

Received March 19, 2020, accepted April 23, 2020, date of publication May 11, 2020, date of current version May 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993730

# Survey on Scenario-Based Safety Assessment of Automated Vehicles

STEFAN RIEDMAIER<sup>1</sup>, THOMAS PONN<sup>ID1</sup>, (Student Member, IEEE), DIETER LUDWIG<sup>2</sup>, BERNHARD SCHICK<sup>3</sup>, AND FRANK DIERMEYER<sup>1</sup>

<sup>1</sup>Institute of Automotive Technology, Technical University of Munich, 85748 Munich, Germany

<sup>2</sup>Highly Automated Driving (HAD), TÜV SÜD Auto Service GmbH, 85748 Munich, Germany

<sup>3</sup>Adrive Living Lab, Kempten University of Applied Sciences, 87435 Kempten, Germany

Corresponding author: Thomas Ponn (ponn@ftm.mw.tum.de)

This work was supported by the TÜV SÜD Auto Service GmbH.

**ABSTRACT** When will automated vehicles come onto the market? This question has puzzled the automotive industry and society for years. The technology and its implementation have made rapid progress over the last decade, but the challenge of how to prove the safety of these systems has not yet been solved. Since a market launch without proof of safety would neither be accepted by society nor by legislators, much time and many resources have been invested into safety assessment in recent years in order to develop new approaches for an efficient assessment. This paper therefore provides an overview of various approaches, and gives a comprehensive survey of the so-called scenario-based approach. The scenario-based approach is a promising method, in which individual traffic situations are typically tested by means of virtual simulation. Since an infinite number of different scenarios can theoretically occur in real-world traffic, even the scenario-based approach leaves the question unanswered as to how to break these down into a finite set of scenarios, and find those which are representative in order to render testing more manageable. This paper provides a comprehensive literature review of related safety-assessment publications that deal precisely with this question. Therefore, this paper develops a novel taxonomy for the scenario-based approach, and classifies all literature sources. Based on this, the existing methods will be compared with each other and, as one conclusion, the alternative concept of formal verification will be combined with the scenario-based approach. Finally, future research priorities are derived.

**INDEX TERMS** Automated vehicles, autonomous vehicles, data analysis, formal verification, intelligent vehicles, key performance indicators, simulation, vehicle safety.

## I. INTRODUCTION

According to the World Health Organization [1], more than one million people died in traffic accidents in 2013. Automated vehicles (Level 3 and higher according to SAE [2]) are expected to make a significant contribution towards considerably reducing this figure in the future. After investing much time and many resources in the implementation of such systems, and due to the existence of various prototypes, the safety issue has received more and more attention in recent years. With SAE Level 1 (Driver Assistance) and Level 2 (Partial Automation), the driver must monitor the system at all times and intervene immediately in the event of a system fault. From Level 3 (Conditional Automation) to Level 5 (Full Automation), safety assessment becomes

particularly important as responsibility is transferred from the driver to the vehicle. Consequently, the safety of Automated Vehicles (AVs) must be thoroughly tested before they can be launched on the market, which is a challenging task [3]–[5].

Various aspects must be taken into account when assessing the safety of AVs. Firstly, safe functionality must be ensured (the so-called Safety of the Intended Functionality, or SOTIF), which focuses on an intended function that could induce hazards due to functional insufficiencies, in the absence of technical system failures [6]. The other aspect is ensuring that the intended function, assuming it is proven safe, does not induce hazards caused by technical failures due to random and systematic faults in the system's hardware or software (functional safety) [7]. The present paper ascribed to the former aspect of safety assessment mentioned. The content of this paper can also be classified under Object and Event Detection and Response (OEDR) by NHTSA [8], [9],

The associate editor coordinating the review of this manuscript and approving it for publication was Malik Jahan Khan .

which is similar to SOTIF but carries a different designation. OEDR examines whether the vehicle is capable of correctly detecting objects and events, and of executing an appropriate response. This must be checked for the entire Operational Design Domain (ODD) of the system. For the rest of this publication, the term “safety assessment” is equivalent to the assessment of SOTIF and OEDR. Additionally, we explicitly consider only the ODD specified by the manufacturer.

The greatest challenge in safety assessment is that road traffic is an open parameter space in which an infinite number of different traffic situations can occur. Absolute proof of safety is therefore not possible, but research is being carried out around various methods, in order to be able to provide the soundest evidence about the system’s safety. Research into this issue started with Advanced Driver Assistant Systems (ADASs) where, e. g. [10] gave an overview of ADAS testing methods in 2015. As early as 2016, Huang *et al.* [11] wrote a short overview of AV test methods, the extent of which was deemed to be unsatisfactory, and which, due to the rapid development in this field, no longer reflected the current state of the art. An overview of the properties of different safety validation methods was published by Junietz *et al.* [12] in 2018. There, the evaluation of the properties is strongly emphasized. However, a comprehensive literature review, similar to the one included in the present paper, is not provided.

Due to the enormous interest in a rapid market introduction of AVs, a large number of publications dealing with the AV safety validation have been published in recent years. Most of them examine scenario-based testing, and therefore we focus strongly on this approach, which is described in more detail in Section II-B.1. The main contributions of this publication are:

- 1) An integration of the scenario-based approach into the overall field of safety assessment
- 2) The definition and clarification of a taxonomy for scenario-based safety assessment
- 3) A comprehensive literature review of approaches on how to define and select scenarios for the scenario-based approach as well as for formal verification (mainly from 2016 to the present)
- 4) A comparison of different methods
- 5) The deduction of necessary research directions for the future

This publication does not focus on standardization activities that can be considered as development guidelines for manufacturers. Examples are the activities of ISO/TC 22<sup>1</sup> and the P.E.A.R.S. initiative [13], [14]. Also beyond the scope of this publication are efforts to develop regulations for the type approval of AVs (e. g. UN-ECE WP.29<sup>2</sup>). Nevertheless, there is a connection between these topics and the content of this paper, because the approaches presented are to be regarded as the basis for the development of regulations

and standards. For a more comprehensive overview on the topic of regulations and standards, interested readers are referred to [15]–[18].

## II. SCENARIO-BASED SAFETY VALIDATION

This chapter first introduces the most important terms and differentiates the scenario-based approach from other safety assessment approaches. Thereafter, a taxonomy for the scenario-based approach is presented and explained.

### A. TERMS AND DEFINITIONS

We start by introducing some key terms, so that a common understanding for this survey paper can be built.

#### 1) SCENARIO TERMINOLOGY

The core methodology of this paper is the scenario-based approach (Section II-B.1) for the assessment of the safety of AVs. It is therefore important to have a common understanding of the term “scenario”. In the context of our paper, we use the definition according to [19], which states that a scenario is a temporal sequence of scene elements, with actions and events of the participating elements occurring within this sequence. “Actions and events” in this respect mean, for example, maneuvers like cut-ins and following a vehicle ahead. Based on this, [20] defines three different categories of scenarios. These are the so-called functional, logical and concrete scenarios. The level of detail and the machine readability increases, beginning with a verbal description of the functional scenarios, continuing to the logical scenarios defined by parameter ranges and distributions, and through to the concrete scenarios defined by exact parameter values.

For logical and concrete scenarios, all parameters that describe the scenario are required. For this purpose, a five-layer model is presented in [21] to structure the parameters. The five layers are as follows:

- Layer 1: Road-level**
- Layer 2: Traffic infrastructure**
- Layer 3: Temporary manipulation of L1 and L2**
- Layer 4: Objects**
- Layer 5: Environment**

Sauerbier *et al.* [22] supplement this layer in the form of a sixth layer (digital information), but as the latter hardly appears in the current literature review, we will use the five-layer model for the remainder of the paper. In order to describe the scenarios uniformly and thus efficiently integrate them into different simulation environments, standards are developed for their description. For the description of static elements, OpenDRIVE<sup>3</sup> is standardized by ASAM, and for dynamic elements, OpenSCENARIO.<sup>4</sup>

Since these definitions were developed and used in the PEGASUS project [23], and are therefore already known to a large part of the community, these definitions are also used in this paper. However, there is not yet a common understanding

<sup>1</sup><https://www.iso.org/committee/46706.html>

<sup>2</sup><https://www.unece.org/trans/main/wp29/introduction.html>

<sup>3</sup><https://www.asam.net/standards/detail/opendrive/>

<sup>4</sup><https://www.asam.net/standards/detail/openscenario/>

of all terms. For example, scenarios are the central element, but there is no common understanding about the duration of a scenario. In the authors' experience, such duration is typically of around 10 seconds.

## 2) MICROSCOPIC AND MACROSCOPIC ASSESSMENT

For the launch of AVs on the market to be socially accepted, it is crucial that they have a lower accident probability than human drivers [24]. In order to be able to make such a macroscopic (statistical) statement about the overall impact of AVs on traffic, a vast amount of data must be made available. Especially in scenario-based safety assessment, on the other hand, individual traffic situations (scenarios) are tested and evaluated. The evaluation of a single scenario is called microscopic evaluation. These definitions of microscopic and macroscopic evaluation are based on [25, Sec. VII]. The transition from a microscopic assessment of a single scenario to a macroscopic assessment of safety is one of the key challenges of the scenario-based approach.

## 3) TESTING, FALSIFICATION AND VERIFICATION

In 2016, Kapinski *et al.* [26] presented a comprehensive overview of simulation-based approaches for assessing embedded control systems, and especially cyberphysical systems. They distinguish between three general techniques: testing, falsification and verification. Since AVs belong to cyberphysical systems, we use this classification as a starting point for our survey.

Suppose there is a model  $M$  of the AV that shows the driving behavior  $\Phi$  and requirements  $\psi$  for the safety of the AV. The model has internal parameters  $p \in P$  and external inputs in the form of a concrete scenario  $u \in U$ . In accordance with the scope of the survey paper and the references therein, the following definitions focus on the ODD / scenario space  $U$ , but can also be extended to the model parameter space  $P$ .

Testing means determining whether the safety requirements are satisfied for a finite set of concrete scenarios  $\hat{U}$ :

$$\text{check } \Phi(M, p, \hat{U}) \models \psi \text{ for } \hat{U} \subseteq U. \quad (1)$$

Falsification is relatively similar to testing, but instead looks for one concrete scenario  $u$  where the AV model violates the requirements. Formally speaking, falsification means to

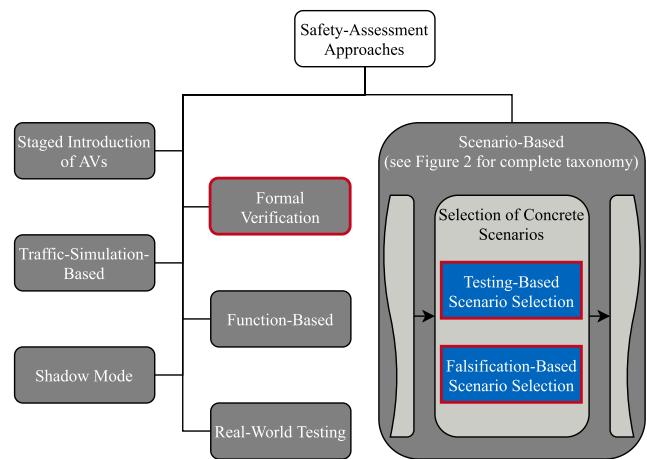
$$\text{find } u \in U \text{ so that } \Phi(M, p, u) \not\models \psi. \quad (2)$$

Even if the model behavior satisfies the requirements in all test scenarios, or no counterexample could be found in the falsification process, safety still cannot be guaranteed across the whole ODD. Formal verification can provide this proof of correctness, but currently lacks scalability to complex systems. Ultimately, verification means to

$$\text{prove } \Phi(M, p, U) \models \psi \text{ for } U. \quad (3)$$

## B. SAFETY-ASSESSMENT APPROACHES

There are multiple approaches available for assessing the SOTIF- or OEDR-related capabilities of AVs. We focus in our paper on the scenario-based approach because it is a very promising method in the current state of the art in science and technology. Nevertheless, since the scenario-based approach is not the only way to assess safety, we also briefly describe alternatives, as shown in Figure 1, and highlight the differences. In general, each of these approaches can be used to assess AVs. In the scenario-based and function-based approaches, a microscopic statement about the safety of the system is first made, which must then be transferred to a macroscopic statement. All other approaches result directly in a macroscopic statement.



**FIGURE 1.** Overview of safety-assessment approaches, with a focus on formal verification and especially the scenario-based approach. Additionally, we highlight the three techniques (outlined in red) from Kapinski [26] from Section II-A, with testing and falsification as scenario-selection methods for the scenario-based approach.

### 1) SCENARIO-BASED APPROACH

By definition, a scenario is a sequence of actions and events. For example, if we consider a typical journey on a highway, there is a significant amount of time in which no actions or events occur. The scenario-based approach, which is also used in large research projects (e.g. in Germany [23], Japan [27] and Singapore<sup>5</sup>), omits the part without significant actions and thus reduces the test scope. In addition, common scenarios, like cut-in situations with large relative distances and a higher velocity of the leading vehicle, which do not provide any relevant information for the safety validation process, can be disregarded. Nevertheless, the issue remains unresolved as to which scenarios need to be considered in scenario-based testing, and how these can be found. This is the central issue addressed in the literature review presented here. According to the definitions given, testing and falsification are techniques for the selection of concrete scenarios ( $u$ ). Therefore, we later distinguish

<sup>5</sup><http://cetransg/>

in our scenario-based taxonomy between testing-based and falsification-based scenario-selection methods.

## 2) FORMAL VERIFICATION

According to the definition given, verification is a mathematical method by which the safety of systems can be formally proven across the whole ODD ( $U$ ). It is not a selection technique to partition the space into scenarios, and is therefore not part of the scenario-based approach. As our conclusion (Section VIII-C) of this paper, we see a combination of the scenario-based approach and verification as a promising approach to efficiently demonstrate the safety of AVs. Therefore, the most important papers in this area are briefly introduced in section VII.

## 3) FUNCTION-BASED APPROACH

In function-based testing, system functions are defined based on requirements, and then tested on the test track or in simulation. This is a widely used procedure for ADAS. Current ISO standards (e.g. ISO 15622 for Adaptive Cruise Control) and UN ECE regulations (e.g. UN ECE R131 for Advanced Emergency Braking Systems) follow a function-based approach and define a few fixed tests for the individual systems, which confirm the basic functionality of the latter.

In order to use the function-based testing method, the functionalities of the system have to be defined. This works for ADAS but is difficult for AVs, because it is impossible to define the required functionality of AVs in every conceivable situation.

Additionally, future standards and regulations should not use a small set of predefined standardized scenarios to test AVs, because this would lead to a performance optimization towards these test cases. Then the assessment result might not correspond to the real driving behavior of the system.

