest to strive for completeness of a scenario database. Although authorities in general do not have the resources or the means for scenario collection at a large scale, testing AVs based on real-world scenarios that cover the complete operational domain of such vehicles is important. For authorities to allow an automated vehicle onto the road (European Commission, 2022), there needs to be sufficient evidence and confidence regarding the safety of the vehicle under all reasonably foreseeable conditions. What is reasonably foreseeable can be linked to the exposure value for a scenario, given that the scenario database is sufficiently complete.

The term 'reasonably preventable' (UNECE, 2022) is often related to the performance of a well-trained, capable and attentive human driver that is put in the same situation. If such a driver is able to prevent a collision, then the 'collision' is called 'reasonably preventable', and an AV should perform at least as good as such a human driver in each of the scenarios.

In addition to being able to prevent reasonably foreseeable and preventable collisions, AVs also need to show driving behaviour fitting for the situation and in agreement what other traffic participants might expect and find acceptable. In (Tejada, Manders, Snijders, Paardekooper, & de Hair-Buijssen, 2020) it is explained how to quantify and characterize of what is called 'safe and social driving'. To be issued a driving license, a human driver needs to behave correctly in traffic according to the locally applicable traffic rules and regulations, and to show safe and acceptable behaviour in the interaction with other road users. It is difficult to quantify what is acceptable or not. One aspect that at least needs to be considered is the ability of a driver to anticipate on behaviour of others and to deploy behaviour that is within the range of expectation of the other traffic participants.

Similar to human drivers, automated systems should not only show safe, but also predictable social behaviour, which is considered good roadmanship. It is for this reason that current research not only focusses on making a scenario database complete, but also on the development of methods to characterize and quantify 'roadmanship' for human drivers, in order to determine criteria for good 'roadmanship' of automated vehicles and appropriate references in the scenarios provided by the databases.

6 RECOMMENDATIONS AND FUTURE RESEARCH

For the safety assessment of AVs, it is recommended to apply the following combination of approaches:

1. A multi-pillar approach: combining different assessment approaches ranging from testing on a test track and in public road trials, the evaluation of virtual simulation results, and the auditing of processes followed by the AV developer;
2. A milestone-based approach: breaking up the large challenge of performing one single safety assessment process to provide a type approval for an AV to be deployed on the public road into smaller more feasible challenges with increasing complexity. The use of milestones leads to the reduction of safety risks in performing the required tests per milestone and reduces the time pressure for the authorities, as the development and implementation of the safety assessment framework can be executed in parallel with the developments of AV solutions by the industry;
3. A scenario-based approach: scenarios are used to describe the situations and conditions that an AV may encounter during operation in the real world, in a structured way. To come to realistic and relevant tests for the safety assessment of the AV, the parameters describing these scenarios are sampled from distributions that are based on real-world data.

Continuous research aims at harmonizing the different scenario-based approaches in Europe, in order to come to a sufficient level of completeness by sharing information from different scenario databases covering different regions. Harmonization includes for instance terminology, scenario definition and parametrization, meta information to be stored along a scenario to enable statistical analysis, and the methods for test case generation out of the relevant scenarios. All these topics are covered in the Horizon Europe project SUNRISE (SUNRISE, 2023).

Research in TNO also addresses methods for estimating a confidence interval for the results of risk quantification (de Gelder & Op den Camp, How certain are we that our automated driving system is safe?, 2023). The level of confidence of a safety risk estimate is important information for road authorities to determine whether or not an ADS can be allowed safely onto the road.

