Thirty-One Challenges in Testing Automated Vehicles: Interviews with Experts from Industry and Research

Felix Beringhoff[1], Joel Greenyer[2], Christian Roesener[1] and Matthias Tichy[3]

Abstract— **There is consensus across the automotive industry that Automated Driving Systems and automated vehicles challenge the way how quality assurance and, particularly, testing must be performed. However, there is a lack of up-to-date empirical studies that substantiate this concern. We conducted interviews with several experts from industry and research to systematically identify challenges as well as improvement opportunities in methods and tools. We report in this paper on 31 challenges that we identified in the areas of scenario- and simulation-based testing, test automation, and test execution. One recurrent challenge expressed by many experts is the problem how to translate a desired condition to be tested into an executable scenario model. This is not alone a question of scripting the scenario, but also of considering a vehicle under test that might try to evade the desired test condition.**

I. Introduction

Testing methods and tools for Automated Driving Systems (ADS) still need improvement: They must comprehensively and transparently ensure their quality and safety, while at the same time the manual effort to design and execute tests should be minimized. This is difficult to achieve due to the amount of possible traffic situations an automated vehicle (AV) faces during its lifetime.

Testing such complex systems put conventional development and testing practices from the automotive industry to the test [1]. To identify the most important challenges and to focus R&D efforts, we conducted an interview study with experts predominantly from industry, but also from academia. Our aim is to provide answers to the following research question: *Which challenges arise with regard to the test implementation and execution for ADS?* Our focus lies on AV equipped with ADS of SAE level four [2].

We would like to obtain a current picture of challenges in testing AV, since the field of automated driving (AD) evolves quickly and a lot of research has been conducted on the verification and validation (V&V) of self-driving vehicles in the last years. Here, scenario-based testing (SBT) [3], virtual testing and test automation are gaining attention as they are considered necessary to cope with the test effort needed for AV [4]. Therefore, these topics also play a big role in our study.

The results, opinions and conclusions expressed in this publication are not necessarily those of Volkswagen Aktiengesellschaft.

[1]Felix Beringhoff and Christian Roesener are with Volkswagen AG, Wolfsburg, Germany [felix.beringhoff, christian.roesener1]@volkswagen.de

[2]FHDW Hannover, Hannover, 30173, Germany joel.greenyer@fhdw.de

[3]Matthias Tichy is with the Institute for Software Engineering and Programming Languages, Ulm University, Germany matthias.tichy@uni-ulm.de

The first contribution of this paper is the description of 31 challenges (26 of those not yet known from two other empirical interview studies [5], [6]), dealing mainly with the right choice of test scenarios and their description, the behavior of the vehicle under test (VuT) during the test execution, and the preparation of virtual test environments. The second contribution of this paper is an overview of existing approaches which aim at addressing these challenges as a starting point for further research. Unsurprisingly, due to the research effort put into the field of AD, most of our identified challenges are already mentioned in other publications. However, our results show that many challenges in the field of SBT and simulation are still unsolved from a practitioners' point of view.

In the next section, we describe the research methodology in more detail. Section III presents the identified challenges. In Section IV, we compare the results to related work. We conclude and give an outlook on future work in Section V.

II. Methodology

We conducted an interview study with 13 experts from industry and research on the topic of testing AV.

Based on the aforementioned research questions, an interview structure was designed to guide the interviewer during the interviews. The interviews were conducted as semi-structured interviews via web conference tools with a duration of 35-65 minutes each. The interviews were recorded and then transcribed and analyzed by the first author. The transcripts were analyzed regarding challenges of testing AV, which were fitted into groups of challenges that were formed iteratively during the analysis.

Experts were selected based on their experience in the field and previous publications. Thirteen experts, ten from industry and three from research institutes, voluntarily participated. Their median of experience amounts to five years in the area of testing driving functions and seven years in the broader field of AD. Every expert has at least three years of experience in either one of the two categories. Each expert works as test manager, test designer, test conductor, tool developer, researcher, or methodologist in the test process. Some of them hold more than one of these positions or gained experiences in multiple positions in the past.

