

RESEARCH ARTICLE

Fidelity of Simulations for Verification & Validation of Automated Driving: A Practitioners' Perspective

Beatriz Cabrero-Daniel^{1,2,*}, M. Cagri Kaya^{1,2,a}, Ali Nouri^{1,2,3}, Mazen Mohamad^{1,2,4,b},
Christian Berger^{1,2,c}

¹Chalmers University of Technology, Gothenburg, Sweden

²University of Gothenburg, Gothenburg, Sweden

³Volvo Cars, Gothenburg, Sweden

⁴RISE Research Institutes of Sweden, Borås, Sweden

^aEmail: cagri.kaya@chalmers.se

^bEmail: mazen.mohamad@ri.se

^cEmail: christian.berger@gu.se

ABSTRACT

Modern advanced driver assistance systems (ADAS) and automated driving (AD) require robust verification and validation (V&V) methods due to their increasing complexity. Simulation-based testing is adopted industry-wide as it provides a controlled, replicable and safe environment with the potential to scale. Yet, fidelity concerns are often faced that still require *in situ* testing in the real world, which is typically considered time-consuming and being dependent on environmental factors which are hardly controllable. This qualitative study contrasts the practitioners' view on the fidelity of state-of-the-art virtual test environments in the context of V&V for ADAS and AD in comparison with recent findings from scientific literature. We conducted semi-structured, in-depth interviews with 11 domain experts from the automotive industry to systematically explore their perception of how simulation technologies can reduce the dependency on real-world testing. A thematic analysis was conducted to deduce codes and to develop key themes from the interview data for comparison with the scientific literature. To enhance the validity of our findings, additional focus group interviews with seven experts were performed to complement the initial round of interviews. Our findings reveal necessary steps in bridging the fidelity gap, including systematically measuring simulation model quality, deciding the necessary amount of realism, and the importance of collaboration and transparency among stakeholders.

ARTICLE DATA

Article History

Received 13 October 2025

Revised 22 January 2026

Accepted 26 January 2026

Keywords

Automotive software
engineering

Digital twins

Functional safety (ISO 26262)

Key performance indicators
(KPIs)

Scenario-based testing

Simulation fidelity

Simulation-to-reality gap

Verification and validation
(V&V)

1. INTRODUCTION

Automated Driving (AD) is fuelling a technological competition in the automotive industry to deploy a technology that may surpass the capabilities of an average yet experienced human driver to handle countless typical driving situations. Expectations include primarily a substantial reduction of fatal traffic accidents [1], which is directly addressing United Nations' Sustainable Development Goal 3.6, followed by a more sustainable use of resources and hence, increased efficiency of the transportation system, and eventually, better comfort for passengers of individual mobility. However, when

today's Advanced Driver Assistance Systems (ADAS) gradually evolve from lower to higher levels of autonomy (towards SAE level 3 [2], already considered AD), verification and validation (V&V) processes during software and system engineering face significant challenges to support traceable and transparent documentation of the system's performance, especially when using simulations as part of V&V to predict real-world performance [3,4]. Although simulations offer a controlled and cost-effective environment to conduct software testing, there remains a fidelity gap between simulated and real-world scenarios [5].

*Corresponding author. Email: beatriz.cabrero-daniel@gu.se

The required amount of V&V when developing AD systems entails finding a balance between safety, the priority and testing resources. Therefore, a large number of tests are conducted both in simulated and under actual real-world conditions [6]. Testing often begins in simulations before transferring to real-world experiments. However, differences in lighting, textures and material properties can alter sensor readings [5], which can also be affected by calibration errors, noise or occlusions, creating discrepancies between simulated and real conditions. Such known and unknown factors contribute to the fidelity gap.

Existing literature emphasizes enhancing the degree of fidelity as a critical factor, which is affecting the reliability and trustworthiness of results from V&V [7,8]. Yet, research lacks a systematic evaluation of practitioners' perspectives to substantiate concerns about using simulations for testing ADAS and AD, and engineers still face challenges in addressing failures, limitations and key functional and non-functional requirements to better bridge the fidelity gap between virtual and real-world V&V.

In this article, we identify key factors to bridge the fidelity gap through an in-depth, semi-structured interview study with experts from industry and academia. We report on expectations, strategies for handling simulation errors and essential features of high-fidelity simulations to enhance V&V for reliable AD systems. The interviews provided broad insights into simulations, toolchains and AD/ADAS features, leading to relevant themes and recommendations for improving simulation-based V&V. We further validated these outcomes with a focus group and comparison with recent literature.

The rest of the article is structured as follows. [Section 2](#) covers background and the related work. [Section 3](#) outlines the research questions and the methodology. [Section 4](#) describes the extracted key themes from the semi-structured, in-depth interviews. We provide our interpretations in [Section 5](#) and conclusions in [Section 6](#), including final remarks and future work.

2. BACKGROUND AND RELATED WORK

Simulations can replicate road networks, the environment (e.g., buildings) [8,9], weather conditions [10,11] and dynamic elements in a scenario [12], using appropriate models from real-world counterparts. The V&V-relevant quality attributes of a simulation cannot be determined independently of its intended use as both quantitative and qualitative performance indicators become relevant. For instance, different elements of an ADAS or AD require different levels of details of models for sensors or traffic participants.

