Comparing Approaches for Prioritizing and Selecting Scenarios in Simulation-based Safety Testing of Automated Driving Systems

Fauzia Khan
Institute of Computer Science
University of Tartu
Tartu, Estonia
fauzia.khan@ut.ee

Hina Anwar
Institute of Computer Science
University of Tartu
Tartu, Estonia
hina.anwar@ut.ee

Dietmar Pfahl
Institute of Computer Science
University of Tartu
Tartu, Estonia
dietmar.pfahl@ut.ee

***Abstract*—Simulation-based safety testing provides a cost-effective method for testing Automated Driving Systems (ADS) across diverse scenarios. However, prioritizing or selecting test scenarios for simulation-based safety testing remains challenging due to the infinite variety of scenarios that ADS may encounter. In this study, we conducted a literature review to identify approaches for selecting or prioritizing scenarios for ADS safety testing. We compare the six identified approaches in a tabular form across various aspects. We discuss one approach in detail, illustrating how it could complement the other selected approaches through an example. Our ongoing work involves extending the comparative analysis to cover all approaches comprehensively.**

Keywords - Automated Driving System (ADS); Simulation-based Testing; Safety Testing, Scenario Selection; Scenario Prioritization.

I. INTRODUCTION

Automated Driving Systems (ADS) represent a much-advanced technology that has the potential to revolutionize the automotive and transportation industry [1]–[3]. An ADS is a safety-critical system that can perform Dynamic Driving Tasks (DDTs) without human driver assistance [4], [5]. An ADS is loaded with advanced technologies that can independently control the vehicle and make decisions regarding steering, acceleration, deceleration, braking, maneuver planning, and object detection [6], [7]. These control decisions are operational and depend upon the level of automation[1]. As automation levels increase, human driver involvement in driving decreases, with complete ADS control achieved at level 5.

Despite many advancements in ADS, thorough testing and validation remain critical for ensuring their safety [9], [10]. Moreover, in the context of safety, even a single failure made by one manufacturer affects its reputation and undermines public trust in the entire automotive industry, diminishing confidence in ADS. Thus, rigorous testing is required to build and maintain trust in ADS.

Generally, there are two common methodologies for testing ADS: real-world testing [11], [12] and simulation-based testing [13]–[15]. In real-world testing, the ADS would have to be driven hundreds of billions of miles to prove that it is safer than a typical human-driven vehicle (HDV) [16], which is costly and risky. In contrast, simulation-based testing is a cost-effective way to assess ADS's safety in various scenarios. [17]–[19]. A scenario refers to a specific situation the ADS under test (typically called an 'ego vehicle') might encounter during its operation. A scenario is defined through various components, including the ego vehicle itself, the static environment (such as roads, buildings, and infrastructure), the dynamic environment (such as other vehicles, pedestrians, and objects), and specific conditions (such as weather light, etc.) [20], [21]. These scenario components have specific elements and sub-elements [22]. For example, in the scenario, "An ego vehicle changes lanes on a highway during the night in rainy weather", the sub-elements associated with the dynamic environment are the "ego vehicle" and "changes lanes". "Highway" is the sub-element associated with the static environment while "rainy" and "night" are sub-elements associated with conditions.

An ADS could encounter infinite scenarios depending upon various factors, such as dynamic objects (pedestrians, vehicles, animals, etc.), static objects (traffic lights, sign boards, etc.), road typologies, weather, and lighting conditions. Combining these factors leads to infinite scenarios, and testing every scenario is not feasible, even in a simulator.

[1] The Society of Automotive Engineers defines six levels of automation for on-road automated vehicles. These levels are described in SAE safety standard J3016 [8]

In this study, we aim to identify and compare the approaches defined in the literature for prioritization or selection of test scenarios to see how these approaches could help select test scenarios for simulation-based safety testing of ADS. This comparison could benefit both the automotive industry and academic research. The contributions of our study are:

- A literature review to identify the studies related to the selection or prioritization of test scenarios for simulation-based safety testing of ADS.
- A comparison of selected approaches w.r.t the purpose and scope of testing, input data type, input data content type, datasets used, number of steps involved, scenario components, supported simulator, tool support, open source, and limitations.
- A preliminary comparison via an illustrative example of one of the selected approaches (SSTSS) could complement other selected approaches. Detailed comparisons among the selected approaches are a work in progress.

The rest of this paper is organized as follows. Section II presents the related work. Section III presents the methodology. Section IV presents the results of the comparative analysis. Section V presents the discussion, and the conclusion and future work are given in Section VI.

II. RELATED WORK

To the best of our knowledge, no similar study exists in the literature whose goal is to compare the approaches of scenario selection or prioritization for safety testing of ADS. However, in this section, we present related secondary studies focused on the testing of ADS. We also highlight how the automotive industry tests ADS safety and compares it to human drivers as a benchmark.