REFERENCES

de Gelder, E., & Op den Camp, O. (2023). How certain are we that our automated driving system is safe? *accepted for Traffic Injury Prevention, Special issue for the ESV 2023 conference, Yokohama*.

de Gelder, E., Elrofai, H., Saberi, A. K., Op den Camp, O., Paardekooper, J.-P., & de Schutter, B. (2021). Risk Quantification for Automated Driving Systems in Real-World Driving Scenarios. *IEEE Access, 9*, 168953 - 168970. doi:10.1109/ACCESS.2021.3136585

de Gelder, E., Op den Camp, O., & de Boer, N. (2020). *Scenario categories for the assessment of automated vehicles.* Retrieved from http://www.cetran.sg/publications

de Gelder, E., Paardekooper, J.-P., Khabbaz, A., Elrofai, H., Op den Camp, O., Kraines, S., . . . de Schutter, B. (2022). Towards an Ontology for Scenario Definition for the Assessment of Automated Vehicles: An Object-Oriented Framework. *IEEE Transactions on Intelligent Vehicles*, 1-1. doi:10.1109/TIV.2022.3144803

de Gelder, E., Paardekooper, J.-P., Op den Camp, O., & De Schutter, B. (2019). Safety assessment of automated vehicles: how to determine whether we have collected enough field data? *Traffic Injury Prevention, 20*, S162-S170.

Donà, R., Ciuffo, B., Tsakalidis, A., Di Cesare, L., Sollima, C., Sangiogi, M., & Galassi, M. C. (2022). Recent Advancements in Automated Vehicle Certification: How the Experience from the Nuclear Sector Contributed to Making Them a Reality. *Energies, 15*. doi:https://doi.org/10.3390/en15207704

European Commission. (2022, August 5). *Implementing regulation - C(2022)5402.* Retrieved from Automated cars – technical specifications: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12152-Automated-cars-technical-specifications_en

ISO 21448. (2021). *Road Vehicles - Safety of the Intended Functionality Standard.* International Organization for Standardization.

ISO 26262. (2018). *Road Vehicles - Functional Safety Standard.* International Organization for Standardization.

Op den Camp, O., de Gelder, E., Kalisvaart, S., & Goossens, H. (2023). *StreetWise: Accelerating Automated Driving with advanced scenario-based safety validation*. Retrieved from TNO Innovation for life: https://www.tno.nl/en/digital/smart-traffic-transport/smart-vehicles/streetwise/

Paardekooper, J.-P., van Montfort, S., Manders, J., Goos, J., de Gelder, E., Op den Camp, O., . . . Thiolon, G. (2019). Automatic Identification of Critical Scenarios in a Public Dataset of 6000 km of Public-Road Driving. *Conference on the Enhanced Safety of Vehicles.* Eindhoven.

SAE On-Road Automated Driving (ORAD) committee. (2018). *SAE J3016, Taxonomy and Definitions for Terms Related too On-Road Motor Vehicle Automated Driving Systems.* SAE International.

SAKURA, SIP-adus and HEADSTART projects. (2021). *White Paper: Towards the Harmonization of Safety Assessment Methods.* Retrieved from https://www.headstart-project.eu/wp-content/uploads/2021/12/211209_HEADSTARTSAKURA_WP_Final-draft.pdf

SUNRISE. (2023). *Safety assurance framework for connected and automated mobility systems.* Retrieved from https://ccam-sunrise-project.eu/

Tejada, A., Manders, J., Snijders, R., Paardekooper, J.-P., & de Hair-Buijssen, S. (2020). Towards a Characterization of Safe Driving Behavior for Automated Vehicles Based on Models of "Typical" Human Driving Behaviour. *23rd IEEE International Conference on Intelligent Transportation Systems, ITSC 2020.* Rhodes, Greece.

UNECE. (2021). *New Assessment/Test Method for Automated Driving (NATM)-Master Document.* Geneva, Switzerland: UNECE.

UNECE. (2021). *UN Regulation No. 157 on uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping System.* Geneva: UN Economic Commission for Europe, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations, ECE/TRANS/WP.29/2020/81.

UNECE. (2022). *Proposal for the 01 series of amendments to UN Regulation No. 157 (ALKS).* Geneva: UN Economic Commission for Europe, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations, ECE/TRANS/WP.29/2022/59/Rev.1.

Westhofen, L., Neurohr, C., Koopmann, T., Butz, M., Schütt, B., Utesch, F., . . . Böde, E. (2023). Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art. *Archives of Computational Methods in Engineering, 30*, 1-35.