Adopting simulation tools are recommended for verification of AD in challenging scenarios by regulations such as UNECE R157 [13] given the reduced safety implications in virtual test environments. Hence, simulations allow for better preparations of system validations on real-world test sites [11]. When considering recent industrial trends such as understanding a car as a *software-defined vehicle* [14] to accelerate releasing new features, integrating simulation tools into the corresponding development and testing processes supported by continuous integration and continuous deployment (CI/CD) pipelines also contributes towards continuous V&V of a vehicle's software features [15].

However, in order to use simulations within early V&V processes, their trustworthiness needs to be assessed [16]. This includes measuring their robustness and accuracy in replicating real-world scenarios (i.e. their fidelity [17]) and studying whether the virtual scenarios in focus are relevant for V&V. This raises questions of (i) how to select relevant scenarios; and (ii) how to assess their fidelity.

2.1. Test Scenario Selection

Well-established V&V approaches for ADs are based on defined scenarios with specific maneuvers [11]. However, the scenario space for testing AD features is intractable [18]. Efforts such as the European New Car Assessment Programme (Euro NCAP) [19] exist to standardize and prioritize test scenarios as they have been typically derived from clusters of real-world accident data to design relevant tests. Such reference scenarios often also serve as a basis for V&V of AD features, both in physical and virtual environments [11,20]. However, as ADAS and AD are supposed to support the driver to avoid ending up in challenging traffic situations in the first place, looking only at such existing test catalogues is insufficient.

2.2. Simulation Fidelity

Correlating simulation outputs with real-world system behaviour (e.g., on test sites or public roads) is essential for V&V and regulatory approval [13,16]. ISO 26262 proposes a detailed process for the qualification of typical categories of software tools including simulation tools [20]. It also addresses the potential impact of tool malfunctions. However, it is challenging to generate realistic synthetic scenarios as virtual sensor input [11] and validate the accuracy of the simulated sensor stack. Schlager et al. [21] propose a classification method for the fidelity of sensor models and discuss their relevance for different stages in the development of AD/ADAS (within the V-model). Magosi et al. [22] present an evaluation methodology

for validating perception sensor models, which involves conducting real-world test drives, replicating the same scenarios in a simulation environment and using statistical techniques to compare the measured and simulated sensor outputs.

3. MATERIALS AND METHODS

For this study, we employed method triangulation, as depicted in Fig. 1, to validate and strengthen our findings and conclusions: We employed dedicated interviews, a literature review and a focus group study, as described below, to address the following research questions (RQs):

- RQ1** What are the expectations of domain experts in the automotive field when it comes to using simulations for V&V of AD features?
- RQ2** What are current practices and challenges in the automotive industry for developing and troubleshooting simulations in the V&V of AD features?
- RQ3** What are the necessary functional and non-functional properties to systematically assess the fidelity gap between simulation and the real world?
- RQ4** How can the fidelity gap be qualitatively and quantitatively evaluated?

In the first phase, a detailed interview protocol was designed to guide the interviewers and to ensure consistency throughout the data collection process according to Strandberg's recommendations about ethically conducting interviews (cf. [23]). This protocol

comprises (i) piloting for interview; (ii) planning phase of the interview and interview questions; and (iii) implementation of the interview protocol.

A pilot interview was held to obtain preliminary data. The research was carried out with selected domain experts predominantly from the automotive industry, for which we adopted convenience sampling [24]. Selected interviewees were expected to have at least 5 years of experience. The interviews were transcribed and any information such as names or affiliations were removed to avoid traceability of an interviewee's identity.

Next, thematic analysis was used to process the interview transcripts. The raw data was analysed to extract patterns and themes related to our research questions [25]. A sequential data analysis process was adopted to familiarize with the content [26], followed by coding, where data segments were annotated with codes indicating key ideas and concepts [27] related to simulation-based V&V of AD. Overlapping codes were organized into potential themes, rigorously reviewed and refined by the authors in multiple rounds to ensure they authentically represent the data [28]. This method distills complex qualitative data into useful insights, creating a nuanced understanding of expectations, limitations and the perceived fidelity gap in simulation tools for V&V of AD.

Interviewees averaged 15 years of domain experience (up to 25 years) and held roles such as simulation engineer, tester, developer, researcher and managing director. Data saturation was reached, as no new themes nor key outcomes emerged in the last interviews.

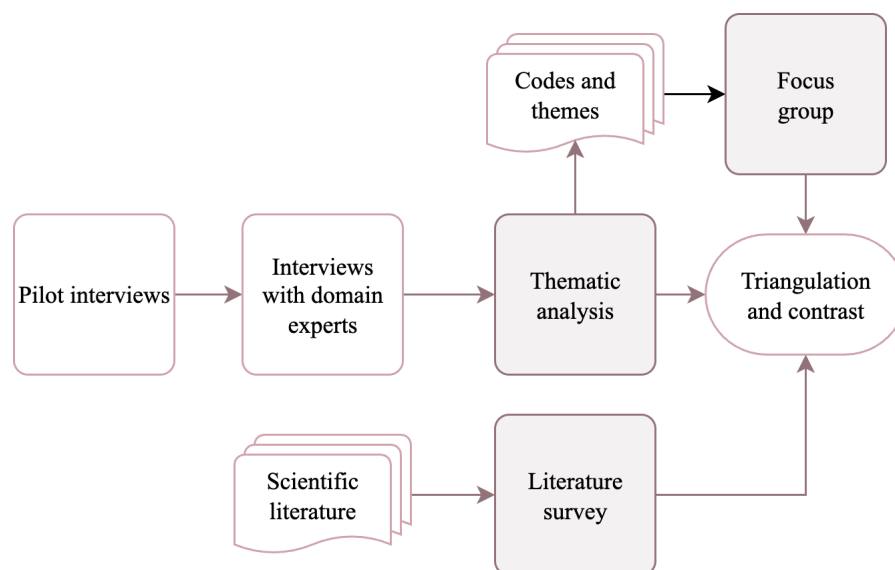


Figure 1. Method triangulation used to contrast the insights from the domain experts with the scientific literature. A focus group with participants from different organizations was then conducted to deepen on the observed challenges.