A. Simulation-based Safety Testing of ADS

Tang et al. [23] presented a survey focused on module and system-level testing of ADS. They analyzed how different modules are affected by various technical factors and highlighted issues during ADS's development or deployment. Khan et al. [5] conducted a systematic literature review to identify safety features, testing methods, tools, and datasets used for the safety testing of ADS. Lou et al. [24] presented a comprehensive study to understand current ADS testing practices and needs by conducting interviews with companies and developers. They further analyzed the gap between ADS research and practitioners' and suggested future directions, such as developing test reduction techniques to accelerate simulation-based testing of ADS. [25] presented a survey focusing on the algorithms and tools used to generate critical scenarios in automated driving. They further identified challenges in existing approaches regarding safety-critical scenario generation approaches.

B. Industry Practices for ADS Safety Testing

In the automotive industry, simulation-based testing of ADS is widely used as an alternative to real-world testing [24]. In a recent study[2], Waymo compared the simulated performance of its autonomous Waymo driver to human drivers involved in fatal crashes in Chandler, Arizona, between 2008-2017. They modeled and simulated a non-impaired human driver (NIEON). The Waymo driver consistently did better than this high benchmark by avoiding 100% of crashes, except when struck from behind. Additionally, in collision simulations, the Waymo driver surpassed the NIEON model, avoiding 75% of crashes and reducing severe injury risk by 93%.

Cruise[3] also evaluated the safety performance of ADS by comparing their collisions with those of human drivers. However, assessing human driving performance remains challenging due to limited data. Particularly, there's a lack of information regarding significant collisions or violations despite the abundance of data on AV safety collected through advanced sensors and data logging. Our study aims to identify and compare scenario selection or prioritization approaches for simulation-based safety testing of ADS.

III. METHODOLOGY

In this section, we present our methodology to identify studies related to scenario selection or prioritization approaches for simulation-based safety testing of ADS and their comparison. We formulate our research question and conduct a literature review to find relevant studies. We explain the formulation of the search string, the execution of the search query, and the exclusion/inclusion criteria. These steps were guided by K. Petersen et al. [26]. Finally, we present the criteria for comparing the approaches developed for test scenario selection or prioritization for simulation-based safety testing of ADS.

A. Research Questions

The goals of the study are to identify relevant literature on scenario selection or prioritization for simulation-based safety testing of ADS and to compare these approaches to identify strengths and weaknesses. To end this, we formulated two research questions (RQs) as follows:

RQ1: *What scenario selection or prioritization approaches exist in simulation-based safety testing?*

RQ2: *How do the identified approaches compare to one another?*

B. Search Query

We searched the online repositories (IEEE, ACM, Springer, Science Direct) to find publications to answer the research questions. To maximize the coverage of the literature, we selected search terms from the research questions. We used the keyword "AND" for the concatenation of search terms and the keyword "OR" for combining synonyms of the search terms to

[2] https://waymo.com/blog/2022/09/benchmarking-av-safety

[3] https://getcruise.com/news/blog/2023/cruises-safety-record-over-one-million-driverless-miles/

maximize coverage. Since we are only interested in the prioritization or selection of test scenarios for simulation-based safety testing of ADS, therefore we used search terms such as "Test scenario selection", "Test reduction", and "Test scenario prioritization". We also used the search terms "Autonomous Driving Systems" and "Self-driving cars" as synonyms to ensure that the search query produces results related to testing scenarios for ADS. We did not explicitly use the term "On-road testing" or "Test scenario generation" with the keyword "NOT" to filter the search query results because we noticed that it would filter out many relevant publications. Search terms related to our research questions, e.g., "Simulation", were also not added to the search query because they make the research query too specific and might not capture relevant publications. We also used "*" with some search terms to capture all variations of search terms. We show the general search query below[4]:

((Test scenario select*) OR (Test reduction) OR (Test scenario prioritize*)) AND ((Autonomous Driving Systems) OR (Self-driving car*))

C. Additional Filters

We entered the search query into the advanced search of the online databases and applied the filters mentioned in Table I. We consider primary studies published in journals and conferences from the year 2000 onwards.

TABLE I: FILTERS APPLIED TO SEARCH QUERY RESULTS

Filter	Value
Year	2000-2024
Content Type	Journal Articles and Conference Papers
Language	English
Discipline	Computer Science, Engineering

D. Exclusion/Inclusion Criteria

To select the publications focused on the selection or prioritization of test scenarios for simulation-based safety testing of ADS, we performed the following steps: (i) We read the titles, abstracts, and keywords to analyze the relevance of the publications. If the decision is not made, we read the introduction, methodology, results, and conclusions and download the relevant ones. (ii) We made an Excel sheet of the downloaded papers and removed the duplicates. If a duplicate is found, we include the most recent and comprehensive versions of duplicated papers. (iii) We included (I1) journal articles or conference papers only whose primary goal is to select, prioritize or reduce test scenarios for simulation-based safety testing of ADS. We excluded all publications meeting at least one of the following criteria: (E1) the publication is written in a language other than English, (E2) the publication is not accessible or unavailable, (E3) the publication lacks sufficient details such as methods, contributions, etc., (E4) the publication falls into categories such as magazine, doctoral symposium paper, thesis, secondary studies position paper, keynote presentation, or abstract.

