SAFETY ASSURANCE FRAMEWORK FOR CONNECTED, AUTOMATED MOBILITY SYSTEMS

D3.1

Report on baseline analysis of existing Methodology

Project short name
SUNRISE

Project full name
Safety assUraNce fRamework for connected, automated mobIlity SystEms

Horizon Research and Innovation Actions | Project No. 101069573
Call HORIZON-CL5-2021-D6-01

Funded by the European Union

ccam-sunrise-project.eu/

Dissemination level	Public (PU) - fully open
Work package	WP3: CCAM V&V Methodology for Safety Assurance
Deliverable number	D3.1: Report on baseline analysis of existing Methodology
Status - Version	Final – V1.1
Submission date	15/05/2024
Keywords	safety assurance, scenario-based testing, automated driving systems.

Authors

Role	Name
Main author:	Anders Thorsén (RISE)
Contributing authors:	Oliver Bartels (BASt), Christian Berger, Tayssir Bouraffa, Cagri Kaya (CHAL), Mirko Muro (CRF), Anastasia Bolovinou, Ilias Panangiotopoulos (ICCS), Daniel Becker (ika), Ashfaq Farooqui, Martin Skoglund, Per Stålberg, Fredrik Warg (RISE), Emmanuel Arnoux (RSA), Olaf Op den Camp (TNO), Ghada Ben Nejma (VED), Antonio Bruto da Costa, Patrick Irvine (WMG)

Quality Control

	Name	Organisation	Date
Peer review 1	Erwin de Gelder	TNO	11/07/2023
Peer review 2	Daniel Becker	ika	04/08/2023

Version history

Version	Date	Author	Summary of changes
0.1	31/10/2022	Anders Thorsén	First draft of the document
0.2	25/04/2023	Anders Thorsén, Ashfaq Farooqui, Martin Skoglund, Per Stålberg (RISE), Daniel Becker (ika), Emmanuel Arnoux (RSA), Olaf Op den Camp (TNO), Patrick Irvine, Antonio Bruto da Costa (WMG)	Update and move to new template draft. Updates from many partners.
0.3	05/05/2023	Anders Thorsén (RISE), Daniel Becker (ika), Olaf Op den Camp (TNO), Patrick Irvine, Antonio Bruto da Costa (WMG)	Clean up for start of analysis phase.

0.4	23/06/2023	Anders Thorsén, Martin Skoglund (RISE), Christian Berger, Tayssir Bouraffa, Cagri Kaya (CHAL), Mirko Muro (CRF), Daniel Becker (ika), Olaf Op den Camp (TNO), Ghada Ben Nejma (VED), Patrick Irvine, Antonio Bruto da Costa (WMG), Oliver Bartels (BASt).	Work focusing on completing existing methodology chapter and analysis chapter.
0.5	30/06/2023	Anders Thorsén, Fredrik Warg (RISE), Anastasia Bolovinou, Ilias Panangiotopoulos (ICCS)	Clean up, restructuring, addition of missing sections, proof reading.
0.6	30/06/2023	Anders Thorsén (RISE)	Version for peer review
0.7	23/08/2023	Anders Thorsén, Martin Skoglund (RISE), Christian Berger, Tayssir Bouraffa, Cagri Kaya (CHAL), Ilias Panangiotopoulos (ICCS), Daniel Becker (ika), Emmanuel Arnoux (RSA), Olaf Op den Camp (TNO), Ghada Ben Nejma (VED), Patrick Irvine, Antonio Bruto da Costa (WMG)	Updates after peer review.
0.8	23/08/2023	Anders Thorsén (RISE)	For coordinator check.
0.9	28/08/2023	Anders Thorsén (RISE)	Final draft.
1.0	29/08/2023	Anders Thorsén (RISE)	Final version
1.1	15/05/2024	Anders Thorsén (RISE), Martin Skoglund (RISE), Jobst Beckmann (ika)	Updates after EC review, minor restructuring of Sections 2.1 and 7.1

Legal disclaimer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

Term:	Description	Source
Abstract scenario	Formalized, declarative description of a traffic scenario focusing on complex relations, particularly on causal relation	SUNRISE Glossary
Automated driving system (ADS)	Hardware and software that are collectively capable of performing the entire dynamic driving task (DDT) on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD)	ISO 34501:2022 [1]
Concrete scenario	Parameterised model of the time sequence of scenes (logical scenario) which begins with an initial scene and defined point in time; the behaviour of the main actor (vehicle under test) is not further specified.	SUNRISE Glossary
Critical scenario	Scenario including one or more risk factors	ISO 34502:2022 [2]
Dynamic driving task (DDT)	All of the real-time operational and tactical functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints	ISO 34501:2022 [1]
Dynamic entity	Entity that experiences state change(s) during a scenario	ISO 34501:2022 [1]
Entity	Element of interest in a scenario	ISO 34501:2022 [1]
Functional scenario	Temporal sequence that describes one of the behaviours of a system during a specific use case, with a nominal scenario and alternative scenarios. It is described in a linguistic way or with a structured language. Functional scenarios are derived from driving functions. They are used to describe the use case at a high level (higher than logical and concrete scenarios)	SUNRISE Glossary
Hazardous scenario	Scenario in which harm occurs unless prevented by an entity other than the ADS	ISO 34502:2022 [2]

Term:	Description	Source
Logical scenario	Beginning with an initial scene, a model of the time sequence of scenes whose parameters are defined as ranges; at a defined point in time, the behaviour of the main actor (vehicle under test) is not further specified	SUNRISE Glossary
Ontology	Container of standardized definitions of concepts that are used in a particular domain of knowledge, such as road traffic. Unlike terminologies and taxonomies, ontologies also describe how the defined concepts relate to and differ from each other. Within SUNRISE the ontology is used for traffic scenarios to enable the translation of the terms to object-oriented code. This, in turn, is used to describe the scenarios in a coding language that can be understood by various software agents, such as databases or simulation tools. The ontology is also used as a schema for development of scenarios databases.	SUNRISE Glossary
Operational design domain (ODD)	The 'boundaries of the operating environment within which the ADS can operate, performing the DDT safely.'	SUNRISE Glossary
Parameter space	Set (e.g., Range) of possible parameter values, e.g., A subset of finite-dimensional Euclidean space	SUNRISE Glossary
Risk factor	Factor or condition of a scenario that, if present, increases either the probability of the occurrence of harm, or the severity of harm, or both	ISO 34502:2022 [2]
Safety test objective	Safety property of the ADS to be shown via a set of tests	ISO 34502:2022 [2]
Scenario	Description of a temporal and spatial traffic constellation	SUNRISE Glossary
Scene	Snapshot of all entities including, but not limited to the automated driving system (ADS) / subject vehicle, scenery, dynamic environment, and all actors and observer's self-representations, and the relationships between those entities	ISO 34501:2022 [1]
Static entity	Entity that does not experience state change(s) during a scenario	ISO 34501:2022 [1]

Term:	Description	Source
Subject vehicle ego vehicle host vehicle	Vehicle under observation in the process of testing, evaluation, or demonstration	ISO 34501:2022 [1]
Surrounding environment	All entities in a scenario, excluding the subject vehicle(s) or ads(s)	ISO 34501:2022 [1]
System under test	ADS that is tested with test scenarios	ISO 34501:2022 [1]
Taxonomy	The study of the general principles of scientific classification	[3]
Terminology	A system of words used to name things in a particular discipline, a vocabulary associated with a certain field of study, profession, or activity.	[4]
Test scenario	Scenario intended for testing and assessing automated driving system(s) (ADS)/subject vehicle(s)	ISO 34501:2022 [1]

ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
AD	Automated Driving
ADAS	Advanced Driver-Assistance System
ADS	Automated Driving System
ALARP	As Low As Reasonably Practicable
ALKS	Automated Lane Keeping System
ASAM	Association For Standardization Of Automation And Measuring Systems
AV	Automated Vehicle
AWI	Approved Work Item (ISO)
BSI	British Standards Institution
CCAM	Cooperative, Connected And Automated Mobility
CoSim	Co-Simulation
DDT	Dynamic Driving Task
DIS	Draft International Standard
EU	European Union
Euro NCAP	European New Car Assessment Programme
FCW	Front Collision Warning
FOT	Field Operational Test
FoV	Field-Of-View
GA	Grant Agreement
GPS	Global Positioning System
GSR	General Safety Regulation
HD	High Definition (High Resolution)

Abbreviation	Meaning
IAMTS	International Alliance For Mobility Testing And Standardization
IMA	Intersection Movement Assist
ISO	International Organization For Standardization
KET	Key Enabling Technology
KPI	Key Performance Indicator
LD	Lane Detection
LSAD	Low-Speed Automated Driving
LTA	Land Transport Authority
MEM	Minimal Endogenous Mortality
MFM	Midlands Future Mobility
MiL	Model in the Loop
MRC	Minimum Risk Condition
MRM	Minimum Risk Manoeuvre
MVTC	Motor Vehicle Test Centre
NCAP	New Car Assessment Program
NTU	Nanyang Technological University
ODD	Operational Design Domain
OEDR	Object And Event Detection And Response
OEM	Original Equipment Manufacturer
PG	Proving Ground
PoC	Proof Of Concept
PRB	Positive Risk Balance
SAF	Safety Assurance Framework
SCDB	Scenario Database

Abbreviation	Meaning
SDL	Scenario Description Language
SDOs	Standards Development Organizations
SiL	Software In the Loop
SPaT	Signal Phase and Timing
SPSD	Safety Pool Scenario Database
SSMS	Synchronised Serial Manoeuvre Sequences
STPA	Systems Theoretical Process Analysis
SuT	System Under Test
SW	Software
TBD	To Be Defined
THW	Time-Headway
TTC	Time To Collision
UC	Use Case
UNECE	United Nations Economic Commission For Europe
V&V	Verification And Validation
V2I	Vehicle-To-Infrastructure
V2V	Vehicle To Vehicle Communication
V2X	Vehicle-To-Everything
VCA	Vehicle Certification Agency
VRU	Vulnerable Road User
VTP	Virtual Testing Platforms
WP	Work Package
XiL	Anything in the Loop (e.g. SiL, HiL)

EXECUTIVE SUMMARY

Safety assurance of cooperative, connected, and automated mobility (CCAM) systems is crucial for their successful adoption in society. To demonstrate that such systems are safe in their complete operational design domains (ODDs) requires robust safety argumentation. The aim of the SUNRISE project is to develop and demonstrate a safety assurance framework (SAF) for the test and safety validation of a varied scope of such systems.

Scenario-based testing methods is believed to become an important part of the safety assessment approach for automated driving systems (ADSs). The SUNRISE project's forerunner project HEADSTART developed a methodology for safety validation of connected and automated vehicles centred around scenario-based testing, a methodology that SUNRISE will develop further and integrate as a part of the SUNRISE SAF. Focus for Work Package 3 of the SUNRISE project is to define and condense an overall methodology to support the safety argumentation using data- and knowledge-driven, scenario-based testing.

This report presents a literature study and baseline tracking of the existing scenario-based methodologies, especially, based on the knowledge and literature review of the HEADSTART project. First, the SUNRISE SAF and scenario-based methodologies are introduced including a suitable taxonomy. Second, the HEADSTART method is summarized in detail. Third, scenario-based methodologies from other projects are described. Fourth, an overview of relevant standardization efforts is presented with a particular focus on the ISO 3450X series "Road vehicles – Test scenarios for automated driving systems". Fifth, other initiatives related to scenario-based safety assessment (mainly outside the EU) are described. Sixth, an extensive analysis is presented comparing the HEADSTART methodology with the other described initiatives. Seventh and final, the findings are summarised in the conclusions.

The SUNRISE methodology will use the HEADSTART methodology as input complemented with other existing best practices documented in this report. For areas that was in focus for the HEADSTART project, such as scenario concept, test scenario selection and test scenario allocation, the HEADSTART method is concluded to be well defined for future development. Important is that the SUNRISE scenario concept need to be versatile and adoptable for scenario concepts used in all relevant existing scenario databases. As far as possible the scenario concept should also be adoptable for possible future relevant scenario concepts. Other areas, like scenario sources, scenario generation, and scenario databases, were not in focus for HEASTART and only conceptually defined. The SUNRISE data framework is essential to solve these parts as SUNRISE, like HEADSTART, relays on external scenario databases. Further, the HEADSTART methodology needs to be complemented with elements like risk assessment, monitoring in order to identify unknown scenarios, and qualitative and quantitative metrics to determine the completeness of a scenario database.

1 INTRODUCTION

1.1 Project intro

Safety assurance of cooperative, connected, and automated mobility (CCAM) systems is crucial for their successful adoption in society, yet it remains being a significant challenge.

CCAM systems need to demonstrate reliability in their complete operational design domains (ODDs), requiring robust safety argumentation. It is generally acknowledged that for higher levels of automation, the validation of these systems by means of real test-drives would be infeasible [5]. In consequence, a carefully designed mixture of physical and virtual testing has emerged as a promising approach, with the virtual part bearing more significant weight in this mixture for cost efficiency reasons. Several worldwide initiatives have started to develop test and assessment methods for automated driving (AD) functions. These initiatives have already moved from conventional validation to a scenario-based approach and combine different test instances (physical and virtual testing) to avoid the million-mile issue.

The initiatives mentioned above provide new approaches to CCAM validation, and many expert groups formed by different stakeholders are already working on CCAM systems' testing and quality assurance. Nevertheless, the fact that there is a lack of a common European validation framework and homogeneity regarding validation procedures to ensure safety of these complex systems, hampers the deployment of CCAM solutions. In this landscape, the role of standards is paramount in establishing common ground and providing technical guidance. However, standardising the whole pipeline of CCAM validation and assurance is in its infancy, as many of the standards are under development or have been very recently published and still need time to be synchronised and established as common practice.

Scenario databases are another issue tackled by several initiatives and projects, providing silo solutions. A single, concrete approach should be used (at least at the European level), dealing with scenarios of any possible variations, including the creation, editing, parameterisation, storing, exporting, importing, in a universally agreed manner.

Furthermore, validation methods and testing procedures still lack appropriate safety assessment criteria in order to build a robust safety case. These must be set and be valid for the whole parameter space of scenarios. Another level of complexity is added, due to regional differences in traffic rules, signs, actors, and situations.

Evolving from the achievements obtained in HEADSTART and taking other initiatives as a baseline, it becomes necessary to move to the next level in the concrete specification and demonstration of a commonly accepted safety assurance framework (SAF) for the safety validation of CCAM systems, including a broad portfolio of use cases [6] and comprehensive test and validation tools. This will be done in SUNRISE, which stands for **S**afety ass**U**ra**N**ce f**R**amework for connected, automated mob**I**lity **S**yst**E**ms.

The SAF is the main element to be developed in the SUNRISE project. This framework takes a central role, fulfilling the needs of different automotive stakeholders with their own interests in using it. The overall objective of the SUNRISE project is to accelerate the safe deployment of innovative CCAM technologies and systems for passengers and goods by creating demonstrable and positive impact towards safety, specifically the EU's long-term goal of moving close to zero fatalities and serious injuries by 2050 (Vision Zero), and the resilience of (road) transport systems. The project aims to achieve this by creating and sharing a European federated database framework centralising detailed scenarios for testing of CCAM functions and systems in a multitude of relevant test cases, with standardised, open interfaces and quality-controlled data exchange.

1.2 Purpose of the deliverable

Work package 3's part in SUNRISE is to define and condense an overall CCAM V&V methodology to support the safety argumentation based on data- and knowledge-driven, scenario-based testing while the overall responsibility for the development of the SAF is handled by Work package 2 "CCAM safety assurance framework".

This deliverable presents a baseline analysis of existing scenario-based testing methodologies for safety assurance of automated vehicles (AVs). The results from SUNRISE's forerunner project HEADSTART [7, 8] is used as starting point and it is complemented with results from other relevant initiatives. Research areas that are investigated include existing scenario concepts, parameter sets and descriptions, scenario databases, test scenario selections, and test scenario allocation concepts and metrics. The outcome of this work will feed into the following work packages to inform the direction of their work.

The partner contributions to this deliverable are summarized in Table 1.

Table 1. Partner contribution to D3.1.

Partner:	Role
RISE	RISE is task leader and main editor for the deliverable.
BASt	BASt has contributed to the "Public authorities and policy makers" chapter.
Chalmers	Chalmers has contributed with analysis related to the output from the Safety Assurance Framework gap analysis in SUNRISE deliverable D2.1 as well as analysing the HEADSTART method vs. ISO 34502.
CRF	Overall review from OEM perspective.
ICCS	Overall check and contribution.
ika	Ika has contributed with describing the PEGASUS family and with analysis.
WMG	WMG has contributed with describing Safety Pool^{TM} and with analysis.
TNO	TNO has described StreetWise, Sakura, CETRAN, CATARC and contributed to the analysis.
RSA	Renault has contributed with the ADScene description.
VED	VED has contributed with the ADScene description.

1.3 Intended audience

The intended audience of this deliverable is primarily the rest of WP3. As it presents a baseline for scenario-based testing methodology for safety assurance of AVs it should also be relevant for the rest of the project consortium as well as for readers outside the consortium.

1.4 Structure of the deliverable and its relation with other work packages/deliverables

The contents of this deliverable are structured as follows: Chapter 2 describes the concept for scenario-based safety argumentation, Chapter 3 summarizes the HEADSTART methodology, Chapter 4 describes other existing methodologies and the projects that developed them, Chapter 5 summarizes the status of standardization organisations' work, Chapter 6 summarizes what is done in consumer testing, by public authorities and policy makers, as well as other relevant academic work (primary outside EU), Chapter 7 presents an analysis of previously presented methodologies and initiatives, and, finally, Chapter 8 summarises the conclusions.

This deliverable takes input from D2.1 [9] and gives output to the subsequent tasks in WP3 and to the other technical WPs (WP2 – WP7).

2 SCENARIO-BASED METHODOLOGY

The introduction of vehicles with higher automation levels, i.e., levels for which the vehicle takes over the responsibility as a driver from the human driver, poses new challenges with respect to the safety-assessment. To solve these different approaches have been proposed. One such approach is the HEADSTART methodology developed in SUNRISE forerunner project HEADSTART [7, 8]. The HEADSTART methodology is an approach for safety validation of connected and automated vehicles centred around scenario-based testing and is described in Chapter. 3.

In [10], Riedmaier et al. present a survey about safety-assessment approaches for CCAM identifying the following seven approaches (including scenario-based):

- Scenario-based (for background see, e.g., [11–17]).
- Real-world testing - distance-based evaluation of safety resulting from field tests. A standard approach for ADAS.
- Shadow mode - AD functions are executed passively in serious production vehicles. Simulation can be used to evaluate the AD functions.
- Staged introduction - The ODD is limited such that safety assessment can be carried out based on real-world testing.
- Formal verification- Mathematical method by which the safety of systems is formally proven across the whole ODD.
- Function based- System functions are defined based on requirements and then tested on test track or in simulation. This is a common procedure for ADAS.
- Traffic simulation based – The whole road network with hundreds of road users is simulated.

The approaches besides scenario-based are not in the scope of this deliverable and therefore not further discussed, but for the interested reader, there are several review and survey papers published, e.g., [18–23].

2.1 Terminology

The standards in the ISO 3450x series "Road vehicles – Test scenarios for automated driving systems" [1, 2, 24, 25] focus on scenario-based testing and are considered a good baseline for SUNRISE WP3 terminology. However, for some terms it has been concluded that SUNRISE needs other definitions, e.g., more generic definitions. These SUNRISE defined terms are referred to as "SUNRISE glossary".

The most important terms used in this deliverable are defined in the "Glossary of Terms" on page 11.

Scenario-based testing is needed as a solution to the billion-mile problem [10] and a common definition of the term scenario is needed in this context. Several definitions exist in the literature, and it is common that each initiative has its own flavour of scenario definition. As the SUNRISE methodology will use the HEADSTART methodology it is reasonable to start with the HEADSTART definition that a scenario is an:

"abstraction and general description of a temporal and spatial traffic constellation without any specification of the parameters"

Since the SUNRISE methodology shall be versatile and able to support different approaches, SUNRISE needs a scenario definition that to reasonable extend capture the use of scenario in other initiatives and projects. Consequently, the HEADSTART scenario definition will be reevaluated during the task. This is further discussed in Section 7.1.

An illustrative view of a scenario useful for understanding the concept is shown in Figure 1. Included are some terms that may be used to describe a scenario.

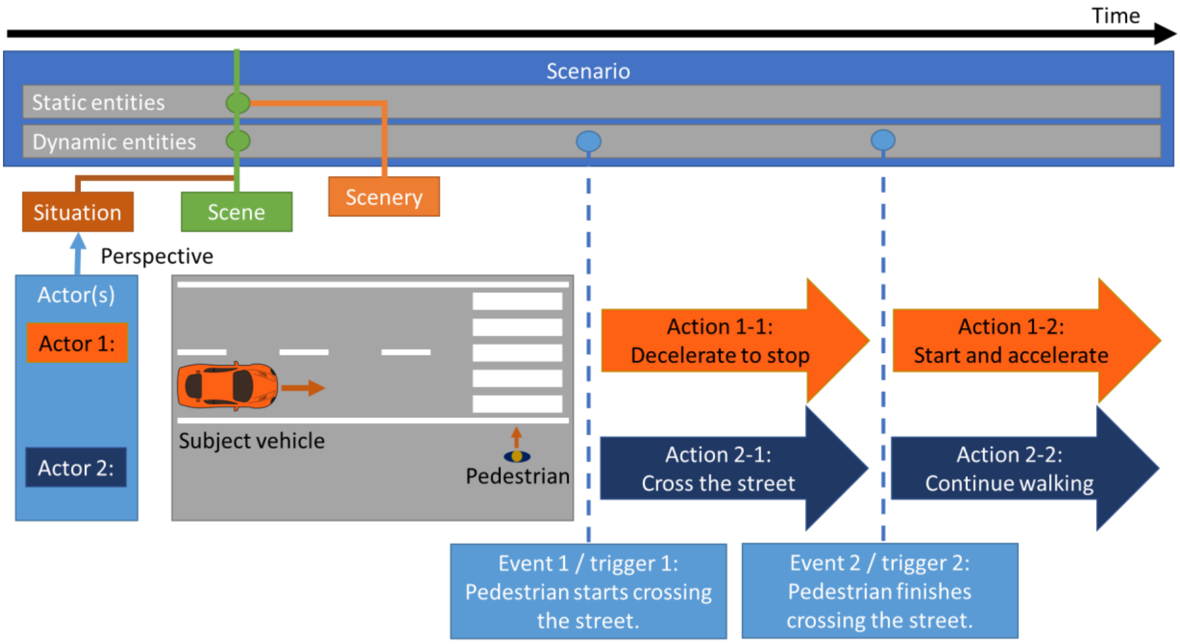


Figure 1. Illustration of relevant terms to describe a scenario (redrawn from [1]).

Further, based on the level of abstraction, the scenarios are defined as functional, abstract, logical, and concrete as shown in Figure 2 [1]. Originally, these were defined by Menzel et al. [26] and later extended with the abstract scenario by Neurohr et al. [27].

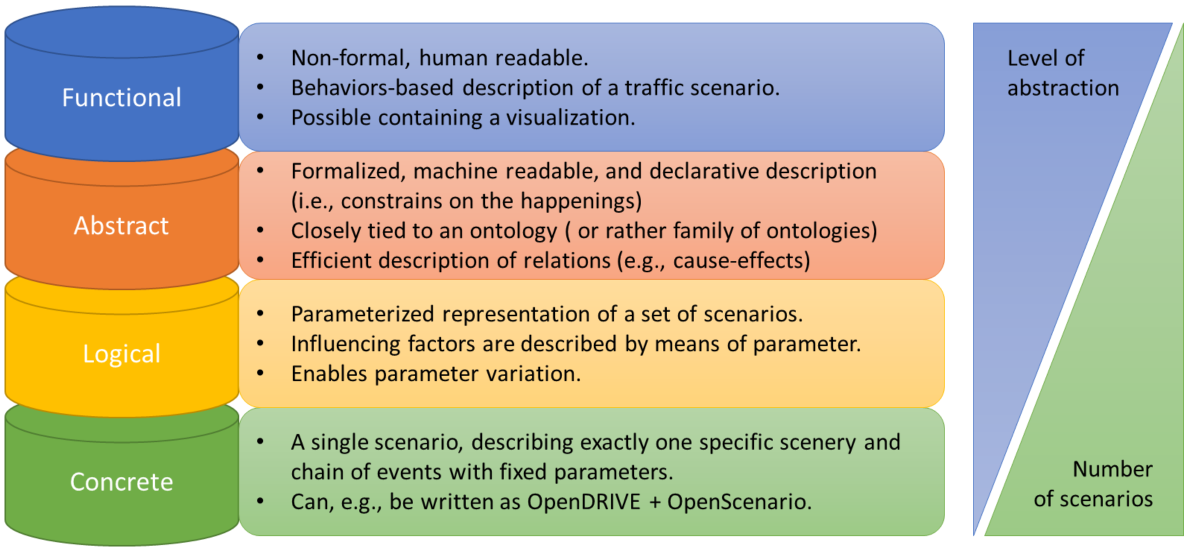


Figure 2. The relationship between functional, abstract, logical and concrete scenarios ([1]).

2.2 Taxonomy

A taxonomy for scenario-based safety assurance is proposed in this section, based on [10] and adopted to align with ISO 34501:2022 [1]. The intention is that it shall be abstract enough to cover and categorize the workflow for most scenario-based approaches, and suitable for SUNRISE. Riedmaier et al.[10] identified scenario generation and test scenario selection[1] as the two most important research topics from a methodology point of view [10], a conclusion that fits well the SUNRISE scope with focus on the second. Figure 3 illustrates this with two outer frames illustrating the input side to the data framework and the output side from the data framework. Blue boxes are in scope for SUNRISE work package 3, dark grey boxes are in scope for other work packages, while light grey boxes are not in scope for SUNRISE.

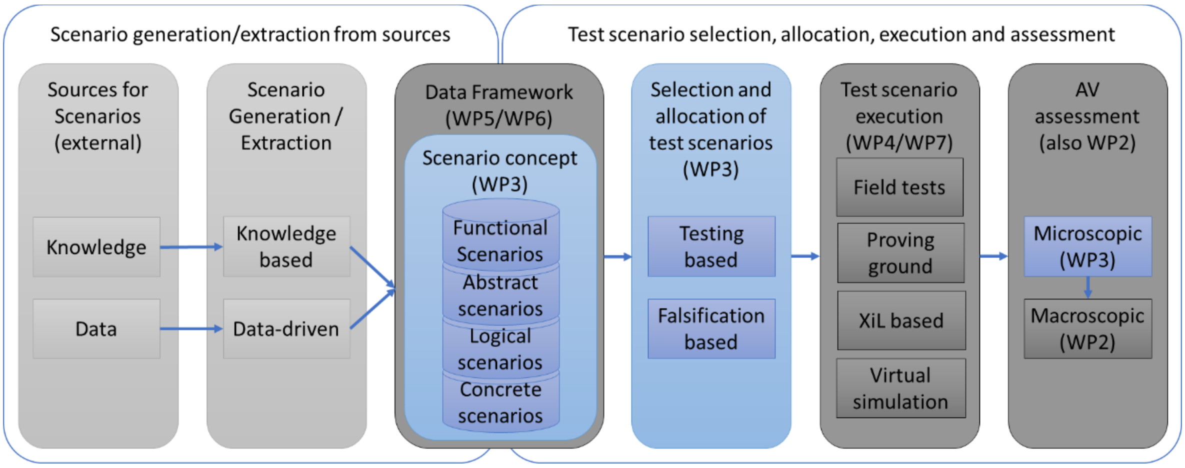


Figure 3. A high-level taxonomy of the scenario-based approach (based on [10]).