## 4) REAL-WORLD TESTING

A solely distance-based evaluation of safety resulting from field tests is no longer economically feasible at higher levels of automation. In order to be able to state with sufficient confidence that the AV is outperforming humans by a defined factor, according to [28] 11 billion miles would have to be driven in the USA. In this context, outperforming means that fewer fatal accidents occur. Analogous statistical considerations exist for Germany, where [29] conclude that a highway chauffeur needs about 6.6 billion test kilometers.

Real world testing is the standard at low levels of automation, but from Level 3, the required scope increases to such an extent that it is no longer economically feasible.

## 5) SHADOW MODE

Wang and Winner [30] present a method, whereby the automated driving function is executed passively in series production vehicles, which is sometimes known as shadow mode. The driving function is provided with the real inputs of

the sensors, but cannot access the actuators of the vehicle. Simulation can be used to evaluate the decisions of the automated driving function and thus determine the safety level. The same approach is used by car manufacturers, e.g. Tesla<sup>6</sup> to test new systems, and new versions of existing systems.

One major disadvantage, however, is that the behavior of the possible conflict partner (other road users) in the simulation does not correspond to reality, because other road users also plan and execute their actions based on the actions of the AV. If the passive driving function in a situation decides differently than the actual (active) driving function in the vehicle, then perhaps another road user would have decided differently, and the results of the simulation have only limited validity.

## 6) STAGED INTRODUCTION OF AVs

The idea is to limit the ODD of the vehicle and thus the number of traffic situations that occur, to the extent that a safety assessment based on real world testing can be carried out in an economically feasible way. A severely limited ODD would be, for example, driving on a certain section of a road of a few hundred meters or a few kilometers, only under good visibility conditions. In addition, a trained safety driver is part of the safety concept, who can intervene immediately if the system makes incorrect decisions. If the vehicle is assessed as safe in this ODD, the ODD can be gradually increased and/or the safety driver can be omitted. Many system manufacturers apply this procedure, mainly in China and the USA. The most recent example is Daimler and Bosch,<sup>7</sup> who are testing their systems in San Jose, California, on a defined section of road.

For the introduction of Level 4 vehicles in a selected downtown area, this procedure can be very promising. In practice, however, it is not suitable for the validation and approval of Level 5 systems.

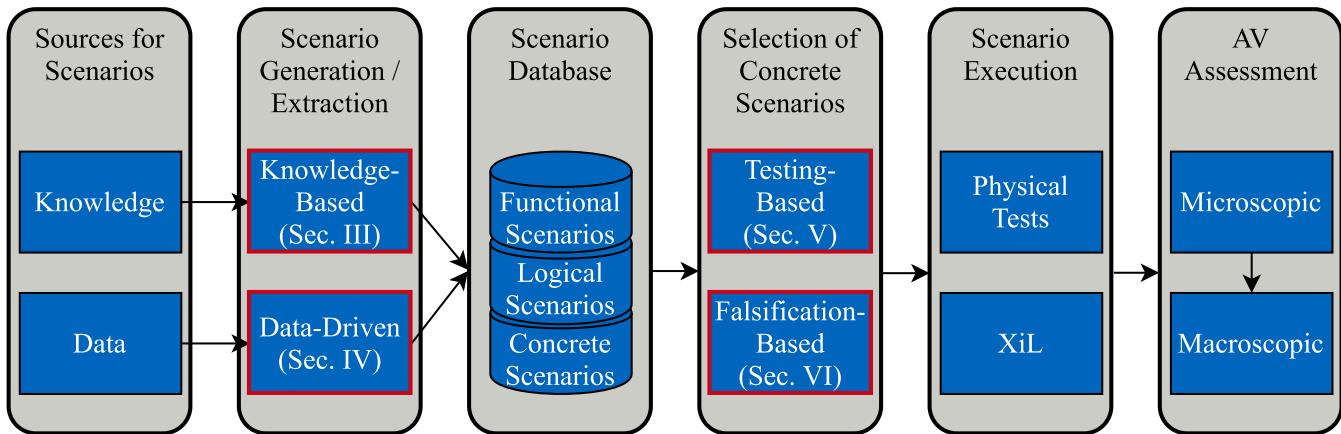
## 7) TRAFFIC-SIMULATION-BASED APPROACH

The concept of traffic simulation is not only to simulate a single scenario, but a whole road network with hundreds of road users (so-called agents). This method is therefore particularly well suited for making a macroscopic statement about the safety of AVs.

Therefore, Kitajima *et al.* [31] develop a multi-agent simulation to estimate the impact of AVs. Hundreds of road users (vehicles and pedestrians) are simulated several times in a 6 km × 3 km road layout in a Japanese city, over a total of 80 000 km. In various simulations, the degree of automation of the vehicles is gradually increased up to Level 4, starting with purely Level 0. Their results show that the number of accidents for the area under consideration decreases from 859 for purely manual driving to 156 accidents at the highest considered automation level (25 Level 2 and 75 Level 4 vehicles).

<sup>6</sup>Tesla Autonomy Day: <https://www.youtube.com/watch?v=Ucp0TTmvqOE> at 2:55:43

<sup>7</sup><https://www.daimler.com/innovation/case/autonomous/pilot-city-san-jose.html>



**FIGURE 2.** Taxonomy of the scenario-based approach. The taxonomy elements reflect a process from left to right.

The introduction of AVs will also change the traffic on our roads. According to [32], this must be taken into account in the safety assessment, by analyzing the change in the frequency of occurrence of scenarios via agent-based simulation. According to the authors, a shift in frequency of occurrence in combination with a change in the severity of damage of scenarios will impact their relevancy.

Saraoglu *et al.* [33] demonstrate a framework called MOBATSIM for the analysis of traffic safety, including AVs, with a focus on the Fault-Error-Failure chain. Thereby, errors such as inaccurate sensor data can be injected and their influence on the safety of the system can be investigated. Thus, the effect of component failures / inaccuracies on the overall traffic safety can be investigated.

The traffic-simulation-based approach can be used to increase the efficiency of the staged introduction of AVs because the whole ODD can be rebuilt in the traffic simulation. For Level 5 systems, this is no longer feasible.

### C. TAXONOMY OF THE SCENARIO-BASED APPROACH

Much literature published recently on the scenario-based approach deals with the question of how to find the set of representative scenarios for the scenario-based approval of AVs. However, the literature is quite diverse. The individual references uncover different strategies and contribute to one or more aspects within the scenario-based approach. Therefore, we developed the taxonomy in Figure 2, which is abstract enough to cover and categorize most approaches, but still reflects the workflow of the scenario-based approach, even for large frameworks [23], [27]. We will refine the taxonomy as the text progresses and place all references within it. If an overlap is detected for an individual reference, we will mention it in the text. In general, the assignment of references is not always unambiguous. Therefore we try to identify the main aspect of the reference and categorize it according to this.

The scenario database is the central element in the taxonomy. According to the definition used in this paper,

the database is just a storage container for all scenario categories. The processing methods are placed externally, and scenarios inserted into the database or scenarios taken from it. The scenario generation/extraction methods use information from different sources to derive different categories of scenarios. On the right-hand side of the database, concrete scenarios are selected based on the database, typically forwarded to a simulation tool-chain for execution, and finally assessed for safety.

We will give a short summary of each block of our taxonomy in order to start from a common understanding. Additionally, we have identified the scenario generation as well as the scenario selection as the most important research topics from a methodological point of view. Therefore, the methods highlighted in red in Figure 2 are described in more detail in Sections III to VI.

#### 1) SOURCES FOR SCENARIOS

Initially, information sources are available which should be used as a basis for the scenario methodology. The information can be in the form of abstract knowledge from experts, standards and guidelines, like the German guidelines for the construction of highways [34] and consumer tests, or in the form of driving or accident data.

Another source of information is data from real world driving (e.g. field operational tests). A prerequisite for the generation of scenarios from driving data is a data set that is as comprehensive as possible. Many institutions and companies have their own, non-accessible data sets, but in recent years, publicly accessible data sets have increasingly been made available by various organizations. An overview of available data sets can be found in [35], [36]. Zhu *et al.* [37] also show an overview of data sets and try to unify them.

Krajewski *et al.* [38] show a novel method for recording real driving data. Here, traffic is recorded with the help of a drone, and the trajectories of the individual road users are extracted from the images using computer vision. This entails

the advantages that no complex and cost-intensive test vehicles with comprehensive sensor technology have to be set up, and that the traffic is not affected by the measurement. However one disadvantage of the presented method is the comparatively small section of about 400 meters that can be recorded by the drone, which is a limitation especially when extracting highway scenarios.

## 2) SCENARIO GENERATION/EXTRACTION

On the one hand, we present knowledge-based approaches in Section III, which take the abstract information from e. g. experts and generate scenarios from it, and on the other hand data-driven approaches in Section III, which extract scenarios from recorded data sets.

## 3) SCENARIO DATABASE

For scenario-based testing, a database with test scenarios is the central element. Due to the number and high dimension of the scenarios, an efficient description and storage of the scenarios is essential. Within the PEGASUS project [23], a database with relevant scenarios for the ODD highway is created [39], [40]. The focus is on a standardized interface for reading in different data sources and processing them into a machine-readable format. A further framework for creating a database, called the Testing Scenario Library, is described in detail in [41], [42]. They also use the definitions for the different scenario types from the PEGASUS project. Another approach to building a scenario database can be found in [43]. Althoff *et al.* [44] introduce the Commonroad framework, including not only scenarios but also models and cost functions, in order to fully reproduce experiments for evaluating motion planners.

## 4) SELECTION OF CONCRETE SCENARIOS

In Section V and VI we will distinguish between scenario-selection approaches, focusing on covering the parameter space with test cases, and approaches focusing on challenging corner cases to find counterexamples. This partitioning is in line with the definitions given for testing and falsification in Section II-A.3.

## 5) SCENARIO EXECUTION

Different testing environments are available for the execution of the selected concrete scenarios. These can either be performed in the real world via field or proving-ground tests, or using a different degree of virtualization via X-in-the-Loop (XIL) simulation [45]. Since the latter has many advantages regarding e. g. costs, expenditure and safety risks, almost all references use simulation for their proof of concept. There are many commercial and free simulators on the market, as well as simulation frameworks developed in the literature [46], [47].

## 6) AV ASSESSMENT

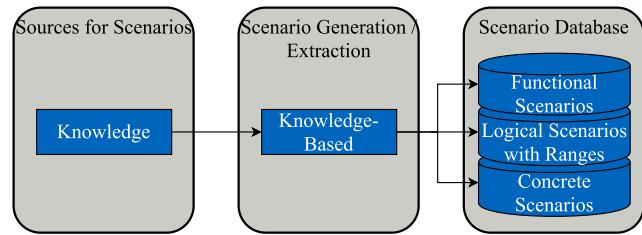
In Figure 2 we distinguish between microscopic and macroscopic safety assessment according to the definitions given

in Section II-A.2. In order to evaluate safety within a microscopic assessment, Key Performance Indicators (KPIs) are needed. Since accidents are rare events, it is beneficial to use criticality metrics as KPIs. The most well-known of these is the Time-to-Collision (TTC) of [48]. There are also many variations of it, for example in [49]–[51]. An overview of criticality metrics can be found in [52]. Hallerbach *et al.* [47] introduce four domains of interest to evaluate the criticality within different spatial areas around the AV. The number of critical situations and accidents occurring in the microscopic assessment can be used for transition to a macroscopic assessment.

The testing-based and the falsification-based scenario-selection methods are both able to assess the safety microscopically for each scenario. The testing-based methods (Section V) focus more on covering the scenario space, whereas the falsification-based methods (Section VI) focus more on finding corner case scenarios. Selecting corner cases is indeed a very efficient way of finding counterexamples, but it is not well suited to a macroscopic assessment. The coverage-based testing approaches include a broader representation of real traffic behavior, and are therefore more suitable for transferring the microscopic results to a statistical macroscopic statement using the parameter distributions (exposure) of microscopically assessed scenarios.

## III. KNOWLEDGE-BASED SCENARIO GENERATION

The knowledge-based approach uses abstract information to create functional, logical or even directly concrete scenarios for the database, as per Figure 3.



**FIGURE 3. Knowledge-based scenario generation.**

Besides standards and guidelines, expert knowledge can be used as a source for scenarios. Ontologies are a widely used method for storing and structuring expert knowledge. Therefore, the focus of knowledge-based scenario generation in this publication is on the use of ontologies.

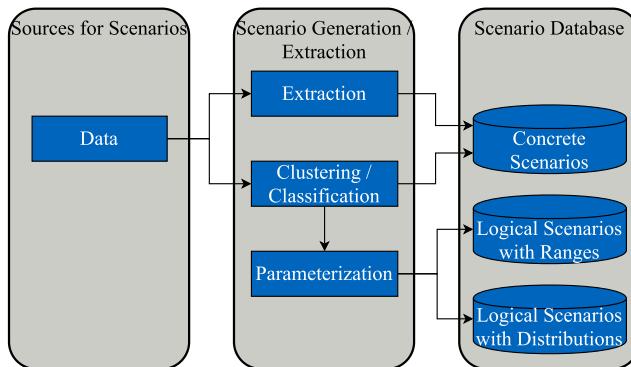
Geyer *et al.* [53] initially proposed a fundamental ontology for AV guidance, and this forms the basis for many of the following references.

Bagschik *et al.* [21] use an ontology for the knowledge-based creation of scenes specifically for German highways, and include all five layers of their five-layer model. According to the authors, the advantage of using ontology as a knowledge base is that it is not necessary to check the resulting comprehensive scene catalogue, only the knowledge base itself. If, for example, different lane markings are defined in

the ontology, it is ensured that the catalog contains scenes with the different markings.

Chen and Kloul [54] combine three ontologies as a knowledge base for the generation of motorway scenarios: a motorway, a weather-based and a vehicle ontology, each connected by relations and effects. Furthermore, traffic rules are expressed as first-order logic. However, according to the authors, the representation of time dependency is still missing in their implementation.

Graz University of Technology [55]–[57] developed an ontology-based framework for the simulation-based testing of AVs. The developed process consists of three steps. First, an ontology is built and used as a input for combinatorial testing (Section V) in the next step. The generated scenarios are automatically transferred to a simulation environment and executed there. In their publications they apply their process with three ontologies, which differ in size and complexity. They also compare two algorithms for converting the ontology into an input model for combinatorial testing. According to the authors, their approach enables an industrial application.



**FIGURE 4.** Data-driven approach to generate scenarios.

#### IV. DATA-DRIVEN SCENARIO EXTRACTION

For the data-driven generation of scenarios according to Figure 4, there are a multitude of different approaches in the literature, which typically use machine-learning or pattern-recognition methods. There are extraction methods that directly filter concrete scenarios without any assignment to predefined logical scenarios or similar clusters. These methods can typically detect novelties in existing data, and sometimes even generate new concrete scenarios, for example with generative neural networks. Alternatively, scenario clustering and classification methods can group the data to also obtain concrete scenarios, but with a kind of group membership. In scenario clustering, the groups are similar clusters and the assignment is made in an unsupervised learning fashion, whereas in scenario classification, the groups are predefined logical scenario classes and the assignment is made in a supervised learning fashion. This has the advantage that the classified data can be used subsequently to

describe the parameters of a logical scenario by ranges or even distributions.