E. Data Extraction

RQ1 can be answered based on the final set of selected studies because each study will describe at least one scenario selection or prioritization approach.

Related to RQ2, we need to extract data items from the selected studies that inform the comparison of existing approaches. The following extracted data items were used as criteria for the comparison:

1) Purpose of Testing: We compare the selected approaches based on their testing purpose as follows: (i) compare ego car vs human-driven car accident, (ii) compare ego car vs human-driven car without accidents, (iii) explore ego car behavior for testing new/improved autonomy software features (for developers), (iv) expose ego car behavior to the scenario typically used in driver tests, (v) expose simulate ego car behavior to provide feedback for the real-world test plan. ***2) Scope of Testing:*** We compare the scope of testing based on whether the selected testing approach is specific to test (i) full ADS, (ii) specific ADS feature or its subset. ***3) Input Data Type of Approach:*** We also extract the information regarding the initial input(s) data type of each selected approach, which includes: (i) text, (ii) video, (iii) image, (iv) audio, (v) combination of any of above. ***4) Input Data Content of Approach:*** We further compare the content of input data used in each selected approach, which includes: (i) human-driven car accident data, (ii) human-driven car data without accidents, (iii) simulator recorded ADS driving, (iv) ADS from real-world recording. ***5) Number of Steps:*** To determine the ease of application of each selected approach, we compare the number of steps involved in each method ***6) Scenario Component:*** We extract the information regarding the scenario components considered as a proof of concept in selected approaches. (i) Static Environment, (ii) Dynamic Environment, (iii) Weather, (iv) Light conditions. ***7) Supported Simulator:*** Considering that each simulator has its unique limitations and capabilities, we extract the information about which simulator each approach supports. ***8) Tool Support:*** We compare the maturity of each approach based on tool availability. ***9) Open Source:*** We also check whether the selected approach is open source, making it easier for researchers and practitioners to access, utilize, and replicate the approach. ***10) Limitation:*** Finally, we present each selected approach's possible limitation(s) for ADS simulation-based safety testing.

To extract data items used in selected approaches, we read each selected publication's abstract, introduction, methodology, and results sections and organized the data in a tabular format. Each table row recorded the data for predefined criteria, as discussed above. During data extraction, conflicts (if any) are discussed and resolved among the authors. We identified the key terms and their synonyms. We removed duplicate terms and

[4] The query was entered in databases on 22 April 2024

applied the bottom-up merging method to merge key terms to get high-level categories.

IV. RESULTS

In this section, we present the results of each RQ. For RQ1, the results of the selection procedure, selected publications, and their approaches are presented. For RQ2, we present the comparison of selected approaches based on predefined criteria (cf. Section III-E).

RQ:1 *What scenario selection or prioritization approaches exist in simulation-based safety testing?*

A. Results of the Selection Procedure

The number of publications after each selection step is given in Table II. We found 1438 publications identified in the initial search at step 1. In step 2, we downloaded 31 publications and found one duplicate in step 3. In step 4, we selected only six publications after applying inclusion and exclusion criteria. The results were recorded in an Excel sheet available online[5]. Table III shows the count of conference papers and journal articles selected from each online repository.

TABLE II: NUMBER OF PUBLICATIONS FOR EACH STEP IN THE SELECTION PROCEDURE

Screening Steps	No. of Publications
Initial Search	1438
Downloaded Publications	31
Duplicate Removal	1
Included Publications	6

TABLE III: DISTRIBUTION OF SELECTED PUBLICATIONS PER REPOSITORY

Database	Conference Paper	Journal Article	Total Publications
IEEE	2	0	2
ACM	1	1	2
Springer	2	0	2
Science Direct	0	0	0

B. Selected Publications and their Approaches

Table IV shows the selected publication and their approaches for selecting or prioritizing scenarios for simulation-based testing for safety testing of ADS. We presented an overview of each approach shown in Table IV.