2.2.1 Workflow overview

This section describes shortly the six boxes in Figure 3 from left to right.

2.2.1.1 Sources for scenarios

The collection of the data is not the scope of SUNRISE, instead SUNRISE mainly relies on available data from external sources.

Information sources for scenarios may be in the form of abstract information from experts, standards, guidelines, or in the form of driving or accident data. Another source is data from real-world driving and synthetic data from simulations [10, 28–32].

2.2.1.2 Scenario Generation/Extraction

Scenario generation and extraction can be divided into:

- *-Purely knowledge-based scenario generation*

  The purely knowledge-based scenario generation approach creates functional, logical or even directly concrete scenarios for the database out of abstract information [10]. Examples are found in [33–39].

[1] Riedmaier et al. [10] uses *scenario selection*, but in this deliverable *test scenario selection* is used to follow the definitions in ISO 34501 [1] and ISO 34502 [2].

- *Partly data-driven scenario generation*
  In the literature, there are a multitude of different approaches used for the partly data-driven scenario generation approach which typically rely on machine learning or pattern recognition methods as described in Table 2 [10].

Table 2. Summary of approaches for data-driven scenario extraction.

Approach	Description	
Extraction:	Concrete scenarios are directly filtered without any assignment to predefined logical scenarios or similar clusters. Examples are available in [40–44].	
Clustering/ Classification:	Data grouped to also obtain concrete scenarios, but with a kind of group membership.	
	Clustering:	Data grouped to similar clusters with assignment made in an unsupervised learning fashion. Examples are available in [45–49]
	Classification:	Data grouped to predefined logical scenario classes using assignment made in a supervised learning fashion. Examples are available in [50–55].
Parameterization	Work regarding parameterization of classified data are available in, e.g., [56–60].	

2.2.1.3 Data Framework

Access to scenario database(s) are essential for scenario-based testing. However, SUNRISE will not establish its own scenario database filled with huge amount of data. Instead, a SUNRISE data framework is developed with a federation layer that allows to retrieve scenarios from external databases. The data framework is developed in WP5 dealing with “Content harmonisation of data framework”, and WP6 with “Data framework design and usage definition”. Work package 3 shall focus on the needs from scenario-based safety assurance methods to store in the database and to retrieve from the database.

2.2.1.4 Selection of test scenarios

There are different possible approaches for selection of scenarios, focusing on covering the parameter space with test cases, and approaches focusing on challenging corner cases to find counterexamples [10].

- *Testing-based test scenario selection*
  Testing-based test scenario selection approaches have in common that a subset of test scenarios is sampled for microscopic assessment of safety in each individual test scenario. The results can then be aggregated into a macroscopic assessment. Two different types of sampling are possible:

- *Sampling within parameter ranges:*
  Sampling is done over the entire parameter range and neglects that scenarios have different probability in real-world. Consequently, it only allows an overall statement based on coverage [61–72].

- *Sampling over parameter distributions:*
  Sampling is done over the parameter distributions that also includes the probability of occurrence of the scenarios. Consequently, weighting of the results is possible for a true statistical statement about the accident probabilities [73–93].

- *Falsification-based test scenario selection*
  Falsification-based test scenario selection aims to find counterexamples violating the safety requirements in microscopic safety assessment. Existing concrete test scenarios, or simple logical test scenarios with parameter ranges, are taken from the scenario database:
  - *Accident-based:*
    Accident data as basis for test scenarios are used in safety assessment of advanced driver-assistance system (ADAS), i.e., for assessment of system with SAE Level ≤2 [94–99].
  - *Criticality:*
    Test scenario selection is done by choosing representative concrete test scenarios based on calculated criticality, possibly followed by adopting them such that the criticality is increased, see [100–102].
  - *Complexity:*
    A third approach is to increase the probability of finding counterexamples by increasing the complexity of the test scenarios [103–111].
  - *Optimizer for simulation-based falsification:*
    Simulation-based falsification is distinguished by an additional feedback loop. Consequently, assessment results of the simulation can be used for optimization to select next concrete test scenarios, data that actually is from the vehicle under test and not from the database [112–130].

2.2.1.5 Test scenario execution

Execution of different relevant test scenarios is part of the methodology. Either the test execution if performed in real-world via field or proving-ground tests, or as different degrees of virtual tests. Most references use simulation for proof of concept (PoC) due to advantages regarding, e.g., cost, expenditure and safety risks [10, 131–133].

2.2.1.6 AV assessment

Riedmaier et al. [10] distinguish between microscopic and macroscopic safety assessment:

- Macroscopic safety assessment, or statistical safety assessment, is about showing that the automated vehicles (AVs) have lower accident probability than human drivers. That type of statement about the overall impact of AVs on traffic will require a huge amount of data.

- Microscopic assessment is based on evaluation of single scenarios, i.e., scenario-based safety assessment.

Figure 3 illustrates this by including the scenario-based microscopic safety assessment of the AV in WP3. The macroscopic safety assessment is shown as part WP2 assuming that complement with non-scenario-based approaches may be needed. The transition from a microscopic safety assessment to macroscopic safety assessments is identified by Riedmaier et al. [10] as one of the key challenges for scenario-based approaches.

3 THE HEADSTART METHODOLOGY

The HEADSTART methodology is an approach for safety validation of connected and automated vehicles. It is centred around scenario-based testing and was published in HEADSTART deliverable D2.1 [134]. As the forerunner for SUNRISE, the HEADSTART methodology is an important input to SUNRISE and an overview of the methodology is shown in Figure 4. In the following the HEADSTART methodology is summarized and mapped versus the taxonomy shown in Figure 3 [135]. More details and examples are found in [134, 136–139].

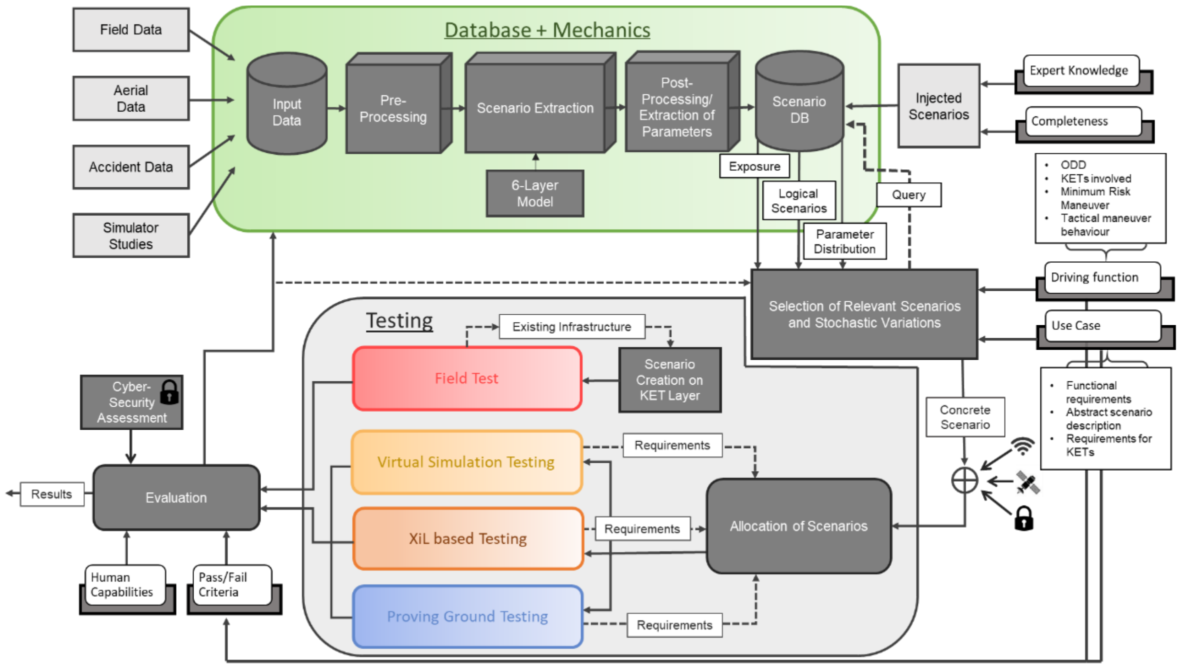


Figure 4. Overview of the overall HEADSTART methodology [135, 138].

Note that HEADSTART focussed on the key enabling technologies (KETs) communication, cyber-security, and positioning. Safety performance validation were addressed, though no definition of safety test objectives is explicitly shown in the overall methodology in Figure 4.

3.1 Scenario concepts, parameter sets and descriptions

As part of the state-of-the-art analysis in HEADSTART, the scenario concepts defined in the projects PEGASUS, MOOVE, StreetWise and SAKURA were analysed (see further [140] and Sec. 4). The chosen scenario layer model for HEADSTART, shown in Figure 5, is based on the PEGASUS layers for infrastructure (Layer 1, 2 & 3), the environment (layer 5) and the digital information (Layer 6). These are combined with separating movable objects in Layer 4a (subject ego vehicle) and Layer 4b (other vehicles) as done in MOOVE. Important to note is that these concepts are always connected to a specific ODD. The existing state-of-the-art projects at the time, usually had an ODD for highway.

The partners in HEADSTART made an evaluation of available scenario formats, and decided that the format should be:

- Compatible with HEADSTART's targeted scenario databases (PEGASUS, MOOVE, StreetWise).
- Compatible with various test instances, especially virtual testing.
- Suitable to describe the KETs defined in the HEASTART project (communication, positioning, and cybersecurity).
- Widely accepted in the industry.

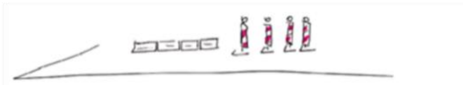
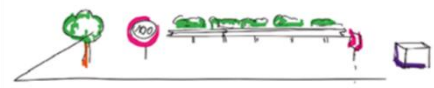
Layer 6: Digital information e.g. V2X communication, Connectivity	
Layer 5: Environment e.g. Weather, lighting and other surrounding conditions	
Layer 4.b: Other objects e.g. Dynamic and static objects	
Layer 4.a: Ego vehicle e.g. Dynamics and behavioural data	
Layer 3: Temporary manipulation of layer 1 & 2 e.g. Geometry, overlaid topology, time frame	
Layer 2: Traffic infrastructure e.g. Structural boundaries, traffic signs, elevated barriers	
Layer 1: Road level e.g. geometry, topology, surface …	

Figure 5. The HEADSTART layer model [134].

After comparing different formats, it was concluded that most suitable for HEADSTART were the ASAM standards:

- ASAM OpenSCENARIO: Description of driving scenarios.
- ASAM OpenDRIVE: Logical description of road networks.
- ASAM OpenCRG: Description of road surfaces.

At the time, the HEADSTART scenario concept was defined, the used abstraction levels were functional, logical, and concrete scenarios with appropriate parameter sets.

3.2 Scenario sources and scenario generation

Figure 4 illustrates the defined workflow in the HEADSTART methodology. Scenario sources include field, aerial, and accident data. These are complemented with data from simulator studies as well as scenarios defined based on expert knowledge to ensure as much completeness as possible. Further, the different sources provide different capabilities and can therefore be used for different purposes. After the input data is collected and uploaded, a pre-processing is done that can reach from simple interpolation of signals to calculation of derived measures and dataset enrichment. However, the identification and characterization of relevant scenarios were not in focus for the HEASTART project. Instead, the assumption was that such scenarios were available from datasets.

Next is the extraction of scenarios based on a logical scenario concept where recorded situations are clustered into clearly defined classes using the HEADSTART scenario concept. After scenarios have been extracted from a recording, parameters describing the scenarios are calculated, including updating exposure calculations, and saving into a database filling up the parameter space of respective logical scenario.

In addition, there may be scenarios that have not occurred in the input data. And there may be injection of scenarios deemed necessary based on expert knowledge.

3.3 Scenario database

Though the methodology shown in Figure 4 illustrates a scenario database, the intention was not to establish a scenario database as part of HEADSTART. Instead, the approach was to rely on external scenario databases like PEGASUS, StreetWise and MOOVE [134, 140].

3.4 Selection of test scenarios

HEADSTART's process for test scenario selection is summarized below and illustrated in Figure 6. The summary is based on the description in HEADSTART deliverable 3.1 [136]. Examples of a few use cases can be found in HEADSTART deliverable 3.2 [137].

1. *Input*
   The required input is the definition of dynamic driving task (DDT), the ODD, and the use case. The driving function's functional capabilities and boundaries are defined, and the test scenario selection process can start.
2. *Define a query*

   Based on the DDT, ODD, and use case, a query for the scenario database is defined.
3. *Extract scenarios from the database*
   The generated query is passed to the scenario database that selects the relevant data and converts it to a requested output format. For HEADSTART, OpenSCENARIO and OpenDRIVE were considered the most promising data formats.
4. *Check if scenarios are inside ODD*
   It is checked whether the scenarios are within the ODD and thus suitable for further processing. Scenarios identified to be outside the ODD are removed from the list.
5. *Check if ODD is sufficiently covered*
   This step checks whether the scenarios from the database cover a sufficient large part of the ODD to ensure an adequate safety evaluation of the AD function (requiring metrics to compare the ODD retrieved scenarios, including the parameter distributions). Alternatively, an evaluation is done based on expert knowledge.

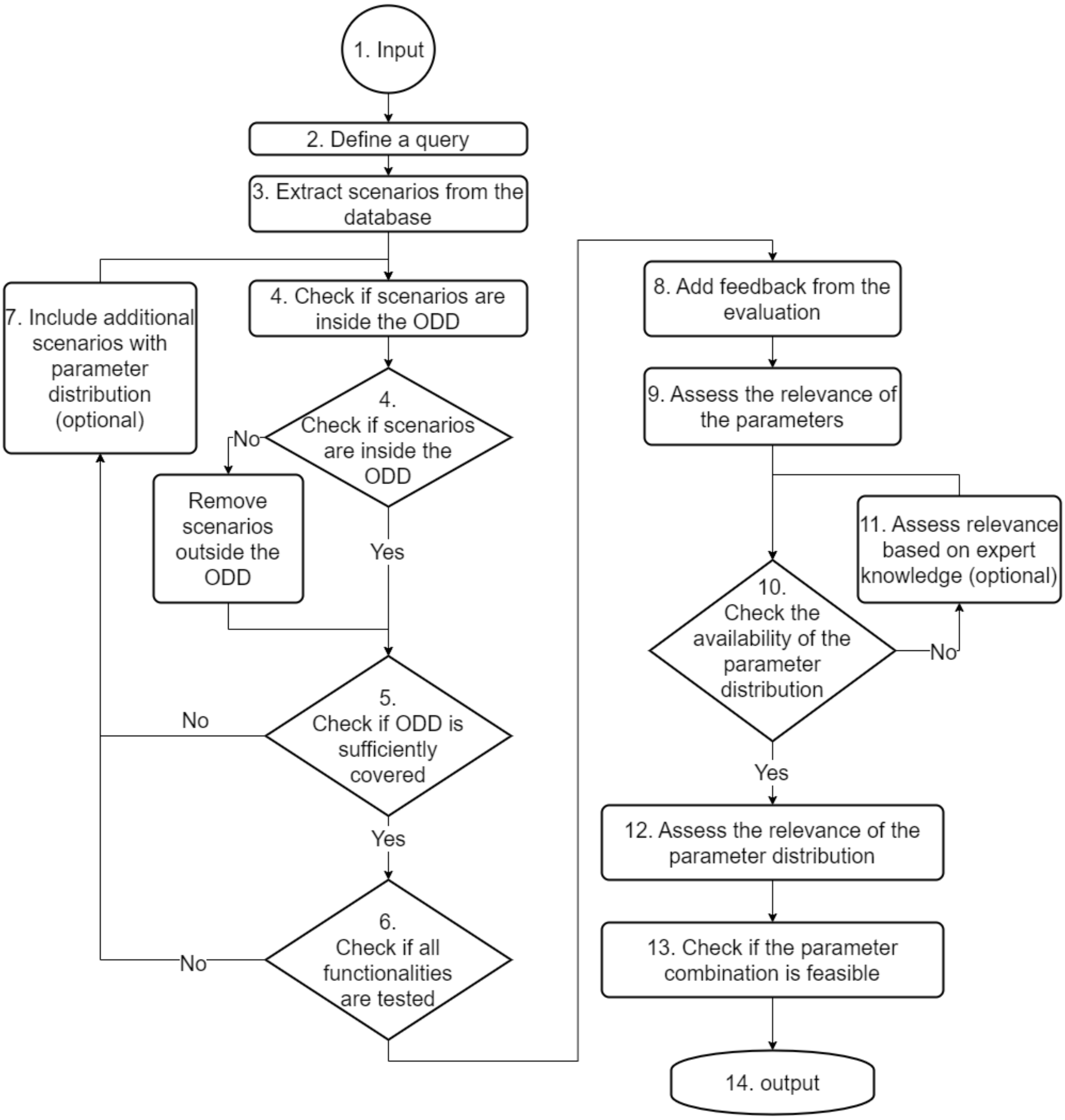


Figure 6. HEADSTART's test scenario selection process.

6. *Check if all functionalities are tested*
   This step checks whether the existing scenarios are sufficient to test all relevant AD functionalities and include testing every basic functionality (equal to every event/response pair defined in the object and event detection response (OEDR)).

7. *Include additional scenarios with parameter distribution (optional)*
   This step is executed if the ODD is not sufficiently covered by the existing scenarios from the database or if not all functionalities of the driving function are tested. In this step, additional scenarios are added based on expert knowledge.

8. *Add feedback from the evaluation*
   Though previous steps have ensured that the scenarios cover all important parts of the driving function, it must be possible to include previously used test cases. One case is when the driving function has passed through the process once, but specific

tests have failed one or more key performance indicators (KPIs). These test cases are rechecked for their data format to ensure compatibility.

9. *Assess the relevance of parameters*
   Analysing each of the logical scenario descriptions extracted from a specific database shows parameters that have higher relevance than others. Especially important is to evaluate what information the AD function extracts from the scenario environment in order to assess this parameter relevance (e.g., the velocity of the cut-in vehicle is more important for a decisive scenario execution than road shoulder width). Parameters not influencing the scenario execution since they lead to the concrete scenario being outside the defined ODD are erased from the scenario descriptions. Examples may be road elevation or a lateral profile on the lane of the respective road network if the driving function is supposed to operate on flat surfaces only.

10. *Check the availability of parameter distribution*
    Combining and varying relevant parameters in a meaningful way requires the corresponding parameter distribution, and this step checks that the information is provided by the database.

11. *Assess relevance based on expert knowledge (optional)*
    Missing parameter distributions for specific scenario descriptions must be defined manually. One way is to do it based on expert knowledge, and another to assign a uniform distribution or a Gaussian distribution.

12. *Assess the relevance of parameter distribution*
    Common parameter values (extracted either from the database or defined by expert knowledge) are usually not the most challenging or safety-critical parameter combinations. To receive a relevance ranking for the concrete scenario descriptions, the occurrence probability of the parameter value can may be combined with its relevance. E.g., the speed of a cut-in vehicle after the actual cut-in is more relevant for the safety evaluation if it is lower because this leads a higher braking demand of the ego vehicle and therefore to more potentially dangerous situations.

13. *Check if the parameter combination is feasible*
    Feasibility checks are needed to ensure that the gathered concrete scenarios with parameter combination are realistically possible.

    - Is the parameter combination physically possible?
    - Is the parameter combination interesting?
    - Are traffic laws obeyed?
      - In case of violation of traffic regulation, the scenario may be removed, or some threshold values are defined that enable some traffic law violation (which is in some cases more realistic).

14. *Output*
    After those feasibility checks, the number of parameter combinations should have been reduced to only the concrete scenario descriptions, which are physically possible, interesting and permitted by traffic regulations.

3.5 Test scenario allocation concepts and metrics

As part of HEADSTAR's procedure for safety validation of driving functions a concept was defined for allocation of test scenarios to the different test methods [134, 136]. The concept is based on every testing method (proving ground, XiL, and virtual testing) having its own capabilities and restrictions. These capabilities and restrictions should be defined and used as input for the test scenario allocation process. Logical scenarios, including parameters, are needed for the allocation of scenarios. The concept is illustrated in Figure 7.

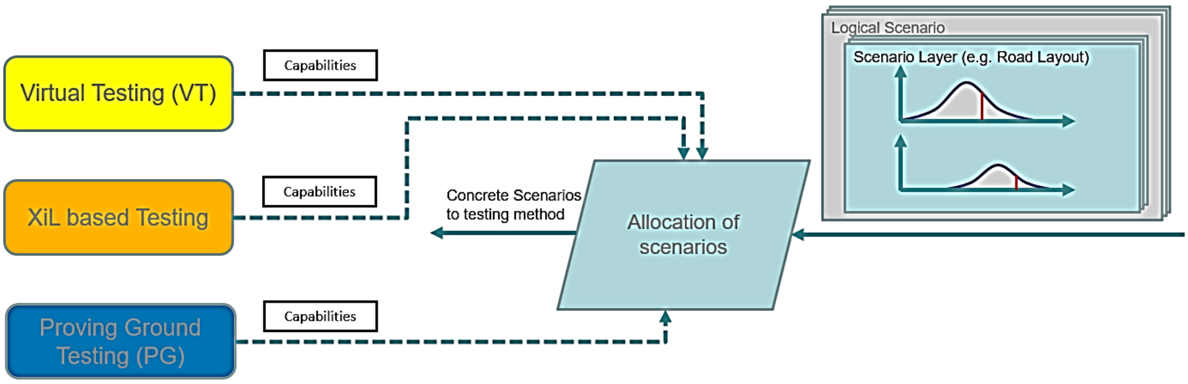


Figure 7. Illustration of the input of testing methods capability to the allocation process [134].

The process includes the following steps:

1. Define capabilities for virtual testing.
2. Define capabilities for XiL testing.
3. Define capabilities for proving ground testing.
4. Compile to a map of capability.
5. Extract logical scenario elements.
6. Match XiL capabilities with a scenario.
7. Align parameter ranges of scenario and XiL.
8. Match proving ground capabilities with a scenario.
9. Align parameter ranges of scenario and proving ground.
10. Match virtual testing capabilities with a scenario.
11. Assign a concrete scenario to a testing method(s).
12. Identify additional requirements that are not covered by scenarios from the scenario database.
13. Get requirement (optional).
14. Assign the requirement to a testing method (optional).
15. Integrate requirement into a scenario data format (optional).
16. Get test case for requirement (optional).
17. Extract information for positioning/communication (optional).
18. Get logical layer for communication with parameter distribution (optional).
19. Get logical layer for positioning with parameter distribution (optional).
20. Select parameter (optional).

The steps are described in detail in HEADSTART deliverable 3.1 [136].

The capabilities are divided into three categories: "*Sensor*", "*Environment*" and "*Vehicle Dynamics*" that either can be represented by simulation models or by the real world, thereby defining a framework providing possibility to categorize every possible testing method in terms of capabilities needed for the allocation of test scenarios. In addition, there are “resource” capabilities like time, cost, and availability (e.g., available area of a proving ground) that can be compared to available meta-data of the concrete scenarios (e.g., the needed area for scenario execution). The terms are further explained in Table 3.

Table 3. Model fidelity vs. capability category [134].

Model Fidelity	Low	Medium	High
Sensor Model	• Object list-based • Based on ground truth data from simulation environment in the FOV	• Based on ideal models • Adding statistical failure rates. • Modified object list entries.	• Based on physical principles of respective sensor type.
Environment Model	• The environment must be able to place different objects in the simulation and update their location and orientation accordingly. • 2D representation could be sufficient.	• 3D representation of objects. • No physics-rendering engine.	• High fidelity sensor models force the most requirements on environment models. Usually, a simulator engine based on ray-tracing is needed. • Potential also real materials and textures.
Vehicle Dynamics Model	• Point mass model.	• Single track vehicle model. • Double track vehicle model.	• Full 6 DOF vehicle model.

3.6 Test scenario execution

HEADSTART deliverable 3.1 [136] describes a method for test method coordination.

- Each selected test scenarios have in the test scenario allocation process been assigned to one of the testing methods (virtual, Xil, proving ground, and field).
- Next step is to decide which of the testing methods should be executed first.
- This decision is made by the target group to take their priorities into account.
- It is also possible to run several testing methods in parallel.
- Once a decision has been made, the scenarios assigned to this testing method are transferred to the testing method responsible.
- After the tests have been completed and evaluated, this phase can be re-entered if further tests are planned.
- The target group has the decision about the further proceeding again, but now also has the results of the tests already performed to determine the further order of the tests.

3.7 AV assessment

After all tests have been executed, the results are evaluated based on pre-defined key performance indicators (KPIs) and pass/fail criteria to conclude if the AD function meets the safety requirements. The results from the evaluation can feed back to the scenario selection process to test certain safety critical scenarios in a more detailed way. Additional formats and criteria are presented in HEADSTART Deliverable D3.4 [139].

4 OTHER EXISTING METHODOLOGIES

Several projects have investigated data-driven, scenario-based testing. Some of them (PEGASUS, MOOVE, StreetWise and SAKURA) were used as input for the HEADSTART method and are described in the HEADSTART Deliverable 1.1 [140]. This chapter presents these four projects, together with some other existing scenario databases and/or projects working with scenarios-based testing. The list is not claimed to be complete but considered to include sufficient relevant twin projects to enable the formulation of a scenario concept to be used in the SUNRISE CCAM SAF.

4.1 PEGASUS Project Family

The research project PEGASUS (**P**roject for the **E**stablishment of **G**enerally **A**ccepted quality criteria, tools, and methods as well as **S**cenarios and **S**ituations) on the release of highly AD functions, addresses the research into new methods for the verification and validation (V&V) of highly AD functions [141]. The exemplary test object is a Level 3 highly AD function for highways (highway chauffeur).

The system under test within PEGASUS is handled as a black box, which means it was not a focus to have a detailed view of the architecture of the complete vehicle or other single components. PEGASUS provides a concept to enhance safety by testing in contrast to safety by design concepts. To come up with a new method for the V&V of AD functions, analysis is performed on different test methods, quality criteria, traffic scenarios, tools, and guidelines in the project. The resulting PEGASUS method for the assessment of highly AD functions is summarized in Figure 8 and will be detailed in the following.