##### A. EXTRACTION OF SPECIAL CONCRETE SCENARIOS

A method for evaluating the uniqueness of a concrete scenario is presented by [58]. The authors use the time signals of a scenario as an input for an autoencoder. According to the authors, by compressing and reproducing the time signals using the auto-encoder, the reproduction error can be used as a novelty indicator of a scenario. This means a high reproduction error is an indication of a rare scenario. The approach can be used to fill a database with representative concrete scenarios. The authors were able to perform an exemplary validation. A global validation of this procedure could however not be shown.

For the authors of [59], so-called corner cases constitute particularly challenging scenarios. In their publication, they consider it as particularly challenging if an object in a position relevant to the driving task cannot be predicted, or only poorly so. The detection of such cases can be performed offline as well as online while driving using camera data. This is done in three steps. First, relevant objects are detected by semantic segmentation. The behavior of the objects in the next time step is predicted using artificial intelligence. In the last step, corner cases are detected by means of a threshold exceeding the deviation between the prediction and the actual behavior of the objects.

Krajewski *et al.* [60] use generative neural networks to model maneuver trajectories from recorded data and to generate new trajectories from them. They compare an extension of the Generative Adversarial Network (GAN) with an extension of the Variational Autoencoder (VAE). By varying the values of the model input parameters, the network can fill a scenario database with many concrete scenarios. Similarly, [61] use Recurrent Neural Networks (RNNs) to model driving data as a sequence and to generate new concrete scenarios from it.

Jesenski *et al.* [62] focus on the variation of the road topology and the position and speed of road users (Layers 1 and 4 of the five-layer model [21]) for scenes. Using Bayesian networks, they model complex road layouts including the relationships between individual road sections. The network is trained on the basis of a real data set and, according to the generated results, it reflects the traffic conditions of the data set. However, this approach can currently only be used to generate scenes and not scenarios, which the authors intend to address in their future work.

##### B. SCENARIO CLUSTERING

Kruber *et al.* [63] and Kruber *et al.* [64] apply an unsupervised clustering technique to find similar situations from measurement data, and group them into clusters. If new measurement data is added, the data is assigned to the existing clusters if they exceed a defined similarity. Otherwise, a new cluster is created for the new measured data.

Watanabe *et al.* [65] define a mixed similarity measure to quantify the distance between scenario signals of different types and to quantify the cluster centrality. Based on the distance and centrality measure, they compare the approximate k-covers algorithm as a Hierarchical Agglomerative Clustering approach with the Partitioning Around Medoids of k-medoids as a Partitional Clustering approach.

Wang and Zhao [66] introduce a four-step approach. In the first step, they extract scenario primitives from time series data without prior knowledge. In the second step, they cluster the primitives and therefore generate primitive templates. They address both steps with a non-parametric Bayesian learning method – a sticky, hierarchical, Dirichlet-process Hidden-Markov-Model. In the last two steps, they plan to model dynamic stochastic relations between the template sets, and sample from them in order to select new concrete scenarios (Section V).

### C. SCENARIO CLASSIFICATION

The definition of logical scenarios based on a potential collision direction between the ego-vehicle and another vehicle is the subject of [67]. This methodology is developed especially for highway scenarios and can also include boundary conditions, such as an action restriction for the ego-vehicle due to other road users. The presented method can thus be seen as a preliminary stage for classification (and also clustering) of real driving data.

The extraction of concrete scenarios from real driving data and their assignment to one of the eight considered categories of logical scenario is shown by [68]. Only parameters of dynamic objects (Layer 4 of the five-layer model [21]) are considered. In their work, the authors use synthetic data with available class labels for training and real data for testing classification algorithms. They achieve the best results with supervised learning methods.

The classification of scenarios from real driving data on motorways and country roads especially, for a Lane-Keeping Assist system is presented by [69] using an end-to-end Deep Learning approach. With the help of Convolutional and Recurrent Neural Networks (CNNs and RNNs, respectively), the scenarios are assigned to one of the 13 considered scenario classes based on ten different sensor channels, which corresponds to the definition of logical scenarios. The developed method can be applied online as well as offline, and their results show that CNNs have a higher accuracy.

Gruner *et al.* [70] represent scenarios in spatiotemporal grid-maps. They address Layer 4 of the five-layer model [21] and aggregate the information regarding all objects within a grid-map, independent of the number of objects. The grid cells without objects are just empty (white pixels). This represents a large difference compared to the time-series representation, where each object requires its own time series. In return, several maps are needed for the time component, and generally for several channels. They distinguish three types of grid-maps. The Velocity Grid (VeG) consists of three grid-maps for the channels occupancy, longitudinal and

lateral velocity. The Stacked Velocity Grid (SVeG) consists of six grid-maps: the three VeG maps, multiplied by two points in time. The History Grid (HiG) consists of the three VeG maps, but incorporates the time component into the occupancy map via a fading effect of gray pixel values. For classification, they use a CNN and show that it currently works best with the SVeG representation.

Dávid *et al.* [71] use artificial intelligence for the online determination of the current risk and classification of the current driving situation. The focus of the authors is on the methodology, and the classified situations are not further used for the safety assessment.

Bach *et al.* [72] propose a two-step approach to derive a test set from recorded data. The first step partitions scenarios into categories based on the system requirements, and pre-selects the scenarios from the partitions using a classification-tree. In the second step, they reduce the amount of scenarios by analyzing them with a coverage criterion and 2D histograms, and by discarding repetitive scenarios.

### D. SCENARIO PARAMETERIZATION

Hartjen *et al.* [73] address multiple steps of our taxonomy at a high level, with an initial proof of concept at an intersection scenario. They define a maneuver catalog for urban vehicular traffic. They extract the maneuvers from data based on a simple rules-based classification. They parameterize the object trajectories with Bezier splines, and learn the parameter distributions from the data. Finally, they sample new trajectories from the distributions in order to select concrete scenarios (Section V).

Zhou and del Re [74] model a lane-change with a hyperbolic tangent function and parameterize it with data from field tests. They conclude that the tail of the parameter distributions can be taken to assess rare critical scenarios (Section VI) and that the common driving behavior focusing on scenario coverage can be used to determine the safety boundary of the AV (Section V). This is in line with our proposed framework.

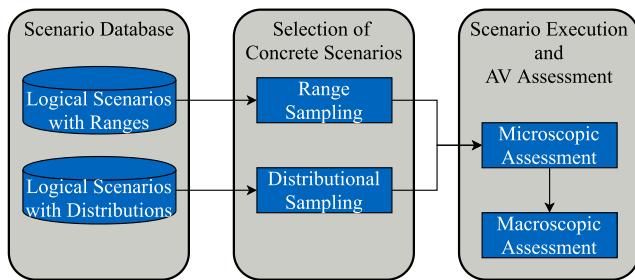
Zofka *et al.* [75] use a particle filter to estimate scenario parameters from field data. They modify the object trajectories with small spatial and temporal translations to select new scenarios and preserve the plausibility of the original scene (Section V).

de Gelder and Paardekooper [76] use Kernel Density Estimation to fit a distribution to the parameters from real-life scenarios. Sampling from the distributions via Monte Carlo simulation techniques, they are able to ensure that the assessed safety level corresponds to the real safety level (Section V). Using Importance Sampling, they are able to generate new scenarios that are particularly critical. The authors demonstrate the applicability of the method by means of the evaluation of an ACC system.

de Gelder *et al.* [77] examine how many scenarios from real data are needed to completely describe the parameter ranges of the activity “braking”. For this purpose, the probability density function (PDF) of the parameters of the activities is determined by kernel density estimation (KDE).

## V. TESTING-BASED SCENARIO SELECTION

What the testing-based approaches for scenario selection have in common is that they sample a subset of concrete scenarios for the microscopic assessment of safety in each individual scenario, and offer the possibility of aggregating the individual results into a macroscopic assessment. Within this section, we distinguish between two types of sampling as shown in Figure 5. Either samples are generated within parameter ranges specified by minimum and maximum values, or from parameter distributions. The latter contains the probability of occurrence (exposure) of the scenarios and therefore allows a weighting of the results for a true statistical macroscopic statement about the accident probabilities. The former intends to cover the entire parameter range, neglecting the significance of the scenarios in real world, and therefore allows just an overall statement to be made, based on the coverage.



**FIGURE 5.** Testing-based scenario selection.

### A. SAMPLING WITHIN PARAMETER RANGES

N-wise sampling is a standard technique where all possible combinations of the parameters are considered. It is only applicable with a coarse discretization of the parameters and for comparable simple systems like SAE Level 1 functions, e.g. Lane-Keeping Assistants [78].

Beglerovic *et al.* [79] use an interactive Design of Experiments (DoE) procedure to generate concrete cut-in scenarios. They identify the system behavior and model it via a robust neural network. After an initial test design, the data-driven model is used for the purpose of optimization within the interactive DoE, in order to generate more concrete scenarios in the area of interest (Section VI-D). They analyze the criticality using KPIs, time plots and Pareto fronts.

An automated framework for regression testing is presented by [80], whereby parameter variations of roads, static and dynamic objects and also of environmental conditions are automatically created and combined. To make sure that all scenarios are physically reasonable, a modified In Parameter Order Generalized (IPOG) algorithm using a nonrecursive backtracking algorithm is used, in combination with a trajectory planner. According to [80], the algorithm works as intended, but needs to become more efficient in the future.

Kim *et al.* [81] generate road networks based on Satisfiability Modulo Theories (SMTs). They define curve coverage criteria and several constraints, and use an SMT-solver to either

generate multiple road segments from a single criterion, or to directly generate the road network from multiple criteria. They further develop the approach in [82]. Majzik *et al.* [83] monitor the AV behavior using Signal Temporal Logic (STL) for the purpose of assessment at the system level. They investigate coverage of an existing test suite with respect to regulations from safety standards, and derive new challenging test cases by increasing coverage with graph generation techniques. Khastgir *et al.* [84] generate concrete scenarios by applying randomization techniques to the cut-in point of a traffic vehicle and to the brake and accelerator pedal values of the AV.

Huang *et al.* [85] focus on surrounding vehicles and create logical scenarios using the example of a Level 2 vehicle. New logical scenarios are created by adding surrounding vehicles, and by varying their parameters such as the starting position and speed, as well as the lateral and longitudinal behavior in the course of the scenario. The combination of these parameters results in a large number of different concrete scenarios, which is why the number is reduced via a assessment of the Scenario Importance. Here it is examined to what extent an action of a surrounding vehicle restricts the desired behavior of the ego-vehicle. If the influence is large, the Scenario Importance is large and vice versa. All scenarios with a Scenario Importance lower than a threshold value are not considered further. The method is implemented using a curve scenario as an example.

Xie *et al.* [86] present a similar approach, which is implemented in three scenarios (curve, following and lane-change). Zhou and Re [87] also focus on surrounding traffic participants and create a structured test catalogue for an Adaptive Cruise Control (ACC) based on the number of participants and their abstract behavior, which according to the authors is sufficient to obtain a sufficient coverage of critical scenarios from real driving data.

Althoff and Dolan [88] use rapidly exploring random trees (RRTs) as a motion-planning algorithm to generate trajectories, in order to check the results of a reachability analysis (VII). They test whether the reachable set of a high-order model is within the reachable set of a low-order model, so that the latter can be used for inexpensive computations. RRTs are impressive due to their good coverage of the state space, yet still guiding the simulation such that uninteresting simulation traces are abandoned (Section VI). Tuncali and Fainekos [89] determine the boundary scenarios where the transition from safe driving to a collision occurs. They define a custom cost function for the RRTs based on the collision surface, velocity and TTC.

### B. SAMPLING FROM PARAMETER DISTRIBUTIONS

Monte Carlo (MC) is a standard technique for sampling from parameter distributions. Since most of the scenarios are not critical for AVs, MC sampling is inefficient. Therefore, most of the paper introduces accelerated approaches which use Extreme Value Theory and can be compared to MC sampling as the baseline.

An accelerated approach using Extreme Value Theory is presented by Åsljung *et al.* [90], [91]. Based on real data and a criticality metric, the safety level of the system is predicted using near misses. They note that the prediction depends considerably on the criticality metric used. Their results in [91] show that the Brake Threat Number is a promising criticality metric. According to them, this approach requires 45 times less measurement data than a statistical approach.

Zhao [92], Zhao *et al.* [93], and Zhao *et al.* [94] develop a statistical model of the behavior of road users based on real data. Subsequently, the behavior is modified in such a way as to provoke more intensive and critical interactions between the automated vehicle and those surrounding it. In addition, Importance Sampling Theory is used to ensure the accuracy of the method. According to them, a safety assessment that is between 300 and 100 000 faster than real-world testing is possible. The method is applied in an exemplary manner for the logical scenarios cut-in and car-following. These publications form the basis for further ones of their research group in the next paragraph.

Huang *et al.* [95] use Piecewise Mixture Distribution Models to model the behavior of the vehicles. The procedure is carried out in an exemplary cut-in situation. They show that this method is 7000 times faster than the crude Monte Carlo method. In [96], [97] the accelerated assessment is performed with a sequential learning approach based on kriging models. Here, a heuristic simulation-based gradient descent procedure is used iteratively to find the best scenario that maximizes an information criterion regarding the accuracy of the conflict probability. According to them, the procedure is more efficient than the random selection of test scenarios, but no quantitative statement is made. There is similar work from Huang *et al.* [98], Huang *et al.* [99], Huang *et al.* [100], and Huang *et al.* [101], [102] as well as from other members of their research group [103], [104].

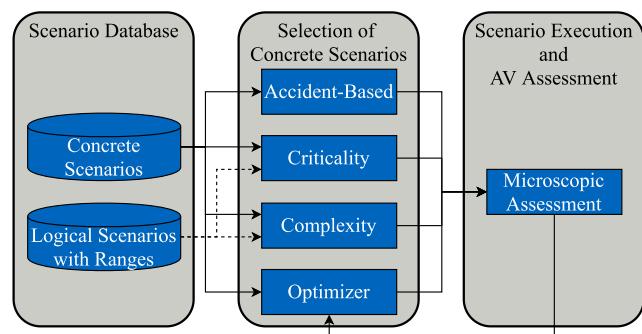
There are further publications by other authors who also use Importance Sampling [105]. Wang *et al.* [106] combine their Importance Sampling-based accelerated approach with a Reachability Analysis (Section VI) and apply the approach to the functional scenario of a pedestrian crossing. The advantage of the presented method is that all generated scenarios are physically feasible and therefore realistic.