1) **SSTSS**: stands for (simulation-based safety testing scenario selection). This approach uses a human driver car crash dataset to prioritize and select test scenarios through the following eight steps: *(i) Scenario catalog selection:* Initially, a publicly available scenario catalog that is comprehensive and published by a reputable organization is chosen such as National Highway Traffic Safety Administration - (NHTSA)[6], Euro NCAP[7]. *(ii) Enumerating ODD of ADS:* enumerate the ODD[8] of Vehicle Under Test (VHT) in terms of spatial, temporal, and environmental conditions[9]. *(iii) Filtering scenarios based on ODD of ADS:* exclude scenarios that do not fall within the ODD enumerated in the previous step. (iv) *Scenario grouping:* categorize the scenarios into groups based on similar critical actions of the ego vehicle or the target object, such as pedestrians, cyclists, animals, etc. (v) *Removing duplicates within a scenario group:* identify the duplicate scenarios within scenario groups. *(vi) Prioritizing scenario groups:* prioritize the scenario based on common crash scenario statistics and assign the highest priority to scenario groups where more accidents occur. *(vii) Filtering scenarios based on simulator limitations:* choose a simulator that can replicate the real world closely and analyze data on performance metrics such as travel time, crashes, etc. After simulator selection, exclude the scenarios that cannot be implemented due to simulator limitations. *(viii) Prioritizing and selecting scenarios for testing ADS* - assign a score based on the scoring technique to each scenario within the scenario group based on the availability of car crash data statistics for scenario each element, such as actors, weather, etc. The final output of the SSTSS process is a ranked list of prioritized scenario groups in descending order of priority. Testing starts with the scenarios from the top-prioritized scenario group in sequence. After all scenarios within the top-prioritized scenario group are tested, the process proceeds to subsequent prioritized scenario groups.

TABLE IV: SELECTED PUBLICATIONS OVERVIEW

Sr.#	Publication	Approach	Online Repository
1	[27]	SSTSS	Springer
2	[28]	SDC-Scissor	IEEE
3	[29]	SDC-Prioritizer	ACM
4	[30]	STRaP	ACM
5	[31]	SPECTRE	Springer
6	[32]	J. Bach et al.	IEEE

2) **SDC-Scissor:** stands for self driving cars-cost-effective test selector. This approach [28] utilizes machine learning (ML) models to select scenarios based on five components, which are as follows: *(i) SDC-Test Generator:* generates random test using Frenetic and AsFault tool [36], [37]. *(ii) SDC-Test Executor:* executes the generated simulation-based tests and records the resultant output to categorize tests as safe or unsafe. *(iii) SDC-Features Extractor:* extracts various road features (distance, turns, angles, etc.) and road statistics from driving scenarios. *(iv) SDC-Benchmarker:* uses these features and corresponding labels to train the ML models and determine the most effective model for predicting test outcomes.

5 https://github.com/ScenarioSelectionApproaches

6 https://rosap.ntl.bts.gov/view/dot/6281/dot_6281_DS1.pdf?

7 Euro NCAP

8 The specific conditions and environments under which a particular driving automation system is designed to operate [33]

9 ODD covers spatial (geography, road types, lanes, speed limits, etc.), temporal (day/night), and environmental conditions (weather). [8], [34], [35]

(v) SDC-Predictor: identifies simulation-based test scenarios that are unlikely to detect faults in the ego vehicle and excludes them from execution.

3) **SDC-Prioritizer**: This approach [29] is similar to SDC-Scissor; however, it uses single-objective and multi-objective genetic algorithms to prioritize test scenarios. The steps are as follows: (i) Initially, a random test scenario is generated using the tool AsFault [37]. Each generated test scenario has information (JSON file) on the start and destination points for the ego car on the map, the entire road network, and the chosen driving path. (ii) Two sets of road features are extracted. i.e., road characteristics (road length and direct distance between the start and destination points, number of turns, angles) and the road segment statistics to identify safe and unsafe scenarios even before executing them. (iii) The extracted features serve as inputs to two black-box scenario prioritization approaches, i.e., SO-SDC-Prioritizer and MO-SDC-Prioritizer, which utilize single and multi-objective GA to prioritize test cases.

4) **STRaP:** stands for Scenario-based Test Reduction and Prioritization. This approach [30] is developed to reduce and prioritize the testing scenarios using data recordings of previous versions of ADS under test. The steps are as follows: (i) the semantic information from each frame of ADS video recordings is extracted via the ADS's communication channels under test and converted into vectors. The semantic information includes static (e.g., traffic lights, crosswalks, etc.) and dynamic objects (e.g., vehicles, cyclists, actions of vehicles, etc.). (ii) The driving recordings are divided into continuous and redundant segments to cut length. (iii) The remaining segments are prioritized based on their coverage of driving scene elements and the rarity of elements such as traffic lights, pedestrians, etc.

5) **SPECTRE:** stands for selection and prioritization of test scenarios for autonomous driving systems. This multi-objective search-based approach [31] was developed to minimize testing costs for newer versions of ADS by utilizing historical test data from previous versions. The steps are as follows: (i) Initially, each scenario is characterized by a set of properties of the ADS under test, such as speed and environmental factors such as weather and obstacles. (ii) Each scenario is executed in simulations, which yields four key output values, resulting in four objective functions: collision occurrence, collision probability, ADS demand, and scenario diversity. (iii) Four optimization objectives are defined based on the results obtained in the previous step. (iv) Finally, multi-objective evolutionary algorithms are used as optimization techniques to prioritize driving scenarios.