The process flow of the overall method is read counterclockwise from bottom left to upper left and consists of five basic elements for the V&V of the highway chauffeur:

1. Data processing
2. Definition of requirements
3. Information storage and processing in a database
4. Assessment of the highly AD function
5. Argumentation

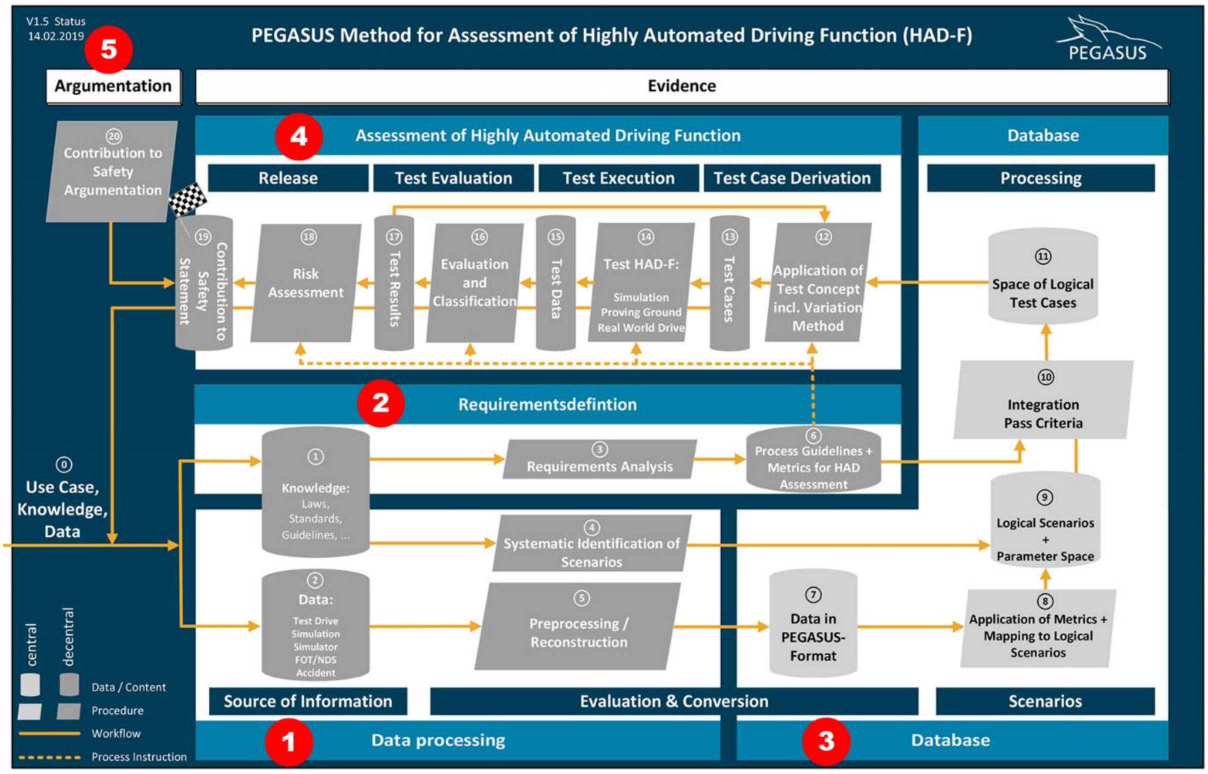


Figure 8. The Pegasus Method.

The main part of the PEGASUS method describes the process for creating evidence supporting the safety argumentation with all the steps and interfaces in between (groups with red numbers 1-4). On the left side is the safety argumentation (red 5). This argumentation is compared at the end of the PEGASUS method with the result of the evidence process to produce a contribution for the safety statement related to the driving function or the test object. This can be used for the overall approval recommendation.

Within the evidence process, the first element is the *data processing*. The input information consists of existing knowledge like regulations and standards as well as given use cases from which logical scenarios (cf. 4.1.1 below) can systematically be derived and transferred to the database. On the other hand, existing recorded scenarios are processed as input data into a common format that is necessary in order to use different types of information sources later in the database. Again, logical scenarios are extracted and fed into the database.

The second part is the *requirements definition* which is executed parallel to the *data processing* and as inputs given knowledge and use cases are utilized. With that, requirements for the AD function or general behaviour requirements for the test object are defined. The requirements are then used in the database to implement evaluation criteria for scenarios and combine them into test cases. In addition, these requirements can be used in the *assessment of the highly AD function* to define process specifications.

The *database* poses the third element. The datasets which have been prepared in a common format in the *data processing* step are used to configure predefined logical scenarios. These are then equipped with parameters and corresponding parameter spaces which are fed from the processed data and knowledge paths shown in Figure 8. Based on this information, the database creates a range of logical test cases with built-in pass and fail criteria from the second element for the different logical scenarios. The underlying process is detailed in Sec. 4.1.2

In the fourth element, the assessment of the highly AD function, is executed. Based on the logical scenarios from the database, test cases are derived and executed in simulation and partly later validated on proving grounds. Systematic field tests will also provide additional findings. The results of the test execution are compared to the pass and fail criteria in order to evaluate them. They are used for a risk assessment to define a safety statement.

Within the last element, the generated evidence is compared with the predefined safety argumentation. The comparison is executed in an external procedure model.

From PEGASUS, two subsequent projects evolved which will detail the developed methodology and operate under the name “PEGASUS project family”. Figure 9 illustrates the orchestration of the different components of the overall methodology that is developed based on PEGASUS, but now with the focus on urban use cases. SET Level was a project in which method and tools for the simulative testing and developing of AD functions were generated. VVMethods wraps an overall methodology around the simulation and enables other test environments as well. The project is still active and will end late 2023.

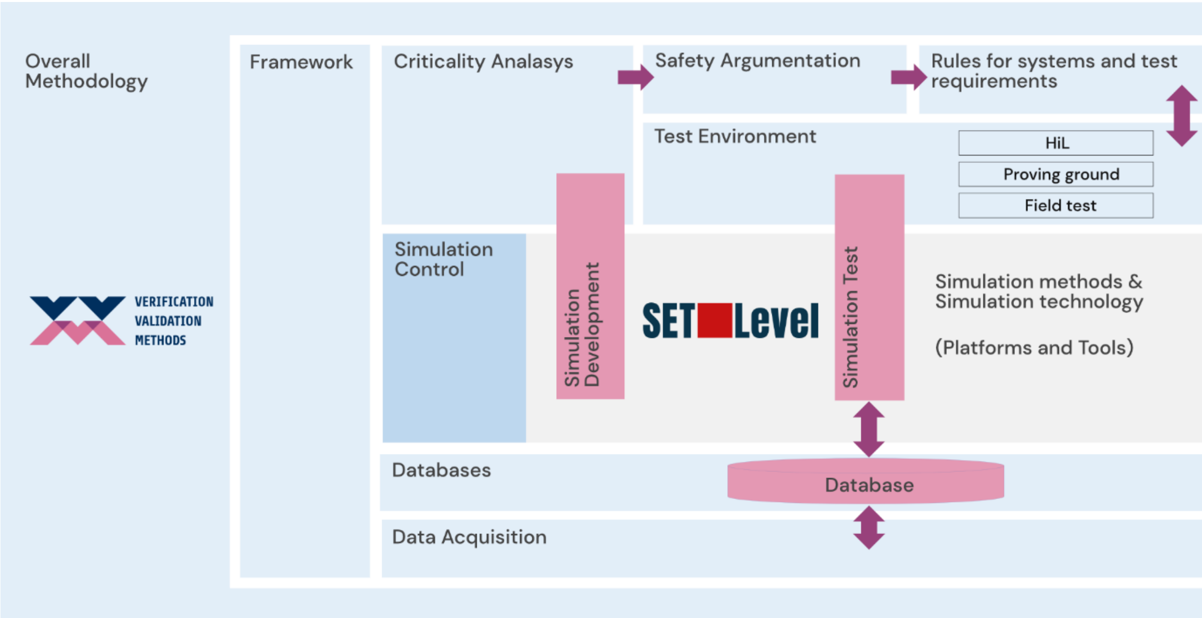


Figure 9. Overview of the scope of PEGASUS successors VVMethods and SET Level.

In the following, aspects of PEGASUS and partly VVMethods are detailed that are relevant to work package 3 of SUNRISE. A full state-of-the-art analysis of the VVMethods project is not possible yet, since the project is still running, and the overall concept is not public. Note that an extensive description of the PEGAUS scenario concept can be found in HEADSTART Deliverable D1.1 [140] .

4.1.1 Scenario concepts, parameter sets and description

The PEGASUS scenario concept had a relevant influence on HEADSTART deliverable 2.1 [4]. Scenarios are structured based on the 6-layer model which has been established during the course of the PEGASUS family projects. The different layers are used to structure a scenario, so it can be discussed on an oral level as well as machine-readable. The latter is done by the ASAM formats OpenDRIVE and OpenSCENARIO. The whole methodology is detailed on the PEGASUS website [142].

Once scenarios can be described, they are collected and fed to the PEGASUS database where they are analysed to transform them into logical scenarios. Within the PEGASUS project, the so called “challenger concept” has been developed to structure the logical scenarios that the AD function has to cope with (cf. Sec. 3.2.1 in HEADSTART Deliverable D1.1 [140]). This concept is used to reduce the scenario space substantially and will be detailed in the following section.

4.1.2 Test Scenario Selection

To create specific scenarios based on field data, a framework for generating a scenario catalogue is necessary [50]. This framework should be able to store all relevant components of the scenario related to Layer 4 of the 6-layer-model (cf. HEADSTART/PEGASUS Methodology). Safety assurance primarily focuses on avoiding collisions between a system-under-test (SUT) and another object, usually another vehicle for the use-case highway chauffeur in PEGASUS. The framework operates on the assumption that a safety-relevant situation can be identified by the need for the SUT to react to avoid a collision with a challenging object. This object is not necessarily the accident perpetrator, but the object with which the SUT would collide if no collision avoidance action was taken. The framework defines a limited number of safety-relevant logical scenarios based on the area of the SUT that the challenging object would collide with and the initial positions of the challenging object. Different initial positions are categorized based on whether the outline of the challenger overlaps with the outline of the SUT in the longitudinal or lateral direction, and the path leading to different types of impact is depicted in Figure 10.

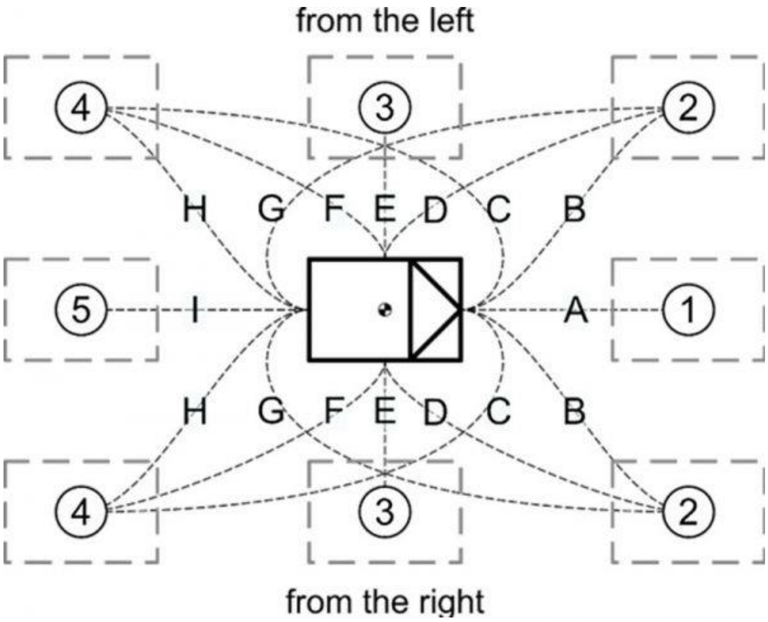


Figure 10. PEGASUS Challenger concept.

Aside from defining safety-relevant logical scenarios, there may be other objects that are important in a scenario. To represent only objects that are relevant for safety, different roles have been identified for additional objects in a scenario. These additional objects either increase the challenge for the SUT to handle the situation safely or are important in the sequence of events of a scenario, so that the SUT may react to an emerging situation even if a collision with a challenging object is not yet imminent.

One group of objects that increases the challenge for the SUT are action restrictions. These objects limit the space for collision avoidance manoeuvres for the SUT. Initially, action constraints can be identified, which are realized by a single object, either located in front, behind, or to the side of the SUT. No distinction is made between positions 2, 3, and 4 in Figure 10, and instead, the positions to the side of the object are treated as a continuous space. Further action restrictions can be identified, which are realized by multiple objects, such as the complete blocking of a side to the ego vehicle or blocking to the side of the SUT with a gap large enough for a collision avoidance manoeuvre. The position and size of this gap can be described as parameters of the scenario instead of describing trajectories for all vehicles contributing to the existence of this gap.

Another role of additional objects in a scenario that increases the challenge for the SUT to handle the situation safely are dynamic occlusions. Objects are represented as dynamic occlusions if they obstruct the vision on the challenging object from the point of view of the ego vehicle. This allows storing scenarios that often are referred to as "Cut-Outs” where another vehicle possibly blocks the view on the back of a traffic jam and then performs a lane change.

In the VVMethods project, the PEGASUS methodology is currently being further developed to cover urban traffic situations. The idea is to structure a traffic situation as several base scenarios and combine them to a complete scenario. This approach will be further investigated and carried out in the SUNRISE project.

4.1.3 Allocation of Test Scenarios to Test Instances

Subsequently, a space of logical test cases is generated which includes pass/fail criteria and additional preparation. A logical test case therefore includes the logical scenario plus evaluation criteria. These criteria are represented by metrics, such as TTC, THW, etc., and thresholds for the respective metrics. The threshold values for the different metrics within the test cases are set which are based on the results of the requirements definition. The results of the process steps integration of pass criteria are stored within the data container of logical test cases. This container includes the test cases, which are relevant for the AD function or general test object based on all available information sources. The test cases are stored in the technical formats OpenDRIVE, OpenSCENARIO, and a format for the parameterization of the logical scenario. The metrics for evaluation of the logical scenarios are stored in external scripts for the application in the following process steps. With this information, it is possible to execute the test cases in suitable simulation environments. Additionally, proving ground tests were conducted to validate the simulations [134].

4.2 StreetWise

TNO StreetWise refers to a methodology for building and maintaining a real-world scenario database, suitable for testing and validating AD functions, that was introduced in 2018 [143]. Starting point for the development was the question how to quantify safety risk for (automated) systems that are to be deployed on public roads. The question results from the need of all stakeholders, policy makers as well as industry, to improve road safety and consequently only deploy systems onto the road that are free of reasonably foreseeable and preventable safety risks, a formulation that is taken from the UNECE R157 regulation for Automated Lane Keeping System [144]. Safety assessment is about determining whether the safety risk of deploying an automated system onto the public road is acceptable or not. Safety assessment procedures aim at quantifying the safety risk by determining the probability that the AV ends up in a collision and addressing the severity of the consequences of such a collision. The safety risk needs to consider each situation that the AV may encounter on the road during its lifetime.

A data-driven, scenario-based approach is proposed for such safety assessment. This approach starts with a description of the large variety of traffic situations by means of scenarios stored in a scenario database. The operational design domain (ODD) of AVs can be described by indicating which scenarios (and variations within the scenarios) are covered within the ODD. For the generation of test cases, scenarios are sampled on the basis of scenario statistics (scenario parameter distributions). It is generally acknowledged that test cases for the safety assessment of AVs should be based on real-world scenarios [145, 146]. It is proposed to use such scenarios to select and generate test cases for the quantitative assessment of an AV through virtual and physical safety validation. Real-world data is used to describe the large variety of different traffic situations on the road, with the manoeuvres of traffic participants, the typical layout of the road and infrastructural elements, and weather and lighting conditions. The collection of scenarios needs to cover the variety of what an automated vehicle can encounter in real traffic during its lifetime. As a result, usually many different scenarios are taken into account to achieve a complete safety assessment.

To serve scenario-based safety assessment, a scenario database is required in which scenarios and their variations are stored with statistics (parameter distributions), also depending on region of scenario occurrence. TNO developed the StreetWise toolchain as shown in Figure 11. The use of scenarios for development and testing puts requirements to scenario databases that need to be established. A scenario database should provide a (complete) view on scenarios (and their variations, also depending on region, traffic rules, and driving culture) that a vehicle can encounter on the road during its lifetime. This includes how scenarios evolve over time with the changes in the mobility system. Scenarios should cover nominal everyday driving and more rare and extreme cases.

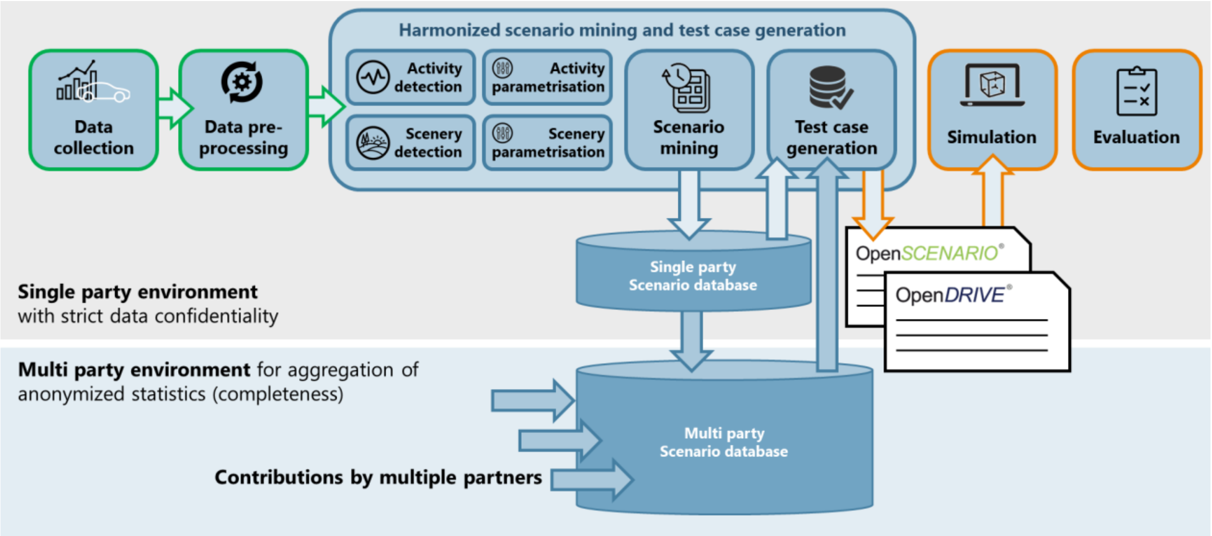


Figure 11. A schematic overview of the TNO toolchain to identify and characterize scenarios and store them in the StreetWise scenario database.

Most important characteristic of StreetWise is the possibility to determine scenario statistics, with metrics such as:

- Exposure: what is the probability of encountering a scenario within certain parameter ranges or given characteristics, e.g., expressed in the number per 100,000 km of driving.
- Completeness: a metric that determines how well the scenarios (and their variations) included in the scenario database cover the occurrence of scenarios in the real world. TNO has published a paper how to estimate completeness from a dataset [60].

4.2.1 Scenario concepts, parameter sets and description

TNO StreetWise uses the following informal definition of the term scenario:

> *A scenario describes any situation on the road including the intent of the ego vehicle, the behaviour of road users, the road layout, and conditions such as weather and lighting. A drive on the road is considered a continuous sequence of scenarios – which might overlap.*

A more formal definition is provided by [147]:

> *A scenario is a quantitative description of the relevant characteristics and activities and/or goals of the ego vehicle(s), the static environment, the dynamic environment, and all the events that are relevant to the ego vehicle(s) within the time interval between the first and the last relevant event. An event corresponds to a moment in time at which a mode transition occurs or a system reaches a specific threshold, where the former can be induced by both internal and external causes.*

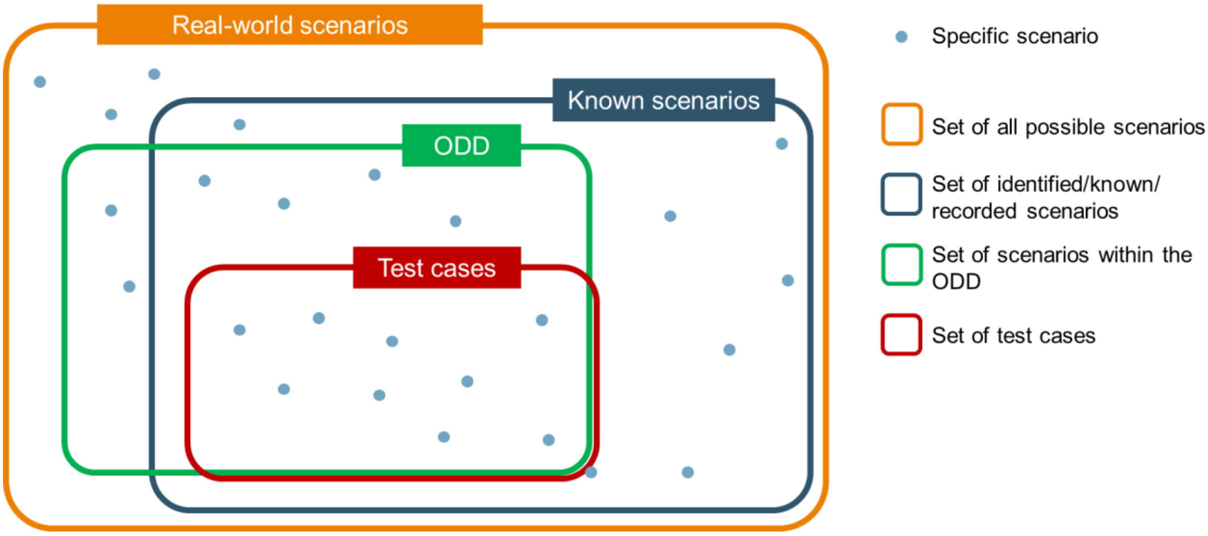


Figure 12. A schematic view on the relation between scenarios, the operational design domain, and test cases.

Figure 12 presents a schematic view on the relation between the scenarios in the real world, the known scenarios that are collected in the StreetWise scenario database, the scenarios in the ODD and the scenarios used to generate the test cases [148]. With this scheme, the important concepts of completeness and coverage can be understood:

- Completeness: how well do the known scenarios cover all possible scenarios in the real-world? How to estimate completeness is shown in [60].
- Coverage: how well do the selected test cases cover the ODD of the system under test? Ideally, there are no unknown scenarios in the ODD and the set of test cases covers at least the complete ODD.

The figure illustrates the importance of striving for completeness in a scenario database. Not only is it difficult to provide a reliable description of the ODD of a function when the scenario database is not sufficiently complete, also the relevance of the selection of test cases is limited in that case. In other words, the function might encounter a scenario in reality for which the function has not been tested.

Currently, in the TNO StreetWise scenario database (with more than 45.000 concrete scenarios identified in approximately 1000 hours of driving data), ten scenario categories are used to describe most situations that occur on highways. The list includes most common scenarios. This does not mean that the list is complete. The applied methodology allows to add scenario categories rather easily to address scenarios that are not covered by one of the categories currently presented in Table 4.

Table 4. Overview of identified scenarios stored in the TNO StreetWise database.

	Scenario category description	Count	Exposure[2] [#/hr]
1	Leading vehicle decelerating	8166	8.4
2	Leading vehicle accelerating	7878	8.1
3	Leading vehicle cruising, while ego and leading vehicle drive at similar speeds	6826	7.0
4	Ego vehicle approaching a slower leading vehicle	2445	2.5
5	Ego vehicle driving in lane, without the presence of a leading vehicle	12896	13.2
6	Vehicle overtaking the ego vehicle	715	0.7
7	Cut-in in front of the ego vehicle	3084	3.2
8	Cut-out in front of the ego vehicle	3760	3.9
9	Ego vehicle performing a lane change while another vehicle approaches from the rear in the target lane	1307	1.3
10	Ego vehicle merging in an occupied lane	375	0.4

Scenario category 7 out of the list above (see Figure 13) refers to a cut-in scenario, which is characterized by the following sequence of activities: ego vehicle is following lane; a target vehicle is cruising outside the ego-vehicle lane at a speed higher than the ego-vehicle; at a certain longitudinal distance between the target vehicle's rear and the ego-vehicle's front, the target vehicle makes a lane change towards the ego-vehicle lane.

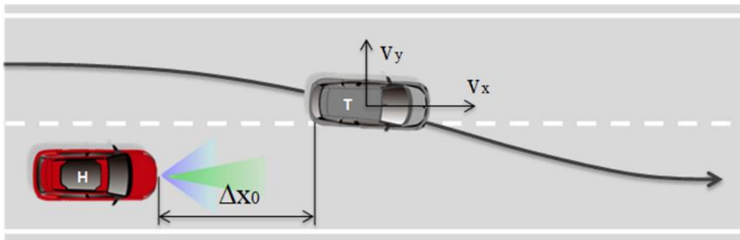


Figure 13. Schematic view of a target vehicle (T) cutting-in in front of an ego vehicle (H).

To describe a cut-in scenario, the following parameter set is currently used:

v_x^H ego initial longitudinal velocity [m/s]

Δv_x^T target initial relative longitudinal velocity with respect to ego vehicle [m/s]

$\overline{v_y^T}$ target average lateral velocity relative to the lane over the duration of the lane change [m/s]

sign v_y target lane change direction [-1: from left to right, 1: from right to left]

THW_{LC} time headway at start of lane change [s] $= \Delta x_0 / v_x^H$

Δx_0 distance between target and ego vehicle when target starts crossing the lane marking.

[2] Here, an exposure metric is given as the number of observations per hour of driving. Different metrics may be considered, e.g., the number of kilometres driven in a certain scenario compared to the total driving distance.