An accelerated method without making a quantitative statement about the overall risk level is presented by [107]. Analogous to the accelerated methods, the parameter distributions are determined on the basis of real data. In addition, a traffic risk index is used to evaluate the scenarios with the corresponding parameters. Subsequently, it is possible to define more critical scenarios automatically with the help of Markov Chain Monte Carlo sampling. A reverse calculation of the tested scenarios back to an overall safety statement – as occurs with the accelerated procedures – is not performed.

Olivares *et al.* [108] focus purely on Layer 1 (road topology) of the five-layer model [21]. The authors develop a road generator that generates road geometry using the Markov

Chain Monte Carlo method. The necessary parameter distributions are extracted from OpenStreetMap.

Åsljung *et al.* [109] highlight the influence of the discretization of the states of traffic participants in the calculation of criticality metrics. The movements of other road users are predicted using a Markov chain model, and the probability of a future collision is calculated. The authors conclude that the discretization of the states has a significant influence on the resulting error of the criticality metric.



**FIGURE 6. FALSIFICATION-BASED SCENARIO SELECTION**  
The aim of the falsification approaches is to find counterexamples violating the safety requirements during microscopic assessment. According to Figure 6, they either take existing concrete scenarios from the database, or simply logical scenarios with parameter ranges. One option for scenario selection is to use an accident database. Another is to take an exemplary concrete scenario and increase its criticality. The third option is to take a logical scenario and find critical scenarios within the defined parameter ranges of this logical scenario. Instead of increasing criticality directly, the probability of finding a counterexample can also be increased by increasing the scenario complexity. Finally, Section VI-D introduces simulation-based falsification. This is distinguished by an additional feedback loop and can therefore use the assessment results of the simulation for optimization to select the next concrete scenarios. In contrast to the database, which mainly consists of fleet data, this is actual data from the vehicle under test.

## VI. FALSIFICATION-BASED SCENARIO SELECTION

The aim of the falsification approaches is to find counterexamples violating the safety requirements during microscopic assessment. According to Figure 6, they either take existing concrete scenarios from the database, or simply logical scenarios with parameter ranges. One option for scenario selection is to use an accident database. Another is to take an exemplary concrete scenario and increase its criticality. The third option is to take a logical scenario and find critical scenarios within the defined parameter ranges of this logical scenario. Instead of increasing criticality directly, the probability of finding a counterexample can also be increased by increasing the scenario complexity. Finally, Section VI-D introduces simulation-based falsification. This is distinguished by an additional feedback loop and can therefore use the assessment results of the simulation for optimization to select the next concrete scenarios. In contrast to the database, which mainly consists of fleet data, this is actual data from the vehicle under test.

### A. SCENARIOS FROM ACCIDENT DATABASES

The use of accident data as a basis for test scenarios arises from the safety assessment of driver assistance systems. These driver assistance systems are permanently monitored by the driver and therefore have to perform well (from a safety point of view) only in those cases where the human driver shows insufficient performance, which would lead to accidents.

Various publications deal with the definition of test scenarios from accident data, focusing on different aspects. Stark *et al.* [110] and Stark *et al.* [111] use the GIDAS

accident database to investigate which requirements have to be met by automated systems in order to avoid as many accidents as possible (with a focus on urban areas). A comparable procedure is carried out by [112] for a motorway chauffeur. Fahrenkrog *et al.* [113], on the other hand, use accident scenarios as a basis for a simulation-based variation of parameters, in order to be able to make a more general statement about the accident-avoidance potential of the system. The derivation of representative logical scenarios from an extensive accident database is carried out by [114] using Big Data techniques.

The exclusive use of accident data does not correspond to a safety assessment of an AV (Level 3 and higher), because accident data only investigates which accidents can be avoided by the system, but does not indicate which accidents / risks will occur in the future due to the system.

#### B. CRITICALITY-BASED SELECTION

In [115] a method is presented with which the risk of real traffic situations can be efficiently determined in order to select critical scenarios for the testing of AVs. The focus is on the behavior of other road users, which corresponds to Layer 4 of the five-layer model [21]. The risk at a spatial location depends on the position and speed of the road users. If the environment of the AV is evaluated as presenting a high-level risk, a critical scenario exists. The procedure is evaluated using the real data from the HighD and the NGSIM data set.

The generation of critical scenarios from complex urban scenarios is the focus of [116], [117]. The criticality of a scenario is calculated based on the area that can be used safely by the AV. Optimization by means of evolutionary algorithms maximizes the criticality of complex scenarios and minimizes the safely passable area, respectively. This is done by adapting the behavior of the surrounding traffic participants. The definition of criticality used in [116], [117] differs from the definition used in other publications, because criticality is not dependent on the performance of the AV, which is similar to complexity in other publications. A validation of whether the generated scenarios in [116], [117] also lead to critical situations is not performed.

#### C. COMPLEXITY-BASED SELECTION

Several references [118]–[120] develop a procedure that combines combinatorial testing and the definition of complex test scenarios. The complexity is described by a complexity index, which assigns a weighting to each parameter of a scenario, using the Analytic Hierarchy Process. In their publication, 16 different parameters are assigned and the process is validated via the evaluation of a lane-departure warning system. The authors can demonstrate that more complex scenarios reveal more system errors.

In [121] the complexity of a scenario is divided into two categories. The first is a Road Semantic Complexity, which describes the complexity of the static environment and is determined using Support Vector Regression. The second is a

Traffic Element Complexity, which describes the complexity of the behavior of up to eight surrounding traffic participants. A validation of the metric is not performed in the publication.

In order to efficiently test the cognitive capabilities of an AV, Zhang *et al.* [122] develops a framework that takes the complexity into account when selecting scenarios. They describe the complexity in terms of cognitive tasks, depending on the road type, semantic content of the road segment and challenging conditions such as fog. In a comparison of two autonomous driving platforms, they can show that there is a negative correlation between performance and the complexity of the scenarios.

Qi *et al.* [123] use a Scenario Character Parameter (SCP) based on the trajectories that lead to a failure. By analyzing the SCP, scenario groups can be created and reduced to one relevant or challenging scenario.

An optimization-based approach (without concrete implementation) for defining challenging scenarios is presented in [124]. Partial aspects of this approach are examined in more detail in [125].

An overview of complexity-influencing factors can be found in [35] and [126].

#### D. SIMULATION-BASED FALSIFICATION

The approaches in this sub-section use an optimizer, which takes the microscopic assessment results of the scenario simulations from the feedback loop visualized in Figure 6. Depending on the optimizer, it processes one or more scenarios in parallel per iteration, e. g. depending on the population size in a genetic algorithm. At the outset, the optimizer needs to be initialized using concrete scenarios and the corresponding assessment results. Based on a cost function that includes an expression of vehicle safety, the optimizer selects the next concrete scenarios and forwards them to the simulation tool for execution and assessment. With the new assessment results, again, the optimizer determines the next concrete scenarios and the next iteration starts. By minimizing the cost function, the optimizer can – with each iteration – determine more and more critical scenarios for the vehicle under test. Normally this approach requires many iterations in a simulation environment.

The research group of Prof. Kochenderfer uses Reinforcement Learning to find the most likely critical scenarios, and calls it Adaptive Stress Testing [127]–[130]. Koren *et al.* [127] use Monte Carlo Tree Search and Deep Reinforcement Learning. They select a scenario with a vehicle approaching pedestrians on a crosswalk as a proof of concept. The Reinforcement Learner generates pedestrian trajectories and sensor noise to consider both actions and sensor failures. The advantage of Reinforcement Learning is that it can even change the time signals during run-time based on the assessment results in the current time step within the scenario execution. Corso *et al.* [128] develop the approach further with a reward-augmentation technique. Both papers build on a predecessor paper [129] from the avionic domain. The latter also provides the basis for a paper [130] addressing

the differential comparison of two simulators. Instead of minimizing safety, the learner tries to maximize the deviation between both.

Beglerovic *et al.* [131] propose an approach with a loop of surrogate modeling and stochastic optimization. They use kriging as the surrogate modeling technique with Differential Evolution Genetic Optimization and Particle Swarm Optimization. The idea is to not execute all optimization iterations on the expensive simulation engine to determine the global minimum, but instead use a cheap surrogate model. As the surrogate model is just an approximation of the simulation behavior, they add an outer loop to refine the surrogate model. A so-called zooming-in algorithm takes new samples in the area of the global minimum, executes them on the simulation engine and updates the surrogate model in this faulty area. Therefore, with each iteration, the surrogate model gets better in the faulty area and the optimizer improves in its determination of the global minimum or faulty area. In the example, the cost function is based on the TTC, and the optimizer controls the coordinates of an object that the AV is approaching. They inject an error into the sensor's field of view to check if the optimizer finds scenarios that exploit this weakness. Beglerovic *et al.* [131] adapt the approach in Abdessalem *et al.* [132] by using the zooming-in algorithm and Kriging instead of neural networks.

Mullins [133] develops the Range Adversarial Planning Tool (RAPT) framework to generate test scenarios. He uses adaptive search-algorithms to iteratively generate new concrete scenarios based on previous results. Related work has already been published by the author in [134], [135]. Gangopadhyay *et al.* [136] use a Bayesian optimization, [137] use a random forest model and Abbas *et al.* [138] use simulated annealing in their test harness capable of testing perception algorithms.

Tuncali *et al.* [139] use formal system requirements, not for formal verification but for falsifying them. They use their automatic test generation tool S-TaLiRo and select simulated annealing for optimization. As a cost function they define a robustness metric that quantifies the gap to falsifying the formal system requirements. The optimization engine and the stochastic sampler try to minimize the robustness metric. In the example, the robustness metric is based on the TTC, and the optimizer controls the target speed for the vehicle in a collision-avoidance setup. It does not directly generate the time signal, but controls a predefined number of control points instead, which are then interpolated. Tuncali *et al.* [140] further develop the approach with a hierarchical framework combining high-fidelity and low-fidelity models. They use a functional gradient descent optimization that outperforms the simulated annealing.

In [141]–[143], covering arrays for combinatorial testing as well as simulated annealing for falsification are applied both in a simulation-based framework to the overall system including sensors. The idea is to use the results of another scenario-selection technique to enhance the initialization of the optimizer. Starting from more promising conditions,

the optimizer converges faster to the global optimum, or perhaps finds a better local optimum. Felbinger *et al.* [144] compare falsification and combinatorial testing for an Autonomous Emergency Braking (AEB) System. Their results show that both methods have been proven to find critical scenarios, but the authors do not assess the efficiency.

Koschi *et al.* [145] divide the falsification of an ACC system into a forward and a backward search. The backward search developed by the authors is based on an accident and is simulated and optimized backwards in time. They conclude that the backward search finds a fault efficiently, even in a highly sophisticated ACC and is therefore superior to common forward-search algorithms.

## VII. SAFETY VERIFICATION

Formal verification as shown in Figure 1 is an alternative to the scenario-based approach presented so far. It requires the traffic rules and generally all specifications, which are just available in prose, to be made available in a formal, machine-readable format. We distinguish three verification methods. In theorem proving, mathematical theorems are usually automatically proven by computer programs. In the reachability analysis, the states that a complex system can reach in the future are calculated. In correct-by-construction synthesis, safe controllers are automatically generated from formal specifications.

### A. FORMALIZATION OF TRAFFIC RULES

Most papers described here address the formalization of traffic rules as one key requirement. A few of them [146]–[148] even focus exclusively on it. Aréchiga [149] defines a set of contracts, with which automated driving can be formally verified, in the formal language Signal Temporal Logic (STL). This means that traffic is guaranteed to be safe if all traffic participants comply with these contracts. The authors highlight that the STL contracts enable multiple formal techniques, like the automatic synthesis of run-time monitors, falsification, formal verification, and parameter synthesis. Compliance with the traffic rules ensures that there are no accidents caused by the AV. This is not to be equated with a macroscopic statement on road safety, because (especially in mixed traffic) new accidents involving human road users may occur due to unexpected behavior of the AV.

### B. THEOREM PROVING

Shalev-Shwartz *et al.* [150] introduce the formal model Responsibility-Sensitive Safety (RSS). They describe safety using numerous axioms and lemmas, and formally verify with worst-case assumptions and mathematical induction that an exemplary planning algorithm satisfies it. They introduce a semantic language including units, measurements, action space and specification as an abstraction of overly detailed maneuver instructions. A Q-learning algorithm serves as the exemplary planner. Outcome is a formal proof stating that no accident can be caused for which the AV is to blame. This model-based approach only targets the planning module

within a typical sense-plan-act architecture. They neglect to address the act module arguing that much research and theory exists from the last decades. For the sense module, they argue that it should be tested with a statistical distance-based approach. Since sensor failures cannot be completely excluded, the probability of such events must fall below a statistically derived threshold. They argue that the threshold can be reduced by several degrees through a triple-redundant design of the sensor system.

Loos *et al.* [151] use a formal proof calculus to verify a distributed car-control system. Multiple cars are equipped with ACC and modeled as distributed hybrid systems which involve both discrete control and the continuous actuation of a cyberphysical system. Their idea is to decompose the verification problem into multiple modular pieces. By proving safety separately for a local and global lane control as well as a local and global highway control, they claim that the distributed car-control system can be certain that every car controller will not cause an accident anywhere at anytime.

Aréchiga *et al.* [152] use the theorem prover KeYmaera to verify the safety of an intelligent cruise controller and a cooperative intersection collision-avoidance system. [153] use the theorem prover Isabelle to verify the safety of a collision-avoidance safety function for autonomous vehicles.

Nilsson *et al.* [154] describe the worst-case performance as an optimization problem, and derive closed-form expressions for it. Thus, they can calculate the set of scenarios where it can be guaranteed that no incorrect decisions are made by the AV.

### C. REACHABILITY ANALYSIS

Reachability analysis aims to determine the states a system can reach from given initial states and possible inputs and parameters. As the exact reachable set cannot be computed for more complex systems, it is typically over-approximated in order to formally guarantee safety. If the reachable set of the AV does not intersect with the predicted occupancy sets of other traffic participants, the AV is safe. In contrast to theorem proving and the scenario-based approach, reachability analysis is mainly performed online during run-time.

Althoff *et al.* [155] introduce reachability analysis arguing safety of AVs. In his dissertation [156], Althoff distinguishes between classical and stochastic reachability analysis. Althoff uses polytopes, zonotopes and multidimensional intervals for over-approximation, as these provide good mathematical characteristics for the calculation of reachable sets. Stochastic reachability analysis calculates with probabilities, but can only provide probabilistic guarantees in contrast to intervals. Althoff further develops the approach in [157], [158]. The dissertation and papers form the basis for further publications of the authors' current research group [159], [160].

O'Kelly *et al.* [161] present their verification tool APEX that internally uses the SMT-solver dReach. They distinguish between the behavioral planner (represented as a formal model in form of a finite transition system) and the

motion planner (represented as a black-box that just provides a trajectory). Their design-time approach can verify the complete trajectory planning and tracking stacks of an AV. They describe unsafe conditions in Metric Interval Temporal Logic and use reachability analysis to guarantee a safe vehicle trajectory. They demonstrate that their approach can identify faulty behavior missed during testing, and how to refine the requirements.