6) **J. Bach et al.:** In this study [32], a two-step selection concept was introduced for selecting scenarios for the safety testing of ADS. The steps are as follows: (i) Initially, scenarios are categorized into abstract groups based on specified system-level requirements. This includes segmentation by geolocation (e.g., different countries with varying traffic rules) or road category (e.g., motorways, rural roads, urban streets), as well as criteria such as the ego vehicle's state and surrounding traffic density. A classification-tree approach is employed to systematically preselect scenarios, ensuring coverage of relevant use cases for the targeted regions. (ii) Repetitive information and situations are removed by looking at the two-dimensional histograms of frames. These histograms are visual sources for how scenarios are spread out and connected. Only scenarios that fill empty areas in the histograms are kept, while others are removed.

RQ2: *How do the identified approaches compare to one another?*

To answer RQ2, we compare the selected approaches w.r.t the purpose and scope of testing, input data type, input data content type, datasets used, number of steps involved, scenario components, supported simulator, tool support, open source, and limitations. Table V compares six selected approaches per the criteria defined in Section III-E. All the selected approaches prioritize test scenarios to reduce the number of testing scenarios in simulation-based testing. However, the testing purpose and the scope of the selected approaches differ. Therefore, it would be interesting to discuss if these approaches could complement one another (by combining testing from different perspectives) for a more comprehensive, effective, and efficient testing of ADS.

In this study, we present the preliminary results of the comparison. Using an illustrative example, we demonstrate how SSTSS could complement or be combined with selected approaches. Detailed comparisons among selected approaches are a work in progress.

Comparison of the SSTSS process (via illustrative example): The selected five approaches differ from SSTSS as they are based on simulation data. The SSTSS process would complement simulation-based approaches in different ways.

(i) The SSTSS process could be used as the first step for the pre-selection of scenarios before employing other approaches. Since the SSTSS process prioritizes statistically significant scenarios using real-world car crash statistics, it could be used to optimize the initial selection of scenarios and driving routes. The STRaP and SPECTRE approaches select testing scenarios from available driving routes within the simulator. However, solely relying on a specific driving loop may overlook various real-world situations, leading to potential gaps in scenario coverage. Furthermore, there could be scenarios of lower significance in a specific driving loop, and allocating resources to test them might not be cost-effective. Once the driving route containing prioritized testing scenarios is selected using the SSTSS process, STRaP or SPECTRE approaches could be used to reduce the test scenarios further. Using such a hybrid approach could reduce the number of test scenarios, making the testing more efficient, cost and resource-effective.

Let's consider an example. Suppose we want to test a BOLT car (ego vehicle) from the Autonomous Driving Lab at the University of Tartu, Estonia. The BOLT car has the following ODD: It can follow the traffic flow, detect pedestrians, and give way to them. Given the ODD, we apply the SSTSS process to select prioritized scenarios for simulation-based testing from a list of 111 scenarios. Given the sorted list of scenarios generated from the SSTSS process (see a sample output in Table VI), a specific driving route could be selected. In the sample output shown in Table VI, the top three prioritized scenario groups are Following Lead Vehicle [10], Crossing Path [11], and Pedestrian Interaction[12]. As a result, the "Tartu loop" could be selected, which includes seven regulated pedestrian tracks, four unregulated pedestrian tracks, two regulated intersections, one right turn, and one roundabout, and multi-lane roads where the ego vehicle can closely follow another vehicle and encounter pedestrians. Furthermore, it includes intersections to encounter crossing paths scenarios. For the selected "Tartu loop", a digital twin is available online[13], which can be used as input for STRaP. The STRaP extracts the semantic information from video recordings of the selected "Tartu loop" to reduce the number of testing scenarios. Then, the driving recordings are segmented to

TABLE V: COMPARISON OF APPROACHES

Sr.#	Features	SSTSS	SDC-Scissor	SDC-Prioritizer	STRaP	SPECTRE	J. Bach et al.
01	Purpose of Testing	compare ego car vs human-driven car accidents	explores ego car behavior to test new/improved autonomy software features	explore ego car behavior to test new/improved autonomy software features	explore ego car behavior to test new/improved autonomy software features	explore ego car behavior to test new/improved autonomy software features	expose simulate ego car behavior to provide feedback for the real-world test plan
02	Scope of Testing	Full ADS	Specific Feature (LDW)	Specific Feature (LDW)	Full ADS	Full ADS	Specific Featur e (PCC)
03	Input DT	Text	Video	Video	Video	Video	Video
04	Input Data Content of Approach	Human-driven car accident data	Simulator recorded ADS driving	Simulator recorded ADS driving	Simulator recorded ADS driving	Simulator recorded ADS driving	Read-world and Simulator recorded ADS driving
05	# of steps	8	5	3	3	4	3
06	Scenario Components	Static & Dynamic Env., Weather & Light condition.	Static Environment (only Roads)	Static Environment (only Roads)	Static & Dynamic Env., Weather	Static & Dynamic Env., Weather & Light conditions	Static Environment (Roads and Location)
07	Sup. Simulator	Carla	BeamNG	BeamNG	LGSVL	LGSVL	-
08	Tool Sup.	No	Yes	No	Yes	No	No
09	OS	Yes	Yes	Yes	Yes	Yes	No
10	Limitations	-Solely relies on the availability of human-driver accident dataset, overlooking ADS data -Requires Manual effort to execute each step.	-Solely relies on simulation driving data for scenario selection, thus potentially missing critical scenarios that arise from real-world human driving behavior -Handles limited driving scenarios, e.g., road shapes only	-Solely relies on simulation driving data for scenario selection, thus potentially missing critical scenarios that arise from real-world human driving behavior -Handles limited driving scenarios, e.g., road shapes only	Solely relies on simulation driving data for scenario selection, thus potentially missing critical scenarios that arise from real-world human driving behavior -Requires manual feature extraction of scenario elements from driving videos	-Solely relies on simulation driving data for scenario selection, thus potentially missing critica l scenarios that arise from real-world huma n driving behavior -Relies on optimization objectives, which could lead to overfitting	Relies on predefined selection criteria, which could overlook critica l scenarios that do not fit within predefined criteria, which may limit the selection of diverse scenarios