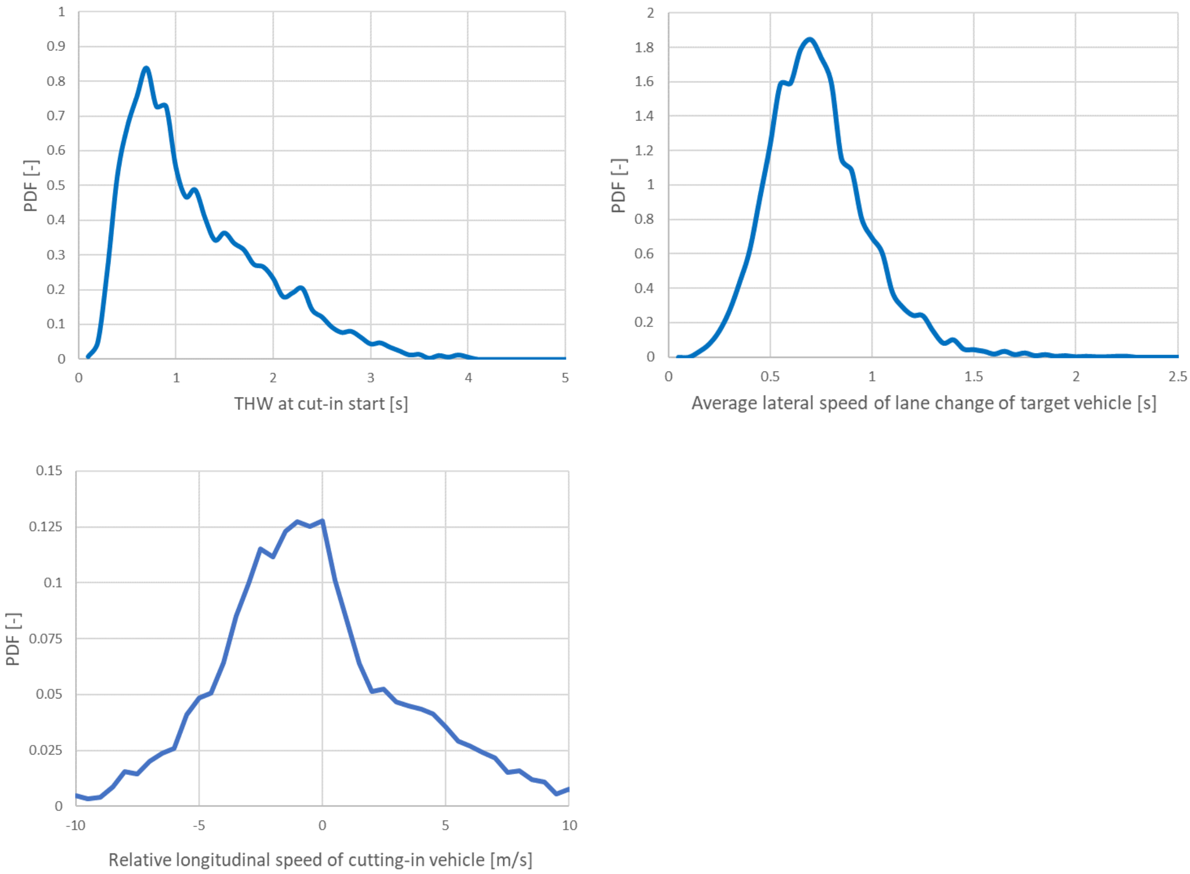


Figure 14. Probability density functions for 3 parameters describing a cut-in.

These parameters, identified for 6316 realizations of a cut-in in a dataset covering more than 110,000 km of highway driving in Europe, provide valuable statistical information. This is illustrated in graphs of the parameter's distributions, or more precisely of the probability density functions (PDF), see Figure 14.

This figure shows the time-headway (THW) between cutting-in vehicle and ego vehicle at start of the cut-in (left upper graph), the average lateral speed of the cutting-in vehicle during the cut-in (right upper graph), and the relative longitudinal speed of the cutting-in vehicle with respect to the ego vehicle (lower graph) – a negative relative longitudinal speed indicates that the gap between ego and cutting-in vehicle decreases, gap closing.

4.2.2 Scenario Selection and Test Case generation

In this subsection, the process of using scenarios for the generation of test cases is shortly described, e.g., for the performance assessment (testing) of an AV that has been developed and for which it needs to be determined whether it can be safely deployed on the road. In the development of such an AV, a functional description needs to be made and the ODD of the vehicle should be described. These descriptions actually result from a discussion between the vehicle developer and the operator of the vehicle (e.g., a shuttle, or a hub-to-hub automated transporter). Where the operator will have requirements regarding functionality and will be able to indicate what the expected ODD looks like, the vehicle developer will indicate to what extent these requirements can be fulfilled by a proposed product.

As vehicle safety assessment considers the vehicle as a whole. The results of tests are influenced by the performance of each of the different subsystems. If the response of the vehicle is not as expected, or the vehicle fails to meet a requirement, it is common practice to investigate the cause and to analyse which of the subsystems has failed. Then, improvements can be implemented in a next development cycle. It is much easier to attribute a failure to a subsystem by performing subsystem tests, where the performance of the perception system, the decision and control logic and the actuation are tested separately. Specific tests are designed for the perception system, and to determine the performance of vehicle dynamics. In a similar way, tests can be designed to challenge the decision and control logic of an automated vehicle. Many of these tests, especially regarding the control and decision logic can be executed in a virtual simulation environment.

Having an overview of scenarios that play a role in operation of an AV facilitates the discussion on the functional specification and the ODD that may be expected. It is for this reason that scenario collection is important. Scenarios are also essential for the generation of test cases. Once the ODD is known, and the scenarios describing what the automated vehicle might encounter are known, test cases can be generated to design tests to evaluate whether the developed AV meets the functional requirements. The process is illustrated in Figure 15.

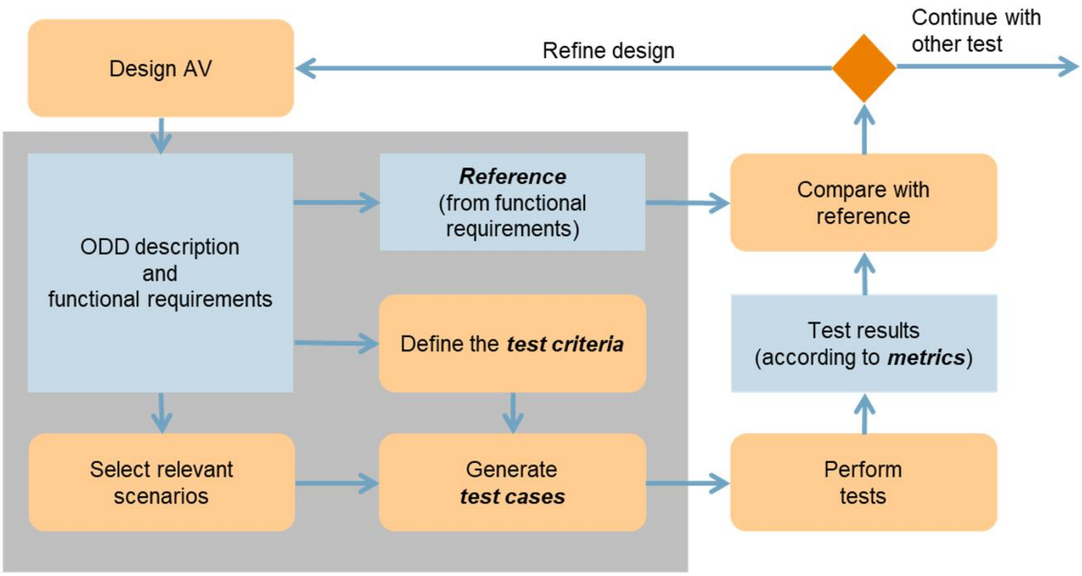


Figure 15. The process of generating test cases.

The design of an AV comes with a range of operating conditions, i.e., the ODD, and the functional requirements. Based on the ODD and the functional requirements, the relevant scenarios can be selected out of the possible scenarios. The ODD and the functional requirements form the basis of the test criteria. Based on the relevant scenarios or scenario categories and the test criteria, the test cases are generated. Next, from the actual performed tests, the test results are obtained according to some metrics that are derived from the test criteria. The test results are compared with a reference that is based on the ODD, the functional requirements, and the metrics.

4.2.3 Allocation of Scenarios to Test Instances

StreetWise does not describe an approach to allocate specific tests to ways of testing. In [149], De Gelder and Op den Camp propose a procedure for safety assessment involving an Applicant, an Assessor and an Authority, to decide on the selection of tests, incl. both proving ground testing and virtual simulation testing.

4.2.4 Validation Metrics for Test Runs and Dynamic Allocation

As indicated in subsection 4.2.1, two concepts are important for validation of the scenario database and for validation of the test cases generated on the basis of the scenario database. The first is completeness, how well in general do the scenarios in the database describe all situations and variations that can happen in the real world. More specifically, given an ODD, how well do the scenarios describe all possible variations within the ODD.

Coverage indicates how well the test cases based on the scenarios within the ODD cover the ODD, and how well these test cases trigger different behaviour of the AV.

4.2.5 Other

One of the challenges in showing compliance with the regulations, is the performance of a safety assessment of such vehicles that provides results that are unambiguous, easily understood by experts in the field, and explainable to authorities and the general public. An important metric in safety assessment is the residual safety risk. Based on the collection of scenarios in the StreetWise scenario database, TNO has proposed a method to determine estimates for the residual safety risk [150], making use of virtual simulations for a wide variety of relevant traffic scenarios. To make such a well-founded estimate of risk, it is important to have reliable information regarding the exposure of scenarios in the real world. It is for this reason that the StreetWise approach is much focused on quantifying exposure of scenarios.

4.3 Safety Pool™

The Safety Pool™ Scenario Database [151], which is developed and maintained by WMG, hosts an estimated ~250,000 scenarios, with this number continuously growing. These scenarios are generated from a variety of methods and sources (which are explained in detail later in this report). The database is used by a variety of stakeholders, (from ADS developers to regulators) and is currently being evaluated by the UK's Vehicle Certification Agency (VCA) as part of the type-approval process for ADS. The platform encourages user contribution and insightful scenario additions through a gamified approach to contribution; a 'tokenised' system values scenarios on uniqueness and assigns a value to each user uploaded scenario, allowing the user to then exchange tokens for additional scenarios which are stored in the database.

The database offers the ability to index the scenarios using ODD and behaviour attribute labels and store the indexed scenarios. The labels used, align with the appropriate standards for ODD (PAS 1883 [152] and ISO 34503 [24]) and are presented in the ASAM OpenLABEL format. Scenarios generated from the Create and Format activities (see Figure 23) are stored in WMG's SDL Level-1 and Level-2 formats (for readability), with the possibility to convert these into ASAM's OpenSCENARIO v1.0 and OpenDRIVE v1.6 scenario description formats.

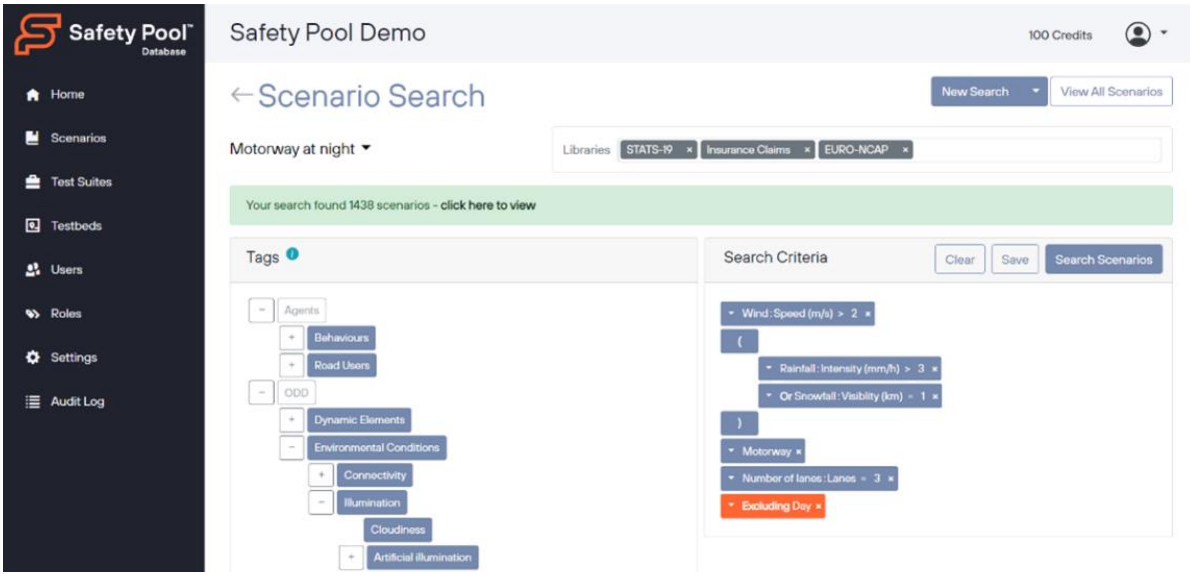


Figure 16. A snapshot of the 'label' searching feature for scenarios in Safety Pool™.

All scenarios hosted on the Safety Pool™ are indexed (see Figure 16) using an extended form of the ASAM OpenLABEL. In the database, every scenario is processed to extract ODD and behaviour attributes as labels which are then organised using a labelling format that uses a JSON schema. This allows users to query the database using these ODD and behaviour labels. This is illustrated in Figure 17.

Safety Pool™ provides a number of useful features, including;

- Searching/Querying scenarios using these labels (as depicted on the right of the figure).
- Saving a complex query using an alias-label which can then be re-used in further queries.
- Hosting real-world route information and associating different sections of the route with labels. This allows users to inspect the route and locate compatible scenarios.
- Providing useful scenario statistics showing a distribution of scenarios according to the ODD taxonomy.
- Extended label structures are used to identify clusters of similar scenarios. This allows us to evaluate the uniqueness of a scenario (compared to others already stored) and identify if a scenario being added to the database, is like existing scenarios and in what ways it is dissimilar to them.

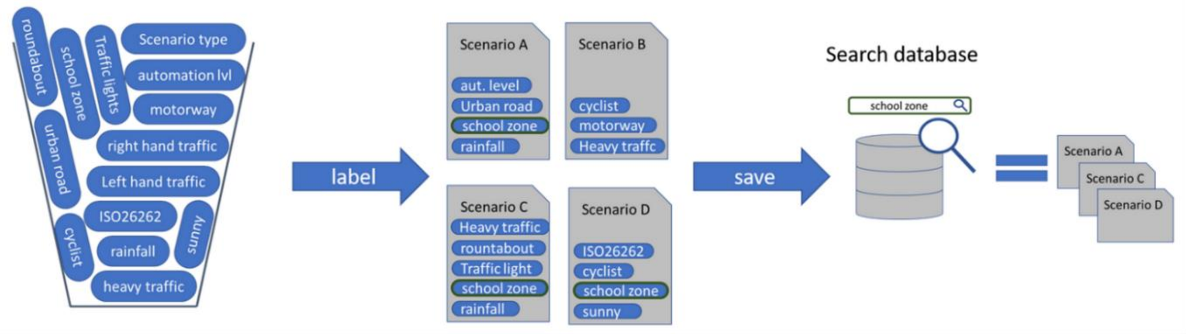


Figure 17. A visualisation of the meta-model and labelling process that feeds into the Scenarios stored in Safety Pool™.

4.3.1 Scenario concepts, parameter sets and description

Scenario concepts

In the V&V life-cycle for an ADS, scenarios are the key assets used to identify failures [39, 153, 154], and may be defined on the basis of the ADS's ODD. [154] define a scenario as a:

> *'temporal development between several scenes in a sequence of scenes. Every scenario starts with an initial scene. Action and events as well as goals & values may be specified to characterise this temporal development in a scenario. Other than a scene, a scenario spans a certain amount of time.'*

The scenario generation methods that form the inputs to the Safety Pool™ database come from a combination of data-based and knowledge-based concepts. There are currently 8 methods used to populate the database with scenarios. Figure 18 below demonstrates this, methods 3-7 are knowledge-based, and 1, 2, 8 are data-based.

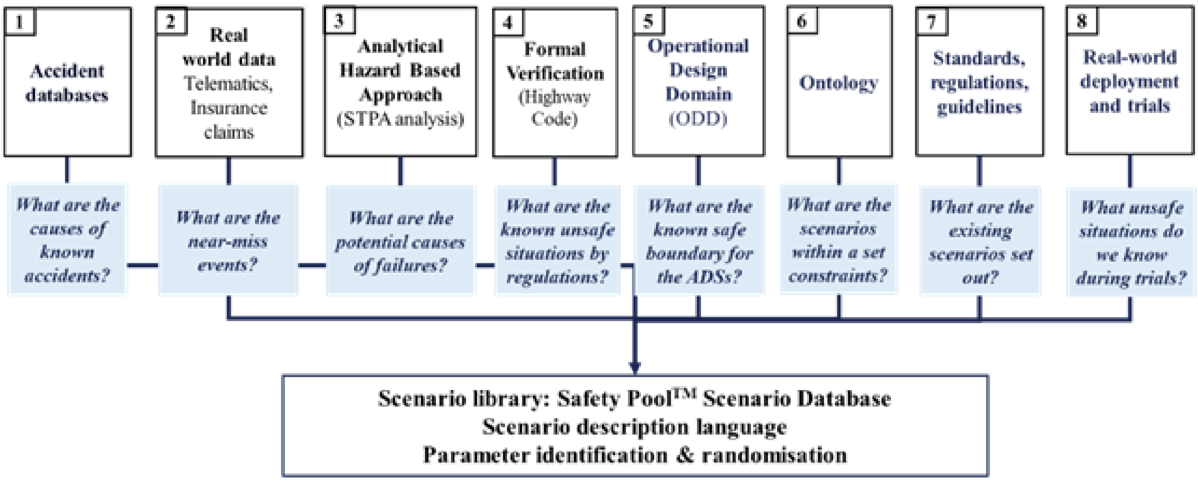


Figure 18. The generation methodologies that contribute to the scenarios stored in Safety Pool™.

Accident databases (1) are an, often publicly available, resource. UK accident databases have been leveraged to form a set of scenarios public in Safety Pool™. Real-world data can come from sources such as anonymised insurance claim records (2). This is an optimal source for collision and near-miss accident trends [99]. Systems Theoretical Process Analysis (STPA) method has been implemented as another scenario source, and has been exemplified in [155] to analyse the characteristics of an ADSs architecture to identify potential system hazards and failures (3). This analysis can be arranged in a format that lends itself to scenario creation.

Formal verification (4) methods are the formal analysis of highway code rules for scenario generation. This work frames each highway code rule as a hypothetical driving scenario with corresponding behaviour and ODD elements. This method focuses on the manoeuvre parameters that are near the boundary of violation and produces scenarios that represent these set of violations. ODD as a generation method (5) uses similar formal representation as of formal verification, but applied to the ODD of the ADS, combining required ODD elements, and limiting the possible scenario combinations based on the ODD restrictions.

Ontology as a generation method (6) is inspired by the works of [34] and [153]. Ontology is used to define all classes within the domain and can be constructed to include all relationships between classes and pre-defined rules. These rules and relationships can be leveraged to ensure the correct instantiation of a scenario is created. If a well-defined ontology is produced, then a high-abstraction scenario can be output at a functional level. Generation of scenarios from standards and regulatory documents (7) is performed as a manual process, contributing a small set of well-defined scenarios, often presented in the form of test cases for a system such as in Euro NCAP documents [156]. Finally, real-world trials and deployments (8) can be leveraged as a useful source of scenarios. This data gives rise to two different broad categories of scenarios which are of equal importance to consider; one being naturalistic driving data, representing a majority of the driving conditions that would be met by an ADS, and a form of edge case data, from scenarios which were difficult for the ADS to navigate during trials for whatever reason.

Scenario description

The native format for the scenarios produced by any of the methods described above is WMG's SDL-Level 2, with possibility to convert to a higher abstraction language representation in SDL-Level 1 through an integrated toolchain.

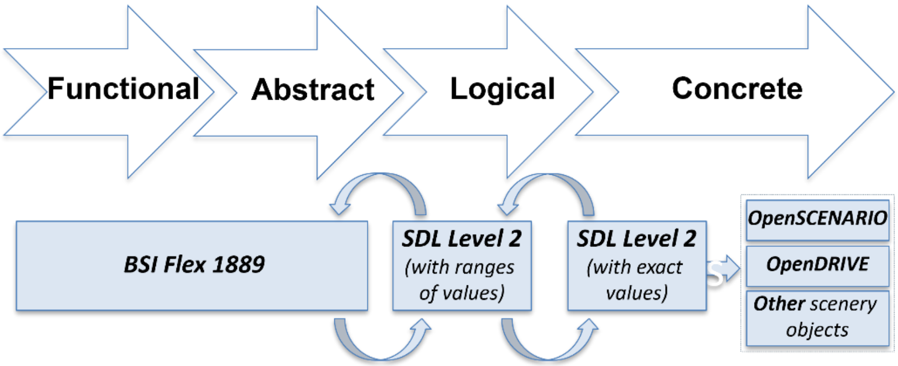


Figure 19. A mapping of Safety Pool™ associated languages to the abstraction levels of scenarios.

WMG's two-level of abstraction language concept is presented in [157]. The language was built on the concept that multiple stakeholders, with varying levels of expertise will be required to interact with scenarios at multiple stages throughout the V&V workflow. Therefore, the expectation that a scenario should be understandable to all using a single representation is flawed. It has two specification forms, an abstract level 1 specification, which is compliant with BSI Flex1889 [158], in which scenarios are expressed in structured natural language, but lack the detail required for simulation. It also provides a (logical scenario) level 2 specification which is concise and readable and allows for more detailed expressions of scenarios. SDL-level 2 scenarios can be made into concrete scenario descriptions using exact values in the description rather than ranges, the mapping of these languages to levels of abstraction can be seen in Figure 19. WMG-SDL Level 2 allows a scenario's scenery and behaviour descriptions to be expressed together, this makes cross-referencing easier. The description itself is clearly segmented into dynamic, scenery and environmental elements and allows for both scripted and unscripted traffic to be defined.

Scenery is described using a textual description that provides all details necessary to depict a road network from road lengths and connection angles to the presence and location of roadside features such as regulatory signage or traffic lights. The scenery description parallels the taxonomy for ODD described in both BSI PAS 1883 [152] and ISO 34503 [24].

In WMG-SDL, all activity is organized as a collection of Synchronised Serial Manoeuvre Sequences (SSMSs). An actor's activity is a sequence of phased manoeuvres belonging to the actor. In Figure 20, an SSMS is represented by the dashed box that groups the various actors' manoeuvre sequences. The phases are numbered from '1' to the number of phases in the SSMS. A WHEN condition triggers an SSMS to begin, which causes all actors' manoeuvre sequences to begin at Phase 1. Phases across different sequences, but within the same SSMS, having identical index values, operate synchronously.

Any two SSMSs taken together may operate asynchronously from each other. For an actor, a phase consists of a manoeuvre, manoeuvre parameters, and a WHILE invariant condition that must hold while that phase is in operation. In Figure 20, the invariant for phase i and actor j is represented by the symbol C_j^i . A phase is considered 'complete' once the active component of all actors' manoeuvres, in the same phase, have completed, or when any WHILE condition linked to that phase is invalidated. Hence, so long as all WHILE conditions of the phase hold, any actor having completed its active manoeuvre component continues with a default drive action until all actors in the same phase complete their respective active manoeuvre components. A drive-only (stopping, accelerating, or braking to reach a target velocity) phase, unlike manoeuvres requiring lateral motion (for instance lane changes) is unique. The completion of a phase containing a drive manoeuvre requires either a WHILE condition (which it is paired with) to become invalid, OR once the active 'speed change' component of the drive action to be completed. In WMG-SDL, each SSMS begins with a triggering condition, the WHEN condition. If this condition is satisfied, then the actors may begin executing the manoeuvres defined within the SSMS [159].

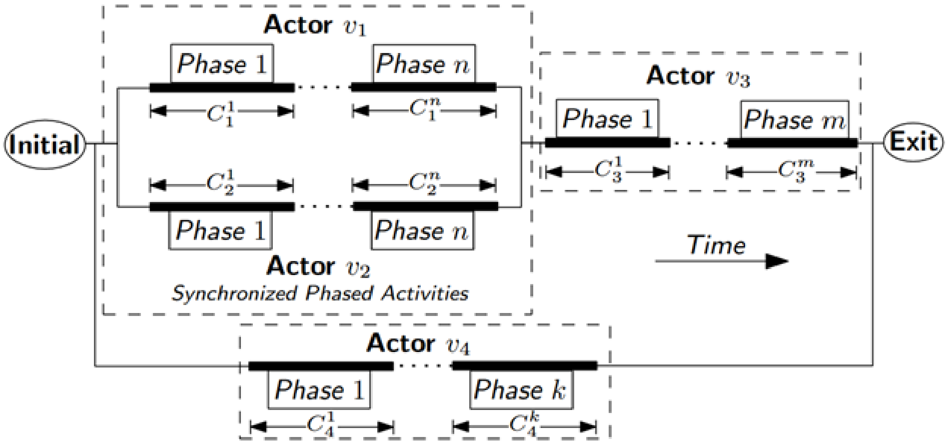


Figure 20. An illustration of the manoeuvre phase implementation within WMG-SDL level 2.

To allow for seamless use of scenarios on Safety Pool™, and to bridge the gap between WMG SDLs ease of specification and readability, and simulation support for ASAM OpenX languages, a methodology for translating scenarios specified in WMG-SDL Level 2 to ASAM OpenX equivalents, has been developed and integrated into Safety Pool™ [160]. All scenarios which have been uploaded to Safety Pool™ have been converted to OpenX equivalents, and any newly uploaded scenarios can be validated and converted into OpenX automatically. The translation itself is a semantic translation which outputs an OSC/ODR file pair for each SDL Level 2 scenario as is depicted in Figure 21. The methodology for this conversion is twofold and involves first parsing the SDL into an object structure that can be searched and manipulated, then a mapping of the WMG-SDL components to their OpenX language equivalent components is performed.

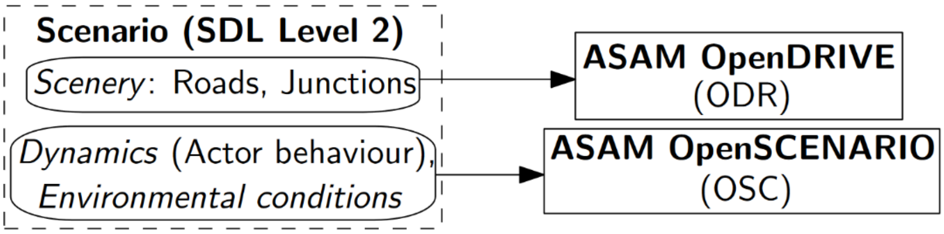


Figure 21. Illustration of the output of the translation from one WMG-SDL Level 2 to ASAM OpenX.

4.3.2 Test Scenario Selection

Test scenario selection in Safety Pool™ is achieved through use of the extended OpenLABEL scenario tagging infrastructure which has been implemented as an automated function of the database. ASAM OpenLABEL [161] provides a standard that covers the content and data structure for adequately tagging scenarios. The items within this standard are heavily inspired by the taxonomy of PAS1883 [152] on the ODD front, and extend this to include behaviour tags that cover global actions performed by vehicles within the scenarios.

Test scenario selection can be performed based on ODD elements, with the ability to include or exclude tagged items from the search. Behaviours included in the ontology of the OpenLABEL standard can be used for the same purpose. Searches can be saved and returned to, across different scenario libraries, or after additional scenarios have been added to the searchable set. A scenario defines a unique combination of the ODD attributes and the additional behaviour characteristics of the dynamic agents. Furthermore, the temporal development of all the conditions is described within a scenario. For example, the system needs to be able to handle those scenarios that fall at or within the ODD boundary, but not necessarily for the scenarios outside the boundary. The ODD boundary and the scenario parameter space are of multi-dimensional level, rather than a 2-D and 1-D representation as shown in Figure 22. Incorporating the ODD and behaviour-based model into the scenario definition, enables a coherent and effective process for the safety assurance of the system.

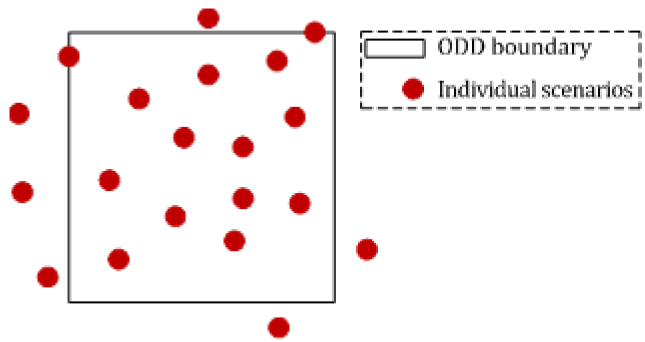


Figure 22. Illustration of the coverage of scenarios in reference to test case selection and ODD.