The threats to validity of the interview study are discussed along the classification scheme from Runeson and Höst [7]: **Construct Validity:** At the beginning of each interview, the interviewer gave an introduction to the interview topic and a definition of terms for the interview. For example it was clarified that all questions refer to all modules of an

ADS of SAE level four, including hardware and software. Additionally the interviewer gave clarification of a question when an interviewee asked for it or it seemed that a question was misunderstood.

Internal Validity: By using an interview guide, the interviewer tried to keep the structure of each of the interviews similar. However, the interviews were conducted in a semi-structured manner which allowed alternations of the order of questions or asking additional questions in order to keep a flow. Due to a change of the question order, order effects on the answers of the experts cannot be excluded. If possible time wise, the interviewees were given the opportunity to talk about topics from their mind at the end of an interview. By doing this, a limitation of topics by the interview guide should be softened to limit the effect of potentially missed topics during the design of the interview guide.

External Validity: To obtain a diverse view on the challenges of testing AV practitioners from different companies and institutes were selected with a maximum of two people coming from the same department. We were able to acquire 13 experts who participated in our study. However due to the amount, we cannot be sure to have gained a holistic view on the challenges of AV testing for the automotive industry as a whole.

Reliability: The execution, transcription and analysis of the interviews were done by the first author only. By conducting all steps by one person, this person knows the interviews inside out, which helped at identifying most of the mentioned challenges. Anyhow, it cannot be guaranteed that all challenges mentioned in the interviews are retrieved. Furthermore, the labeling of statements from the interviews and the categorization of statements is to some extent dependent on subjective judgements of the author which is a threat to the reliability of this study.

III. Results

The following section presents the identified technical challenges of testing AV. The challenges are grouped in the categories scenario-based testing, simulation-based testing, test automation, and test execution. For each challenge, at least one statement from one of the interviews is shown in italics. Since the interviews were conducted in German language, the statements have been translated into English.

A. Scenario-based Testing

During our interviews, most of the mentioned challenges are related to the topic of scenario-based testing (SBT).

CH-A1 terminology: Since the SBT is not established in the automotive testing process as of now, the words used to describe SBT processes are currently used differently. Some of the interviewed experts mentioned that it needs to be checked whether the terminology of SBT fits to the common practice. One expert said: *The word scenario is a commonplace term, everyone uses it to refer to something else. (ID13)* Some practitioners use the word *scenario* to describe simulation scenarios, test cases or some use it to describe sequences of control commands, which hinders a precise communication. But not only the terminology used for talking about SBT needs clarification, also **(CH-A2 consistent scenario description)** standard formats for the scenario descriptions are seen as helpful to avoid misunderstandings, since *people discuss the same thing but describe it differently and vice versa. (ID09)* One example mentioned by an expert refers to the harmonization of a definition for scenario starting and ending points. However the creation of a consistent scenario description format which copes with the complexity of possible driving scenarios is considered challenging.

CH-A3 complexity of scenarios: Speaking of the complexity of the scenarios, one interviewee compared test scenarios for ADS features with those for driver support features. According to [2], driver support features can only support a human driver in a subset of the dynamic driving task. The interviewee concludes that *the complexity of the scenario is usually higher (ID02)* to be suitable for testing each layer, e.g., maneuver planning, of an ADS. Therefore, the scenario description and execution will be more demanding, so the expert.

CH-A4 multifunctional scenario description: Scenarios are not only used during AV testing, moreover *scenarios are needed on all layers of the development process. (ID13)* They might be used for specification purposes as well as for testing purposes. Each of which has different requirements on the scenario description. With currently known manual processes to create layers of individual scenarios, it is hardly feasible to create a traceability along all layers of the development and testing process, according to an interviewed expert.

CH-A5 coupling of scenarios and test cases: For traceability purposes, the different artifacts of scenario and test case need to be linked. For this, unambiguous identifier for each artifact are needed, as stated by one of the experts. However, one interviewed expert states that the challenge is *to find this link between the test case and the scenarios, so to speak, to find out which scenarios actually represent this test case and are relevant for it. (ID11)* Another expert mentioned the relationship between test case and scenario as a question where they are currently stuck. The mentioned issue is that classically test cases, for example for hardware in the loop (HiL) plants, are on a signal level, while scenarios in the context of AD are on a traffic level. This makes it difficult to determine how these two artifacts play together and whether they even have to.