To contrast the domain experts' input with recent insights and recommendations as suggested in scientific literature, we surveyed recent studies addressing the fidelity gap for simulations. Moreover, we conducted a focus group with 7 participants representing 5 different organizations from all parts of the automotive value chain (1 original equipment manufacturer, 3 test site operators, 1 academic, 1 sensor supplier and 1 provider); all details can be found in [29]. The session included multiple brainstorming activities aimed to identify and discuss simulation challenges for V&V and to align interview insights with observed challenges.

4. RESULTS

In this section, we provide themes extracted from the interviews, as illustrated in Fig. 2, Fig. 3, Fig. 4 and Fig. 5. Interviewees' insights are contrasted with the recent scientific literature and put in context with the help of the focus group study.

4.1. Expectations for V&V Simulations

Interviewees expect to find errors in the simulation rather than on the test track, or even worse, after deployment in the real world. As a participant stated "what is expected to be seen in the real world must be in the simulation and vice versa." Successful simulations increase awareness so that dangerous situations can be understood, so the corresponding AD features are tested in the test track driverless, or not tested at all. This involves covering a wide range of Operational Design Domains (ODDs) and particularly focusing on the planning aspects of AD systems where issues are most commonly occurring.

Quantitative metrics are essential to compare simulations with real-world data and ensure test repeatability and reliability. Interviewees emphasized the need for KPIs to evaluate simulations, these metrics should reflect specific system components, associate sensor outputs to properties like velocities and account for human tester error margins in real-world tests. KPIs were a recurrent topic in the interviews, leading to a number of related codes under the umbrella of "expectations," as seen in Fig. 2.

Multiple experts stated the importance of establishing trust in simulations. Transparency is considered to help, including being clear about the limitations of the simulation tools as well as ideally sharing models or algorithms while respecting intellectual property and adopting common standards.

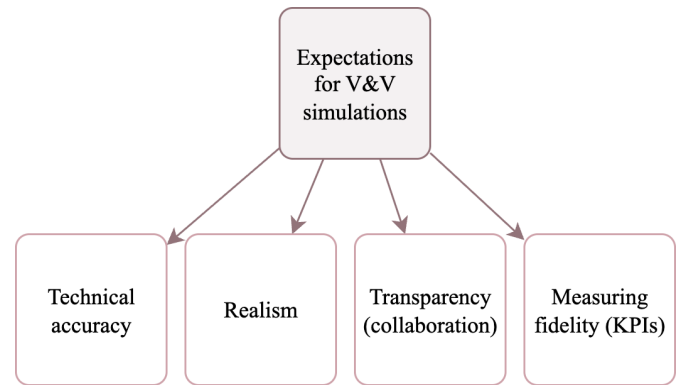


Figure 2. Themes arising during the interviews under the parent theme "expectations for V&V simulations". All the themes and codes under this parent are related to the expectations of domain experts in the automotive field when it comes to using simulations for V&V of AD features, addressing RQ1.

4.1.1. Striving for Technical Accuracy and Realism

Accurate sensor outputs and realistic environmental interactions are key to high-fidelity, especially for models integrated into AD simulators [30]. This quest for "technical accuracy" indicates the need for realistic models, particularly for sensors and scenarios, to bridge the fidelity gap. Digital Twins, which strive for greater realism in environmental and vehicle modeling, are gaining increasing attention [31,32,33,34]. Hinton et al. [35] suggest the knowledge distillation technique to increase accuracy, especially for sensor models to improve efficiency in the context of AD as also reported by Saputra et al. [36] and Zhang et al. [37]. This conclusion was reiterated by the focus group: Simulation models need to be correlated with real-world data at full-system scale and statistically significant thresholds to determine the realism of the simulations must be defined.

4.1.2. Fostering Transparency and Standardization

Interviewees expect the benefits of increased transparency when using simulation by, for instance, sharing data, models, algorithms and scenario pools. It is expected that the overall quality and reliability of simulation-based V&V for AD can be increased when enhancing transparency through collaboration and sharing. This is supported by standardized test catalogs like Euro NCAP and the adoption of industry standards such as OpenSCENARIO [38]. Besides, increasing the usage of open-source software contributes to standardization and transparency [39]. In addition, data-driven V&V and artifact interchange can enhance collaboration across the value chain (e.g., manufacturers, suppliers, regulators). Yet, intellectual property and ownership

issues hinder progress in competitive fields like AD. Participants agreed that sharing or granting access to data, models and tools is essential, but they observed recurring barriers – especially the effort of integrating proprietary tools and data formats despite existing standards.

4.1.3. KPIs to Measure Fidelity

The establishment and use of KPIs for systematically assessing the fidelity of a simulation is essential [3]. Widner et al. [40] highlight validation metrics as a part of their framework for vehicle dynamics model validation, emphasizing their importance in quantifying the model's credibility.