4.3.3 Allocation of Test scenarios to Test Instances

Test allocation is a key step within the environment element of the V&V workflow which feeds scenarios into Safety Pool, outlined in Figure 23. This step entails the allocation of test scenarios to be executed in different environments. Once the allocation, or the test plan has been created, the next step is to execute and analyse the scenario.

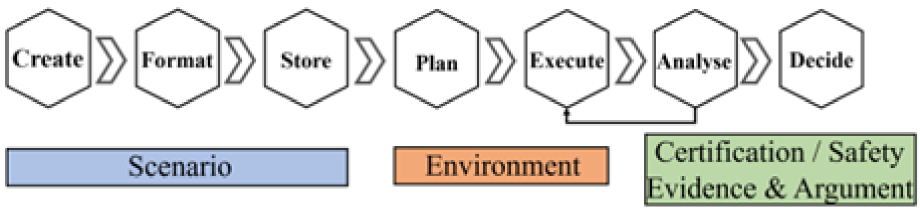


Figure 23. Multi-modal scenario-based V&V process visualised linearly, from scenario creation to analysis and decision of performance in simulation.

The certification/safety evidence & argument element contains analyse and decide. Analyse can be further divided into 3 separate stages:

1) Correct execution - whether the intended test case has been correctly executed.
2) Pass/fail assessment – monitoring the execution of the scenario and assessing the runtime output against a set of pre-defined pass/fail criteria/metrics.
3) Scenario parameter space exploration – based on the current and past concrete parameters (e.g., speed, acceleration) and the pass/fail criteria, a test case generator such as an optimisation algorithm that can be applied to introduce a new set of test case parameters with the aim of violating the scenario pass criteria.

The output from the test case generator will result in the creation of new test cases and can then be fed back into the execution module, and it forms a closed feedback loop within the workflow as shown in Figure 24. This allows the increase of scenario coverage, the decrease of the *'unknown unsafe'* region and the addition of new scenarios into the database. The final stage is the decide stage, based on whether the intended test cases have occurred, the assessment on the pass/fail criteria and the scenario coverage were achieved. This stage will determine the outcome of the whole V&V process.

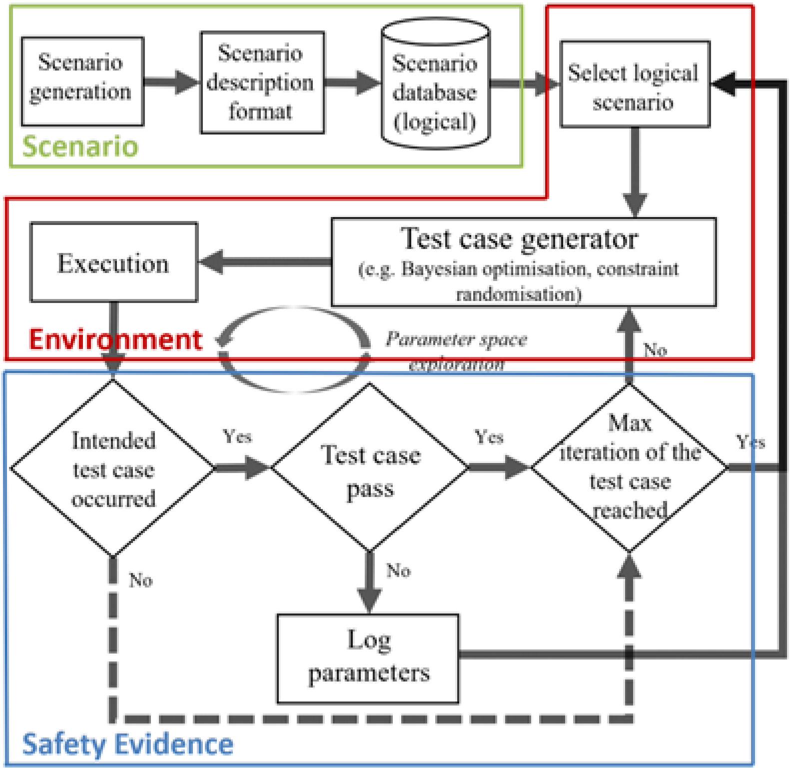


Figure 24. WMG/Uni. of Warwick's Scenario based testing framework at a logical level, which feeds into the Safety Pool^{TM} database.

By underpinning the whole workflow with ODD and behaviour elements, a coherent workflow is achieved, from scenario creation all the way to safety evidence, ensuring that all the scenarios, execution, and analysis are relevant to the testing of the system. Methodologies for test case selection, including Bayesian optimization in [121] and constraint randomization in [67], have been proposed as part of a testing framework for automated test-case generation in [162].

4.4 ADScene

ADScene is an open initiative undertaken by the two OEMs Renault Group and Stellantis. The purpose is to industrialize and complement research assets coming from research projects conducted in the French Institutes of Technologies, in the VEDECOM Institute (MOOVE projects) [163], and in SystemX (SVA, 3SA, SVR projects) [164] as well as in and the SAM Project. The latter is the French automated mobility initiative funded by the French Ministry of Economy and the Ministry of Transport under coordination or the PFA (French Automotive Platform). The initiative It started back in 2019, with the publication of a PFA position paper on the safety of ADS [165] where French automotive industry recommends to capitalize all relevant safety scenarios for ADS design & validation.

Four sources of scenarios are considered for the completeness of the scenario database as shown in Figure 25.

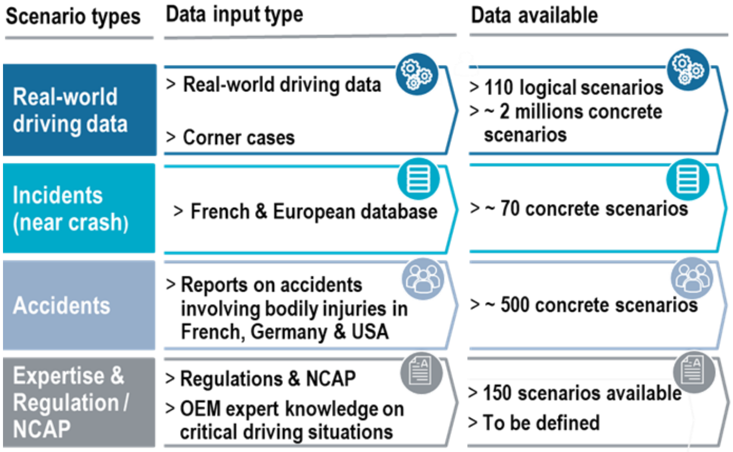


Figure 25. ADScene database content.

Scenario-based assessment is a key component of the safety argument for next generation of AD functions. However, automotive industry, regulatory bodies and researchers need to share a common view of what is a state-of-the-art scenario database. Renault Group and Stellantis are convinced that a joint, multi-partner precompetitive approach is required to generate such an industrial database. For that reason, Renault Group and Stellantis initiated the ADScene project: a scenario library for AD and ADAS leveraging research projects.

ADScene is a platform that includes (illustrated in Figure 26):

- scenarios mining from driving data,
- scenarios manager & data analysis,
- scenarios for other tools (export scenarios to other simulation or MBSE[3] software's).

[3] Model-based systems engineering

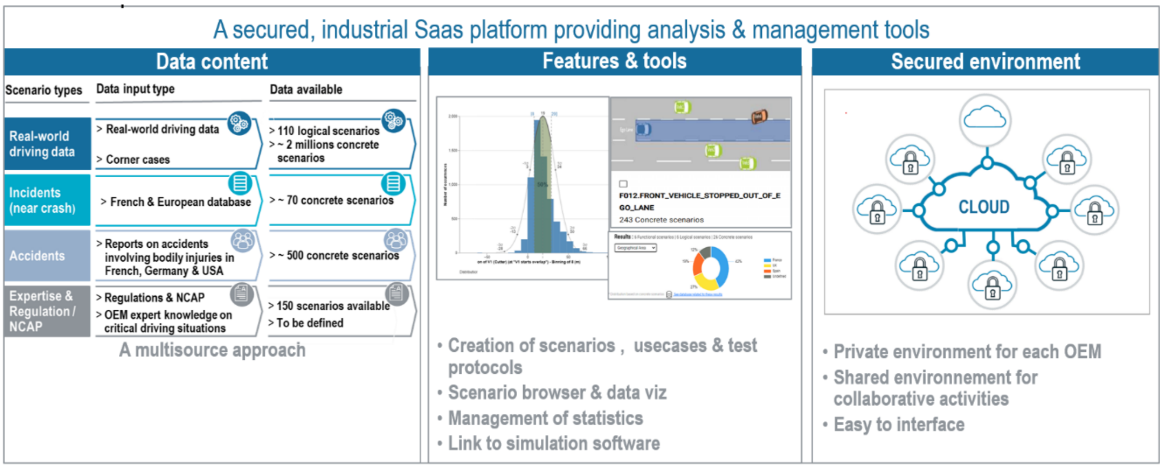


Figure 26. ADScene: A platform providing analysis and management tools.

4.4.1 Scenario concepts, parameter sets and description

In ADScene, there are “scenarios”, “use cases”, “test protocols” and “reference data”. Concerning scenarios, there are “functional, logical, and concrete scenario” descriptions, and there are simulable logical scenarios for normal driving scenarios at the moment. Associated definitions come mainly from Ulbrich et al. [154] at the moment, and ISO definitions will become the reference. “Reference data” are all relevant data to describe “scenarios”, “use cases”, and “test protocols”.

The scenario data model (the baseline data model MOOVE with its 4 layers has been developed to be compliant with Pegasus 6 layer model, and normative description of the ODD) can be represented via the following diagram shown in Figure 27 and as explained in the text.

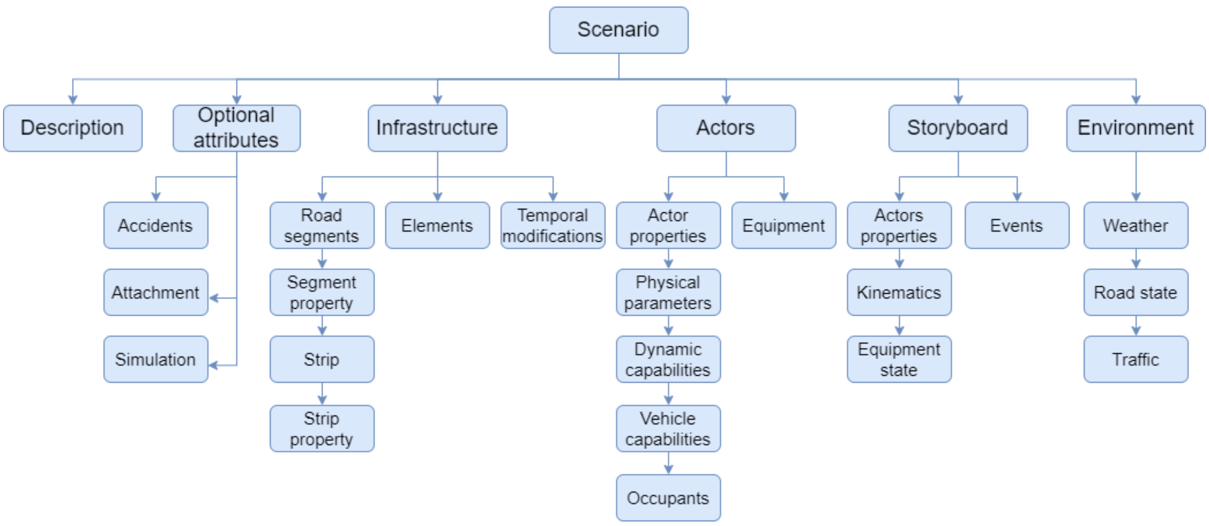


Figure 27. ADScene data model.

Because functional scenarios are classes of families of scenarios, gathered under a common name, they are only described using a picture, a description and tags for easy filtering or triggering.

For logical and concrete scenarios, a more precise description is needed to depict the temporal development between several scenes of a driving situation. That is why for this kind of scenarios, after defining its “road infrastructure” and relevant “actors”, it is necessary to describe the deferent scenes, actions and event by “steps” to produce a storyboard.

Because scenarios are often used to design or validate ADSs in different ways (depending on the companies), APIs have been created to export or import scenarios to/from other software. Examples on supported tools include simulation tools and system engineering tools. Further, ADScene includes functionalities to create use cases and test cases or test protocols.

A use case is the usage of a scenario to specify a functional behaviour or a functional limitation of a system in a particular situation. A use case is often linked to a system requirement. A test case is the instantiation of a use case with specific parameters. A Test protocol is the gathering of several test cases. For example: Euro NCAP proposes well known test protocols every 3 years. An Illustration of the relationship between these components is shown in Figure 28.

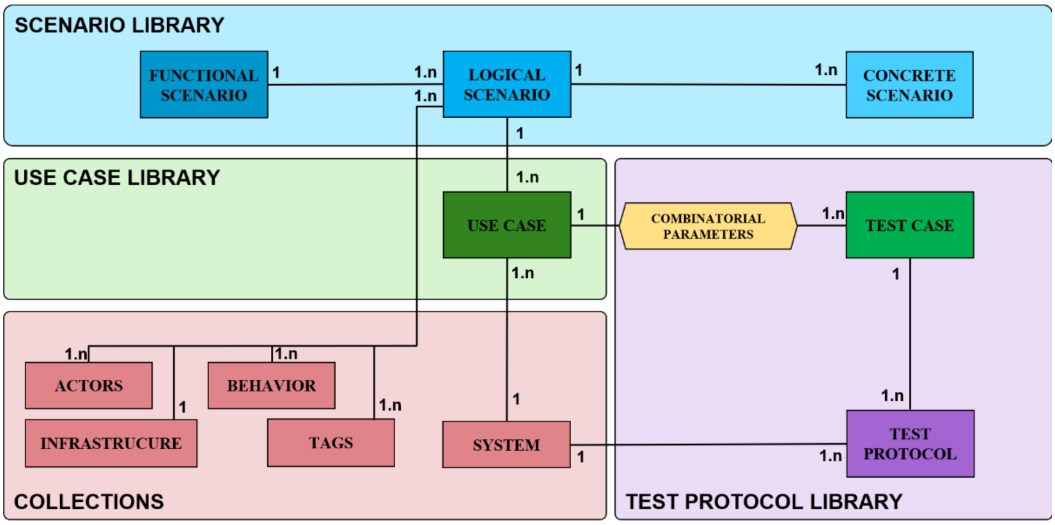


Figure 28. Illustration of the relationship between ADScene libraries.

4.4.2 Test Scenario Selection

Scenario Selection is based on several criteria that can be chosen. In ADScene selection interface, any parameter of a scenario can be used as a selection criterion, even the edition status (draft/ready/validated/obsolete), or the author of a scenario can allow scenario selection.

An important use of ‘scenario selection’ is to get interesting scenarios for an ODD, and to calculate a coverage of a scenario database. Research is still needed to identify the coverage of a scenario database selection.

Scenario Selection occurs twice: a first time to choose for a particular ADS the relevant scenarios, and a second time for a logical scenario to choose the optimized set of concrete test scenarios to be used for testing.

A research work still needs to be done to identify how to complement “ODD Based” selection which is not enough to reach completeness in term of safety design & validation for a particular automated system. The full coverage of the ADS interaction with its environment and of its responses shall be also included and is not present in the ODD. More, depending on the safety analysis, safety demonstration, and safety quantitative targets, the relevant scenarios to be considered for a particular ADS are different. Ongoing research work are conducted for scenario selection in ADScene.

Collaboration with domain experts, deep learning experts, and system engineers can greatly contribute to the success of scenario selection in ADAS/AD development. The scenario selection process should adapt to evolving technologies, regulations, and user requirements. In collaboration with VEDECOM, ADCScene has defined an approach where a new technique for scenario selection is proposed that is both goal-oriented and respects the data distribution, while optimizing the selection based on a predefined criterion. As a result, the selection varies depending on the choice made, but representativeness is always ensured.

Subspace Creation is done in different way depending on the kind of scenarios.

For “nominal” logical scenarios, concrete scenarios (about 400,000 at the moment), parameters are automatically (cf. the workflow in Figure 30) extracted from real-world driving and subspaces are defined. “Nominal" scenarios are scenarios usually well managed by human drivers in nominal conditions. PFA has published a document [164] to define and give a raw description of these nominal scenarios.

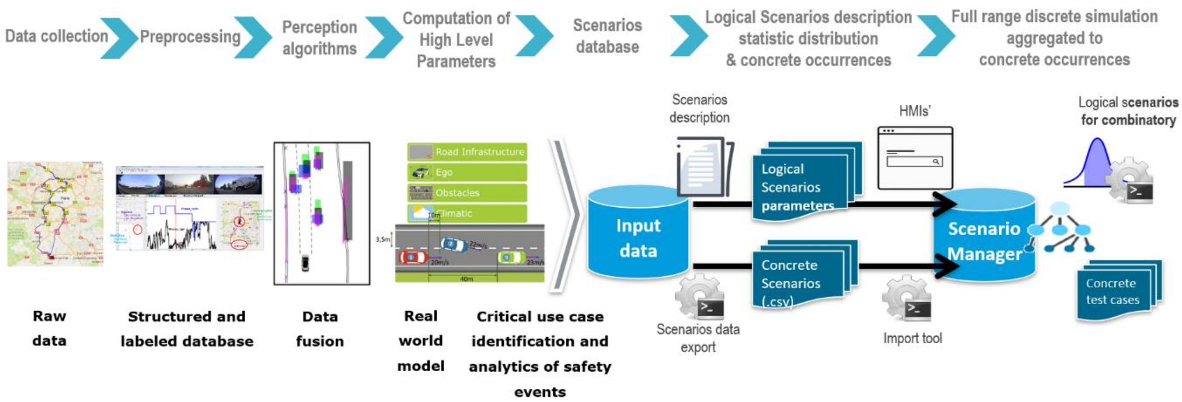


Figure 29. ADScene workflow.

For logical accident scenarios, concrete accident scenarios (250 on French highways, and about 200 on peri-urban and urban areas) are manually extracted from accident databases (VOIESUR in France, and soon other databases) to get the relevant accident scenarios. These concrete accident scenarios are characterized by occurrence, severity, and cause with the definition of the traffic feared event. This manual process is used for the majority of logical scenarios in ADScene.

ADScene continuously monitor and update the subspace boundaries and analysis techniques based on real-world feedback and performance evaluation. This allows ADScene to enhance the ability to detect and respond to critical scenarios. Different algorithms and techniques may be employed to define and analyse subspaces based on the requirements of the driving scenarios and objectives of the ADAS system. In ADScene, ongoing research use tensor form of data and variational autoencoder techniques, to find a two-dimensional space that presents all the structure of the different data.

4.4.3 Allocation of Scenarios to Test Instances

In ADScene, there are scenario descriptions, simulable scenarios and “test protocol descriptions” to help ADScene users preparing their physical or digital testing plan. Allocation of scenarios to test instance, i.e., testing environment (MiL, SiL, ViL, Tests on Tracks, Tests on open roads, etc …) can be done by using a specific tag and using ADScene API to connect ADScene to an MBSE (Model Based System Engineering) tool or to a validation plan management tool. The same API can also connect ADScene to physical or digital testing tools to import the relevant scenarios.

4.4.4 Validation Metrics for Test Runs and Dynamic Allocation

The term validation metrics can include several aspects. Validation metrics can be pass/fail criteria used in test protocols and cases. ADScene contains pass/fail criteria for some test scenarios like test scenarios from EuroNCAP or extracted from Regulations (e.g. UN Regulation).

ADScene proposes an interface and a process to create its test cases from scenarios, first by linking a scenario to a “System Under Test” and to requirements, and in a second step to define precise “concrete” parameters (e.g., speed = 15kph), and pass/fail criteria.

To measure the progress and the quality of different ADScene libraries (scenario library, use case library, test library), validation metrics have been established. Several Metrics already exist in ADScene like coverage rate, precision, recall for scenarios detection from real world driving. In addition to that, research efforts are currently underway to measure indicators such as remaining distance to cover a predefined percentage and establish confidence intervals for various quantities using the Monte Carlo method.

4.5 Hi-Drive

The Hi-Drive project [9] is addressing challenges toward the deployment of higher automation. It advances the European state-of-the-art from SAE L3 ‘Conditional Automation’ further up towards SAE L4 ‘High Automation’ by demonstrating in large-scale trials. Further, it shall enhance the robustness and reliability of AD functions in demanding and error-prone conditions. In this context, the key aim of the Hi-Drive project is to focus on testing and demonstrating AD, by improving intelligent vehicle technologies, to cover a large set of traffic environments, not currently achievable. Therefore, the Hi-Drive project enables testing of a variety of AD functions, from motorway chauffeur to urban chauffeur, explored in diverse scenarios with heterogeneous driving cultures across Europe. The project’s ambition is to extend considerably the ODD from the present situation, which frequently demands that a human driver resume control of the vehicle.

The Hi-Drive project is an ongoing project why limited information so far is available.

4.5.1 Scenario concepts, parameter sets and description

For generating the Hi-Drive use case catalogue, three types of information are provided:

- Description of Hi-Drive AD function instances: here the SW/HW of the system under test and the targeted ODD is briefly described by each AD function owner.
- Description of use cases for testing a particular AD function instance: here an abstract description of the interaction between the Hi-Drive AD function and its environment in order to reach a particular goal is provided that serves the purpose of Hi-Drive AD function instance testing.
- Description of a set of test scenarios applicable for each use case: details about actors, actions, and events for each use case are provided through a test scenario format.

Hi-Drive AD function instances are grouped per operational road environment forming four subsections, namely Motorway, Urban, Rural, and Parking AD functions.

In this deliverable (SUNRISE D3.1), the focus is on what Hi-Drive calls "Hi-Drive Use Case and Test Scenario catalogue". In its proposed methodology for generating use cases and associated test scenarios, two types of ODD conditions have been considered for testing the integration of a Hi-Drive enabler technology in the AD function instance under test: a) the ODD for testing "AD performance", in which the test is if higher AD performance and prolonged AD usage can be achieved under nominal ODD conditions, and b) an extended ODD for testing "AD availability", where additional challenging operating conditions are tested to assess the AD robustness under conditions beyond the nominal ODD of the AD system under test.

4.6 SAKURA

The SAKURA project (Safety Assurance KUdos for Reliable Autonomous vehicles)[4] is one of the coordinated initiatives funded by the Ministry of Economy, Trade and Industry (METI) of Japan, under the strategies defined by the Committee on Business Discussions on Autonomous Driving Technologies. SAKURA proposes a so-called 'Physical Principles Approach' for performing safety evaluation of ADSs. The SAKURA safety assurance methodologies largely rely on a scenario-based approach, with an emphasis on this Physics Principles Approach and a focus on developing a complete scenario generation process and tools, including a scenario database.

In this scenario-based approach, the DDT of the ADS is decomposed in:

- "Perception" – information on the vehicle equipped with the ADS under assessment in its environment as perceived by the ego-vehicle sensor system,

[4] SAKURA website: https://www.sakura-prj.go.jp/project_info/
which also includes explanations of the database tools:
https://www.sakura-prj.go.jp/project_info/tabid84.html

- “Judgement” – the control and decision logic of the ADS to provide instructions for the path and speed planning of the ego vehicle, and

- “Operation” – the instructions for the actuators for achieving the path and speed planning. In this way, scenarios are decomposed and structured in consideration of the physics of the ADS.

4.6.1 Scenario concepts, parameter sets and description

The SAKURA Physical Principles Approach leads to the distinction of the following three types of scenarios:

- Perception Scenarios with perception disturbances on the ADS sensor system, to evaluate the conditions in which the sensor system may fail.

- Traffic Disturbance Scenarios that consider traffic conditions (road geometry, ego-vehicle behaviour, and surrounding vehicle location and activities) that may lead to safety hazards.

- Vehicle Stability Disturbance Scenarios that refer to situations in which perception and judgement work correctly, but in which the ego vehicle fails to control its own dynamics correctly.

In this document (SUNRISE D3.1), the focus is on what SAKURA calls Traffic Disturbance Scenarios. SAKURA proposes a methodology for structuring scenarios as combinations of the road geometry, the ego vehicle behaviours and the positions and motions of the surrounding vehicles. A matrix results, that contains 40 possible combinations in total, among which 32 combinations correspond to scenarios that can occur in real traffic, see Figure 30.

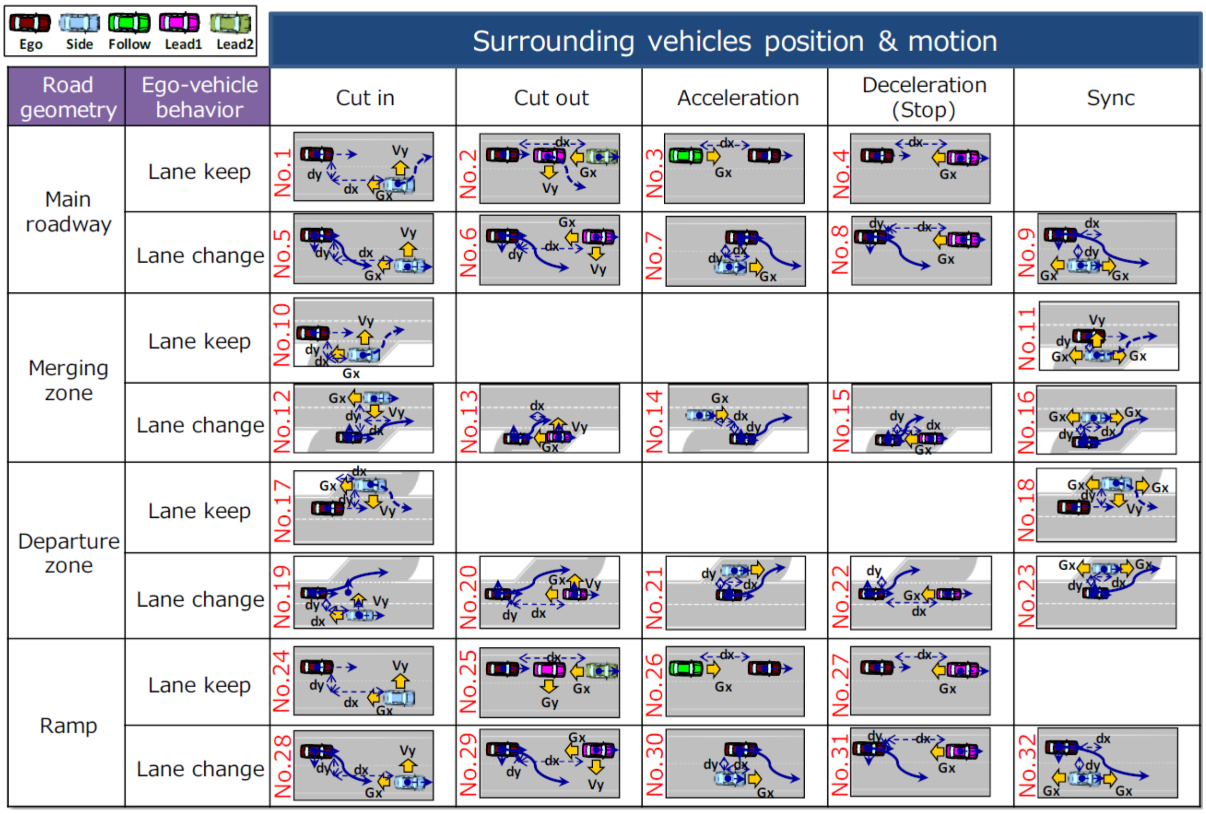


Figure 30. General vehicle traffic disturbance scenarios according to SAKURA.