CH-A6 consistent scenario interpretation: A common scenario description format for simulation tools does not guarantee a consistent interpretation of the format by simulation tools. Nevertheless, the transferability of scenarios between projects and simulation tools with guaranteed repeatability of the exact same scenario is desired, as mentioned by some of the interviewed experts. Although there are first releases of standard formats like OpenSCENARIO [8] available, the scenario format *is interpreted differently which can lead to problems. (ID08)*

CH-A7 scenario elicitation: A big talking point in many interviews was the topic of scenario elicitation. This topic deals with questions as raised by one of the experts: *Which scenarios do I test? That's the key question. Where do I get* *them from? (ID05)* According to the interviewees it needs to be determined which are the relevant scenarios and how they can be retrieved. This is considered as a continuous process, because with the release of AV, the relevant scenarios might change. Those new scenarios may only be identified after a release of AV to the public roads and need to find their way back in the continuous testing process. One expert mentioned that fleet testing to detect those new scenarios has a big potential in the future. However, for the data processing in the background, a lot of work needs to be done. One of the interviewees mentioned that it would be welcomed to have independent review processes for their scenario-based processes, including the scenario selection, which certify that one did all humanly possible to guarantee the safety of the vehicle. According to one of the interviewed experts, there are currently concurring approaches for the scenario selection and description, which can be seen in different research projects. That is why it may still be a long way until scenario sets for regulatory use cases are published.

CH-A8 scenario space and use of corner cases: Other than extracting scenarios from observing the traffic, also the possibility of artificially creating scenarios by combining several scenario aspects exists. Since the parameter space for scenarios is sheer endless, some of the interviewees mentioned that tools which intelligently scan the scenario space to identify as few scenarios as necessary to cover the scenario space are required. One of the interviewed experts mentioned the combination of scenario aspects to create corner case scenarios as a challenge: *In my opinion the challenge is to intelligently create those combinations and to test them automatically. (ID06)*

CH-A9 scenario generation: After a conceptual model of the scenario is created, the scenario needs to be translated into a scenario description language, which can be executed, e.g., by simulation tools. Due to the amount of scenarios needed for testing purposes, one expert does not think *that one will come to a point where all necessary scenarios can be created by hand. (ID10)* The scripting of a scenario includes calculating trajectories and determining actions of all actors which lead to desired traffic situations. One interviewed expert shared his experience on this matter and mentions that it is especially challenging when not only relative conditions between actors need to be taken into account but also absolute conditions like lane changes, which should happen at a certain point on a track. This is especially required for SBT on proving grounds (**CH-A10 SBT on proving grounds**) *which gets more and more important and needs to be expanded. (ID05)* This topic comes with several challenges regarding the precise control of the traffic participants and realization of test conditions. Additionally, the safety during the test execution in the real world requires additional attention.

CH-A11 scenario concatenation: Regarding the matter of tests on proving grounds, concatenating several scenarios to a continuous chain can be extremely beneficial for increasing test time efficiency, as stated by one of the experts. However, the development of a tool that determines the concatenated trajectories of each actor is considered challenging, but would be *a cracker (ID05)* due to its benefits.

CH-A12 test end criteria: On the matter of concluding the test activities, one expert formulated this as follows: *At some point one has to say: we tested enough. [. . .] We need to formalize, when enough has been tested. (ID05)* To find reliable test end criteria is mentioned as an essential challenge that is yet to be solved and is seen as a current research topic by one of the interviewed experts.