Interviewees noted that sensor outputs are key to assessing simulation reliability and identifying successful simulations, with KPIs helping to determine and increase reliability. The focus group agreed on using meaningful metrics to verify performance across simulated and real environments. They recommended analyses at the component level before scaling to full-system evaluation.

Insight for RQ1: The recommendation is to systematically strive to establish and measure individual simulation model quality with respect to their required realism, as well as collaborating towards standardization and increasing transparency through sharing being able to systematically assess the fidelity of the simulation contribute to better bridging the fidelity gap.

4.2. Practices and Challenges for Developing Simulations

Interviewees emphasized the inclusion of generic scenarios and scenarios driven by regulations such as Euro NCAP. Adding randomization to generic or standardized scenarios, employing scenario generators and being proactive in searching for challenging situations were also highlighted to add variety. The interviewees also emphasized the systematic search for edge-case scenarios.

The other theme that arose, as seen in Fig. 3, is concerned with troubleshooting whether the cause of issues originates from the simulation or the system-under-test. Systematic approaches are recommended, including examining the scenarios and the input and output models. Furthermore, conducting root cause analysis (RCA) and clarifying testing objectives were emphasized.

4.2.1. Comprehensive Scenario Design

Interviewees agreed on designing diverse simulation scenarios including generic and edge-case scenarios.

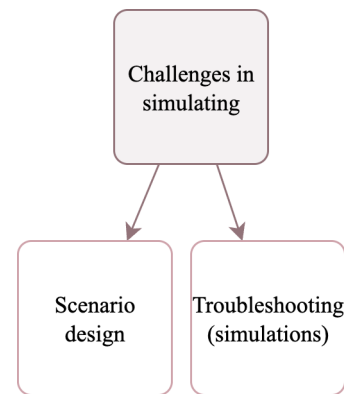


Figure 3. Themes arising during the interviews under the parent theme “challenges in developing simulations”. All themes and codes under it are related to the current practices and challenges in the automotive industry for developing and troubleshooting simulations in the V&V of AD features, addressing RQ2.

Doing so in a methodological way is important to aim for variety and quality. The domain experts' input is aligned with Tobin et al.'s recommendation of adopting the technique of “domain randomization” to test models in different environments [41]. Kontes et al. and Pouyanfar et al. report about this technique being applied in the context of AD [42,43].

The focus group participants also highlighted the need to have a real-world setup in which to gather real data to compare the simulations against and balancing a large number of scenarios and “cherry picking” good quality scenarios.

4.2.2. Systematic Troubleshooting

Interviewees underlined that systematic troubleshooting helps analysing discrepancies between simulation and real-world results effectively. They recommended methods like RCA to isolate issues by breaking systems into subsystems and examining components independently, and analysing simulation components such as sensor and actuator models separately without the system under test. This conclusion is aligning with the work of Maier et al. [44] suggesting causal methods for scenario-based testing of ADAS. Breaking down the system into its components for better performance analysis is emphasized by the focus group participants. Moreover, they highlighted the needs to have “easy debugging,” especially when it comes to nuances of the physical sensors.

Insight for RQ2: Test scenarios need to have extensive coverage. Methodologically isolating all modules of a simulation environment allows for systematically narrowing down issues affecting the V&V environment to identify root causes. It has been recommended to start with generic and regulatory-compliant components and scenarios before specific, tailored components and scenarios are investigated.

4.3. Properties to Systematically Assess Fidelity

Interviewees emphasized that simulations must accurately replicate the functional aspects of AD systems to ensure their safety and reliability. This includes adherence to safety standards like ISO 26262. The role of requirements in the test definition was highlighted by multiple interviewees as measurable, objective and clear requirements leading to the development of useful tests. A systematic assessment was suggested to be possible by breaking down requirements into technical specifications and using them as a benchmark to evaluate simulations. This would then also help to effectively assess the fidelity of the simulation. These concerns were coded as seen in Fig. 4.

Since requirements are defined earlier than tests, developers must consider the time gap between specification and testing stages. However, interviewees noted that exact real-world replication of simulations is maybe impossible, but such efforts are vital to evaluate fidelity gaps and derive acceptable deviations between simulated and real-world results.

4.3.1. Functional Correctness and Safety

Simulations serve as tools to test safety-critical scenarios and build confidence before real-world trials but require rigorous development, validation and maintenance. Interviewees emphasized ensuring simulations have functional correctness for the system-under-test in order to replicate the vehicle's safe (or unsafe) behaviour, aligning with Neurohr et al. [45]. The focus group added that clear, measurable requirements (which might evolve over time) are essential to assess simulation confidence before real-world testing.

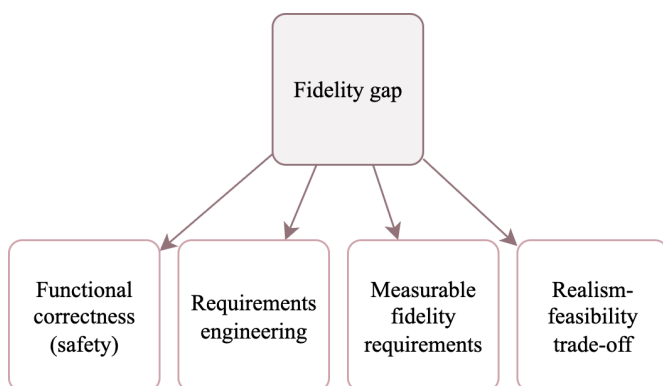


Figure 4. Themes arising during the interviews under the parent theme “fidelity gap” and its measurement and assessment. All themes and codes under it are related to the necessary functional and non-functional properties to systematically assess the fidelity gap between simulation and the real world, addressing RQ3.