Figure 31 presents the SAKURA project scenario generation and safety evaluation process using definitions for functional, logical, and concrete scenarios developed initially by the German PEGASUS project.

The three disturbance categories mentioned above, describe a systematic approach that defines all safety-relevant elements of a scenario and their combinations which represent the structure of functional scenario development.

To define logical scenarios, the assignment of parameter ranges in the functional scenarios is made. It is preferable to define these ranges by enabling a data-driven approach to extract and process vehicle trajectories from traffic monitoring data systematically. Nevertheless, the traffic data will not contain enough critical situations and crashes to address statistically significant results in the use of test scenarios that are based on traffic disturbance scenarios only. Thus, the SAKURA project has developed complementary methodologies, such as the generation of synthetic scenarios with safety-critical conditions obtained by the extrapolation of collected data [166].

Lastly, the definition of tests out of concrete scenarios is obtained by using the logical scenario parameter search engine to select concrete values from the parameter distribution. Application of other methodologies may be applied for this purpose, notwithstanding that the SAKURA project has investigated and developed several of them [93, 167, 168]. After the definition of tests out of concrete scenarios, it becomes necessary to discriminate between safe and unsafe conditions by means of safety criteria. Corresponding authorities shall define the safety criteria.

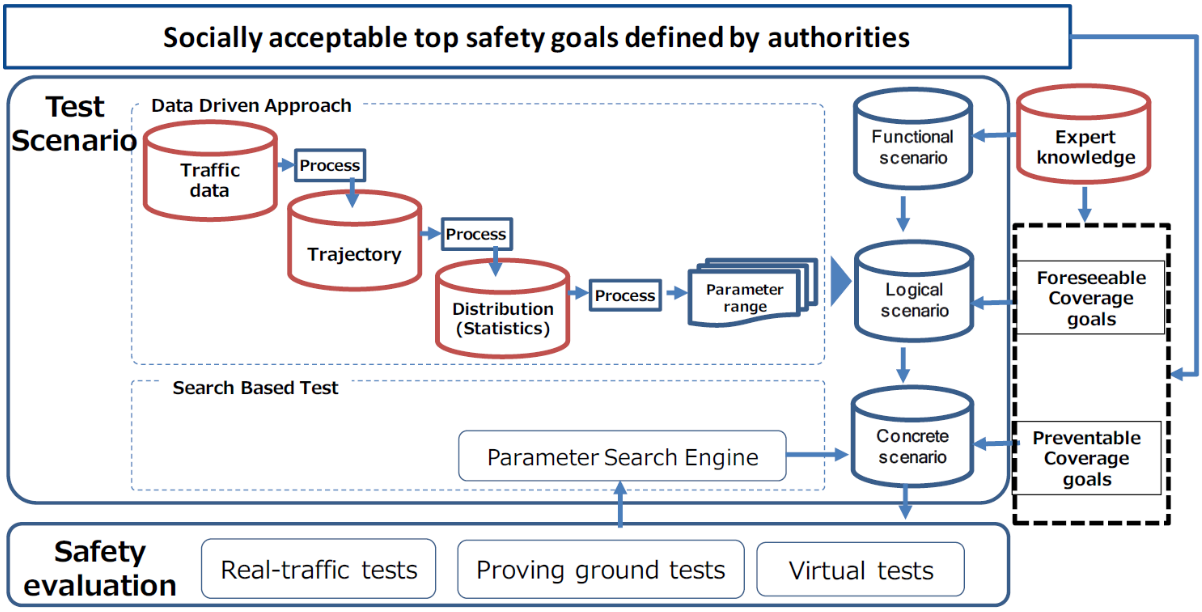


Figure 31. SAKURA project scenario generation and safety evaluation process.

4.7 CETRAN[5]

In 2016, the Land Transport Authority (LTA) of Singapore asked the Nanyang Technological University (NTU) for support in setting up a safety assessment framework to evaluate the performance of Autonomous Vehicles and to set realistic and reasonable requirements before allowing new AV solutions onto the public road. This has led to the establishment of CETRAN (Centre of Excellence for Testing and Research of Autonomous vehicles @ NTU) in the same year. CETRAN established a partnership with several knowledge institutes such as TNO, TÜV SÜD, and IRT SystemX. A partnership was also established with car manufacturer BMW Singapore.

To aid developers of autonomous vehicles, the Singaporean government initiated four working groups to draft technical references (TRs). Starting in November 2017, this led to TR68 part 1 (Singapore Standards Council's (SSC's) Manufacturing Standards Committee, "TR 68-1: 2019 Technical Reference for autonomous vehicles – Part 1: Basic behaviour," Enterprise Singapore, 2019.) till part 4 at the end of 2018. The participants to these working groups were all based in Singapore and ranged from vehicle developers, academia, and research bodies (incl. CETRAN), industry partners and various authorities.

The TR68 describes the processes to follow and the aspects to consider. They also provide a first indication of the behaviour that is expected from autonomous vehicles. They do not provide an overview of the scenarios to consider, nor the criteria for passing or failing a test. CETRAN developed a scenario-based assessment approach that works according to milestones.

[5] Centre of Excellence for Testing and Research of Autonomous Vehicles @ NTU, Singapore

Basic requirements were set for the Singapore safety assessment framework for autonomous vehicles:

- Passing the tests should allow the safe performance of a trial with the approved autonomous vehicles on public roads in Singapore, without the need for a safety driver on-board the vehicle. All aspects of operational safety need to be considered, incl. Functional Safety, Safety-of-the-Intended-Functionality, and behavioural safety on the road.
- The method should consider the specific situation on the roads in Singapore. This concerns the road and infrastructure layout, the traffic rules, the road users, and the specific Singapore weather and lighting conditions.
- The applied testing methods should be agnostic of the applied technology, and it should treat the solution provided by the vehicle developer as a black box, as the solution and the technologies applied in the solution are proprietary to the developer.
- The test methods should be optimized towards the lowest possible testing effort, for both the applicant (the vehicle developer) as well as for the assessor (CETRAN under the aegis of the LTA).
- The applied testing methods should appreciate the state of the art of fast innovation and technology development, considering that a large technology gap exists until large scale implementation on Singaporean roads.

Based on these requirements, a Singapore Milestone Testing Regime for autonomous vehicles was drafted in which three clear milestones were defined (https://www.lta.gov.sg/):

- Milestone 1: to pass M1, the developer should provide a safety demo in a closed environment such as the CETRAN AV Test Centre. It is checked whether a safety operator can take over control from the autonomous vehicle at any moment in time.
- Milestone 2: small more complex trial on the public road in areas designated by the authorities. A safety operator must be present and be able to take over full control of the autonomous vehicle at any moment in time, in case the autonomous vehicle shows any unexpected or unsafe behaviour.
- Milestone 3: the trial for passing M3 is more complex and involves public roads in residential areas. Also, for this milestone, a safety operator must be present and be able to take over control of the autonomous vehicle in case the vehicle does not behave according to expectation.

After passing M3, (small) trials on public roads may be conducted where an operator is not expected to take over control, or the operator might even be completely absent.

At CETRAN, the following scenario concept is used: A scenario provides a description of a situation that can happen or has happened in the real world. In other words, scenarios are used to describe any type of situation that a vehicle in operation can encounter during its lifetime. This includes:

- *Dynamic environment*: the manoeuvres of other actors such as vehicles, cyclists and pedestrians. The dynamic environment usually refers to those moving traffic participants that are relevant for the ego vehicle, the vehicle from whose perspective the scenario is perceived.
- *Static environment*: The static environment refers to the part of a scenario that does not change during a scenario. This includes geospatially stationary elements, such as the

infrastructure layout, the road layout, the type of road, lane markers and the road edge. Also, the presence of road furniture and buildings near the roadside that act as a view-blocking obstruction are considered part of the static environment.

- *Conditions*: Important for the description of a scenario are the conditions, such as the weather and lighting conditions, as these also have an influence on the ego vehicle. For instance, precipitation can have a large influence on sensor performance and vehicle dynamics. Lighting conditions also influence sensor performance. Cameras, for instance, might have difficulty in detecting and classifying objects during night-time in the absence of artificial light.

The definition of tests starts with the selection of scenarios that are relevant for the autonomous vehicle under assessment. An overview of possibly relevant scenario categories in Singapore is provided in [148].

Though the scenario collection is far from complete (based on specific data of Singaporean traffic was collected from the traffic rule handbook, accident statistics and observation studies on bus routes), it was demonstrated successfully that the test cases resulting from this generic test case generation method are specific for Singapore and can be tailored easily to the respective ODDs of autonomous vehicles for deployment of specific routes in Singapore.

4.8 CATARC[6]

CATARC Automotive Data of China Co., Ltd. is the leading third party, simulation & validation organization in China. It has the biggest scenario database in China for validation and verification purposes and provides related toolchain and datasets for ADS testing.

At the end of 2018, Automotive Data of China Co.,Ltd carried out a data collection campaign in China. The topics include the collection of China's naturalistic scenario data, lane-changing scenarios analysis and modelling of the lane-changing paths. The research includes the collection of driving scenario data over 100,000 kilometres in China's normal driving environment, including highways, city expressways and urban roads. All collected data were labelled to extract different scenarios, mainly lane changes. The data acquisition car was equipped with the hardware such as HD video camera, millimetre wave radar, GPS, IMU, screen, and the data acquisition software.

CATARC joined IAMTS - International Alliance for Mobility Testing and Standardization, located in Vienna. IAMTS is a non-profit association (https://www.iamts.org/). It is meant to successfully impact the automotive industry through widespread adoption of industry practices and procedures and maintaining the world-wide testbed database. CATARC is leading WG1 in IAMTS on “Global Test Scenario Library”.

[6] China Automotive Technology & Research Center Ltd, https://www.iamts.org/storage/app/media/Publications/IAMTS0002202301.pdf.

4.9 U.S. Department of Transportation

The National Highway Traffic Safety Administration (NHTSA) of U.S. Department of Transportation in 2018 published a report “A Framework for Automated Driving System Testable Cases and Scenarios” [169]. They identified conceptual ADS features and grouped them in seven generic categories aligned with the SAE levels of driving automation [170]. Out of these conceptual ADS features an ODD taxonomy was defined. For three of these ADS features, an evaluation of normal driving scenarios that each ADS feature may encounter was performed, including expected hazards and sporadic/fluctuating events. Together with identified baseline ODDs, the scenario analysis was used to identify important OEDR functional capabilities.

Existing test methods and tools were identified and evaluated, to define a testing framework resulting in three main components of a testing architecture for ADS:

- Modelling and simulation
- Closed-track testing.
- Open-road testing

A test scenario framework suitable for the test architecture was developed, that can be viewed as a multidimensional test matrix including following elements:

- Tactical manoeuvre behaviour
- ODD elements
- OEDR behaviour
- Failure mode behaviour

Test scenarios are defined at a high level by these elements. Each element can be seen as a checklist of sorts to identify the manoeuvres, ODD, OEDR, and failure mode behaviours that outlines the test setup and execution. Test procedures for sampling of defined scenarios were then developed including, among other things, information on potential test personnel, test facilities, test execution, data collection, performance metrics, and success criteria that are translated from collected data and results.

An ongoing project is the Virtual Open Innovation Collaborative Environment for Safety (VOICES) proof of concept (PoC) project [171, 172].VOICES enables their stakeholders to simultaneously interact in synchronized use cases and scenarios through the VOICES platform to research, develop, and assess transportation solutions in a distributed virtual environment producing a high-fidelity representation of the transportation ecosystem. VOICES is defined as a traffic scenario-based system that integrates a set of tools [173]. Expected built-in capabilities includes emulating and simulating infrastructure; integrating live data from real devices; traffic and vehicle simulators; replaying collected real data; and more. [173] includes figures for the VOICES high-level scenario manager architecture (including SCDBs) and examples how to use the scenario manager to begin testing. However, no more detailed information is found.

5 STANDARDISATION ORGANISATIONS

An overview of global standardisation activities relevant to the safety assurance frameworks is listed in [174] and in Table 5. Focus is placed on the ISO 3450x series, ISO 21448 SOTIF and ASAM OpenX Standards in the context of scenario-based testing. This focus stems from their direct relevance to this testing approach and alignment with the SUNRISE goals. ISO 26262 [175] is the functional safety standard for road vehicles and must always be considered, but it does not focus on scenario-based testing and therefore is left to the interested reader. Other standards still have value in different contexts or for broader applications.

Table 5. Standards relevant for safety assurance frameworks (based on [174]).

Standard	(Series) Name	Comment	Rel.*	References
ISO 3450x series	Road vehicles — Test scenarios for automated driving systems	See Sec. 5.1.	TD PM CL	[1, 2, 176, 25, 177]
ISO/SAE PAS 22736:2021	Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles	Aligned with SAE J3016:2021	TD	[170, 178]
ISO 21448:2022	Road vehicles — Safety of the intended functionality	See Sec.5.2.	TD PM	[179]
ISO 26262:2018	Road vehicles — Functional safety		TD PM	[175]
ASAM OpenX standards		See Sec. 5.3.	TD PM	[180]
ANSI/UL 4600:2022	Standard for Safety for the Evaluation of Autonomous Products		CL	[181]
BSI Flex 1889	Natural language description for abstract scenarios for automated driving systems – Specification	See Sec. 5.3.6.	TD	

* TD = Terms and definition, PM = Process and methods, CL=.Check list

5.1 ISO 34500 Test scenarios for automated driving systems

Starting during 2022, ISO is publishing the ISO 34500 standard series "Road vehicles – Test scenarios for automated driving systems". An overview of the series is shown in Figure 32.

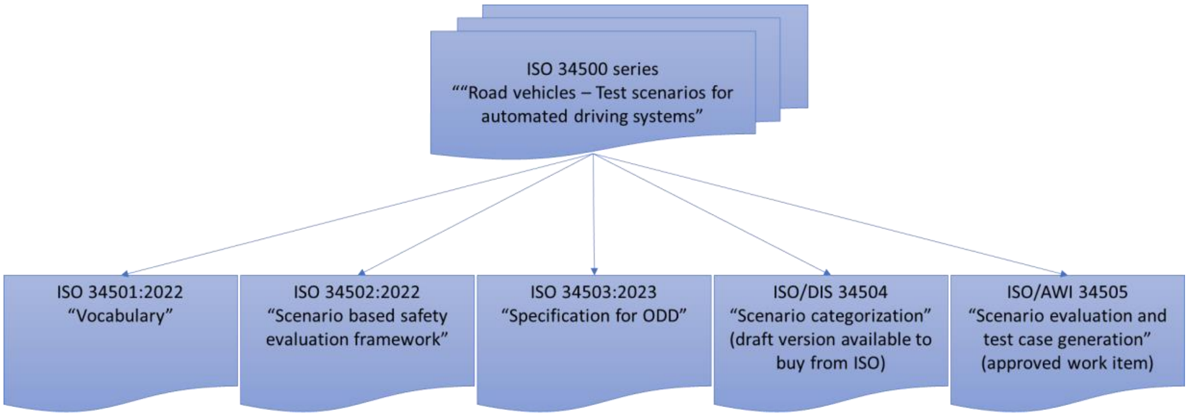


Figure 32. Overview of the ISO 34500 series [1, 2, 24, 25, 177].

5.1.1 ISO 34501 Vocabulary (Terminology)

ISO 34501:2022 [1] defines terms in the context of test scenarios for ADS. The defined terminology is closely associated with standards like ISO 21448:2022 [179] and ISO 26262:2018 [175] but in ISO 34501 the terms are interpreted in the context of test scenario description.

5.1.2 ISO 34502 Scenario-based safety evaluation framework

A scenario-based safety evaluation framework is defined in ISO 34502:2022 [2]. The overall safety task "Identification and risk evaluation of hazardous scenarios of the ADS" is illustrated in Figure 33 and the scenario-based strategy is further broken down to subtasks in Figure 34.

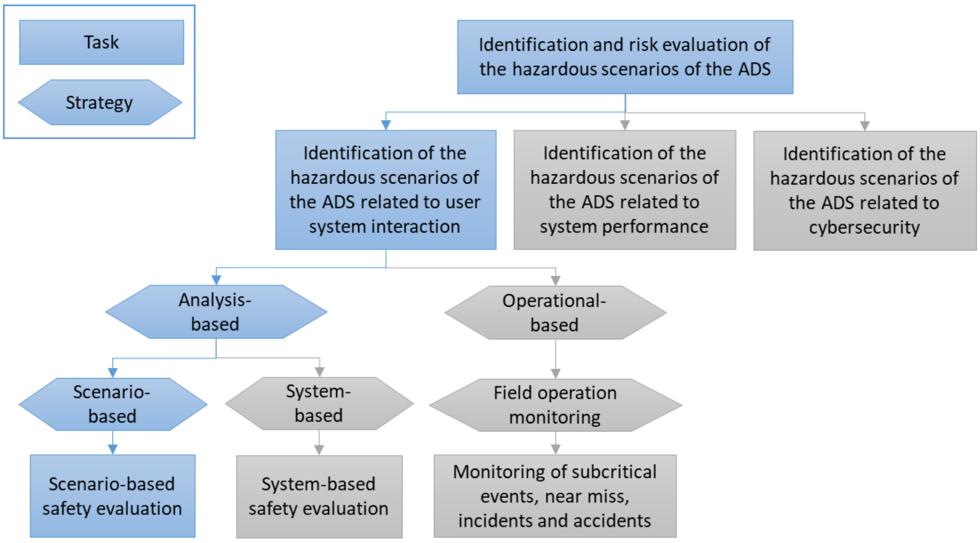


Figure 33. Illustration of the breakdown of the overall safety task to subtasks (based on [2]).

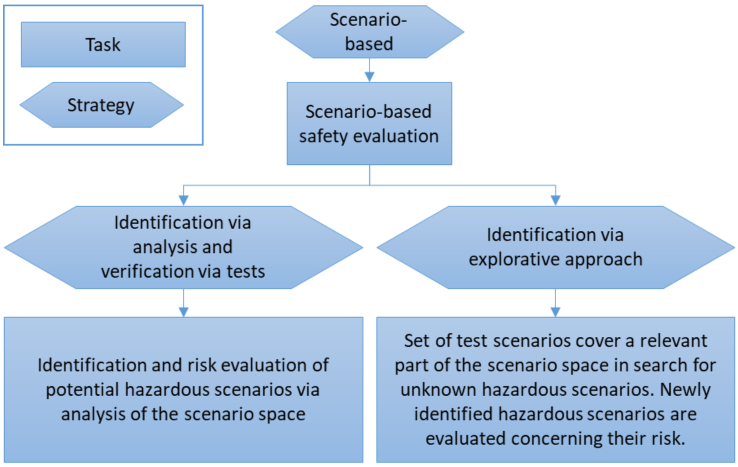


Figure 34. The scenario-based strategy proposed by ISO 34502:2022 [2].

Using the presented approach, the relevant scenario space is analysed to identify risk factors. Considered are general physical limitations like that a sensor has a field of view based on the physics of the systems. Other implementation specific issues like limitations of the ML algorithm to correctly classify objects sensor failures due to random hardware faults are neglected. The overall process of ISO 34502:2022 is shown in Figure 35. The grey parts are not covered by ISO 34502:2022.

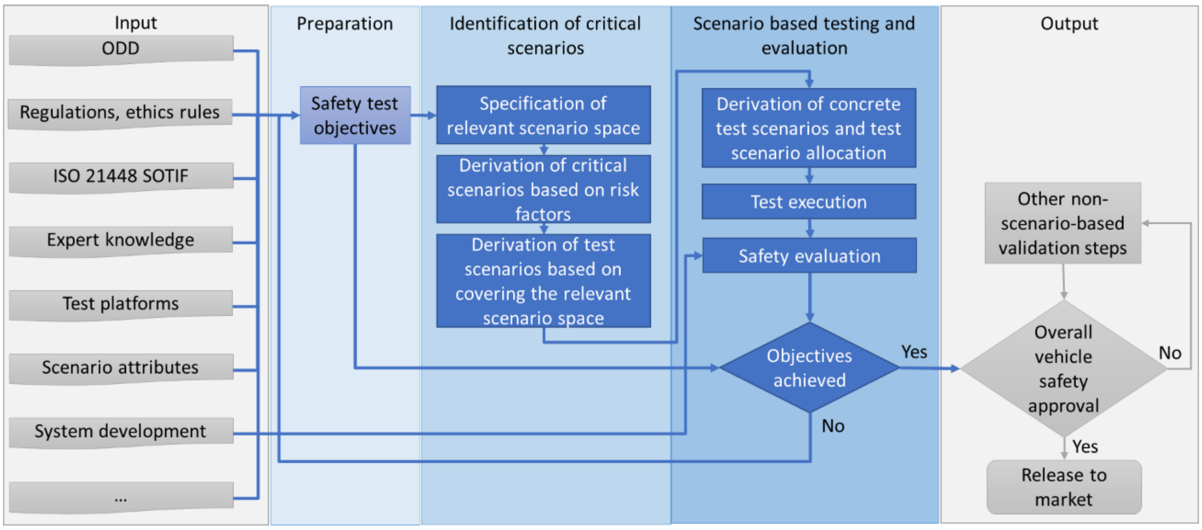


Figure 35. The process steps defined by ISO 34502:2022 [2].

5.1.2.1 Preparation: Safety test objectives

An important part of the ISO 34502 process is to define the safety test objectives based on the input information shown in the grey box on the left in Figure 35. The safety objectives are derived using general risk acceptance criteria principles like ‘as low as reasonable practicable’ (ALARP), ‘minimal endogenous mortality’ (MEM), ‘positive risk balance’ (PRB), and applicable regulations. Either they are derived from or provided by external sources like ISO 21448 [179] or by related regulations like UNECE regulation 157 [144]. The safety objective should be defined such that the overall safety argumentation of the ADS is supported by their fulfilment.

5.1.2.2 Identification of critical scenarios

Specification of relevant scenario space

Next step is to specify the relevant scenario space, i.e., the possible scenarios the ADS can encounter considering the specified ODD and possible manoeuvres of the ADS. Functional, abstract, logical, and concrete scenario definitions can be used.

Derivation of critical scenarios based on risk factors

The defined relevant scenario space is analysed to identify risk factors, that are used to determine critical scenarios.

Derivation of test scenarios based on covering the relevant scenario space

After critical scenarios have been identified, a set of test scenarios is derived such that the relevant scenario space is sufficiently covered.

5.1.2.3 Scenario-based testing and evaluation

Derivation of concrete test scenarios and test scenario allocation

In this step, general requirements are defined for testing concrete scenarios. Further, guidance shall be provided for the allocation of test scenarios to different test platforms and general capability requirements for V&V tolls are defined. The latter includes, e.g., simulation/virtual testing platforms (VTP), test-track platforms and real-world platforms, all having different capabilities related to accuracy, repeatability, and traceability.

- Parameter ranges and their combinations shall be defined for testing in order to achieve sufficient test coverage of the scenario space for the required safety argumentation.
- Concrete parameter values and combinations shall be defined based on relevant safety test objectives.

All identified test scenarios shall be allocated to at least one test platform based on their suitability, results shall be traceable, and VTP and test-track platforms shall deliver the same repeatable and reproducible results within reasonable tolerances.

Test execution

All test scenarios defined in previous steps shall be executed and sufficient coverage, according to the previously defined safety test objectives, shall be ensured.

- Used VTP tools shall fulfil capability requirements.
- Test scenarios executed on test tracks shall be replicated with sufficient accuracy, all tools shall fulfil capability requirements, measurement equipment shall be qualified, and all behaviour of dynamic entities shall be sufficiently documented.
- Real-world tests should be set up considering, e.g., guidelines and limitation for route selection, weather, surrounding conditions.

5.1.2.4 Safety evaluation

General requirements should be defined to evaluate each test scenario as well as for overall risk evaluation. Pass/fail criteria shall be adopted form the safety objectives.

5.1.3 ISO 34503:2023 Specification for ODD

ISO 34503 is about specification of the ODD in the context of test scenarios for ADSs [24]. As the standard is scheduled to be published in June 2023 all information is taken from the Draft International Standard (DIS) version [182].

The ODD attributes and the definition of the attributes play a key role in scenario-based testing. The ODD definition needs to be testable, should be precise and detailed. It is up to the end user, to do any abstraction for the own test scenario.

The ODD must be readable for both experts and non-experts. At the top level, the ODD shall be classified into the following attributes: scenery, environmental conditions, and dynamic elements.

These top-level attributes are then divided into more detailed attributes:

- The scenery elements consist of following attributes:
    - Zones, geo-fenced areas, zone type, region.
    - Drivable area: type, geometry, lane specification, signs, edge, surface.
    - Junctions: roundabouts, intersection.
    - Road structure: building, bollards, streetlight.
    - Special structures: tunnel, toll plaza, bridge.
    - Temporary structures: roadwork, detour.

- The environmental conditions consist of:
    - Air temperature.
    - Weather: wind, rainfall, snow.
    - Particulates: smoke, dust.
    - Illumination: day, night, cloudiness, artificial.
    - Connectivity: V2V, V2Iinfrastructure, V2Pedestrian, V2Network, Positioning GPS.
- The dynamic element consists of traffic agents and subject vehicle.

Finally,each attribute is given a specific value like light air 0.3 m/s – 1.5 m/s, number of lanes or position of the sun.

5.1.4 ISO/DIS 34504 Scenario categorization

ISO 34504 aims to define scenario categorization in the context of test scenarios for automated road vehicles. The standard is currently in Draft International Standard (DIS) version ISO/DIS 34504 [25]. A scenario category refers to a set of scenarios that share one or more characteristics like exemplified in Figure 36. The scenario categorization can be used to structure various test cases.

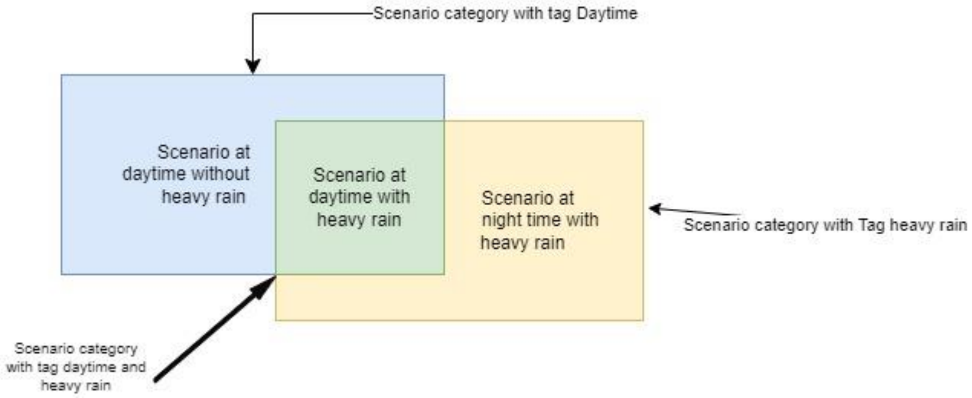


Figure 36. Simple example of relation between scenarios and scenario category (based on [25]).