A combination of simulation-based optimization (Section VI-D) prior to a reachability analysis is also possible [162], [163]. They call it a robustness-guided verification technique. As the reachability analysis is computationally expensive and limited to not too complex systems, at first they identify interesting areas via optimization. That improves the efficiency or even enables a reachability analysis of more complex systems.

Tuncali *et al.* [164] use barrier certificates for proof of safety. According to [156], they are similar to Lyapunov functions, but focus on safety instead of stability. The idea is that if the barrier separates initial states from unsafe states, the system safety is proven. The difficulty lies in finding the barrier certificate function. Barrier certificates are also similar to reachability analysis, but calculate the upper bound for reaching an unsafe state, rather than being in one.

### D. CORRECT-BY-CONSTRUCTION SYNTHESIS

Another formal verification approach is the synthesis of correct-by-construction controllers that are automatically generated from formal specifications.

Johnson *et al.* [165] describe the specification in the Linear Temporal Logic language. They represent the AV behavior with a probabilistic model, and formally verify the system properties using the model checker PRISM. They demonstrate their approach with a real, full-scale AV that drives on a road network of a parking lot controlled by the formally synthesized controller. They use the experimental data to validate the formal analysis of the probabilistic model.

Wongpiromsarn *et al.* [166] also use Linear Temporal Logic specifications for the automatic synthesis of a trajectory planner and continuous controller. They use a receding horizon framework to split the synthesis into smaller problems. Nilsson *et al.* [167] synthesize an ACC, guaranteeing safe trajectories of the closed-loop system. The two presented methods are performing the computations either on the continuous state space or on a finite-state abstraction.

## VIII. COMPARISON

In this section, we derive criteria for the evaluation of the different approaches presented in the survey paper. We then compare these approaches by applying the evaluation criteria in order to identify possible research gaps.

### A. EVALUATION CRITERIA

We have derived the following ten evaluation criteria based on expert judgment, which serve to analyze the approaches presented in this survey paper. In order to ensure traceability in

**TABLE 1.** Rating system for the evaluation criteria.

Criteria	Ratings			
	0	1	2	3
Scenario Representativeness	Not applicable	Unrealistic scenarios	Without distributions	With distributions
Parameter Compatibility	Not applicable	Only one parameter type	Multiple parameter types	All parameter types
Corner Case Identification	Not applicable	Cases unintentionally	One case very efficient	All cases
Scenario Space Coverage	Not applicable	Exploitation-driven samples	Exploration-driven samples	Full coverage
Scenario Space Expansion	Not applicable	Extremely difficult expansion	Very difficult expansion	Difficult expansion
System Applicability	Not applicable	Only one component	Multiple components	Overall system
Computational Feasibility	Not applicable	Complex and online	Complex and offline	Efficient calculations
Black-box Compatibility	Not applicable	White-box	Gray-box	Black-box
Statement Reliability	Not applicable	Negative counterexample	Statistical statement	Proof of correctness
Assessment Transferability	Not applicable	Hardly transferable	Coverage statement	Statistical statement

this evaluation process, we introduce a multi-level evaluation system in Table 1, which contains descriptions for each rating which are as generic as possible. We do not evaluate at the level of individual papers, but rather categories of approaches. We select the rating that best represents the entire category to the best of our knowledge. Individual papers may differ from this. If a criterion is not applicable, it is graded as 0. This rating should not necessarily be considered in a “negative” light: it is merely that the corresponding approach does not take this criterion into account. The order of the criteria is related to the scenario-based process. In some cases, the criteria influence each other. For example, good scenario coverage also allows for more reliable macroscopic statements.

### 1) SCENARIO REPRESENTATIVENESS

The scenarios should reflect the real world as closely as possible. In the best case, parameter distributions are determined and used to consider the probabilities of occurrence in the real world. At the very least, the scenarios must comply with the laws of physics.

### 2) PARAMETER COMPATIBILITY

Different scenario parameter types require different representations. Most parameters are real, continuous, time- and/or location-dependent values (e. g. velocities, road friction, etc.) that might – for the sake of simplicity – be assumed to be constant during the scenario-generation and selection process. On the other hand, there are parameters like traffic lights that have only discrete states and are therefore easier to include in the safety-assessment process.

### 3) CORNER CASE IDENTIFICATION

Efficiency in identifying corner cases is especially crucial for system developers and for spot-checking by testing organizations. They can often be found in the tail of parameter distributions. If a failure occurs during safety assessment, it can be used to both improve the system and provide testing organizations with a quick insight into the system’s performance capabilities.

### 4) SCENARIO SPACE COVERAGE

Since infinite situations occur in the real world, the scenario methods must cover a large number of permutations within the physically possible parameter space. To ensure sufficient coverage, concrete scenarios should be sampled within the entire space. In the best case a formal coverage can even be achieved without sampling.

### 5) SCENARIO SPACE EXPANSION

Currently, many papers validate their proposed methodology with a simple proof of concept to ensure quick traceability for the reader. For industrialization, however, safety assessment approaches must scale to a full range of logical scenarios of the ODD, taking into account all layers of the five-layer model [21], the correct representation of all parameter types, and generally more complex scenarios.

### 6) SYSTEM APPLICABILITY

Many approaches focus merely on the planning module, and only a few on perception. Ultimately, the safety of the overall system must be assessed. Therefore, the methods should be extended to cover the overall system with all its sub-modules.

### 7) COMPUTATIONAL FEASIBILITY

The methods differ in their computational complexity. For offline methods that assess safety during design, it should be feasible to execute them within a reasonable time frame. For online methods that are executed during vehicle runtime, efficient calculations are a decisive factor for real-time capability.

### 8) BLACK-BOX COMPATIBILITY

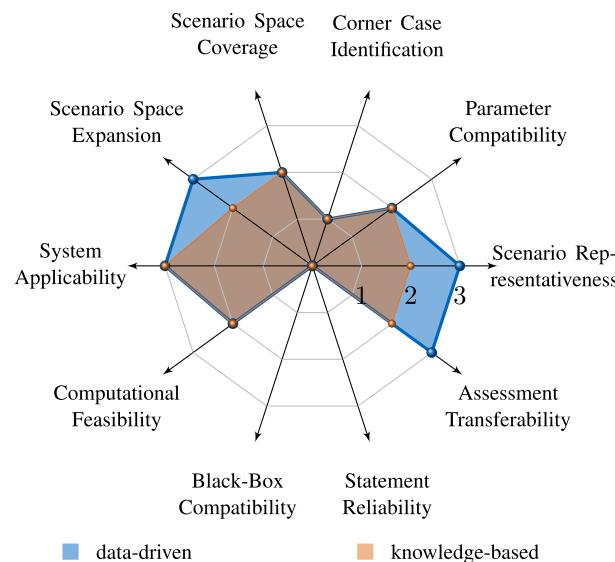
Some methods require white-box models to calculate a gradient or to apply special verification techniques. Black-box approaches are more flexible and respect the intellectual property provided by suppliers and simulation-tool manufacturers.

## 9) STATEMENT RELIABILITY

For a credible decision on whether AVs are safe enough to be launched on the market, the reliability of a statement is crucial. At best, the safety of AVs can be formally guaranteed. If this is not possible, a statistical statement should be sought as a minimum.

## 10) ASSESSMENT TRANSFERABILITY

To ensure the responsible introduction of AVs into traffic, accident rates should first be compared with societal expectations. Since the scenario-based approach starts with a microscopic assessment, the transferability of its results is crucial for a macroscopic statement. The probabilities of occurrence of individual scenarios enable a genuine statistical assessment of overall accident rates.

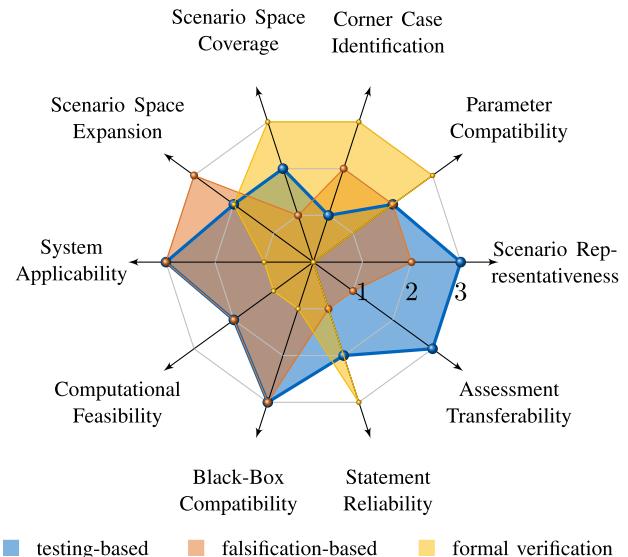


**FIGURE 7.** Evaluation and comparison of the data-driven and knowledge-based approach for scenario generation/extraction.

## B. COMPARISON OF CATEGORIZED APPROACHES

We now analyze and compare the methods presented in sections III to VII against the evaluation criteria derived in the previous section. We compare the knowledge-based approach with the data-driven approach in the Kiviat diagram in Figure 7, as the purpose of both is to fill the database. In addition, we compare testing-based scenario selection, falsification-based scenario selection and formal verification in Figure 8. As shown in Figure 1 we again would like to point out that the formal-verification approach is an alternative to the scenario-based approach (Figure 2) including the testing- and falsification-based scenario selection. This should be taken into account when viewing the table. Nevertheless, it makes sense to include formal verification, since the objective of all approaches is to assess safety.

The data-driven approach covers slightly more area in the Kiviat diagram than the knowledge-based approach. The advantage of the knowledge-based approach is that an initial



**FIGURE 8.** Evaluation and comparison of testing-based and falsification-based scenario selection and formal verification for the safety assessment of AVs.

catalog of scenarios can be quickly created based on existing knowledge such as standards and guidelines. However, it is difficult to extend the catalog to completeness. On the other hand, the data-driven approach requires many prerequisites such as fleet vehicles equipped with additional high performance measurement systems and data processing pipelines. If the prerequisites are met, it is characterized by representative real-world scenarios, high coverage and the derivation of occurrence probabilities.

The testing-based scenario selection covers slightly more area in the Kiviat diagram than the falsification-based scenario selection and the formal verification. The latter currently lacks scalability to complex systems and is computationally expensive, but may provide guaranteed statements without the need for testing. Falsification is a helpful tool for the developer to efficiently identify weaknesses in his system. However, falsification lacks coverage and the ability to make macroscopic safety statements. Testing can provide these macroscopic statements, which are important for comparison against human drivers, but requires an enormous amount of tests.

## C. IDENTIFICATION OF RESEARCH GAPS

In this section, a summary of the research gaps and future challenges are given, based on the evaluation results from the previous section (Figure 7 and 8).

### 1) COMBINATION OF APPROACHES TO COMPENSATE FOR DRAWBACKS

Currently there exists no approach that stands out with regard to all evaluation criteria.

A combination of approaches seems necessary to compensate for the disadvantages of the individual approaches, both inside as well as outside of the scenario-based approach.

For example, in terms of scenario generation, the lack of parameter distribution in the knowledge-based approach can be compensated by extracting parameter distributions from the data-driven approach. Li *et al.* [168] suggest a combination of scenario-based and functionality-based testing as improvement. Since formal verification techniques make the most reliable statements in Table 8, but are usually only applicable to the planning module, they should be applied whenever possible, but combined with the scenario-based approach at a system level.

## 2) CONSISTENT TERMINOLOGY AS A BASIS

There is a lack of common terminology, and certain terms are used in various publications with differing meanings.

The terms 'Scene', 'Situation' and 'Scenario' as defined in [19] as well as the 'five-layer model' as defined in [21] seem to be promising in terms of standardization. The terms 'critical', 'complex' and 'challenging' will be taken here as examples of terms with different meanings in literature. In our opinion, 'complex' and 'challenging' can be used as synonyms, but they can be clearly distinguished from 'critical'. The delimitation is based on the intended evaluation. If the performance of an AV is evaluated, we speak of a critical scenario. If the scenario itself is evaluated, it can be a complex or challenging scenario. The following definitions can be used as a basis for further discussion:

*Critical*: Describes an assessment of the performance of the ego-vehicle behavior in a concrete scenario. The criticality is only determinable after test-case execution, and the behavior of different AV functions lead to different criticality-results for the same concrete scenario.

*Challenging or Complex*: Describes an assessment of a concrete scenario itself. Determinable before test-case execution and independent of the AV performance. Whether a concrete scenario is challenging / complex depends on the chosen parameter values. Therefore, the difficulty for the AV to master the concrete scenario without the occurrence of a critical situation can be seen as challenging or complex.

## 3) DATABASE FOR BENCHMARKING

For proof of concept, most papers use completely different scenarios, parameters, ODD and simulation models, such that a detailed comparison of approached is rendered almost impossible.

The Commonroad Framework [44], which includes not only scenarios but also models and cost functions, seems to be a good starting point for benchmarking. Standards like OpenDrive and OpenScenario can be used as common interface formats. For industrialization, it is necessary to transition from simple, exemplary proof of concepts to complex scenarios using a dedicated use case.

## 4) FUNCTIONAL DECOMPOSITION FOR

### SCENARIO REDUCTION

Even for simulation, there are still many scenarios required for a macroscopic safety assessment.

Therefore, further methods to increase efficiency should be considered in the future. A possibility to reduce the number of necessary scenarios is the functional decomposition of the system [169], [170]. One use case of this procedure can comprise the revision of a sub-module, so that only scenarios which require this module must be re-validated. Then there must be approaches for each module. It can even be helpful to have approaches that are tailor-made for special modules. Most verification approaches focus on the planning module. [164] can be seen as a starting point for verifying the perception system with barrier certificates. Even in scenario-based testing, there are not many publications that focus particularly on the evaluation of perception [171], [172]. However, the perception is a very important module of AVs [173].

## 5) EXPOSURE FOR MACROSCOPIC ASSESSMENT

There is currently a big gap between the microscopic assessment of single scenarios and the macroscopic assessment. Without this, no general statement on the introduction of AVs can be made.

The work of [24], [25], [174], [175] can be used as a starting point. Further research is required to transfer the results by means of exposure or traffic-simulation-based techniques from Section II-B.

## 6) MODEL VALIDATION TO ENABLE SIMULATION

The shift from real world tests to simulation is a huge challenge [176]. Even though if there are a few scenario-based methods that could be executed on a proving ground with restrictions, almost all approaches currently use a physical or mathematical model to assess safety. However, almost none address model validation. Without the latter, virtual assessment results have no credibility in terms of their use in decision making. This applies both to simulation in the scenario-based approach and to the formal models of verification. The latter is usually referred to as conformance testing, which checks the conformance between a model and the real system in terms of obtaining formal properties [159].

Our work [45], [177] and the work of [178]–[182] can be used as a starting point for future simulation model validation. Additionally, [183] can be seen as a starting point for simulation tool qualification.