4.3.2. Clear and Measurable Requirements

Aligning requirements and testing is a common endeavor in software development for every industry. For the automotive context, the work of Sippl et al. [46] is an example that shows this connection explicitly. Supporting this, interviewees emphasize the importance of having clear, objective and measurable requirements for developing effective tests. Therefore, to pave the way for better-assessing simulations, requirements should include metrics and rationales for measurement values. The focus group participants had in-depth discussions on the need for having clear and measurable requirements, bringing in examples and experiences from different projects. Only agreeing and refining these requirements allows for appropriate planning of tests and V&V.

4.3.3. Realism-Feasibility Trade-Off

Defining acceptance criteria for a “perfect simulation” would require accurately modeling all physical principles affecting sensors and actuators with computationally feasible algorithms. The focus group raised concerns about using “ideal sensors” that omit the noise or distortions when balancing simulation realism and feasibility, and discussed setting limits for the V&V efforts, mostly connecting them to the scope and phase of the project.

Interviewees recognized the need to balance simulation realism with practical limitations regarding the realism versus feasibility trade-off: Realism for its own sake is inefficient and having just the right amount of realism is the key [47]. Other studies addressed the replicability of simulation tests in the real-world. For example, Fremont et al. [48] reported that 62.5% of the unsafe cases in the simulation resulted in unsafe behavior on the test track and 93.3% of the safe cases in the simulation resulted in safe behavior on the track. On the other hand, Reway et al. identified hardly controllable factors such as weather conditions as contributing said gap [49].

While combining simulations with real-world testing is beneficial, improving simulation realism increases complexity and execution time. Thus, there's a trade-off between feasibility and accuracy. Clearly defining the simulation tool's scope is essential, as it guides architecture, component detail and signal precision.

Insight for RQ3: Ensuring functional correctness, having clear initial requirements that are ideally easy to convert into test scenarios and balancing realism with feasibility, are factors to improve bridging the fidelity gap.

4.4. Scenario Benchmarks

Interviewees emphasized that vehicle and sensor models must reflect real-world behaviour and real data. For that, broad scenario coverage is important. However, while there is no limit to conducting “enough” tests, real-world completeness is impossible. Some interviewees stated that Euro NCAP tests can be a “minimum requirement” or an “absolute must” allowing more complex tests or corner cases be built upon. These two themes from the interviews can be seen in Fig. 5.

4.4.1. Establishing Benchmark Standards

Standards offer a consistent way of evaluating simulation performance and fidelity, enabling comparable tests across stakeholders, as seen with Euro NCAP benchmarks [19]. This aligns with ongoing efforts like ASAM’s open standards (e.g., OpenSCENARIO), which are used in simulators such as CARLA [9] for scenario design. Even though it has been acknowledged that it is not possible to always prefer open-source tools over proprietary or in-house applications, using open-source tools contributes to transparency and encourages increased interactions among stakeholders [50]. The focus group participants agreed on the need to have test scenario benchmarks and referred Euro NCAP as a well-established option. At the same time, standard test sets, including Euro NCAP, have drawbacks such as “blank backgrounds” making them trivial compared to real-world scenarios.

4.4.2. Prioritizing Scenario Quality Over Quantity

Interviewees reflected on the need for clear criteria for whether enough simulations have been performed. Interviewees agreed that the “interestingness” and quality of tested scenarios are better indicators than total

kilometers driven, aligning with safety assurance protocols based on key principles rather than distance metrics alone [11,45,51]. Similarly, focus group participants discussed how to ensure that the selected scenarios and test cases are appropriate for V&V of AD systems. The participants highlighted the need to cover a large proportion of the ODD, always making sure that the test cases are representative of real-world scenarios.

Insight for RQ4: Adopting standardized benchmarks and focusing on the quality of simulation scenarios are important factors in evaluating simulations for AD, ensuring both consistent performance assessment and a deeper understanding of their real-world applicability.

5. DISCUSSION

This study is a step towards improving V&V processes for AD and ADAS by identifying steps to enhance simulation fidelity. Findings show that achieving effective simulation-based V&V requires rigorous methodologies, industrial standards, and further research and engineering effort.

An observation from the interviews and the focus group study, supported by scientific literature, is the notable number of studies on the simulation-reality gap and simulation fidelity, particularly in the context of perception sensor models (e.g., [21,22,52]). This focus may stem from their critical role in AD and ADAS systems, as discrepancies in sensor models can dramatically affect system reliability, performance and safety. Additionally, perception systems may be seen as fine-grained parts of the simulation stack since real-world measurements are easier to compare with simulated sensor output, especially within a limited ODD, making them an ideal starting point for improving simulation fidelity. However, other simulation components, such as vehicle dynamics models, are also subject to research on the simulation-reality gap [40].