(LDW = Lane Departure Warning, PCC = Predictive Cruise Control, DT = Data Type, Env.= Environment, Sup.= Support, OS = Open Source)

[10] The Following Lead Vehicle scenario group includes the situations where the ego vehicle is driving behind another vehicle referred to as the "lead vehicle"

[11] The Crossing Path scenario group includes the situations where the paths of ego vehicle or more vehicles intersect or cross each other at an intersection or junction.

[12] The Pedestrian Interaction scenario group includes the situation where ego vehicle encounters pedestrians while driving.

[13] https://adl.cs.ut.ee/lab/simulation

eliminate redundancy and excessive length. The remaining segments are prioritized based on the coverage and the rarity of driving scenes to select the crucial test scenarios for effective and efficient testing.

(ii) The SSTSS process could be combined with other approaches to test specific features of an ADS. Since the SSTSS process prioritizes scenario groups, and each scenario group corresponds to specific ADS feature(s), indicating a need to focus on testing these features. The prioritized scenario groups using SSTSS could suggest which ADS features to focus on, and further thorough testing of these features can be performed using feature-specific approaches such as SDC-Scissor, SDC-prioritizer, etc. For example, in the lane change [14] scenario group, approaches like SDC-Scissor or SDC-prioritizer could be further used to select or prioritize safe or unsafe scenarios for testing ADS's Lane Departure Warning (LDW) feature.

In ADS testing, one potential question is determining how safe is safe enough. One answer could be that if an ADS is safer than a human driver, it is secure enough to deploy. In line with this, the SSTSS process uses human driver car crash datasets (as one of the inputs) to select test scenarios. Testing the ADS in these scenarios can directly compare its safety performance to that of human drivers. Also, it could be helpful for stakeholders to confirm if the ADS meets or exceeds the safety standards set for human drivers. Therefore, the SSTSS process could complement other approaches for the initial selection of test scenarios.

The main strength of the SSTSS approach is that it uses real-world car crash statistics for scenario selection, which prioritizes statistically significant yet often overlooked scenarios. However, the reliance on the availability and quality of car crash datasets may restrict the applicability in scenarios not well-represented in the data or geographic regions with sparse data collection. If the data is not up-to-date, infrastructure, technological developments, or other changes in driving behavior might not be reflected in the data and could affect the output of the process.

The SSTSS process selects and prioritizes scenarios based on the ODD of ADS under test (as one of the inputs). By considering ODD, on the one hand, SSTSS ensures that the selected scenarios reflect the intended use cases and environments for which the ADS under test is designed. This increases the probability of identifying potential safety-critical issues in those specific operating conditions. On the other hand, if manufacturers do not clearly define or convey the ODD, the selection of relevant scenarios may be less accurate.

V. Threats to Validity

In this section, we discuss the possible threats to the validity of our study and the measures taken to mitigate them.

TABLE VI: A SAMPLE SSTSS PROCESS OUTPUT

Sr. #	Prioritized Scenario Groups	Selected Scenario Identifier	Scenario Priority
01	Following Lead Vehicle	S17	1
		S18	2
		S19	3
		S20	3
02	Crossing Path	S25	4
		S29	4
		S28	5
		S30	6
		S31	6
		S33	6
		S34	6
		S26	7
03	Pedestrian Interaction	S27	8
		S32	8
		S36	9
		S37	9
		S41	9
		S45	9
		S39	10
		S40	10
		S42	11
		S43	11
		S38	12
		S44	13

The description of scenario identifiers is available online[15]

Researcher bias: The first two authors chose the keywords and selected publications, which might have introduced subjective bias. We tried to mitigate this threat by explicitly defining inclusion and exclusion criteria for selecting studies to select publications objectively.