CH-A13 comparison to distance metrics: Since the SBT approach is not the classical approach for testing in the automotive industry, comparison metrics to currently used safety statements could be needed. Currently, vehicles are often tested on the road by driving the vehicles for a certain mileage. Additionally accident statics use mileage to describe the likeliness of accidents. To find a conversion between tested scenarios and mileage for comparison to accident statistics is considered challenging by one of the interviewed experts: *I think the biggest challenge is to find a conversion between scenarios and mileage. (ID07)*

B. Simulation-based Testing

Simulation-based testing and its technologies were also a big talking point during the interviews. Several experts mentioned that the increasing test demand cannot be satisfied by real world tests alone and that the testing must be shifted more to a virtual environment. However, this comes with several challenges:

CH-B1 purely simulation-based sign-off not yet possible: A key issue is that environment simulations are far from corresponding to reality. Several experts mentioned that it is not yet possible to completely renounce real world tests. One of the interviewed experts did not notice *a breakthrough [...] lately (ID09)* that would allow to perform the testing as a whole, and especially the final approval, in a simulation environment. CH-B1 describes the existing gap between the real world and simulation environments as of today.

CH-B2 determination of trustworthiness of simulation results: To be able to shift testing to a larger extent to virtual environments, a question still to answer is: *When can we trust a simulation result? (ID07)* Some interviewees mentioned that additionally to the flaws of the simulation, it must be possible to determine a probability statement about how good the simulation represents the real world performance of the VuT. According to one expert, this is necessary to determine the point in time when testing with simulation is sufficient and can replace real world testing. CH-B2 describes the difficulty of determining how good a simulation environment is at representing reality.

CH-B3 certification of simulation tools: While the preceding two challenges deal with the technical difficulties of comparing simulation to reality, for ADS developer the difficulty might be more on finding tool vendors who certify that their simulation tools provide fully trustworthy results. One expert is not aware of anyone, *who can prove or who dares to guarantee that their simulation software is so good that you don't need to perform real vehicle tests. (ID07)* Another expert thinks that certified simulation tools are necessary to be sure that simulation-based tests are performed correctly. CH-B1 and CH-B2 lead to the consequence of not having certified tools which is described by CH-B3.

CH-B4 identification of simulation-suitable tests: Due to the aforementioned challenges, not all of the tests can be conducted virtually. It needs to be analyzed which tests can be performed in simulation and which tests need to be conducted in the real world: *You need to determine what can be tested with an environment simulation as of today. (ID03)* Key points are to determine which details are required in a simulation for certain tests, according to an interviewed expert.

CH-B5 unknown effects: However, there might still be effects which occur in the real world which were not considered in simulation. One expert therefore mentioned that one needs to raise the question *whether there are effects that I do not yet consider. (ID06)* This applies as well to the fidelity of the simulation as to scenario aspects.

CH-B6 choice of simulation method: One of the interviewed experts stated that together with answering the question whether a test is suitable to be performed in a simulated environment, it also needs to be decided *how do I want to simulate? (ID08)* There are different possibilities to perform the simulations: open-loop testing, e.g., replay-to-sim, closed-loop testing, or a mixture of both.

CH-B7 lack of simulation models: For closed loop testing in the simulation, new simulation models are needed which are not yet available from the shelf. Those models *will be the great challenge (ID06)*, according to one interviewee.

CH-B8 validation of simulation models: Those models and the whole simulation environment also need to be validated. However, validation of simulation models is a challenging task, as stated for example by one of the interviewed experts: *Actually one would need fully validated simulation environments. And this is incredibly difficult to achieve. (ID07)*

Two types of models were highlighted during the interviews: driver models and models for the sensor simulation.

CH-B9 driver models: Driver models for the surrounding traffic are considered as *totally complex. (ID05)* and are lacking description standards.

CH-B10 sensor simulation: Models for sensor simulation purposes require high quality environment models, including reflectance of surfaces for radars and high quality visualization for camera sensors. As one interviewed expert stated: *There is still a lot of room for improvement. As far as quality is concerned [of] physical sensor simulation. (ID05)* Another expert sees sensor simulation as a current topic in research.

CH-B11 weather effects: Related to the CH-B10 is the topic of weather effects, like storm, rain, snow, ice or fog which have an influence on the perception modules. One expert even mentions this as *the biggest challenge of HiL testing (ID03)*, since the weather influences several sensors, like cameras, radar, lidar, ultrasonic sensors, and laser scanners, for which individual stimuli need to be generated.