## IX. CONCLUSION

In order for automated vehicles to be launched on the market, an assessment of their safety is essential. Since this is a huge challenge, much research is currently being carried out into new approaches for evaluating safety. We focus on the safety of the intended functionality, and thereby on the scenario-based approach. Based on a newly developed taxonomy for the scenario-based approach, this paper summarizes the most important publications of recent years. Subsequently, the methods are compared with each other and formal verification is integrated as an alternative concept. Based on the comparison of the different approaches,

we propose the use of formal verification techniques for the planning module, and the scenario-based approach at the overall system level to ensure the safety of automated vehicles. Within the scenario-based approach, all of the sub-methods investigated have the potential to contribute to the safety assessment of automated vehicles by identifying the most relevant scenarios. So far, however, all methods are purely exemplary, i. e. of limited scope and low complexity. They all, without exception, fall short of proof of industrialization. Therefore, the biggest challenge for the future is to implement the methods in a scope and level of detail that allows to obtain a reliable safety statement.

## ACKNOWLEDGMENT

The authors would like to thank Christian Gnandt, Christoph Miethaner, and Benjamin Koller for proofreading the article and for enhancing the content through their critical remarks. (*Stefan Riedmaier and Thomas Ponn contributed equally to this work.*)

## CONTRIBUTION

Stefan Riedmaier and Thomas Ponn (corresponding author) initiated and wrote this paper. They were involved in all stages of development, and primarily developed the concept as well as the whole content of this work. Dieter Ludwig contributed to the structure of the paper and improved the content. In addition, Dieter Ludwig and Bernhard Schick contributed in comprehensive discussions, thanks to their vast experience in automotive safety assessment. Frank Diermeyer contributed to the conception of the research project and revised the paper critically for important intellectual content. He gave final approval of the version to be published and agrees to all aspects of the work. As a guarantor, he accepts responsibility for the overall integrity of the paper.

## REFERENCES

- [1] World Health Organization. (2013). *Number of Road Traffic Deaths*. [Online]. Available: [https://www.who.int/gho/road\\_safety/mortality/traffic\\_deaths\\_number/en/](https://www.who.int/gho/road_safety/mortality/traffic_deaths_number/en/)
- [2] SAE International, “Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles,” SAE Tech. Paper J3016, 2018.
- [3] P. Koopman and M. Wagner, “Challenges in autonomous vehicle testing and validation,” *SAE Int. J. Transp. Saf.*, vol. 4, no. 1, pp. 15–24, 2016.
- [4] P. Koopman and M. Wagner, “Autonomous vehicle safety: An interdisciplinary challenge,” *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 90–96, Spring 2017.
- [5] P. Koopman and M. Wagner, “Toward a framework for highly automated vehicle safety validation,” in *Proc. WCX World Congr. Exper.* 2018, pp. 1–13, doi: [10.4271/2018-01-1071](https://doi.org/10.4271/2018-01-1071).
- [6] I. O. F. Standardization, “ISO/PAS 21448: Road vehicles — Safety of the intended functionality,” 2019.
- [7] ISO 26262: *Road Vehicles—Functional Safety*, International Organization for Standardization, Geneva, Switzerland, 2018.
- [8] NHTSA. (2016). *Federal Automated Vehicles Policy: Accelerating the Next Revolution In Roadway Safety: Sep. 2016*. [Online]. Available: <https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20g%20uidance20PDF.pdf>
- [9] (2018). *A Framework for Automated Driving System Testable Cases and Scenarios*. [Online]. Available: [https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automat-eddrivingsystems\\_092618\\_v1a\\_tag.pdf](https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automat-eddrivingsystems_092618_v1a_tag.pdf)
- [10] J. E. Stellet, M. R. Zofka, J. Schumacher, T. Schamm, F. Niewels, and J. M. Zollner, “Testing of advanced driver assistance towards automated driving: A survey and taxonomy on existing approaches and open questions,” in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 1455–1462.
- [11] W. Huang, K. Wang, Y. Lv, and F. Zhu, “Autonomous vehicles testing methods review,” in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 163–168.
- [12] P. Junietz, W. Wachenfeld, K. Klonecki, and H. Winner, “Evaluation of different approaches to address safety validation of automated driving,” in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 491–496.
- [13] Y. Page, F. Fahrenkrog, A. Fiorentino, J. Gwehenberger, T. Helmer, M. Lindman, O. Op den Camp, L. van Rooij, S. Puch, and M. Franzle, “A comprehensive and harmonized method for assessing the effectiveness of advanced driver assistance systems by virtual simulation: The PEARS initiative,” in *Proc. 24th Int. Tech. Conf. Enhanced Saf. Vehicles (ESV)*, 2015, pp. 1–12.
- [14] P.E.A.R.S. (2020). *Prospective Effectiveness Assessment for Road Safety*. [Online]. Available: <https://pearsinitiative.com/>
- [15] S. E. Shladover and C. Nowakowski, “Regulatory challenges for road vehicle automation: Lessons from the California experience,” *Transp. Res. A, Policy Pract.*, vol. 122, pp. 125–133, Apr. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0965856417300952>
- [16] A. Takacs, D. A. Drexler, P. Galambos, I. J. Rudas, and T. Haidegger, “Assessment and standardization of autonomous vehicles,” in *Proc. IEEE 22nd Int. Conf. Intell. Eng. Syst. (INES)*, Jun. 2018, p. 000.
- [17] M. Wood, P. Robbel, M. Maass, R. D. Tebbens, M. Meijis, M. Harb, J. Reach, K. Robinson, D. Wittmann, T. Srivastava, and M. E. Bouzouraa, “Safety first for automated driving,” Aptiv Services US, LLC, AUDI AG, Bayrische Motoren Werke AG, Beijing Baidu Netcom Sci. Technol., Continental Teves AG Co oHG, Daimler AG, FCA US LLC, HERE Global B.V., Infineon Technol. AG, Intel, Volkswagen AG, White Paper, 2019.
- [18] UL 4600: *Standard for Safety for the Evaluation of Autonomous Products*, UL, Northbrook, IL, USA, 2019.
- [19] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, “Defining and substantiating the terms scene, situation, and scenario for automated driving,” in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 982–988.
- [20] T. Menzel, G. Bagschik, and M. Maurer, “Scenarios for development, test and validation of automated vehicles,” in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1821–1827.
- [21] G. Bagschik, T. Menzel, and M. Maurer, “Ontology based scene creation for the development of automated vehicles,” in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1813–1820.
- [22] J. Sauerbier, J. Bock, H. Weber, and L. Eckstein, “Definition of scenarios for safety validation of automated driving functions,” *ATZ worldwide*, vol. 121, no. 1, pp. 42–45, Jan. 2019.
- [23] German Aerospace Center. (2019). *PEGASUS-Project*. [Online]. Available: <https://www.pegasusprojekt.de/en/home>
- [24] P. Junietz, U. Steininger, and H. Winner, “Macroscopic safety requirements for highly automated driving,” *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 3, pp. 1–10, Mar. 2019.
- [25] P. M. Junietz, “Microscopic and macroscopic risk metrics for the safety validation of automated driving,” Ph.D. dissertation, Tech. Univ. Darmstadt, Darmstadt, Germany, 2019.
- [26] J. Kapinski, J. V. Deshmukh, X. Jin, H. Ito, and K. Butts, “Simulation-based approaches for verification of embedded control systems: An overview of traditional and advanced modeling, testing, and verification techniques,” *IEEE Control Syst.*, vol. 36, no. 6, pp. 45–64, Dec. 2016.
- [27] J. Antonia-Makoshi, N. Uchida, K. Yamaaki, K. Ozawa, E. Kitahara, and S. Taniguchi, “Development of a safety assurance process for autonomous vehicles in Japan,” in *26th Int. Tech. Conf. Exhib. Enhanced Saf. Vehicles (ESV)*. Washington, DC, USA: NHTSA, 2019.
- [28] N. Kalra and S. M. Paddock, “Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?” *Transp. Res. A, Policy Pract.*, vol. 94, pp. 182–193, Dec. 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0965856416302129>
- [29] W. Wachenfeld and H. Winner, “The release of autonomous vehicles,” in *Autonomous Driving: Tech., Legal Social Aspects*, M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, Eds. Berlin, Heidelberg: Springer, 2016, pp. 425–449, doi: [10.1007/978-3-662-48847-8\\_21](https://doi.org/10.1007/978-3-662-48847-8_21).
- [30] C. Wang and H. Winner, “Overcoming challenges of validation automated driving and identification of critical scenarios,” in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 2639–2644.

- [31] S. Kitajima, K. Shimono, J. Tajima, J. Antona-Makoshi, and N. Uchida, "Multi-agent traffic simulations to estimate the impact of automated technologies on safety," *Traffic Injury Prevention*, vol. 20, no. 1, pp. S58–S64, Jun. 2019.
- [32] C. Roesener, M. Harth, H. Weber, J. Josten, and L. Eckstein, "Modelling human driver performance for safety assessment of road vehicle automation," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 735–741.
- [33] M. Saraoglu, A. Morozov, and K. Janschek, "MOBATSim: MOdel-based autonomous traffic simulation framework for Fault-Error-Failure chain analysis," *IFAC-PapersOnLine*, vol. 52, no. 8, pp. 239–244, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896319304100>
- [34] *Richtlinien für die Anlage von Autobahnen*, Forschungsgesellschaft für Straßen- und Verkehrswesen, Cologne, Germany, 2008.
- [35] J. Guo, U. Kurup, and M. Shah, "Is it safe to drive? An overview of factors, metrics, and datasets for driveability assessment in autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–17, 2019.
- [36] Y. Kang, H. Yin, and C. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 2, pp. 171–185, Jun. 2019.
- [37] J. Zhu, W. Wang, and D. Zhao, "A tempt to unify heterogeneous driving databases using traffic primitives," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2052–2057.
- [38] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2118–2125.
- [39] A. Pütz, A. Zlocki, J. Bock, and L. Eckstein, "System validation of highly automated vehicles with a database of relevant traffic scenarios," in *Proc. 12th ITS Eur. Congr.*, Strasbourg, France, 2017, pp. 1–8.
- [40] A. Pütz, A. Zlocki, J. Küfen, J. Bock, and L. Eckstein, "Database approach for the sign-off process of highly automated vehicles," in *Proc. 25th Int. Tech. Conf. Enhanced Saf. Vehicles (ESV) Nat. Highway Traffic Saf. Admin.*, 2017, pp. 1–8.
- [41] S. Feng, Y. Feng, C. Yu, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles, part I: Methodology," 2019, *arXiv:1905.03419*. [Online]. Available: <http://arxiv.org/abs/1905.03419>
- [42] S. Feng, Y. Feng, H. Sun, S. Bao, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles, part II: Case studies," 2019, *arXiv:1905.03428*. [Online]. Available: <http://arxiv.org/abs/1905.03428>
- [43] D. Zhao, Y. Guo, and Y. J. Jia, "TrafficNet: An open naturalistic driving scenario library," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–8.
- [44] M. Althoff, M. Koschi, and S. Manzinger, "CommonRoad: Composable benchmarks for motion planning on roads," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 719–726.
- [45] S. Riedmaier, J. Nesenson, C. Gutenkunst, T. Düser, B. Schick, and H. Abdellatif, "Validation of X-in-the-loop approaches for virtual homologation of automated driving functions," in *Proc. 11th Graz Symp. Virtual Vehicle (GSVF)*, 2018, pp. 1–12.
- [46] J. C. Kirchhof, E. Kusmenko, B. Rumpe, and H. Zhang, "Simulation as a service for cooperative vehicles," in *Proc. ACM/IEEE 22nd Int. Conf. Model Driven Eng. Lang. Syst. Companion (MODELS-C)*, Sep. 2019, pp. 28–37.
- [47] S. Hallerbach, Y. Xia, U. Eberle, and F. Koester, "Simulation-based identification of critical scenarios for cooperative and automated vehicles," *SAE Int. J. Connected Automated Vehicles*, vol. 1, no. 2, pp. 93–106, 2018.
- [48] J. C. Hayward, "Near miss determination through use of a scale of danger," *Highway Res. Rec.*, vol. 384, no. 384, p. 24, 1972.
- [49] W. Wachenfeld, P. Junietz, R. Wenzel, and H. Winner, "The worst-time-to-collision metric for situation identification," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 729–734.
- [50] P. Junietz, F. Bonakdar, B. Klammann, and H. Winner, "Criticality metric for the safety validation of automated driving using model predictive trajectory optimization," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 60–65.
- [51] M. Schreier, V. Willert, and J. Adamy, "An integrated approach to maneuver-based trajectory prediction and criticality assessment in arbitrary road environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2751–2766, Oct. 2016.
- [52] S. M. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS Res.*, vol. 41, no. 4, pp. 153–163, Dec. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0386111217300286>
- [53] S. Geyer, M. Kienle, B. Franz, H. Winner, K. Bengler, M. Baltzer, F. Flemisch, M. Kauer, T. Weißgerber, R. Bruder, S. Hakuli, and S. Meier, "Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance," *IET Intell. Transp. Syst.*, vol. 8, no. 3, pp. 183–189, May 2014.
- [54] W. Chen and L. Kloul, "An ontology-based approach to generate the advanced driver assistance use cases of highway traffic," in *Proc. 10th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage.*, 2018, pp. 1–11. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01835139>
- [55] F. Klueck, Y. Li, M. Nica, J. Tao, and F. Wotawa, "Using ontologies for test suites generation for automated and autonomous driving functions," in *Proc. IEEE Int. Symp. Softw. Rel. Eng. Workshops (ISSREW)*, Oct. 2018, pp. 118–123.
- [56] F. Wotawa and Y. Li, "From ontologies to input models for combinatorial testing," in *Testing Software and Systems*, I. Medina-Bulo, M. G. Merayo, and R. Hierons, Eds. Cham, Switzerland: Springer, 2018, pp. 155–170.
- [57] Y. Li, J. Tao, and F. Wotawa, "Ontology-based test generation for automated and autonomous driving functions," *Inf. Softw. Technol.*, vol. 117, Jan. 2020, Art. no. 106200. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950584918302271>
- [58] J. Langner, J. Bach, L. Ries, S. Otten, M. Holzapfel, and E. Sax, "Estimating the uniqueness of test scenarios derived from recorded Real-World-Driving-Data using autoencoders," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1860–1866.
- [59] J.-A. Bolte, A. Bar, D. Lipinski, and T. Fingscheidt, "Towards corner case detection for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 438–445.
- [60] R. Krajewski, T. Moers, D. Nerger, and L. Eckstein, "Data-driven maneuver modeling using generative adversarial networks and variational autoencoders for safety validation of highly automated vehicles," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2383–2390.
- [61] I. R. Jenkins, L. O. Gee, A. Knauss, H. Yin, and J. Schroeder, "Accident scenario generation with recurrent neural networks," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3340–3345.
- [62] S. Jesenski, J. E. Stellet, F. Schiegg, and J. M. Zollner, "Generation of scenes in intersections for the validation of highly automated driving functions," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 502–509.
- [63] F. Kruber, J. Wurst, and M. Botsch, "An unsupervised random forest clustering technique for automatic traffic scenario categorization," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2811–2818.
- [64] F. Kruber, J. Wurst, E. S. Morales, S. Chakraborty, and M. Botsch, "Unsupervised and supervised learning with the random forest algorithm for traffic scenario clustering and classification," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 2463–2470.
- [65] H. Watanabe, T. Malý, J. Wallner, T. Dirndorfer, M. Mai, and G. Prokop, "Methodology of scenario clustering for predictive safety functions," in *9. Tagung Automatisiertes Fahren*. Munich, Germany: Technical Univ. of Munich, 2019.
- [66] W. Wang and D. Zhao, "Extracting traffic primitives directly from naturally logged data for self-driving applications," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 1223–1229, Apr. 2018.
- [67] H. Weber, J. Bock, J. Klimke, C. Roesener, J. Hiller, R. Krajewski, A. Zlocki, and L. Eckstein, "A framework for definition of logical scenarios for safety assurance of automated driving," *Traffic Injury Prevention*, vol. 20, no. 1, pp. S65–S70, Jun. 2019.
- [68] A. Erdogan, B. Ugranli, E. Adali, A. Sentas, E. Mungan, E. Kaplan, and A. Leitner, "Real-world maneuver extraction for autonomous vehicle validation: A comparative study," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 267–272.
- [69] H. Beglerovic, T. Schloemich, S. Metzner, and M. Horn, "Deep learning applied to scenario classification for lane-keep-assist systems," *Appl. Sci.*, vol. 8, no. 12, p. 2590, 2018.
- [70] R. Gruner, P. Henzler, G. Hinz, C. Eckstein, and A. Knoll, "Spatiotemporal representation of driving scenarios and classification using neural networks," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1782–1788.