Search string validity: The search query string either generates too few results (false negatives) or too many (false positives). For both conditions, we tried to mitigate threats. We used wild card (*) to widen the coverage and the keyword "AND" to minimize the false positive. We tried to keep the search query the same across all online repositories, but we slightly varied it due to the constraints of each repository. We reviewed the search query results to see if they contained studies we knew already. We were aware of 4 relevant papers, and the search string retrieved 3 of them. However, we found the missing paper's conference version in another searched online repository. Furthermore, the results of the search strings were also manually checked, and false positives were removed. We did not manually add the missing publications to the results because we wanted to follow the procedure defined for strictly reviewing the literature. Furthermore, the studies using the

[14] The lane change scenario includes the situation where the ego vehicle or other adjacent vehicle is merging or switching lanes in the same direction without maintaining an appropriate distance and speed with adjacent vehicle.

[15] https://github.com/ScenarioSelectionApproaches

synonyms of the keywords used in the search query in the title might have been missed. We mitigate this threat by using synonyms of the keywords used.

VI. CONCLUSION AND FUTURE WORK

ADS requires exhaustive testing before its deployment on real roads. Simulation-based testing provides a cost-effective approach to test ADS and requires test scenarios. In this study, we identify the approaches used to select or prioritize test scenarios for simulation-based safety testing of ADS and compare them. We also illustrate an example of how one approach complements or could be combined with other approaches for improving testing effectiveness and efficiency. In the future, we aim to demonstrate via examples how the remaining five selected approaches can complement or combine with each other to enhance testing effectiveness and efficiency.

ACKNOWLEDGMENT

The research reported in this paper has been partly funded by BMK, BMAW, and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE [(FFG grant no. 892418)] part of the FFG COMET Competence Centers for Excellent Technologies Programme, as well as by the European Regional Development Fund, grant PRG1226 of the Estonian Research Council and by Bolt Technology OU.

REFERENCES

[1] C.-Y. Chan, "Advancements, prospects, and impacts of automated driving systems," International journal of transportation science and technology, vol. 6, no. 3, pp. 208–216, 2017.

[2] B. Schoettle and M. Sivak, "A survey of public opinion about autonomous and self-driving vehicles in the us, the uk, and australia," tech. rep., University of Michigan, Ann Arbor, Transportation Research Institute, 2014.

[3] R. Pfeffer, G. N. Basedow, N. R. Thiesen, M. Spadinger, A. Albers, and E. Sax, "Automated driving-challenges for the automotive industry in product development with focus on process models and organizational structure," in 2019 IEEE International Systems Conference (SysCon), pp. 1–6, IEEE, 2019.

[4] X. Zhang, J. Tao, K. Tan, M. Torngren, J. M. G. S¨ anchez, M. R. Ramli,´ X. Tao, M. Gyllenhammar, F. Wotawa, N. Mohan, et al., "Finding critical scenarios for automated driving systems: A systematic literature review," arXiv preprint arXiv:2110.08664, 2021.

[5] F. Khan, M. Falco, H. Anwar, and D. Pfahl, "Safety testing of automated driving systems: A literature review," IEEE Access, 2023.

[6] E. Zablocki, H. Ben-Younes, P. P´ erez, and M. Cord, "Explainability of´ deep vision-based autonomous driving systems: Review and challenges," International Journal of Computer Vision, vol. 130, no. 10, pp. 2425–2452, 2022.

[7] J. M. Anderson, K. Nidhi, K. D. Stanley, P. Sorensen, C. Samaras, and O. A. Oluwatola, Autonomous vehicle technology: A guide for policymakers. Rand Corporation, 2014.

[8] SAE, "3016—taxonomy and definitions for terms related to on-road motor vehicle automated driving systems." https://www.sae.org/standards/ content/j3016 202104/, 2021. Accessed: 15-07-2023.

[9] K. Bimbraw, "Autonomous cars: Past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology," in 2015 12th international conference on informatics in control, automation and robotics (ICINCO), vol. 1, pp. 191–198, IEEE, 2015.

[10] Rodionova, I. Alvarez, M. S. Elli, F. Oboril, J. Quast, and R. Mangharam, "How safe is safe enough? automatic safety constraints boundary estimation for decision-making in automated vehicles," in 2020 IEEE Intelligent Vehicles Symposium (IV), pp. 1457–1464, IEEE, 2020.

[11] D. Sportillo, A. Paljic, and L. Ojeda, "On-road evaluation of autonomous driving training," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 182–190, IEEE, 2019.

[12] X. Zhao, V. Robu, D. Flynn, K. Salako, and L. Strigini, "Assessing the safety and reliability of autonomous vehicles from road testing," in 2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE), pp. 13–23, IEEE, 2019.

[13] S. Baltodano, S. Sibi, N. Martelaro, N. Gowda, and W. Ju, "The rrads platform: a real road autonomous driving simulator," in Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 281–288, 2015.

[14] Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in Conference on robot learning, pp. 1–16, PMLR, 2017.