CH-B12 integration of the test object in the simulation environment: Another topic is the integration of the VuT in the simulation environment. This comes with *a huge effort because of delay times in the communication of VuT and simulation, (ID03)* according to one interviewed expert.

C. Test Automation

Test automation usually comes to play for the test execution of repetitive tests but is not limited to this step of the test process. Test automation enables reproducible testing together with a reduction of costs and time effort [9].

CH-C1 lack of test automation: Due to the expected increasing test effort for AV, test automation is considered as a must-have. However, the current use of automation in the testing process is considered as insufficient to cope with the needs of tests for ADS of level four or five. According to one expert, test automation needs to be further adapted to all layers of the V-model, from software unit testing to vehicle testing and concerns test case generation as well as test execution. From the interviewees experience, the test activities *are done manually or are, to a small extent, automated. But this is not as advanced in terms of complexity and to the extent which is necessary for level four and five. (ID03)*

CH-C2 interaction between test automation, object and infrastructure: Increasing the extent of test automation and moving from manual testing to test automation is challenging. Test automation tools need to be integrated in the test environment and interfaces to the test object and other simulation tools like environment simulations need to be created. Integrating the test object, environment simulation, and test automation software is technically possible but it is a challenging task until everything plays together as intended (also see CH-B12) and might *crunch in some places until you got everything synchronized and fitted together. (ID04)* Extending the test automation idea on the test implementation step might also be beneficial because **(CH-C3 test implementation effort)** test implementation as of today often is costly, tool-specific, and needs to be performed manually every time details of the systems are changed. Test cases are written by humans and therefore *test cases cannot be translated automatically (ID01)*, e.g., to control inputs for a test automation software.

D. Test Execution

The challenges categorized under this headline are relevant to scenario-based, simulation-based, and real world tests and therefore are grouped separately. In the following, test execution is understood as all actions during run-time of a test, which can be the run of a simulation or a test drive on a test track.

CH-D1 omission of the test driver: When testing an AV with engaged ADS, the vehicle is not under the control of a human test driver anymore. Therefore the actions of the VuT depend on the algorithms of the ADS and cannot easily be modified by a test driver. One of the interviewed experts stated the following example: *As a test driver, I can't perform the cut-in a little closer because it is not me anymore who controls the vehicle. It is the vehicle who is in control. (ID13)*

CH-D2 realization of test conditions: Having in mind that an AV with engaged ADS controls the vehicle by itself without a human test driver at the wheel, the challenge of realizing test conditions arise and was mentioned by many experts during the interviews. During the test execution it might be necessary to realize test (pre-)conditions, like a certain critically of a traffic situation, a certain scene of a scenario, or internal system states of the VuT. The focus of the mentioned challenge lies on the realization of certain traffic situations or scenes, *because the vehicle may actively try to avoid this scene (ID07)*, or does not behave as expected by the test designer, which makes the test execution even more challenging or even infeasible.

CH-D3 coordination of the surrounding traffic: The movement of other traffic participants extends the parameter space of a scenario. The question is how the movement of the surrounding traffic can be used for creating variations of scenarios and which movements are relevant to consider. An additional challenge arises in finding control methods that allow for realistic as well as scenario-compliant behavior of the surrounding traffic. As an interviewed expert states: *One difficulty is [. . .] how to coordinate the movement of the other traffic participants. (ID11)*

IV. Comparison with Related Work

For the discussion of the results of the interview study, we set the explored challenges in relation to related publications. First of all, we compare our study to two similar interview studies and highlight mutual results. Thereafter, we report on research approaches that address identified challenges and identify gaps.