The proposed approach for categorization of the scenarios is by providing tags that carry information about the scenarios:

- A tag should indicate the purpose of the tag like dynamic entities, scenery, environmental conditions, scenario assignment, intended test usage.
- The tags shall also be structured into trees where each layer represent a different abstraction level.
- There should also be a topic for the tag.

5.1.5 ISO/AWI 34505 Scenario evaluation and test case generation

ISO 34505 is a standard under development that will define a methodology to evaluate the test scenarios and provides a procedure extending test scenarios to test cases.

5.2 ISO 21448 Road vehicles - Safety of the intended functionality

ISO 21448:2022 Road vehicles - Safety of the intended functionality [179] addresses the concept of Safety Of The Intended Functionality (SOTIF), which pertains to mitigating unreasonable risks due to hazards arising from functional deficiencies or foreseeable misuse. This standard establishes principles for evaluating the performance of the intended functionality, defines crucial safety-related terms and concepts, and strives to enhance consistency and transparency in the development and assessment of ADSs. The concepts outlined for evaluating performance are particularly pertinent to the subject of SAF. One important part of the SOTIF process is the identification of unknown scenarios.

5.3 ASAM OpenX Standards

ASAM stands for the Association for Standardisation of Automation and Measuring Systems, the ASAM OpenX® series of standards in the domain "Simulation" aims to provide a complete set of standards for simulation-based testing of automated driving functions [180]. The standards are developed to provide for a wide range of use cases for virtual development, addressing scenario description format, ODD content, and scenario labelling, all to enable hybrid testing approaches for virtual and physical components in ADS testing. ASAM has identified a need for tools to support the implementation, training, and use of the above standards. As a result, ASAM has implemented an open-source platform to host and share ASAM compatible tooling that help to better understand the standards and facilitate their usage.

5.3.1 ASAM OpenSCENARIO

ASAM OpenSCENARIO defines a data model and format for describing scenarios used in driving and traffic simulators, as well as in automotive virtual development, testing, and validation. The primary use case of ASAM OpenSCENARIO is to describe complex, synchronized manoeuvres that involve multiple entities, like vehicles, pedestrians, and other traffic participants.

ASAM OpenSCENARIO exist currently as a two-language concept, ASAM OpenSCENARIO V1.X has existed since early 2021 and as a result has seen a great deal of tool support from industry and simulation environment developers. ASAM claim that the current language capability, following the V1.2, release is covering both concrete and logical scenario descriptions. The language format defined in the standard is Extensible Markup Language (XML) and the core of the specification is an XML schema file.

ASAM OpenScenario 2.0 is a domain specific language (DSL) that has been developed as a superset of the existing OpenSCENARIO 1.x language concept. It has been extended to include the support for defining abstract scenarios.

Concrete and logical scenarios can be expressed in both versions, however ASAM OpenSCENARIO 2.x offers additional features that enrich the scenario description. As ASAM OpenSCENARIO 1.x does not support abstract descriptions, it has no direct equivalent to

ASAM OpenSCENARIO 2.x. ASAM have planned to converge the language concepts in 2024, as can be seen in Figure 37

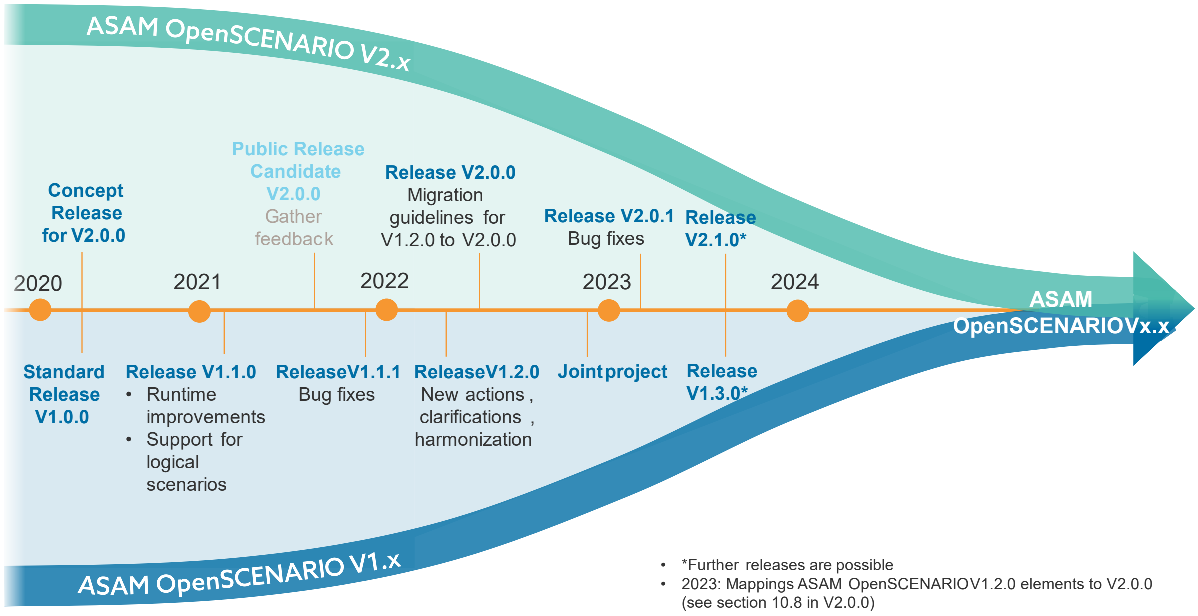


Figure 37. ASAM OpenSCENARIO convergence roadmap [180].

5.3.2 ASAM OpenDRIVE

Utilizing XML syntax with the file extension xodr, the ASAM OpenDRIVE format serves as a universal foundation for representing road networks. Information within an ASAM OpenDRIVE file encompasses road, lane, and object geometry, encompassing road markings and roadside features such as signals. The described road networks can be either artificially generated or based on actual data. The primary aim of ASAM OpenDRIVE is to supply road network descriptions for simulation, facilitating the development and validation of ADAS and AD functionalities. This allows seamless interchange of these descriptions among different simulators. By offering a standardized road description format, the industry can minimize expenses related to creating and converting these files for developmental and testing purposes. Road data may originate from road network editors, map data conversions, or scanned real-world road conversions.

The ASAM OpenDRIVE format captures static road network elements essential for realistic vehicle simulation. ASAM OpenSCENARIO defines dynamic simulation content like vehicle manoeuvres.

5.3.3 ASAM OpenLABEL

ASAM OpenLABEL provides an annotation format and labelling methods for scenarios and objects. Developed with a focus on multi-sensor data tagging and scenario tagging, the user guide provides the method for which the standard should be used. The categorisation and description of the building blocks of any ADSs perception stack was taken as an input to the development of the concept as it’s through this lens that an understanding of the status of its surroundings can be inferred. Using a standardized format helps cut costs and save resources used in creating, converting, and transferring annotated and tagged data. ASAM OpenLABEL is represented in a JSON format and can therefore be easily parsed by tools and applications.

The scope of the standard is to:

- specify the annotation schema for which any ASAM OpenLABEL compliant annotation will contain,
- represent the annotation schema as a JSON schema for ease of use and transfer,
- explain the relationship between defined elements of the schema, for example, actions, objects, events, contexts, relations, frames, tags,
- and to give guidance on the use of the standard.

The advent of scenario databases for storing multi-sensor data, annotated multi-sensor data, simulation scenarios, and test scenarios has resulted in vast stores of data which need to be meaningfully organised. The sensor data and scenarios stored in these databases must be organized and tagged using semantic, meaningful tags. These tags refer to the content of the data, from its ODD and the high -level behaviour of the dynamic agents to basic administrative information. Tagging scenarios as a concept means the enrichment of raw data by adding metadata. Scenario tagging based on the OpenLABEL standard addresses the issue of the difficulty of labelling from scenario artifacts in various scenario definition languages.

Scenario tagging in OpenLABEL allows for: standardised clustering of test scenarios in a scenario database, facilitating scenario storage systems that are separate to the format of the scenario itself (definition language), enabling the efficient search and filtering of test scenarios in scenario databases, enable the sharing of scenarios between systems that may not have the ability to inspect the scenario definition or underlying scenario data, improving the maintainability of scenarios and scenario data, additionally enabling specific machine-learning classification tasks to be performed on scenario data.

The ASAM OpenLABEL tags are organized into three categories which can be used to describe different aspects of a scenario.

- Operational Design Domain (ODD) tags: ODD tags describe the environmental conditions and road features present in a scenario, such as rainfall and junction. The ASAM OpenLABEL ODD tags are aligned with and share their definitions with the BSI PAS 1883 ODD Taxonomy [10].
- Behaviour tags: Behaviour tags describe the types of road users and the behaviours exhibited by them in a scenario, such as a pedestrian who is walking.
- Administration tags: Administration tags describe the qualities of a scenario which cannot or may not easily be derived from a scenario, such as the creation date of a scenario.

5.3.4 ASAM OpenODD

ASAM OpenODD is not yet a finalized standard but a concept, which serves as basis for a future standard which is currently in development. The aim is to provide a format that is capable of representing a defined ODD for connected and automated vehicles (CAV). Specifically, the project would like to create a machine-interpretable format to represent ODD specification. This format allows an ODD description to be exchangeable, processable, and comparable.

During the concept project, the following aspects were addressed:

1. Attributes**:** Provision of a base set of relevant attributes for the ASAM OpenODD format.
2. Specification: Development of semantics and syntax for the ASAM OpenODD description language, also enabling the use of different ontologies/taxonomies for the definition of ODDs.
3. Metrics: Evaluation of the possibility of measurable metrics and what the ODD needs to be able to represent, so any application can perform analysis on the ODD.
4. Representing Uncertainty**:** Representation of uncertainty with the goal to enable the ODD format to handle rare events and misuse.

5.3.5 ASAM OpenXOntology

ASAM OpenXOntology is a project in the concept stage which presents an ontology to provide a common foundation for definitions, properties and relations of central concepts used throughout the ASAM OpenX® series of standards. Its relationships to these are demonstrated in Figure 38.

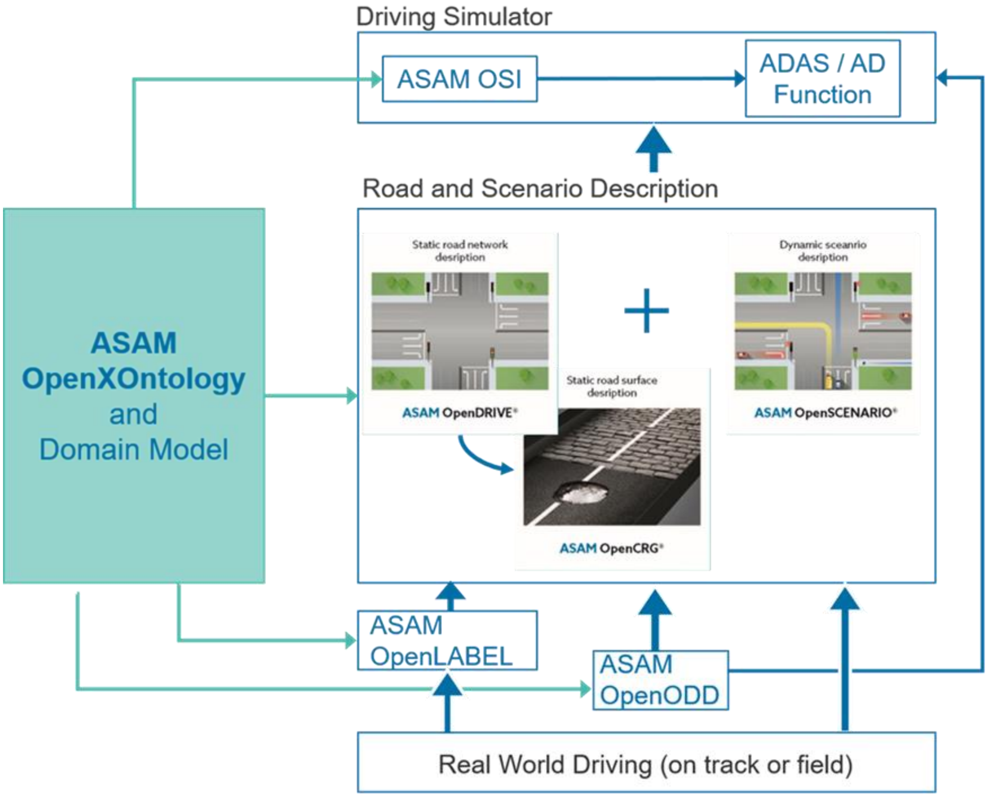


Figure 38. ASAM OpenXOntology and its relation to other ASAM standards.

The outcome of the project is the demonstration of what an ontology should contain and the format that an ontology should take for best representing the domain of simulation. This resulted in an Ontology file for the existing concepts of the domain in OWL format, which is openly accessible, reference documentation which accompanies the ontology, and concrete examples of using, extending, and integrating the application to demonstrate the capability of the ontology (contained within the specification and user guide).

5.3.6 BSI Flex 1889

BSI Flex 1889 v1.0 outlines structured natural language requirements for defining test scenarios of ADS at Level 3 and higher. It establishes mandatory and optional attributes for abstract scenario descriptions and presents scenario-based testing concepts. The standard details the aims of using a natural language in scenario definition to attain the abstract classification for scenarios. It includes the syntax and semantics for a natural language representation that should be followed to comply with the standard, as long with a guide for creating an abstract scenario representation.

This standard aims to establish a shared language for test scenario definition, promoting efficient communication among organizations and aiding automated vehicle manufacturers and developers in gathering evidence to enhance product safety confidence. It's intended for test, audit, and assurance organizations, as well as ADS manufacturers, suppliers, and various professionals like regulators, development engineers, test engineers, scenario editors, and more, serving as a valuable tool in the ADS testing ecosystem.

6 OTHER RELATED INITIATIVES

6.1 Consumer testing - Euro NCAP

Euro NCAP (European New Car Assessment Programme) faces challenges in providing meaningful advice to consumers as safety equipment becomes more prevalent on the market. Rapid technological advances, such as AI and over-the-air software updates, challenge established safety testing traditions. Euro NCAP believes it holds the potential to improve vehicle safety in the next decade further to support Vision Zero.

Euro NCAP has set a vision for 2030, including several ambitious goals to improve vehicle safety across Europe. The primary aim of Euro NCAP is to reduce the number of road fatalities and serious injuries, and the vision reflects this objective.

The key elements of Euro NCAP's vision [156] are:

- Zero road fatalities and/or severe injuries in new cars sold in Europe by 2030.
- An increased focus on VRUs, including pedestrians, cyclists, and motorcyclists.
- Integrating new and emerging vehicle technologies into the assessment process, such as ADSs and electric powertrains.
- A more comprehensive and holistic approach to vehicle safety includes crashworthiness, active safety features, and driver assistance technologies.
- Improved testing methods better reflect real-world crash scenarios, including more diverse crash tests and evaluations of occupant protection in different seating positions.
- Greater transparency and accessibility of safety ratings help consumers make informed purchasing decisions and encourage the adoption of safer vehicles.

Scenario-based testing is essential in the Euro NCAP vision of testing and evaluating a vehicle's performance in different real-world crash scenarios. The goal of the Euro NCAPs testing vision is to evaluate a vehicle's safety features and performance in a range of different scenarios, including:

- Intersection collisions: This includes testing how the vehicle performs in collisions with other vehicles at intersections, such as T-junctions or roundabouts.
- Pedestrian and cyclist collisions: Euro NCAP also evaluate a vehicle's ability to detect and avoid collisions with pedestrians and cyclists, which are the most vulnerable road users.
- Lane-change and overtaking types of testing assess how well the vehicle can respond to unexpected situations, such as a sudden lane change or an overtaking manoeuvre by another vehicle.
- Rear-end collisions: This involves testing a vehicle's performance in rear-end collisions, the most common type of accident on the road.

While traditional crash tests are designed to assess a vehicle's performance in specific controlled conditions, scenario-based testing is intended to replicate more complex and diverse scenarios likely to occur in real-world crashes. From what data source the scenarios will be derived is unclear, but the general sentiment of scenario-based testing, based or derived from a database in the vision is very much in line with the spirit of the SUNRISE project. By evaluating a vehicle's performance in different real-world situations, Euro NCAP endeavours to provide a more accurate reflection of how well it protects its occupants and avoids accidents in the real world. The increased fidelity and complexity in the testing are also envisioned, to encourage the development of new safety technologies and help to drive improvements in vehicle safety standards.

6.2 Public authorities and policy makers - UNECE

The UN Regulation 157 [144] is the first regulatory step for an ADS in traffic and therefore it provides innovative provisions aimed at addressing the complexity related to the evaluation of the system safety. It contains administrative provisions suitable for type approval, technical requirements, audit and reporting provisions and testing provisions. This regulation includes general requirements regarding the system safety and the failsafe response for an ALKS (Automated Lane Keeping System). The regulation also contains a guidance on traffic disturbance critical scenarios for ALKS in Annex 3 which clarifies the derivation process to define conditions under which the ALKS vehicle shall avoid a collision. In Annex 4 of this regulation, the special requirements to be applied to the functional and operational safety aspects of ALKS are described, and in Annex 5, the specifications for track testing of ALKS vehicles are described. In this annex, the track tests with the purpose to verify the technical requirements on ALKS are defined. Additional specifications for public road testing of ALKS are defined in Annex 6 of this Regulation if the tests of the technical requirements are successfully passed. This regulation with its testing and assessment method works very well for a system like ALKS with a clearly defined ODD in a not too complex traffic environment, where all situations and requirements to the system can be discretely described.

New Assessment/Test Method (NATM)

The automotive industry is rapidly evolving with the increasing adoption of connected and automated, autonomous, and connected vehicles. This technological transformation is creating a need for a new assessment method that can effectively relate individual test results to the remaining tests and results. The UNECE Working Party on Automated/Autonomous and Connected Vehicles (GRVA) is on the way to introduce a "multi-pillar" approach in response to this challenge.

Validation Method for Automated Driving (VMAD) working group was instructed to develop the New Assessment/Test Method for Automated Driving (NATM) [183] guidelines that could provide direction to developers and contracting parties of the 1958 and 1998 UN vehicle regulations agreements on recommended procedures for validating the safety of ADSs.

The document provides a high-level framework for the NATM, outlining the scope and general overviews of the scenario catalogue and each pillar (simulation/virtual testing, test track, real-world testing, audit/assessment, and in-use monitoring) as the overall process of the NATM.

There is still much work left to make the NATM practically useful as it only covers the high-level goals and methods many details are still to be addressed. Therefore, VMAD continues to develop the elements of the NATM, and FRAV (Functional Requirements for Automated and Autonomous Vehicles) continues to develop safety requirements for ADS. Progress and alignment will be reflected in future releases of the NATM. Once the NATM has matured to include evaluation criteria based on performance requirements, it is anticipated to support the validation process with guidelines and/or regulations/requirements.

Under these new envisioned guidelines, a more comprehensive set of processes and requirements are expected than in traditional safety assessments of vehicles. The focus will be on a range of items, including a scenario catalogue, simulation and virtual testing, track testing, real-world testing, audit/assessment, and in-service monitoring and reporting. Moreover, it is expected that topics can only be addressed by combining the above items. For example, trustworthiness requires simulation/virtual testing, track testing, and real-world testing, while the qualification of tools necessitates audit/assessment, testing, and in-service monitoring. Coverage of hazardous scenarios, on the other hand, calls for audit/assessment of representative scenario test catalogues, coverage of the intended traffic environment, and sufficient exploration of unknown hazardous scenarios, needs real-world testing and in-service monitoring. The complex method will need refinement and is anticipated to be supported by the efforts of the SUNRISE project. The advancements made by the project, aim to help to provide the technical service that is crucial in making the assessment of ADS feasible and practical, in order to ensure that vehicles on the road meet the highest safety and reliability standards.

6.3 Other academic works

Young-Min Baek et al. [11] perform a review to analyse and identify conceptual variables related to scenario methods, which are used to capture and communicate specifications to better understand problems among different stakeholders. Through a semi-systematic literature review, the study collects data, concepts, and values of scenarios or scenario methods to define and classify scenarios based on the maturity level. For this, they define Scenario Variables (SVs) as the basic building block of any scenario.

SVs are any concepts related to scenarios or scenario methods that can have a concrete value (or a set of values). They are used to provide overall information and understanding of a scenario method and are classified into four levels of constructs (method level, suite level, scenario level, and event level).

Based on their classification of SV, Young-Min Baek et al. proposes a Conceptual Scenario Model (CSM). The CSM is a metamodel for specifying scenarios that includes four key meta-classes to satisfy scenario requirements: Goal/Hypothesis, Path/Flow (i.e., course of events, possibilities), Context, and Constituent Events. In addition, it contains the following meta-classes: World-of-Interest (WoI), Situation, Scene, and Dynamics. The CSM is a framework for understanding and evaluating different scenarios that can be used in decision making. It consists of a set of conceptual variables that are used to describe a scenario, such as the number of participants, the goals of the scenario, the strategies and tactics used, the expected outcomes, and the possible impacts. These variables are then used to analyse a scenario and make decisions about how to best execute it. The model can also be used to identify potential areas of improvement within a given scenario and to suggest alternative scenarios that could be better suited to a given situation. This model provides a basis for developing a scenario specification method for a specific engineering purpose or application domain.

Sun et al. [184] reviewed fifty well-received works on test automation of highly automated vehicles. The reviewed methods were categorised into three major groups. First, coverage-oriented methods aimed to maximise testing coverage. Secondly, unsafe-scenario-oriented methods are capable of generating/finding high-risk, boundary, collision, and worst-case scenarios to provoke fault detection. And lastly, naturalistic assessment methods generate scenarios per naturalistic distributions and can estimate safety indicators such as injury rate, conflict rate and collision rate. In the reviewed works, the unsafe-scenario-oriented test automation research accounts for the most significant proportion of contributions, followed by naturalistic assessment-oriented test automation research. The research on coverage was relatively meagre. With HighD [185] as a data source, Sun et al. extracted one simple car-follow scenario and one slightly more complex cut-in scenario to be used as test vectors for the methods under evaluation. A weakness in the evaluation is the simple function used and the lack of ODD interaction. Based on the evaluation of seven methods effectiveness and efficiency to attain three constructed test purposes, Sun et al. give recommendations on method use, connected to the complexity of the Scenario, in conjunction with the test purpose. In general, complete enumeration or T-wise methods are recommended for coverage purposes, and adaptive search methods to find test vectors that find faults by provoking the system. Recommendations regarding evaluating high-level safety indicators depend on occurrence frequency. Monte Carlo for higher, and importance sampling for lower occurrence frequencies.

Re et al. [186] has tested and compared two different lane departure warning systems currently on the market and found that even though the systems adhere to the exact system-level requirements, the implemented systems differ substantially in a real-world setting. To enable comparison between systems, a robustness index is introduced. Although the results cannot be directly translated to systems with higher automation levels, they reveal that care must be taken when defining scenarios selected as performance indicators for assessing system-level performance.

Rajabli et al. [187] have conducted a structured literature review on automotive-relevant V&V approaches applicable to AVs. They categorized the review material into 8 main topics and analysed them to define answers to 3 questions. Namely, what are the common requirements, main challenges, open issues, and opportunities when verifying and validating automated vehicles. The focus is stated to be functional safety, i.e. behaviour in the presence of failure, and not as in [10] and [188] safety of the intended functionality [179]. Systems that are required to adhere to the strictest integrity level according to ISO 26262, tolerate a very restrictive maximum failure rate over time. A failure rate that is not only difficult to quantify but also difficult to adhere to. This one of the most difficult challenges to overcome, considering the huge amount of code and hardware that must be free from faults in an automated vehicle. A summary of tools that support the endeavour is presented. The categories are simulation environments and test scenarios, test case definition and generation, corner cases and adversarial examples, fault injection, mutation testing, software safety cages, cyber-physical systems techniques, and formal methods. No single approach is identified as universally effective. Combining approaches is a must, where scalability and combining contributions need investigation.

Berger and Birkemeyer [189] present an event identification approach based on encoding multi-dimensional data with space-filling curves (SFC) to obtain single dimensional representations. Characteristics stripes emerge on these single dimensional representations that correlate semantically with events in the original, multi-dimensional space with respect to distribution, spread, and temporal occurrence. Due to the fixed value range of the single dimensional representation, this approach can provide valuable insights into analysing the scenario space and identifying hazardous scenarios systematically while remaining computationally efficient.

7 ANALYSIS

This chapter presents the analysis of existing methodologies. The purpose is to define an initial draft for the condensed SUNRISE methodology to be part of the SUNRISE SAF by starting with the HEADSTART methodology and identifying similarities and differences with the other described methods and initiatives.

It should be noted that the analysis is not done versus any requirements. Instead, the aim is to identify requirements for the SUNRISE methodology to make it versatile and compatible with all the methodologies described in this report.

The chapter is structured such that first, the definition of the term scenario is analysed, then the HEADSTART methodology is analysed versus (1) the SAF gap analysis published in SUNRISE D2.1 [9], (2) other existing scenario-based methodologies described in Sec. 4, (3) the ISO 34503 process described in Sec. 5.1.2, and (4) other related initiatives described in Sec. 6.

7.1 Scenario definition

For scenario-based testing, the term scenario is important. Used definitions for scenarios for initiatives and projects described in this deliverable are listed in Table 6. Reading the different definitions reveals similarities and differences.

The SUNRISE methodology shall be versatile and able to support different approaches and consequently needs a scenario definition compatible with all definitions listed in Table 6. Further, it should capture all uses of the word “scenario” in the SUNRISE project. For example, it should cover what is shown in Figure 1 and Figure 2, but not exclude other definitions.

To meet above, together with SUNRISE task 3.2 it has been concluded that the SUNRISE scenario definition should aim to meet following:

- It should be simple and exclude any terms that can be left to the scenario description,
- It should be formulated such that it is compatible with different abstraction levels including functional, abstract, logical, and concrete scenarios.

A definition that meets this is a simplified version of the HEADSTART definition from which “without any specification of the parameters” has been removed to make it fully compatible with, e.g., logical, and concrete scenarios.

The resulting SUNRISE scenario definition is formulated as:

> *scenario:* *a "description of a temporal and spatial traffic constellation."*

Thereby the scenario definition is condensed into its most elementary form, to both cover all the different definitions from the sources, and still enabling scenario-based testing using this definition.

Note that using this simple definition, it is important that the term "scenario" is used in a context.

For the scenario description, it will be necessary to include other elements, e.g., by using terms as shown in Figure 1. Requirements on the scenario description will, to the extent that is needed, be defined in later tasks of SUNRISE.