- [71] B. David, G. Lacz, and G. Hunyady, "Highway situation analysis with scenario classification and neural network based risk estimation for autonomous vehicles," in *Proc. IEEE 17th World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2019, pp. 375–380.
- [72] J. Bach, J. Langner, S. Otten, E. Sax, and M. Holzapfel, "Test scenario selection for system-level verification and validation of geolocation-dependent automotive control systems," in *Proc. Int. Conf. Eng., Technol. Innov. (ICE/ITMC)*, Jun. 2017, pp. 203–210.
- [73] L. Hartjen, R. Philipp, F. Schuldt, F. Howar, and B. Friedrich, "Classification of driving maneuvers in urban traffic for parametrization of test scenarios," in *9. Tagung Automatisiertes Fahren*. Munich, Germany: Technical Univ. of Munich, 2019.
- [74] J. Zhou and L. del Re, "Identification of critical cases of ADAS safety by FOT based parameterization of a catalogue," in *Proc. 11th Asian Control Conf. (ASCC)*, Dec. 2017, pp. 453–458.
- [75] M. R. Zofka, F. Kuhnt, R. Kohlhaas, C. Rist, T. Schamm, and J. M. Zöllner, "Data-driven simulation and parametrization of traffic scenarios for the development of advanced driver assistance systems," in *Proc. 18th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2015, pp. 1422–1428.
- [76] E. de Gelder and J.-P. Paardekooper, "Assessment of automated driving systems using real-life scenarios," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 589–594.
- [77] E. de Gelder, J.-P. Paardekooper, O. Op den Camp, and B. De Schutter, "Safety assessment of automated vehicles: How to determine whether we have collected enough field data?" *Traffic Injury Prevention*, vol. 20, no. 1, pp. S162–S170, Jun. 2019.
- [78] T. Ponn, D. Fratzke, C. Gnandt, and M. Lienkamp, "Towards certification of autonomous driving: Systematic test case generation for a comprehensive but economically-feasible assessment of lane keeping assist algorithms," in *Proc. 5th Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2019, pp. 333–342.
- [79] H. Beglerovic, A. Ravi, N. Wikström, H.-M. Koegeler, A. Leitner, and J. Holzinger, "Model-based safety validation of the automated driving function highway pilot," in *Proc. 8th Int. Munich Chassis Symp. (Proceedings)*, P. P. E. Pfeffer, Ed. Wiesbaden, Germany: Springer, 2017, pp. 309–329.
- [80] E. Rocklage, H. Kraft, A. Karatas, and J. Seewig, "Automated scenario generation for regression testing of autonomous vehicles," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 476–483.
- [81] B. Kim, A. Jarandikar, J. Shum, S. Shiraishi, and M. Yamaura, "The SMT-based automatic road network generation in vehicle simulation environment," in *Proc. 13th Int. Conf. Embedded Softw. (EMSOFT)*, Oct. 2016, pp. 1–10.
- [82] B. Kim, T. Masuda, and S. Shiraishi, "Test specification and generation for connected and autonomous vehicle in virtual environments," *ACM Trans. Cyber-Phys. Syst.*, vol. 4, no. 1, pp. 1–26, Jan. 2020.
- [83] I. Majzik, O. Semerath, C. Hajdu, K. Maruszy, Z. Szatmari, Z. Micskei, A. Voros, A. A. Babikian, and D. Varro, "Towards system-level testing with coverage guarantees for autonomous vehicles," in *Proc. ACM/IEEE 22nd Int. Conf. Model Driven Eng. Lang. Syst. (MODELS)*, Sep. 2019, pp. 89–94.
- [84] S. Khastgir, G. Dhadyalla, S. Birrell, S. Redmond, R. Addinall, and P. Jennings, "Test scenario generation for driving simulators using constrained randomization technique," in *Proc. SAE World Congr. Exper. Warrendale, PA, USA*, 2017.
- [85] L. Huang, Q. Xia, F. Xie, H.-L. Xiu, and H. Shu, "Study on the test scenarios of level 2 automated vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 49–54.
- [86] F. Xie, T. Chen, Q. Xia, L. Huang, and H. Shu, "Study on the controlled field test scenarios of automated vehicles," *SAE Tech. Paper* 2018-01-1633, 2018, doi: [10.4271/2018-01-1633](https://doi.org/10.4271/2018-01-1633).
- [87] J. Zhou and L. D. Re, "Reduced complexity safety testing for ADAS & ADF," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5985–5990, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896317317755>
- [88] M. Althoff and J. M. Dolan, "Reachability computation of low-order models for the safety verification of high-order road vehicle models," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2012, pp. 3559–3566.
- [89] C. E. Tuncali and G. Fainekos, "Rapidly-exploring random trees for testing automated vehicles," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019.
- [90] D. Åsljung, J. Nilsson, and J. Fredriksson, "Comparing collision threat measures for verification of autonomous vehicles using extreme value theory," *IFAC-PapersOnLine*, vol. 49, no. 15, pp. 57–62, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896316309855>
- [91] D. Åsljung, J. Nilsson, and J. Fredriksson, "Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles," *IEEE Trans. Intell. Vehicles*, vol. 2, no. 4, pp. 288–297, Dec. 2017.
- [92] D. Zhao, "Accelerated evaluation of automated vehicles," Ph.D. dissertation, Univ. Michigan, Ann Arbor, MI, USA, 2016.
- [93] D. Zhao, H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan, "Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 595–607, Mar. 2017.
- [94] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. LeBlanc, "Accelerated evaluation of automated vehicles in car-following maneuvers," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 733–744, Mar. 2018.
- [95] Z. Huang, D. Zhao, H. Lam, D. J. LeBlanc, and H. Peng, "Evaluation of automated vehicles in the frontal cut-in scenario—An enhanced approach using piecewise mixture models," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May/Jun. 2017, pp. 197–202.
- [96] Z. Huang, H. Lam, and D. Zhao, "Sequential experimentation to efficiently test automated vehicles," in *Proc. Winter Simulation Conf. (WSC)*, Dec. 2017, pp. 3078–3089.
- [97] Z. Huang, H. Lam, and D. Zhao, "Towards affordable on-track testing for autonomous vehicle—A Kriging-based statistical approach," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
- [98] Z. Huang, Y. Guo, M. Arief, H. Lam, and D. Zhao, "A versatile approach to evaluating and testing automated vehicles based on kernel methods," in *Proc. Annu. Amer. Control Conf. (ACC)*, Jun. 2018, pp. 4796–4802.
- [99] Z. Huang, H. Lam, D. J. LeBlanc, and D. Zhao, "Accelerated evaluation of automated vehicles using piecewise mixture models," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 9, pp. 2845–2855, Sep. 2018.
- [100] Z. Huang, H. Lam, and D. Zhao, "An accelerated testing approach for automated vehicles with background traffic described by joint distributions," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 933–938.
- [101] Z. Huang, M. Arief, H. Lam, and D. Zhao, "Synthesis of different autonomous vehicles test approaches," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2000–2005.
- [102] Z. Huang, M. Arief, H. Lam, and D. Zhao, "Evaluation uncertainty in data-driven self-driving testing," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1902–1907.
- [103] S. Zhang, H. Peng, D. Zhao, and H. E. Tseng, "Accelerated evaluation of autonomous vehicles in the lane change scenario based on subset simulation technique," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3935–3940.
- [104] M. Arief, P. Glynn, and D. Zhao, "An accelerated approach to safely and efficiently test pre-production autonomous vehicles on public streets," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2006–2011.
- [105] M. O'Kelly, A. Sinha, H. Namkoong, R. Tedrake, and J. C. Duchi, "Scalable end-to-end autonomous vehicle testing via rare-event simulation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9827–9838.
- [106] X. Wang, H. Peng, and D. Zhao, "Combining reachability analysis and importance sampling for accelerated evaluation of highly automated vehicles at pedestrian crossing," in *Proc. Dyn. Syst. Control Conf.*, 2019.
- [107] Y. Akagi, R. Kato, S. Kitajima, J. Antona-Makoshi, and N. Uchida, "A risk-index based sampling method to generate scenarios for the evaluation of automated driving vehicle safety\*," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 667–672.
- [108] S. P. Olivares, N. Rebernig, A. Eichberger, and E. Stadlober, "Virtual stochastic testing of advanced driver assistance systems," in *Advanced Microsystems for Automotive Applications*. Cham, Switzerland: Springer, 2016, pp. 25–35.
- [109] D. Åsljung, M. Westlund, and J. Fredriksson, "A probabilistic framework for collision probability estimation and an analysis of the discretization precision," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 52–57.
- [110] L. Stark, M. Düring, S. Schoenawa, J. E. Maschke, and C. M. Do, "Quantifying vision zero: Crash avoidance in rural and motorway accident scenarios by combination of ACC, AEB, and LKS projected to German accident occurrence," *Traffic Injury Prevention*, vol. 20, no. 1, pp. S126–S132, Jun. 2019.
- [111] L. Stark, S. Obst, S. Schoenawa, and M. Düring, "Towards vision zero: Addressing white spots by accident data based ADAS design and evaluation," in *Proc. IEEE Int. Conf. Veh. Electron. Saf. (ICVES)*, Sep. 2019, pp. 1–6.