Related Empirical Studies: In 2017, Knauss et al. identified challenges in testing functionality and safety of AV through focus groups and interviews with overall 26 participants from industry and research. They present challenges in five categories among which simulation- and virtual-testing-related challenges are as well as challenges concerning scenario complexity and the amount of test cases. In comparison to our study, Knauss et al. focused on level three ADS and above instead of solely focusing level four ADS. Still, our challenges CH-A3, A10, B3, and B10 are also identified by Knauss et al. [5]. Song et al. conducted an exploratory study on the the testing of automated systems in general, not only focusing on AV like we did, including literature review, focus group discussion and interviews with industry practitioners. Four of the identified challenges are the unpredictable environment, complexity, data accessibility, missing standards and guidelines. Among these challenges are also CH-A1 and A3 from our study [6]. While we identified 26 additional challenges to those mentioned in [5], [6], we also report five mutual challenges and show that these challenges are relevant to our interviewed experts, too.

Existing Approaches and Research Directions: In Table I, we list existing approaches that address the 31 challenges presented previously. Four challenges deal with the inadequate state of the art (CH-B1, C1), a lack of artifacts (CH-B7) or changed circumstances while testing (CH-D1) which cannot directly be linked to research approaches. Instead, we refer to related challenges and their approaches which contribute to those four challenges.

Scenario-based testing is a broad field of research with many existing publications. Regarding the scenario elicitation (CH-A7), [15], [16] present overviews on data-driven and knowledge-driven approaches which can be seen as starting points for further research. For coping with the scenario complexity (CH-A3), a six-layer-model for the scenario description [13] and three abstraction levels, functional, logical and concrete scenarios [12], are proposed. Also standardization groups work on consistent scenario description languages (CH-A2) for simulation purposes [8]. As one of our interviewees stated, the definition of a test end criteria (CH-A12) is a current research topic which, e.g., is addressed by [20] and [21]. Further related publications to CH-A1, A4–A6, A8, and A9 are listed in Table I. For the challenges of converting tested scenarios to a distance metric (CH-A13) as well as the concatenation of scenarios (CH-A11) we did not find suitable literature.

Regarding *simulation-based testing*, the automotive community is well aware of the discrepancy between virtual tests and reality which hinders purely virtual testing (CH-B1). Nevertheless, an existing approach for simulation-supported homologation tests for electronic stability control systems could be transferred to ADS homologation [24]. Additional suggestions for the validation of simulation models (CH-B8) are reported in [22], [24], [27]. For sensor models, [29] gives an overview on publications for modelling approaches (CH-B10, B11). [16] addresses the topic of missing out relevant

TABLE I
Challenges as Mentioned in Related Work

Challenge:	Identified in:	Addressed by:
scenario-based testing		
A1	[6]	[10], [11]
A2, A6		[8]
A3	[5], [6]	[12], [13]
A4		[14]
A5		[11]
A7		[3], [15], [16]
A8		[17]
A9		[18]
A10	[5]	[19]
A12		[20], [21]
A11, A13		no publications found
simulation-based testing		
B2, B4		[22]
B3	[5]	[23], [24]
B5		[25]
B6		[22], [26]
B8		[22], [24], [27]
B9		[28]
B10	[5]	[29]
B11		[29]
B12		[30]
B1; B7		see B3, B8; see B9, B10
test automation		
C1		see C2, C3
C2		[30]
C3		[31], [32]
test execution		
D1		see D2
D2	[1]	no publication found
D3		[33], [34]

scenario aspects (CH-B5) due to *unknown unknowns* and refers to [25] and ISO/PAS 21448. Apart from that, Table I presents approaches to CH-B2–B4, B6, B7, B9, B12.

Test automation (CH-C1–C3) is a relevant topic in the automotive community. Approaches regarding standardized interfaces and test implementation exist, e.g., in [31], [32].

Regarding the *test execution*, the challenge of realizing test conditions (CH-D2) stands out as it was often-mentioned (11 out of 13 interviewees). The problem of the AV avoiding certain behavior, which is necessary for an intended test, can also be found in [1]. However to the best of our knowledge, solutions to overcome this problem are sparsely discussed in literature and research projects. An approach presented in [34] for the coordination of platoons of connected and automated vehicles could be transferable to CH-D3 of controlling the surrounding traffic of an VuT, although it does not appear to have been developed with the aim of being used in the testing process of an AV. In contrast, an approach dedicated for testing motion planning of an AV using evolutionary algorithms for the optimization of trajectories for other traffic participants can be found in [33].