5.1. A Way Forward: The Uncertainty-Aware Model

When comparing simulation results with real-world measurements, addressing uncertainty is critical. For example, when assessing the accuracy of a simulated braking event by comparing the longitudinal acceleration of a vehicle and its simulated counterpart in the same ODD, it is essential to account for all possible sources of real-world uncertainty. These uncertainties stem from factors such as environmental conditions, road and tire types, and different braking system components. A thorough understanding of these factors can reduce the fidelity gap, leading to more realistic simulations. Moreover, by having a clear comprehension of the real system’s output range, discrepancies between the real and simulated systems can be identified and

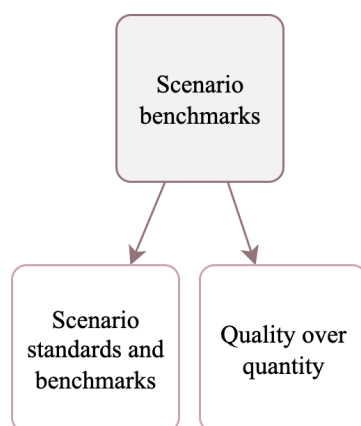


Figure 5. Themes arising during the interviews under the parent theme “scenario benchmarks” and their construction. All themes and codes under it are related to how to qualitatively and quantitatively evaluate the fidelity gap, addressing RQ4.

addressed more accurately. For instance, by calculating the possible output range of the vehicle's longitudinal acceleration value using statistical techniques (e.g., bootstrapping and Monte Carlo simulation), the simulation output for the same scenario can be assessed more accurately and the reasons for discrepancies (if any) better understood.

5.2. Threats to Validity

While focusing on specific toolchain components or AD/ADAS features could have unveiled technical particularities, we could have also introduced bias by narrowing the perspective. Thus, we prioritized a holistic understanding over tool-specific conclusions and interview questions were designed to capture broad practitioner insights on improving simulation fidelity.

It could be argued that having a larger group of participants might enable new perspectives to emerge, however Crouch and McKenzie [53] showed that already smaller sized sets can be sufficient for interview-based qualitative research. To further ensure saturation, we complemented the study with a focus group and contrasted those insights with the interviews. Convenience sampling to find said experts still poses a risk to validity, as selected participants may not fully represent the broader population of experts. This risk is mitigated by cross-referencing our findings with recent literature and assessing their alignment with the focus group study.

Regarding the analysis, the coders worked collaboratively on theme extraction and discussed the results until consensus was reached.

6. CONCLUSION

V&V activities for software and system testing supported by simulations allows for a controlled, replicable and safe test environment. We conducted in-depth, semi-structured interviews with a broad range of domain experts to extract thematic clusters to identify factors influencing the fidelity of simulations in the context of V&V for ADAS and AD. We contrasted their views with recent findings from scientific literature to provide a broad and deep understanding of expectations, limitations and challenges. We complemented these activities with a focus group that represented a typical value chain during the development of ADAS systems to validate the insights from the initial interviews that we contrasted with literature.

We see a clear need for systematic approaches to evaluate the fidelity gap qualitatively *and* quantitatively using agreed KPIs. However, our focus group unveiled that reaching consensus is challenging. Strong consensus, though, was identified on the need for increasing transparency in sharing data, models, algorithms and

scenario pools, while considering confidentiality. The necessity of improving standardization is also clearly highlighted.

To have realistic simulations is a common goal. However, balancing expectations about the level of realism with practicality and feasibility is needed, which is referred to as the realism-feasibility trade-off. Hence, striving for transparent and quantitative KPIs to assess the simulation success and fidelity in the context of AD is an agreed challenge.

Conflict of Interest

To the best of the authors' knowledge there are no financial and personal relationships with other people or organizations that could inappropriately influence or bias the present work. The authors therefore declare that they have no conflicts of interest.

Data Availability

In the case of this study, data cannot be made available given the agreement reached with the interviewed practitioners to protect them and their respective companies. The text however describes and aggregates the information they provided in an anonymized way.

Funding

This work was supported by Sweden's innovation agency, Vinnova, under Grant No. 2021-05043 entitled "Enabling Virtual Validation and Verification for ADAS and AD Features (EVIDENT)".

Authors' Contribution

The first four authors conducted the interviews, as described in this article. All five authors contributed equally to all other research tasks in the study leading to the current report.

Disclaimer

The authors' views and opinions do not necessarily reflect Volvo Cars' official policy or position.

REFERENCES

- [1] World Health Organization (WHO). Global Status Report on Road Safety 2023. World Health Organization, Technical Report, 2023. URL: <https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023>.