- [112] P. Feig, V. Labenski, T. Leonhardt, and J. Schatz, "Assessment of technical requirements for level 3 and beyond automated driving systems based on naturalistic driving and accident data analysis," in *Proc. 26th Int. Tech. Conf. Enhanced Saf. Vehicles (ESV)*, ed. Washington, DC, USA: National Highway Traffic Safety Administration, 2019.
- [113] F. Fahrenkrog, L. Wang, T. Platzer, A. Fries, F. Raisch, and K. Kompaß, "Prospective effectiveness safety assessment of automated driving functions—From the method to the results," in *Proc. 26th Int. Tech. Conf. Enhanced Saf. Vehicles (ESV)*, ed. Washington, DC, USA: National Highway Traffic Safety Administration, 2019.
- [114] J. J. So, I. Park, J. Wee, S. Park, and I. Yun, "Generating traffic safety test scenarios for automated vehicles using a big data technique," *KSCE J. Civil Eng.*, vol. 23, no. 6, pp. 2702–2712, Jun. 2019, doi: [10.1007/s12205-019-1287-4](https://doi.org/10.1007/s12205-019-1287-4).
- [115] A. Pierson, W. Schwarting, S. Karaman, and D. Rus, "Learning risk level set parameters from data sets for safer driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 273–280.
- [116] M. Althoff and S. Lutz, "Automatic generation of safety-critical test scenarios for collision avoidance of road vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1326–1333.
- [117] M. Klischat and M. Althoff, "Generating critical test scenarios for automated vehicles with evolutionary algorithms," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 2352–2358.
- [118] F. Gao, J. Duan, Y. He, and Z. Wang, "A test scenario automatic generation strategy for intelligent driving systems," *Math. Problems Eng.*, vol. 2019, pp. 1–10, Jan. 2019.
- [119] Q. Xia, J. Duan, F. Gao, T. Chen, and C. Yang, "Automatic generation method of test scenario for ADAS based on complexity," *SAE Tech. Paper 2017-01-1992*, 2017, doi: [10.4271/2017-01-1992](https://doi.org/10.4271/2017-01-1992).
- [120] Q. Xia, J. Duan, F. Gao, Q. Hu, and Y. He, "Test scenario design for intelligent driving systems ensuring coverage and effectiveness," *Int. J. Automot. Technol.*, vol. 19, no. 4, pp. 751–758, Aug. 2018.
- [121] J. Wang, C. Zhang, Y. Liu, and Q. Zhang, "Traffic sensory data classification by quantifying scenario complexity," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1543–1548.
- [122] C. Zhang, Y. Liu, Q. Zhang, and L. Wang, "A graded offline evaluation framework for intelligent vehicle's cognitive ability," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 320–325.
- [123] Y. Qi, Y. Luo, K. Li, W. Kong, and Y. Wang, "A trajectory-based method for scenario analysis and test effort reduction for highly automated vehicle," in *Proc. SAE Tech. Paper Series*, Warrendale, PA, USA, 2019.
- [124] T. Ponn, C. Gnandt, and F. Diermeyer, "An optimization-based method to identify relevant scenarios for type approval of automated vehicles," in *Proc. 26th Int. Tech. Conf. Enhanced Saf. Vehicles (ESV)*, Washington, DC, USA: National Highway Traffic Safety Administration, 2019.
- [125] T. Ponn, A. Schwab, F. Diermeyer, C. Gnandt, and J. Záhorský, "A method for the selection of challenging driving scenarios for automated vehicles based on an objective characterization of the driving behavior," in *9. Tagung Automatisiertes Fahren*. Munich, Germany: Technical Univ. of Munich, 2019.
- [126] P. Koopman and F. Fratrik, "How many operational design domains, objects, and events?" in *Proc. SafeAI*, 2019, pp. 1–4.
- [127] M. Koren, S. Alsaif, R. Lee, and M. J. Kochenderfer, "Adaptive stress testing for autonomous vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1898–1904.
- [128] A. Corso, P. Du, K. Driggs-Campbell, and M. J. Kochenderfer, "Adaptive stress testing with reward augmentation for autonomous vehicle validation," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 163–168.
- [129] R. Lee, M. Kochenderfer, O. Mengshoel, G. Brat, and M. Owen, "Adaptive stress testing of airborne collision avoidance systems," in *Proc. IEEE/AIAA 34th Digit. Avionics Syst. Conf. (DASC)*, Sep. 2015, pp. 1–24.
- [130] R. Lee, O. Mengshoel, A. Saksena, R. Gardner, D. Genin, J. Brush, and M. J. Kochenderfer, "Differential adaptive stress testing of airborne collision avoidance systems," in *Proc. AIAA Model. Simul. Technol. Conf.*, Jan. 2018, pp. 1–11.
- [131] H. Beglerovic, M. Stoltz, and M. Horn, "Testing of autonomous vehicles using surrogate models and stochastic optimization," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
- [132] R. B. Abdessalem, S. Nejati, L. C. Briand, and T. Stifter, "Testing advanced driver assistance systems using multi-objective search and neural networks," in *Proc. 31st IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE)*, D. Lo, S. Apel, and S. Khurshid, Eds. New York, NY, USA: ACM Press, 2016, pp. 63–74.
- [133] G. E. Mullins, "Adaptive sampling methods for testing autonomous systems," Ph.D. dissertation, Univ. Maryland, College Park, MD, USA, 2018.
- [134] G. E. Mullins, P. G. Stankiewicz, and S. K. Gupta, "Automated generation of diverse and challenging scenarios for test and evaluation of autonomous vehicles," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 1443–1450.
- [135] G. E. Mullins, P. G. Stankiewicz, R. C. Hawthorne, and S. K. Gupta, "Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles," *J. Syst. Softw.*, vol. 137, pp. 197–215, Mar. 2018.
- [136] B. Gangopadhyay, S. Khastgir, S. Dey, P. Dasgupta, G. Montana, and P. Jennings, "Identification of test cases for automated driving systems using Bayesian optimization," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1961–1967.
- [137] M. Nabhan, M. Schoenauer, Y. Tourbier, and H. Hage, "Optimizing coverage of simulated driving scenarios for the autonomous vehicle," in *Proc. IEEE Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2019.
- [138] H. Abbas, M. O'Kelly, A. Rodionova, and R. Mangharam, "Safe at any speed: A simulation-based test harness for autonomous vehicles," in *Proc. 7th Int. Workshop Design, Modeling, Eval. Cyber Phys. Syst. (CyPhy)*, 2017, pp. 94–106.
- [139] C. E. Tuncali, T. P. Pavlic, and G. Fainekos, "Utilizing S-TaLiRo as an automatic test generation framework for autonomous vehicles," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1470–1475.
- [140] C. E. Tuncali, S. Yaghoubi, T. P. Pavlic, and G. Fainekos, "Functional gradient descent optimization for automatic test case generation for vehicle controllers," in *Proc. 13th IEEE Conf. Autom. Sci. Eng. (CASE)*, Aug. 2017, pp. 1059–1064.
- [141] C. E. Tuncali, G. Fainekos, H. Ito, and J. Kapinski, "Simulation-based adversarial test generation for autonomous vehicles with machine learning components," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018.
- [142] C. E. Tuncali, G. Fainekos, D. Prokhorov, H. Ito, and J. Kapinski, "Requirements-driven test generation for autonomous vehicles with machine learning components," *IEEE Trans. Intell. Vehicles*, early access, Nov. 25, 2019, doi: [10.1109/TIV.2019.2955903](https://doi.org/10.1109/TIV.2019.2955903).
- [143] C. E. Tuncali, "Search-based test generation for automated driving systems: From perception to control logic," Ph.D. dissertation, Arizona State Univ., Tempe, AZ, USA, 2019.
- [144] H. Felbinger, F. Kluck, Y. Li, M. Nica, J. Tao, F. Wotawa, and M. Zimmermann, "Comparing two systematic approaches for testing automated driving functions," in *Proc. IEEE Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2019, pp. 1–6.
- [145] M. Koschi, C. Pek, S. Maierhofer, and M. Althoff, "Computationally efficient safety falsification of adaptive cruise control systems," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 2879–2886.
- [146] A. Rizaldi and M. Althoff, "Formalising traffic rules for accountability of autonomous vehicles," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 1658–1665.
- [147] C. Pek, P. Zahn, and M. Althoff, "Verifying the safety of lane change maneuvers of self-driving vehicles based on formalized traffic rules," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1477–1483.
- [148] B. Vanholme, D. Gruyer, B. Lusetti, S. Glaser, and S. Mammar, "Highly automated driving on highways based on legal safety," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 333–347, Mar. 2013.
- [149] N. Arechiga, "Specifying safety of autonomous vehicles in signal temporal logic," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 58–63.
- [150] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," 2017, *arXiv:1708.06374*. [Online]. Available: <http://arxiv.org/abs/1708.06374>
- [151] S. M. Loos, A. Platzer, and L. Nistor, "Adaptive cruise control: Hybrid, distributed, and now formally verified," in *Formal Methods*, M. Butler and W. Schulte, Eds. Berlin, Germany: Springer, 2011, pp. 42–56.
- [152] N. Arechiga, S. M. Loos, A. Platzer, and B. H. Krogh, "Using theorem provers to guarantee closed-loop system properties," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2012, pp. 3573–3580.
- [153] H. Täubig, U. Frese, C. Hertzberg, C. Lüth, S. Mohr, E. Vorobev, and D. Walter, "Guaranteeing functional safety: Design for provability and computer-aided verification," *Auto. Robots*, vol. 32, no. 3, pp. 303–331, Apr. 2012.

- [154] J. Nilsson, A. C. E. Odbom, and J. Fredriksson, "Worst-case analysis of automotive collision avoidance systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1899–1911, Apr. 2016.
- [155] M. Althoff, O. Stursberg, and M. Buss, "Reachability analysis of linear systems with uncertain parameters and inputs," in *Proc. 46th IEEE Conf. Decis. Control*, 2007, pp. 726–732.
- [156] M. Althoff, "Reachability analysis and its application to the safety assessment of autonomous cars," Ph.D. dissertation, Tech. Univ. Munich, Munich, Germany, 2010.
- [157] M. Althoff and J. M. Dolan, "Set-based computation of vehicle behaviors for the online verification of autonomous vehicles," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1162–1167.
- [158] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 903–918, Aug. 2014.
- [159] B. Schurmann, D. Hes, J. Eilbrecht, O. Stursberg, F. Koster, and M. Althoff, "Ensuring drivability of planned motions using formal methods," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–8.
- [160] F. Gruber and M. Althoff, "Anytime safety verification of autonomous vehicles," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1708–1714.
- [161] M. E. O'Kelly, H. Abbas, S. Gao, S. Kato, S. Shiraishi, and R. Mangharam, "APEX: Autonomous vehicle plan verification and execution," in *Proc. SAE World Congr. Exhib.*, Warrendale, PA, USA, 2016, pp. 1–14.
- [162] H. Abbas, M. E. O'Kelly, A. Rodionova, and R. Mangharam, "A driver's license test for driverless vehicles," *Mech. Eng.*, vol. 139, no. 12, pp. S13–S16, 2017.
- [163] M. O'Kelly, H. Abbas, and R. Mangharam, "Computer-aided design for safe autonomous vehicles," in *Proc. Resilience Week (RWS)*, Sep. 2017, pp. 90–96.
- [164] C. E. Tuncali, J. Kapinski, H. Ito, and J. V. Deshmukh, "Reasoning about safety of learning-enabled components in autonomous cyber-physical systems," in *Proc. 55th Annu. Design Autom. Conf.*, Jun. 2018, pp. 1–6.
- [165] B. Johnson, F. Havlak, H. Kress-Gazit, and M. Campbell, "Experimental evaluation and formal analysis of high-level tasks with dynamic obstacle anticipation on a full-sized autonomous vehicle," *J. Field Robot.*, vol. 34, no. 5, pp. 897–911, Aug. 2017.
- [166] T. Wongpiromsarn, U. Topcu, and R. M. Murray, "Receding horizon temporal logic planning," *IEEE Trans. Autom. Control*, vol. 57, no. 11, pp. 2817–2830, Nov. 2012.
- [167] P. Nilsson, O. Hussien, A. Balkan, Y. Chen, A. D. Ames, J. W. Grizzle, N. Ozay, H. Peng, and P. Tabuada, "Correct-by-Construction adaptive cruise control: Two approaches," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 4, pp. 1294–1307, Jul. 2016.
- [168] L. Li, W.-L. Huang, Y. Liu, N.-N. Zheng, and F.-Y. Wang, "Intelligence testing for autonomous vehicles: A new approach," *IEEE Trans. Intell. Vehicles*, vol. 1, no. 2, pp. 158–166, Jun. 2016.
- [169] C. Amersbach and H. Winner, "Functional decomposition: An approach to reduce the approval effort for highly automated driving," in *8. Tagung Fahrerassistenz*. Munich, Germany: Technical Univ. of Munich, 2017.
- [170] C. Amersbach and H. Winner, "Functional decomposition—A contribution to overcome the parameter space explosion during validation of highly automated driving," *Traffic Injury Prevention*, vol. 20, no. 1, pp. S52–S57, Jun. 2019.
- [171] T. Ponn, F. Müller, and F. Diermeyer, "Systematic analysis of the sensor coverage of automated vehicles using phenomenological sensor Models\*," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 1000–1006.
- [172] H. Martin, B. Winkler, S. Grubmüller, and D. Watzenig, "Identification of performance limitations of sensing technologies for automated driving," in *Proc. IEEE Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2019, pp. 1–6.
- [173] M. Berk, O. Schubert, H.-M. Kroll, B. Buschardt, and D. Straub, "Reliability assessment of safety-critical sensor information: Does one need a reference truth?" *IEEE Trans. Rel.*, vol. 68, no. 4, pp. 1227–1241, Dec. 2019.
- [174] C. Amersbach and H. Winner, "Defining required and feasible test coverage for scenario-based validation of highly automated Vehicles\*," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 425–430.
- [175] F. Hauer, T. Schmidt, B. Holzmüller, and A. Pretschner, "Did we test all scenarios for automated and autonomous driving systems?" in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 2950–2955.
- [176] A. Knauss, J. Schroder, C. Berger, and H. Eriksson, "Paving the roadway for safety of automated vehicles: An empirical study on testing challenges," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1873–1880.
- [177] B. Danquah, S. Riedmaier, J. Ruhm, S. Kalt, and M. Lienkamp, "Statistical model verification and validation concept in automotive vehicle design," in *Proc. 30th CIRP Design*, 2020.
- [178] K. Groh, T. Kuehbeck, B. Fleischmann, M. Schiemetz, and C. C. Chibelushi, "Towards a scenario-based assessment method for highly automated driving functions," in *8. Tagung Fahrerassistenzsysteme*. Munich, Germany: Technical Univ. of Munich, 2017.
- [179] K. Groh, S. Wagner, T. Kuehbeck, and A. Knoll, "Simulation and its contribution to evaluate highly automated driving functions," in *Proc. SAE Tech. Paper*, Warrendale, PA, USA, Apr. 2019.
- [180] S. Wagner, K. Groh, T. Kuehbeck, and A. Knoll, "Towards cross-verification and use of simulation in the assessment of automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 1589–1596.
- [181] D. Notz, M. Sigl, T. Kuehbeck, S. Wagner, K. Groh, C. Schutz, and D. Watzenig, "Methods for improving the accuracy of the virtual assessment of autonomous driving," in *Proc. IEEE Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2019, pp. 1–6.
- [182] S. Wagner, K. Groh, T. Kuehbeck, M. Dorfel, and A. Knoll, "Using Time-to-React based on naturalistic traffic object behavior for scenario-based risk assessment of automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1521–1528.
- [183] H. Abdellatif and C. Gnandt, "Use of simulation for the homologation of automated driving functions," *ATZelectron. worldwide*, vol. 14, no. 12, pp. 68–71, Dec. 2019.



**STEFAN RIEDMAIER** received the B.Eng. degree in electrical engineering and information technology, and the M.Sc. degree in advanced driver assistance systems from the Kempten University of Applied Sciences, Kempten (Allgäu), Germany, in 2015 and 2016, respectively. He is currently pursuing the Ph.D. degree in mechanical engineering with the Technical University of Munich (TUM), Munich, Germany. From 2017 to 2019, he was a Research Assistant with the Research Center Allgäu, and has been holding the same position at the Institute of Automotive Technology, TUM, since 2019. His research interests address an objective quality assessment of automated driving simulations to enable virtual-based safety assessment and homologation.



**THOMAS PONN** (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in mechanical engineering from the Technical University of Munich (TUM), Munich, Germany, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree in safety validation of automated vehicles. Since the beginning of 2017, he has been a Staff Member of the Institute of Automotive Technology, TUM. His research interests are focused on the scenario-based approach, with a special emphasis on the selection of test scenarios.



**DIETER LUDWIG** was born in Pretoria, South Africa, in 1969. He received the Diploma degree in mechanical engineering from the Trier University of Applied Sciences, Trier, Germany. From 2001 to 2017, he was a Cross-Domain Consultant for functional safety (FuSa) in the automotive industry, specializing in safety analyses and assessment/audit. He is currently responsible for the development of the FuSa area at HAD at TÜV SÜD headquarters in Munich. He is also the Technical Representative of the company in specialist circles, working groups, and committees for topics of safe functionality (ISO/PAS21448) and functional safety (TR4804, GRVA FRAV). His work focuses on the development of new innovative methods and approaches to functional safety for the homologation and approval of highly automated and networked driving functions.



**FRANK DIERMEYER** received the Diploma and Ph.D. degrees in mechanical engineering from the Technical University of Munich (TUM), Munich, Germany, in 2001 and 2008, respectively. Since 2008, he has been a Senior Engineer with the Chair of Automotive Technology, TUM, and has been the Leader of the Automated Driving Research Group. His research interests include teleoperated driving, human-machine interaction, and safety validation.

• • •



**BERNHARD SCHICK** received the degree in mechatronic engineering from the University of Applied Science Heilbronn. From 1994, whilst at TÜV SÜD, he built up his expertise in the field of vehicle dynamics and advanced driver assistance systems, in various positions up to a General Manager. He joined IPG Automotive, in 2007, as the Managing Director, where he worked in the field of vehicle dynamics simulation. Since 2014, at AVL List in Graz, he has been responsible—as the Global Business Unit Manager—for calibration and virtual testing technologies. Since 2016, he has been a Research Professor with the University of Applied Science Kempten, and the Head of the Adrive LivingLab institute. His research focus is automated driving and vehicle dynamics.