V. Conclusion and Future Work

By identifying 31 challenges in the fields of scenario- and simulation-based testing, test automation and test execution, we can show that many of the known challenges of AV testing still are present, and are also recognized as challenges from our interviewed experts.

Our comparison of identified challenges and other publications indicates that approaches for several of the named challenges exist. However, future work on almost all challenges is necessary and should be tackled in the near future. Additionally, we recognise a lack of approaches concerning the realization of test conditions for automated vehicles due to the limited control over, and possible non-determinism of, the AV during the test. This challenge needs further solutions and tool support, which we plan to address in future work.

ACKNOWLEDGMENT

We thank all participants of our interview study.

References

[1] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

[2] On-Road Automated Driving (ORAD) committee, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, apr 2021. [Online]. Available: https://doi.org/10.4271/J3016_202104

[3] C. Roesener, F. Fahrenkrog, A. Uhlig, and L. Eckstein, "A scenario-based assessment approach for automated driving by using time series classification of human-driving behaviour," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 1360–1365.

[4] A. Knauss, J. Schroder, C. Berger, and H. Eriksson, "Software-related challenges of testing automated vehicles," in *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*, 2017, pp. 328–330.

[5] A. Knauss, J. Schröder, C. Berger, and H. Eriksson, "Paving the roadway for safety of automated vehicles: An empirical study on testing challenges," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1873–1880.

[6] Q. Song, E. Engström, and P. Runeson, "Concepts in testing of autonomous systems: Academic literature and industry practice," in *2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN)*, 2021, pp. 74–81.

[7] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical software engineering*, vol. 14, no. 2, pp. 131–164, 2009.

[8] ASAM e. V., "Asam openscenario." [Online]. Available: https://www.asam.net/standards/detail/openscenario/

[9] F. Witte, "Testautomatisierung," in *Testmanagement und Softwaretest*, F. Witte, Ed. Wiesbaden: Springer Fachmedien Wiesbaden, 2019, pp. 231–241.

[10] DIN SAE Consortium, "Terms and definitions related to testing of automated vehicle technologies," 2019. [Online]. Available: https://www.sae.org/standards/content/dinsaespec91381/

[11] M. Steimle, T. Menzel, and M. Maurer, "Toward a consistent taxonomy for scenario-based development and test approaches for automated vehicles: A proposal for a structuring framework, a basic vocabulary, and its application," *IEEE Access*, vol. 9, pp. 147 828–147 854, 2021.

[12] T. Menzel, G. Bagschik, and M. Maurer, "Scenarios for development, test and validation of automated vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1821–1827.

[13] J. Bock, R. Krajewski, L. Eckstein, J. Klimke, J. Sauerbier, and A. Zlocki, "Data basis for scenario-based validation of had on highways," in *27th Aachen colloquium automobile and engine technology*, 2018, pp. 8–10.

[14] F. Bock, C. Sippl, A. Heinz, C. Lauerz, and R. German, "Advantageous usage of textual domain-specific languages for scenario-driven development of automated driving functions," in *2019 IEEE International Systems Conference (SysCon)*, 2019, pp. 1–8.

[15] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick, and F. Diermeyer, "Survey on scenario-based safety assessment of automated vehicles," *IEEE Access*, vol. 8, pp. 87 456–87 477, 2020.

[16] C. Neurohr, L. Westhofen, T. Henning, T. d. Graaff, E. Möhlmann, and E. Böde, "Fundamental considerations around scenario-based testing for automated driving," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 121–127.

[17] D. Bogdoll, J. Breitenstein, F. Heidecker, M. Bieshaar, B. Sick, T. Fingscheidt, and M. Zollner, "Description of corner cases in automated driving: Goals and challenges," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1023–1028.

[18] A. Heinz and J. S. Wolfram Remlinger, "Track-/scenario-based trajectory generation for testing automated driving functions," in *8. Tagung Fahrerassistenz*, 2017.