[15] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Mozeiko,ˇ E. Boise, G. Uhm, M. Gerow, S. Mehta, et al., "Lgsvl simulator: A high fidelity simulator for autonomous driving," in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6, IEEE, 2020.

[16] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," Transportation Research Part A: Policy and Practice, vol. 94, pp. 182–193, 2016.

[17] T. D. Son, A. Bhave, and H. Van der Auweraer, "Simulation-based testing framework for autonomous driving development," in 2019 IEEE International Conference on Mechatronics (ICM), vol. 1, pp. 576–583, IEEE, 2019.

[18] D. J. Fremont, E. Kim, Y. V. Pant, S. A. Seshia, A. Acharya, X. Bruso, P. Wells, S. Lemke, Q. Lu, and S. Mehta, "Formal scenario-based testing of autonomous vehicles: From simulation to the real world," in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–8, IEEE, 2020.

[19] Z. Zhong, Y. Tang, Y. Zhou, V. d. O. Neves, Y. Liu, and B. Ray, "A survey on scenario-based testing for automated driving systems in highfidelity simulation," arXiv preprint arXiv:2112.00964, 2021.

[20] H. Elrofai, D. Worm, and O. Op den Camp, "Scenario identification for validation of automated driving functions," in Advanced Microsystems for Automotive Applications 2016: Smart Systems for the Automobile of the Future, pp. 153–163, Springer, 2016.

[21] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, "Defining and substantiating the terms scene, situation, and scenario for automated driving," in 2015 IEEE 18th international conference on intelligent transportation systems, pp. 982–988, IEEE, 2015.

[22] E. de Gelder, O. O. den Camp, and N. de Boer, "Scenario categories for the assessment of automated vehicles," CETRAN, Singapore, Version, vol. 1, 2020.

[23] S. Tang, Z. Zhang, Y. Zhang, J. Zhou, Y. Guo, S. Liu, S. Guo, Y.-F. Li, L. Ma, Y. Xue, et al., "A survey on automated driving system testing: Landscapes and trends," ACM Transactions on Software Engineering and Methodology, vol. 32, no. 5, pp. 1–62, 2023.

[24] G. Lou, Y. Deng, X. Zheng, M. Zhang, and T. Zhang, "Testing of autonomous driving systems: where are we and where should we go?," in Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 31–43, 2022.

[25] W. Ding, C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao, "A survey on safety-critical driving scenario generation—a methodological perspective," IEEE Transactions on Intelligent Transportation Systems, 2023.

[26] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in 12th international conference on

evaluation and assessment in software engineering (EASE), BCS Learning & Development, 2008.

[27] F. Khan, H. Anwar, and D. Pfahl, "A process for scenario prioritization and selection in simulation-based safety testing of automated driving systems," in International Conference on Product-Focused Software Process Improvement, (Austria), pp. 89–99, Springer, 2023.

[28] C. Birchler, N. Ganz, S. Khatiri, A. Gambi, and S. Panichella, "Costeffective simulation-based test selection in self-driving cars software," Science of Computer Programming, vol. 226, p. 102926, 2023.

[29] C. Birchler, S. Khatiri, P. Derakhshanfar, S. Panichella, and A. Panichella, "Single and multi-objective test cases prioritization for self-driving cars in virtual environments," ACM Transactions on Software Engineering and Methodology, vol. 32, no. 2, pp. 1–30, 2023.

[30] Y. Deng, X. Zheng, M. Zhang, G. Lou, and T. Zhang, "Scenario-based test reduction and prioritization for multi-module autonomous driving systems," in Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 82–93, 2022.

[31] C. Lu, H. Zhang, T. Yue, and S. Ali, "Search-based selection and prioritization of test scenarios for autonomous driving systems," in International Symposium on Search Based Software Engineering, pp. 41–55, Springer, 2021.

[32] J. Bach, J. Langner, S. Otten, E. Sax, and M. Holzapfel, "Test scenario¨ selection for system-level verification and validation of geolocationdependent automotive control systems," in 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 203– 210, IEEE, 2017.

[33] Iso, "Pas 21448-road vehicles-safety of the intended functionality," Int. Organization for Standardization, 2019.

[34] Mercedes-Benz, "Safety first for automated driving (safad)." https://group.mercedes-benz.com/innovation/case/autonomous/ safety-first-for-automated-driving-2.html, 2019. Accessed: 19-072023.

[35] Waymo, "Waymo safety report: On the road to fully self-driving." https: //g.co/kgs/N6nC77u, 2018. Accessed: 18-07-2023.

[36] E. Castellano, A. Cetinkaya, and P. Arcaini, "Analysis of road representations in search-based testing of autonomous driving systems," in 2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS), pp. 167–178, IEEE, 2021.

[37] Gambi, M. Muller, and G. Fraser, "Asfault: Testing self-driving car¨ software using search-based procedural content generation," in 2019 IEEE/ACM 41st International Conference on Software Engineering:Companion Proceedings (ICSE-Companion), pp. 27–30, IEEE, 2019.