- [2] SAE International Recommended Practice. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE Standard J3016_202104. Warrendale (PA), United States: SAE International, 2021 (revised April 2021).
- [3] R. Donà, B. Ciuffo. Virtual Testing of Automated Driving Systems. A Survey on Validation Methods. *IEEE Access*, 2022, 10: 24349–24367.
- [4] S. Tang, Z. Zhang, Y. Zhang, J. Zhou, Y. Guo, S. Liu, et al. A Survey on Automated Driving System Testing: Landscapes and Trends. *ACM Transactions on Software Engineering and Methodology*, 2023, 32(5): 1–62.
- [5] X. Hu, S. Li, T. Huang, B. Tang, R. Huai, L. Chen. How Simulation Helps Autonomous Driving: A Survey of Sim2real, Digital Twins, and Parallel Intelligence. *arXiv Preprint. arXiv:2305.01263*, 2023.
- [6] S. Jesenski, J.E. Stellet, W. Branz, J.M. Zöllner. Simulation-Based Methods for Validation of Automated Driving: A Model-Based Analysis and an Overview about Methods for Implementation. *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand: IEEE, 2019, pp. 1914–1921.
- [7] C. Goodin, D.W. Carruth, L. Dabbiru, M. Hedrick, Z.S. Aspin, J.T. Carrillo, et al. Fidelity Requirements for Simulating Sensor Performance in Autonomous Ground Vehicles. *Proceedings of the SPIE Conference on Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications*. Orlando (FL), United States: SPIE, 2023, pp. 78–85.
- [8] S. Shah, D. Dey, C. Lovett, A. Kapoor. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. In: M. Hutter, R. Siegwart (Eds.), *Field and Service Robotics*. Springer Proceedings in Advanced Robotics, Vol. 5. Cham, Switzerland: Springer, 2018, pp. 621–635.
- [9] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, V. Koltun. CARLA: An Open Urban Driving Simulator. *Proceedings of the 1st Conference on Robot Learning (CoRL)*. Mountain View, California, United States: PMLR, 2017, pp. 1–16.
- [10] NVIDIA. NVIDIA Announces Digital Twin Platform for Scientific Computing. *NVIDIA Newsroom*, 2022. URL: <https://nvidianews.nvidia.com/news/nvidia-announces-digital-twin-platform-for-scientific-computing>.
- [11] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, M. Maurer. Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving. *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*. ITSC. Gran Canaria, Spain: IEEE, 2015, pp. 982–988.
- [12] W. van Toll, F. Grzeskowiak, A.L. Gandía, J. Amirian, F. Berton, J. Bruneau, et al. Generalized Microscopic Crowd Simulation using Costs in Velocity Space. *Proceedings of the ACM Symposium on Interactive 3D Graphics and Games (I3D)*. New York (NY), United States: ACM, 2020, pp. 1–9.
- [13] UNECE. UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS). Geneva, Switzerland: United Nations Economic Commission for Europe, 2021. URL: <https://unece.org/transport/documents/2021/03/standards/un-regulation-no-157-automated-lane-keeping-systems-alks>.
- [14] SOAFEE. Scalable Open Architecture for Embedded Edge. SOAFEE, 2024. URL: <https://www.soafee.io/>.
- [15] I. Raghupatruni, S. Karjee, A. Gupta, V. Naik, T. Huber. Towards Establishing Continuous-X Pipeline Using Modular Software-in-the-Loop Test Environments. *SAE Technical Paper*, 2021.
- [16] EUR-Lex - 52021PC0206 - EN - EUR-Lex. Brussels: European Commission, 2021. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.
- [17] C. Johnson, E. Graupe, M. Kassel. A Literature Review of Simulation Fidelity for Autonomous-Vehicle Research and Development. *SAE International Journal of Aerospace*, 2023, 16(3): 253–261.
- [18] C. Li, C.-H. Cheng, T. Sun, Y. Chen, R. Yan. ComOpT: Combination and Optimization for Testing Autonomous Driving Systems. *Proceedings of the 2022 International Conference on Robotics and Automation (ICRA)*. Philadelphia (PA), United States: IEEE, 2022, pp. 7738–7744.
- [19] Euro NCAP. How to Read the Stars. Euro NCAP, 2024. URL: <https://www.euroncap.com/en/about-euro-ncap/>.
- [20] ISO 26262:2018. Road Vehicles. Functional Safety - Guidelines on ISO 26262. Geneva, Switzerland: International Organization for Standardization (ISO), 2018.
- [21] B. Schlager, S. Muckenhuber, S. Schmidt, H. Holzer, R. Rott, F.M. Maier, et al. State-of-the-Art Sensor Models for Virtual Testing of Advanced Driver Assistance Systems/Autonomous Driving Functions. *SAE International Journal of Connected and Automated Vehicles*, 2020, 3(3): 233–261.
- [22] Z.F. Magosi, C. Wellershaus, V.R. Tihanyi, P. Luley, A. Eichberger. Evaluation Methodology for Physical Radar Perception Sensor Models Based on On-Road Measurements for the Testing and Validation of Automated Driving. *Energies*, 2022, 15(7): 2545.
- [23] P.E. Strandberg. Ethical Interviews in Software Engineering. *Proceedings of the 2019 ACM/IEEE*