[19] H.-P. Schöner and W. Hurich, "Testing with coordinated automated vehicles," in *Handbook of Driver Assistance Systems*, H. Winner, S. Hakuli, F. Lotz, and C. Singer, Eds. Cham: Springer International Publishing, 2016, pp. 261–276.

[20] F. Hauer, T. Schmidt, B. Holzmuller, and A. Pretschner, "Did we test all scenarios for automated and autonomous driving systems?" in *2019 IEEE Intelligent Transportation Systems Conference - ITSC*, 2019, pp. 2950–2955.

[21] W. Wang, C. Liu, and D. Zhao, "How much data are enough? a statistical approach with case study on longitudinal driving behavior," in *IEEE Transactions on Intelligent Vehicles*, 2017, pp. 85–98.

[22] M. Wood, P. Robbel, M. Maass, R. D. Tebbens, M. Meijs, M. Harb, J. Reach, K. Robinson, D. Wittmann, T. Srivastava, *et al.*, "Safety first for automated driving," *Aptiv, Audi, BMW, Baidu, Continental Teves, Daimler, FCA, HERE, Infineon Technologies, Intel, Volkswagen*, 2019.

[23] Y. Page, F. Fahrenkrog, A. Fiorentino, J. Gwehenberger, T. Helmer, M. Lindman, O. op den Camp, L. van Rooij, S. Puch, M. Fränzle, *et al.*, "A comprehensive and harmonized method for assessing the effectiveness of advanced driver assistance systems by virtual simulation: The pears," in *International Technical Conference on the Enhanced Safety of Vehicles*. ., 2015, pp. 1–12.

[24] H. Abdellatif and C. Gnandt, "Use of simulation for the homologation of automated driving functions," *ATZelectronics worldwide*, vol. 14, no. 12, pp. 68–71, 2019.

[25] H. Lakkaraju, E. Kamar, R. Caruana, and E. Horvitz, "Identifying unknown unknowns in the open world: Representations and policies for guided exploration," in *Thirty-first aaai conference on artificial intelligence*, 2017.

[26] M. Steimle, N. Weber, and M. Maurer, "Toward generating sufficiently valid test case results: A method for systematically assigning test

cases to test bench configurations in a scenario-based test approach for automated vehicles," *IEEE Access*, vol. 10, pp. 6260–6285, 2022.

[27] S. Riedmaier, J. Nesensohn, C. Gutenkunst, T. Düser, B. Schick, and H. Abdellatif, "Validation of x-in-the-loop approaches for virtual homologation of automated driving functions," in *11th Graz Symposium Virtual Vehicle*, 2018.

[28] J. Klimke, D. Becker, and L. Eckstein, "System design of an agent model for the closed-loop simulation of relevant scenarios in the development of ads," in *29th Aachen Colloquium Sustainable Mobility 2020*, 2020, pp. 1493–1530.

[29] M. Hoss, M. Scholtes, and L. Eckstein, "A review of testing object-based environment perception for safe automated driving," *Automotive Innovation*, pp. 1–28, 2022.

[30] T. Fricke, F. Schuck, J. Vogelpohl, J. Ax, R. Stolpe, and H. Amelunxen, "Collaborative development of a test environment for automated driving," in *20th International Congress and Exhibition*, ser. VDI-Berichte, 2021, vol. 2384, pp. 5–16.

[31] ASAM e. V., "Asam xil." [Online]. Available: https://www.asam.net/standards/detail/xil/

[32] C. Karthaus, A. Bick, B. Douglas, and F. Hanke, "X-in-the-loop-a universal testing methodology," *ATZelectronics worldwide*, vol. 16, no. 11, pp. 46–51, 2021.

[33] M. Klischat and M. Althoff, "Generating critical test scenarios for automated vehicles with evolutionary algorithms," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 2352–2358.

[34] J. Wu, S. Ahn, Y. Zhou, P. Liu, and X. Qu, "The cooperative sorting strategy for connected and automated vehicle platoons," *Transportation Research Part C: Emerging Technologies*, vol. 123, p. 102986, 2021.