- International Symposium on Empirical Software Engineering and Measurement (ESEM). Porto de Galinhas, Brazil: IEEE, 2019, pp. 1–11.
- [24] I. Etikan, S.A. Musa, R.S. Alkassim. Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 2015, 5(1): 1–4.
- [25] L.S. Nowell, J.M. Norris, D.E. White, N.J. Moules. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 2017, 16(1): 1–13.
- [26] G. Guest, K.M. MacQueen, E.E. Namey. *Applied Thematic Analysis*. Thousand Oaks (CA), United States: Sage Publications, 2012.
- [27] J. Saldaña. *The Coding Manual for Qualitative Researchers*. Thousand Oaks (CA), United States: Sage Publications Ltd, 2021.
- [28] V. Braun, V. Clarke. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 2006, 3(2): 77–101.
- [29] B. Cabrero-Daniel, M. Mohamad. Challenges of Virtual Validation and Verification for Automotive Functions. In: D. Taibi, D. Smite (Eds.), *Software Engineering and Advanced Applications*. SEAA 2025. Lecture Notes in Computer Science, Vol. 16081. Cham, Switzerland: Springer, 2026, pp. 183–200.
- [30] M. Jasiński. A Generic Validation Scheme for Real-Time Capable Automotive Radar Sensor Models Integrated into an Autonomous Driving Simulator. *Proceedings of the 2019 24th International Conference on Methods and Models in Automation and Robotics (MMAR)*. Miedzyzdroje, Poland: IEEE, 2019, pp. 612–617.
- [31] B. Cabrero-Daniel, A.Y. Abdelkarim, A. Broberg. Digital Twins for Early Verification and Validation of Autonomous Driving Features: Open-source Tools and Standard Formats. *Proceedings of the 2024 IEEE Intelligent Vehicles Symposium (IV)*. Jeju Island, South Korea: IEEE, 2024, pp. 2477–2482.
- [32] B. Yu, C. Chen, J. Tang, S. Liu, J.-L. Gaudiot. Autonomous Vehicles Digital Twin: A Practical Paradigm for Autonomous Driving System Development. *Computer*, 2022, 55(9): 26–34.
- [33] H. Xiong, Z. Wang, G. Wu, Y. Pan. Design and Implementation of Digital Twin-Assisted Simulation Method for Autonomous Vehicle in Car-Following Scenario. *Journal of Sensors*, 2022, 4879490: 1–12.
- [34] A. Niaz, M.U. Shoukat, Y. Jia, S. Khan, F. Niaz, M.U. Raza. Autonomous Driving Test Method Based on Digital Twin: A Survey. *Proceedings of the 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*. Quetta, Pakistan: IEEE, 2021, pp. 1–7.
- [35] G. Hinton, O. Vinyals, J. Dean. Distilling the Knowledge in a Neural Network. *arXiv Preprint*. arXiv:1503.02531, 2015.
- [36] M.R.U. Saputra, P.P. De Gusmao, Y. Almalioglu, A. Markham, N. Trigoni. Distilling Knowledge from a Deep Pose Regressor Network. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, South Korea: IEEE, 2019, pp. 263–272.
- [37] L. Zhang, R. Dong, H.-S. Tai, K. Ma. PointDistiller: Structured Knowledge Distillation Towards Efficient and Compact 3D Detection. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver (BC), Canada: IEEE, 2023, pp. 21791–21801.
- [38] Association for Standardization of Automation and Measuring Systems. *ASAM OpenSCENARIO Version 2.0*. ASAM, 2024. URL: <https://www.asam.net/standards/detail/openscenario/v200/>.
- [39] K.-R. Hase. “Open Proof” for Railway Safety Software - A Potential Way-Out of Vendor Lock-in Advancing to Standardization, Transparency, and Software Security. In: E. Schnieder, G. Tarnai (Eds.), *FORMS/FORMAT 2010*. Berlin, Heidelberg: Springer, 2011, pp. 5–38.
- [40] A. Widner, V. Tihanyi, T. Tettamanti. Framework for Vehicle Dynamics Model Validation. *IEEE Access*, 2022, 10: 35422–35436.
- [41] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vancouver (BC), Canada: IEEE, 2017, pp. 23–30.
- [42] G.D. Kontes, D.D. Scherer, T. Nisslbeck, J. Fischer, C. Mutschler. High-Speed Collision Avoidance using Deep Reinforcement Learning and Domain Randomization for Autonomous Vehicles. *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. Rhodes, Greece: IEEE, 2020, pp. 1–8.
- [43] S. Pouyanfar, M. Saleem, N. George, S.-C. Chen. ROADS: Randomization for Obstacle Avoidance and Driving in Simulation. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Long Beach (CA), United States: IEEE, 2019, pp. 1267–1276.
- [44] R. Maier, L. Grabinger, D. Urlhart, J. Mottok. Causal Models to Support Scenario-Based Testing of ADAS. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(2): 1815–1831.
- [45] B. Neurohr, T. Koopmann, E. Möhlmann, M. Fränzle. Determining the Validity of Simulation

- Models for the Verification of Automated Driving Systems. *IEEE Access*, 2023, 11: 102949–102960.
- [46] C. Sippl, F. Bock, C. Lauer, A. Heinz, T. Neumayer, R. German. Scenario-Based Systems Engineering: An Approach Towards Automated Driving Function Development. *Proceedings of the 2019 IEEE International Systems Conference (SysCon)*. Orlando (FL), United States: IEEE, 2019, pp. 1–8.
- [47] P. Koopman, M. Wagner. Toward a Framework for Highly Automated Vehicle Safety Validation. *SAE Technical Paper*, 2018.
- [48] D.J. Fremont, E. Kim, Y.V. Pant, S.A. Seshia, A. Acharya, X. Brusio, et al. Formal Scenario-Based Testing of Autonomous Vehicles: From Simulation to the Real World. *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. Rhodes, Greece: IEEE, 2020, pp. 1–8.
- [49] F. Reway, A. Hoffmann, D. Wachtel, W. Huber, A. Knoll, E. Ribeiro. Test Method for Measuring the Simulation-to-Reality Gap of Camera-Based Object Detection Algorithms for Autonomous Driving. *Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV)*. Las Vegas (NV), United States: IEEE, 2020, pp. 1249–1256.
- [50] R. Xu, H. Xiang, X. Han, X. Xia, Z. Meng, C.-J. Chen, et al. The OpenCDA Open-Source Ecosystem for Cooperative Driving Automation Research. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(4): 2698–2711.
- [51] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick, F. Diermeyer. Survey on Scenario-Based Safety Assessment of Automated Vehicles. *IEEE Access*, 2020, 8: 87456–87477.
- [52] A. Ngo, M.P. Bauer, M. Resch. A Multi-Layered Approach for Measuring the Simulation-to-Reality Gap of Radar Perception for Autonomous Driving. *Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. Indianapolis (IN), United States: IEEE, 2021, pp. 4008–4014.
- [53] M. Crouch, H. McKenzie. The Logic of Small Samples in Interview-Based Qualitative Research. *Social Science Information*, 2006, 45(4): 483–499.