Further, the SUNRISE methodology aims to be future-proof and should be defined such that it, e.g., can be adopted to scenario descriptions based on existing and coming standards like ASAM OpenX (see Section 5.3).

Table 6. List of scenario definitions from other projects and initiatives.

Source	Definition	Original source
HEADSTART	*"abstraction and general description of a temporal and spatial traffic constellation without any specification of the parameters"*	DIN SAE SPEC 91381:2019-06 [190]
ISO 21448	*"description of the temporal relationship between several scenes in a sequence of scenes, with goals and values within a specified situation, influenced by actions and events"*	ISO 21448:2022 (SOTIF) [179]
ISO 34501	*"Sequence of scenes usually including the automated driving system(s) (ADS) / subject vehicle(s), and its/their interactions in the process of performing the dynamic driving task (DDT)"*	ISO 34501:2022 [1]
StreetWise	*Informal:* *A scenario describes any situation on the road including the intent of the ego vehicle, the behaviour of road users, the road layout, and conditions such as weather and lighting. A drive on the road is considered a continuous sequence of scenarios – which might overlap.*	StreetWise
	Formal: *'A scenario is a quantitative description of the relevant characteristics and activities and/or goals of the ego vehicle(s), the static environment, the dynamic environment, and all the events that are relevant to the ego vehicle(s) within the time interval between the first and the last relevant event. An event corresponds to a moment in time at which a mode transition occurs or a system reaches a specific threshold, where the former can be induced by both internal and external causes'*	[147]
Safety Pool™	*'temporal development between several scenes in a sequence of scenes. Every scenario starts with an initial scene. Action and events as well as goals & values may be specified to characterise this temporal*	[154]

	development in a scenario. Other than a scene, a scenario spans a certain amount of time'	

7.2 SAF gap analysis

The HEADSTART methodology, as depicted in Figure 4, has been analysed versus the identified gaps in the SAF gap analysis presented in SUNRISE D2.1 [9]. Following points are considered relevant for safety argumentation based on knowledge and data-driven scenario-based testing (WP3) and are valuable input for the overall development of the SAF (WP2).

1. The HEADSTART methodology does not fully cover qualitative and quantitative metrics as part of a process to determine the completeness of a scenario database (i.e., not the test scenarios, cf. point 7 below). To ensure a comprehensive validation process, the methodology should include qualitative and quantitative metrics that evaluate the coverage and representativeness of the scenarios in the database. These metrics would contribute to the overall safety argumentation by ensuring that a wide range of scenarios, relevant to the SUT, are adequately considered and tested.
2. To establish a robust validation chain, the SUNRISE methodology should cover a systematic approach for scenario definition and quality evaluation. By incorporating a systematic approach, the generated scenarios can be well-documented and reproducible. Also, this approach can ensure that the extracted scenarios are representative of the real-world driving situations and contribute effectively to the safety argumentation process, which encompasses virtual testing, confined areas testing, and public road testing, accompanied by relevant standard safety metrics.
3. The HEADSTART methodology falls short in providing a systematic and standardizable approach for querying scenario databases and managing scenarios therein. Developing guidelines and tools for efficient scenario retrieval and management are crucial for conducting scenario-based testing, which would enhance the overall methodology's effectiveness.
4. To ensure a holistic safety approach, the SUNRISE methodology should expand the test scenarios selection process to a broader perspective that covers various driving environments and situations. By sourcing from a wide range of scenarios, including urban and highway areas, the methodology can address diverse real-world driving conditions and better identify any relevant associated risks.
5. Besides supporting existing scenario databases, the methodology should include a feedback loop process for potentially new and unsafe scenarios discovered during V&V of an AD function. By including these arising scenarios, the methodology can address and mitigate potential safety risks that may not have been initially considered. This iterative process can then help to enhance the overall safety argumentation by incorporating real-world observations and challenges into the testing and validation process. The SUNRISE methodology should allow for including such scenarios to improve system robustness continually.

6. The HEADSTART methodology lacks a standardized process for input and output interfaces with scenario databases. Implementing standardized formats would facilitate interoperability, data sharing, and consistent integration with different testing platforms and tools to seamlessly communicate and exchange scenarios. By adhering to common formats, it becomes easier to integrate scenarios-based testing from different sources, ensuring compatibility and reducing the risk of errors or inconsistencies in the safety argumentation.

7. Recognizing the significance of field data in determining real system performance (linked to point 6 above), the SUNRISE methodology should address relevant quality metrics as part of a process to determine systematically the quality of such data. Additionally, standardized input formats for incorporating field data into the validation process shall be defined.

8. Including a well-defined methodology for the homologation procedure is essential. This procedure would ensure that AVs adhere to safety regulations and undergo thorough certification processes and thus, promoting public trust.

9. Explicitly addressing the needs and roles of various stakeholders is crucial for a successful implementation. The SUNRISE methodology should encompass a systematic approach to elicit requirements of regulators, policymakers, industrial practitioners, and the general public, fostering collaboration and transparency.

7.3 Existing methodologies

In this section, the HEADSTART methodology is compared with the methodologies described in Sec. 4. Note that the analysis focuses on available information, and if information is missing for one methodology, it is left out. The purpose is to identify requirements for the SUNRISE method to make it versatile and compatible with all the methodologies described in this report.

7.3.1 Scenario concepts, parameter sets and descriptions

The HEADSTART scenario concept is described in Sec. 3.1, with its layer model shown in Figure 5. The abstraction levels are functional, logical, and concrete with corresponding scenario space and the ODD was usually highway for state-of-the-art project existing at the time of defining the HEADSTART scenario concept. Later, the abstract scenario level was introduced by Neurohr in [27] as part of the PEGASUS Family project VVMethods, and also is included in ISO 34501 [1] as shown in Figure 2.

Further, the HEADSTART scenario structure is based on the six-layer model form the German PEGASUS project [191] combined with the scenario structure from the French MOOVE project.

VVMethods utilizes a holistic scenario concept for urban traffic. The first step of the scenario concept is to categorize (and describe) traffic from the perspective of an ego vehicle in relation to the infrastructure (Layer 1-3 of the 6 Layer Model). Instances of this categorization are called enveloping scenarios. Within these enveloping scenario, bilateral Interactions are described through base scenarios, which themselves are compositions of scenario classes such as Ego-manoeuvre, object-manoeuvre, and conflict-type. The base scenarios can be chained to sequences, forming more complex traffic scenarios. Specific compositions of base scenarios, which require further details or other concepts, are described through focus scenarios. The parameters are then extracted from the processed real-world data.

StreetWise uses scenarios to provide a structured approach to capture all situations and conditions that a vehicle can encounter on the road. Scenario categories are defined to provide structuring of the vast amount of (concrete) scenarios. A scenario category is defined as a set of scenarios that share a common characteristic, e.g., all highway scenarios that have a target vehicle change lane from an adjacent lane into the ego-vehicle lane in front of the ego vehicle is the scenario category “cutting-in vehicle”. The ten distinguished scenario categories for highway in StreetWise are parameterized, in order to be able to perform statistical analysis using the probability density functions for each of the parameters. This is in line and compatible with the approach in HEADSTART.

Actually, TNO has shown how to include V2X communication (as key enabling technology) into the scenario description. It appeared that the communication itself is not added to the scenario, but merely the conditions that might disturb V2X communication. This was done based on StreetWise scenario descriptions and showed the compatibility with the HEADSTART approaches.

SAKURA In the whitepaper written in the international collaboration between SAKURA, SIPadus and HEADSTART [192], it became clear that the scenario concept that SAKURA uses, is compatible with HEADSTART, considering the SAKURA traffic disturbance scenarios.

CETRAN uses an approach inspired by and similar to TNO StreetWise, making the CETRAN approach being compatible with HEADSTART.

Safety Pool™: The scenario concepts that feed into the Safety Pool™ scenario database have been developed by WMG, University of Warwick as outputs of the V&V methodology implemented in the research group. The scenario-based approach is similar to the methodology developed throughout the HEADSTART project and in other European projects involving the validation and verification of ADS. Scenarios are arranged into libraries by source, this can mean different methods of generation or as outcomes of different projects. Each scenario is tagged individually rather than into ‘themes', ODD and behaviour tags enable ODD based scenario searching and filtering to find the right scenarios for a particular ODD query. The ODD and Behaviour method of classifying the structure of a scenario (as presented in ISO 34503 [146])is used rather than the six-layer model used in HEADSTART.

In **ADScene,** there are “scenarios”, “use cases”, “test protocols” and “reference data”. The abstraction levels of the scenario description follow the functional, logical and concrete descriptions used in HEADSTART. Note also that as part of ADScene, the MOOVE 4 layer model has been further developed to be compliant with Pegasus 6 layer model (see Sec. 4.4.1).

To conclude, it is essential that the SUNRISE scenario concept is versatile and compatible with all the methodologies described in this report. The SUNRISE scenario concept may build on the HEADSTART concept as it is compatible with several scenario databases, e.g., PEGASUS, StreeWise, ADScene, and SAKURA. In addition, it must be adapted to support the Safety Pool^{TM} concept using ODD and Behaviour method of classifying the structure of a scenario, as well as easily adaptable to possible new scenario concepts in the future.

7.3.2 Scenario sources and scenario generation

HEADSTART's approach to scenario sources and scenario generation is described in 3.2. As this part was not in focus for HEADSTART, the description is conceptual, and scenarios were assumed available from other scenario databases.

VVMethods uses real-world data of captured traffic from vehicles as a source for scenarios. A Scenario-Engine can automatically detect scenarios and extract their parameters from an object list-based trajectory data set, including map information using the above-mentioned PEGASUS scenario concept.

StreetWise uses object-level driving data, captured by a state-of-the-art sensor set onboard vehicles as input to the scenario identification and characterization pipeline. In the pipeline, the object-level data is interpreted, events and activities are detected, and scenarios are identified based on sequences of such events and activities. The parameters describing the identified concrete scenarios, are stored with the scenario in the StreetWise scenario database. To validate the scenario pipeline, and the identification and characterization algorithms, also context cameras are used in the vehicles.

The StreetWise architecture allows for different data sources as well, e.g., road side units equipped with a sensor suite, or drones [193].

CETRAN: uses an approach inspired by and similar to TNO StreetWise.

Safety Pool^{TM}: The V&V methodology behind the scenarios on Safety Pool^{TM} acknowledges the significance of both data-based and knowledge-based scenario generation. Scenarios are generated from various data sources ranging from real-world naturalistic driving data, to safety critical cases in the form of accident records and near-miss insurance records, to knowledge-based methods such as using STPA to access scenarios from a systems safety perspective.

Scenarios on Safety Pool^{TM} are represented as abstract scenarios, in SDL level 1, Logical scenarios in the SDL level 2 and concrete scenarios in the OpenX representations. Though concrete SDL level 2 scenarios can also be uploaded where necessary.

ADScenes uses four sources of data for scenarios in order to achieve completeness: real-world driving data, incident reports (near crash), accidents reports, and expert knowledges/regulations. From nominal logical scenarios, concrete scenarios are extracted from real world data, and from logical accident scenarios, concrete accident scenarios are extracted from accident databases.

Hi-Drive project defines a set of basic driving scenarios from which more complex test scenarios can be composed. The driving scenario concept is based on the Layer 4 of the PEGASUS 6-Layer-Model, which defines the driving scenarios, while the other layers define parameters/conditions to further specify the scenarios (unpublished work, work in progress). The driving scenario concept is used to analyse real-world logged data from large scale public road trials for which tools for automatic scenario extraction from data are developed. The driving scenario concept is also used in AD function impact assessment simulations including perspective simulations per ISO/TR 21934 [194] focusing on prospective assessment of traffic safety for vehicle-integrated technologies acting in the pre-crash phase by means of virtual simulation . The project at the end will also deliver a scenario database.

To conclude, like HEADSTART, SUNRISE targets through external databases multiple scenarios sources such as field data, ariel data, accident data, and simulator studies. The databases contain real-world data from, e.g., vehicles, event data recorders, road-side units, and drones, incident reports, accident reports, expert knowledge/regulations, and data from simulations. An open issue is how to determine the completeness of a given ODD for scenarios available from the external databases through the data framework. Qualitative and quantitative metrics for this need to be defined.

7.3.3 Scenario database

The approach used in HEADSTART for scenario databases is described in 3.3 and was basically to rely on external databases.

In VVMethods, an abstraction of the real world into a scenario database is utilized. Therefore, a sophisticated scenario concept is utilized to map the whole ODD (including its edges) into such a scenario database. The scenario database is filled with real-world data in a common format, also known as the OMEGA format. This makes sure, that certain qualities are met. Moreover, scenarios are generated and enriched from knowledge based and hybrid approaches.

StreetWise, with data-driven scenario identification and characterisation, makes statistical analysis of the scenarios possible. One important aspect is the exposure for each scenario, which is computed based on the probability density functions, and the total amount of data (expressed in nr. Of hours of driving, or total covered distance in km) in the scenario database. Such statistics are also used to determine the evolution in completeness with every dataset from which the scenarios are added to the StreetWise database.

In **Safety Pool™**, scenarios can be written directly into the database via user interfaces (UIs), upload can also be achieved in bulk for a proposed scenario-set or library. All upload requests take place through this UI rather than through use of a dedicated API, parsing of scenarios for correctness against grammar files is performed within the database, with live feedback available for the user. New scenarios are compared against existing scenarios in the database and given a score for similarity which determines their assigned worth as an addition to the platform. ODD is the basis on which scenarios are given a similarity score, and this also can be used as a metric to assess completeness of a scenario set from a given ODD.

ADScene is a platform providing analysis and management tools including data content (scenarios), features and tools, and secured environment, all illustrated in Figure 26.

To conclude, like HEADSTART, SUNRISE will not develop its own scenario database. Instead, a Data framework is developed to support retrieving scenarios from external databases. Compared with HEADSTART, SUNRISE's scope is more extensive like, e.g., the federation layer developed in WP5 and WP6. SUNRISE needs to define queries and format for responses while in HEADSTART was just conceptual. A challenge is designing the data framework so versatile that all relevant scenario databases can be connected.

7.3.4 Test scenario selection and subspace creation methodology

The HEADSTART's method for test scenario selection is described in 3.4. The idea is that based on the definition of the driving function a query for the scenario database is defined to extract test scenarios suitable for safety evaluation of the vehicle with the driving function. However, as the database solution is not part of HEADSTART, formulation of the queries remains an open issue.

VVMethods: PEGASUS defines a limited number of safety-relevant logical scenarios based on the area of the SUT that the challenging object would collide with, and the initial positions of the challenging object. This is called the PEGASUS challenger concept shown in Figure 10. VVMethods develops the PEGASUS challenger methodology to further cover urban traffic situations besides highway. Further, the proposed scenario concept in VVMethods can function as subspace creation technique to structure and limit the number of scenarios.

StreetWise categorizes scenarios into scenario categories as shown in Table 4. The selection of scenarios for test case generation is based on these scenario categories. To match the ODD, the ranges of the parameters to be considered, can be selected for each parameter independently. Also, a functionality is provided to make selections based on predefined tags, which makes the method scalable.

Safety Pool™'s generation methods utilise data-based approaches (using real-world, insurance and accident data) along with knowledge-based approaches (ontology, ODD, behaviour, rules of the road, and STPA). The scenario generation and storage methodology fits into a wider testing framework which is detailed in Figure 24. Scenario selection takes place within the section titled environment in the figure. The test case generator iterates through the ranges of the logical scenarios within simulation, this iteration can be informed by a multitude of approaches, for example Bayesian optimisation or constraint randomisation. The test case pass/fail criteria module can consist of different types: 1) utilising generic pass/fail criteria which are appliable to a wide range of use cases, or alternatively 2) utilising use case – specific and system-specific pass/fail criteria.

Subspaces of scenarios can also be created by ODD and behaviour-based tag filtering of existing scenarios within the database. These can then be used for training or testing for a certain ODD.

ADScene: Scenario selection is currently based on several criteria that can be chosen. Each parameter or tag of a scenario can be chosen, but research work is still needed.

To conclude, SUNRISE needs to develop a scenario selection concept that is compatible with relevant existing scenario databases and adaptable to relevant new databases.

HEADSTART did propose a scenario selection process, but as the scenario databases were not part of the project, queries were not defined and tested. If compatible queries can be defined in the SUNRISE data framework, compatibility with external databases is expected. Subspace creation methodologies used to further structure and limited the number of scenarios was not part of HEADSTART and should, for SUNRISE, be developed based on, e.g., work done in VVMethods and by WMG for Safety Pool^{TM}.

7.3.5 Test scenario allocation concepts and metrics

The method used in HEADSTART for scenario allocation is described in 3.5. The allocation is based on that the capabilities for each test method are defined, and thereafter used to identify which test methods can best be used for each concrete test scenario. This process must account for the specified ODD and DDT.

PEGASUS family: The PEGASUS approach for scenario allocation is the basis for the HEADSTART approach. In the VVMethods context, completeness of the scenario concept is discussed by the coverage on three levels:

- Coverage of Concept:
  The scenario concept needs to make sure, that all possibly relevant entities and actions within the ODD are representable → *Logical Scenario Classes*
- Coverage of Data:
  The recorded and generated scenarios create a distribution of its parameters. These distributions must map the real world correctly. → Logical Scenario Instances
- Coverage of Test:
  The sampling into concrete scenarios for tests must cover the distributions sufficiently. → Concrete Scenario

ADScene has scenario descriptions, simulatable scenarios and “test protocol descriptions” to help ADScene users preparing their physical or digital testing plans.

StreetWise does not distinguish scenarios in the scenario database for specific ways of testing (proving ground testing, virtual testing, or testing on the public road). Hence, StreetWise does not describe an approach to allocate specific tests to the way of testing.

Safety Pool^{TM}: WMG states that test allocation is a key step within the V&V workflow, but the concept is under development and not yet published.

To conclude, building on the HEADSTART scenario allocation method, the SUNRISE method should be compatible with the PEGASUS family. As the test capabilities in HEADSTART are connected to the test method and not the scenario, it is general and should also be compatible with StreetWise and other databases. If test related information is available in the database, like for ADSscene, the query response with scenarios must include that to the extent SUNRISE can use it. ADScene, e.g., includes information which scenarios that are simulatable, and test protocol descriptions.

7.4 ISO 34502

Analysing the HEADSTART methodology versus the process of ISO 34502 illustrated in Figure 34 reveals two important aspects that should be considered for enhancing the HEADSTART methodology. These aspects, labelled as A and B in the following, have implications for several topics as outlined before in Sec. 7.2:

A. "Identification and risk evaluation of potential hazardous scenarios via analysis of the scenario space"

This aspect has direct implications for the HEADSTART methodology, aligning with the previously discussed topics of scenario database completeness (point 1 in Sec. 7.2), scenario quality evaluation (point 2 in Sec. 7.2), scenario querying (point 3 in Sec. 7.2), and systematic risk assessment. However, the HEADSTART methodology, as depicted in Figure 4, does not explicitly cover a systematic risk assessment element. To address this, it is crucial to incorporate a robust risk assessment process that evaluates potential hazardous scenarios within the scenario space. By doing so, the methodology can identify and prioritize scenarios that pose higher risks to enhance safety measures.

B. "Set of test scenarios covers a relevant part of the scenario space in search for unknown hazardous scenarios. Newly identified hazardous scenarios are evaluated concerning their risk."

This aspect has also significant implications for the HEADSTART methodology, relating to continuous monitoring of CCAM, feeding back potentially new and unsafe scenarios (point 5 in Sec. 7.2), and field data from event data recorders (point 7 in Sec. 7.2). However, Figure 4 does not explicitly address them. Hence, adaptations to the methodology are necessary to better support aspect B. This includes incorporating mechanisms for continuous monitoring to identify unknown hazardous scenarios, evaluating their associated risks, and updating the scenario database accordingly. Additionally, a systematic method for estimating quality properties of the existing scenario space within the scenario database should be included.

7.5 Other initiatives

In comparison to the scenario-based methodology for safety assurance in HEADSTART, the analysis in Section 6.3 have uncovered some noteworthy enhancements that warrant consideration. Some insights touch on non-scenario-based approaches, which fall beyond the scope of SUNRISE.

A key takeaway from the analysis in Section 6.3 is the absence of a one--fits-all methodology or technique. This underscores the necessity for additional input data in the process steps to guide users of a general framework, such as the one outlined in HEADSTART.

- The scenario selection process step particularly requires supplementary data to inform the choice of the selection method. This could include factors like test purpose, scenario data encoded through space-filling curves, and scenario complexity. These considerations should correlate with insights into the efficacy of various methods.

- For the process step of scenario allocation, additional data could help formulate effective strategies for distributing scenarios across test environments.

- Regarding evaluating and formulating test goals, caution is crucial when selecting scenarios as performance indicators for assessing system-level performance. There might be a need to reduce the ambiguity in the specifications of certain scenarios to enable good evaluation results.

The aim is to presents a concise compilation of the analysis that provide actionable suggestions, addresses obstacles. For a more comprehensive insight, kindly refer to Section 6.3 and the original works cited therein.

8 CONCLUSIONS

In summary, the analysis of existing scenario-based methodologies for safety assurance with the HEADSTART method has provided valuable insights for developing the SUNRISE methodology. The analysis has identified similarities and differences, which can guide the integration of best practices and address specific challenges or gaps.

The following list of changes to the HEADSTART methodology are proposed to be integrated into the SUNRISE methodology:

1. For SUNRISE, a “scenario” is defined as a “description of temporal and spatial traffic constellation” (Sec. 7.1). It intentionally kept simple to capture all the use of the term in the scope of scenario-based testing. Details of the scenario are left to the description.

2. SUNRISE’s scenario concept should be versatile and able to support different approaches (Sec. 7.3). HEADSTART’s scenario concept is a good starting point as it is compatible with, e.g., PEGASUS, StreetWise, and ADScene, but should extended to also support other methods like the ODD and behaviour model used in Safety Pool™. Furthermore, it should be flexible and easily adaptable to new concepts in the future. This is relevant not only for task 3.2 but also for tasks 5.1 and 5.2.

3. Like HEADSTART, SUNRISE targets multiple data sources such as field data, aerial data, accident data, and simulator studies and relies on external scenario databases. Databases analysed in this report includes the PEGASUS project family, StreetWise, Safety Pool™, ADScene, and Hi-Drive (Sec. 4, Sec. 7.3.2, and Sec. 7.3.3). Together, they cover multiple data sources including, real-world data form, e.g., vehicles, roadside units and drones, incident reports, accident reports, expert knowledge/regulations, and from simulations. However, qualitative and quantitative metrics to determine the completeness for a given ODD of the federated scenario database are missing and need to be developed (Sec. 7.2 point 1). This is considered relevant for tasks 5.1 and 5.3.

4. HEADSTART’s scenario selection process should be suitable to SUNRISE considering, that proper queries for scenario searches are defined (Sec. 7.3.4, Sec. 7.2 point 3). Further, metrics are needed for the quality evaluation of the selected scenarios (Sec. 7.2 point 2). Methods for further structure and limit the number of scenarios using, e.g., subspace creation techniques should be included based on works done in, e.g., VVMethods and by WMG for Safety Pool™ (Sec. 7.3.4). This is considered relevant for task 3.3.

5. SUNRISE’s test scenario allocation process and metrics can be based on HEADSTART’s process: First, the capabilities of each test method are analysed, and then the test scenarios are allocated to suitable test methods (Sec. 7.3.5). As long as point 4 above is solved, the test scenario allocation process should be compatible with all supported scenario databases. This is relevant for tasks 3.4 and 3.5.

6. SUNRISE should include mechanisms for identifying unknown scenarios (Sec. 7.2, point 5, Sec. 5.2). This is considered relevant for task 4.1.

7. The HEADSTART methodology does not explicitly include a systematic risk assessment element. For SUNRISE, it is crucial to incorporate support for identification and risk evaluation of potential hazardous scenarios (Sec. 7.4). This should involve thoroughly analysing the scenario space, including identified unknown scenarios (see point 6 above), within the context of a specified SUT, ODD, and DDT. This is relevant for task 2.2, 3.3, and 3.5.

9 REFERENCES

1. ISO (2022) ISO 34501:2022 - Road vehicles -- Test scenarios for automated driving systems -- Vocabulary

2. ISO (2022) ISO 34502:2022 - Road vehicles -- Test scenarios for automated driving systems -- Scenario based safety evaluation framework

3. (2023) Definition of TAXONOMY. https://www.merriam-webster.com/dictionary/taxonomy. Accessed 16 Aug 2023

4. Terminology - Definition, Meaning & Synonyms. In: Vocabulary.com. https://www.vocabulary.com/dictionary/terminology. Accessed 16 Aug 2023

5. ECE/TRANS/WP.29/2022/58 Secretariat (GRVA) New Assessment/Test Method for Automated Driving (NATM) Guidelines for Validating Automated Driving System (ADS) – amendments to ECE/TRANS/WP.29/2022/58 | UNECE. https://unece.org/transport/documents/2022/05/informal-documents/grva-new-assessmenttest-method-automated-driving. Accessed 15 Aug 2023

6. Ilias Panagiotopoulos, Bernhard Hillbrand, Patrick Weissensteiner, et al (2023) SUNRISE D7.1 CCAM use cases validation requirements

7. HEADSTART project is on web. In: Headstart Project. https://www.headstart-project.eu/. Accessed 30 Jun 2023

8. HARMONISED EUROPEAN SOLUTIONS FOR TESTING AUTOMATED ROAD TRANSPORT | HEADSTART Project | Fact Sheet | H2020. In: CORDIS | European Commission. https://cordis.europa.eu/project/id/824309. Accessed 19 Jun 2023

9. Anastasia Bolovinou, Christina Anagnostopoulou, Vasilis Roungas, et al (2023) Hi-DriveD3.1 Use cases definition and description

10. Riedmaier S, Ponn T, Ludwig D, et al (2020) Survey on Scenario-Based Safety Assessment of Automated Vehicles. IEEE Access 8:87456–87477. https://doi.org/10.1109/ACCESS.2020.2993730

11. Baek Y-M, Cho E, Shin D, Bae D-H (2022) Literature Review to Collect Conceptual Variables of Scenario Methods for Establishing a Conceptual Scenario Framework

12. Ding W, Xu C, Arief M, et al (2022) A Survey on Safety-Critical Driving Scenario Generation -- A Methodological Perspective

13. Ma J, Che X, Li Y, Lai EM-K (2021) Traffic Scenarios for Automated Vehicle Testing: A Review of Description Languages and Systems. Machines 9:342. https://doi.org/10.3390/machines9120342

14. Nalic D, Mihalj T, Bäumler M, et al (2020) Scenario Based Testing of Automated Driving Systems: A Literature Survey. In: FISITA World Congress 2020 - Technical Programme. FISITA, p 1

15. Sun J, Zhang H, Zhou H, et al (2022) Scenario-Based Test Automation for Highly Automated Vehicles: A Review and Paving the Way for Systematic Safety Assurance.

IEEE Transactions on Intelligent Transportation Systems 23:14088–14103. https://doi.org/10.1109/TITS.2021.3136353

16. Zhang X, Tao J, Tan K, et al (2022) Finding Critical Scenarios for Automated Driving Systems: A Systematic Mapping Study. IEEE Transactions on Software Engineering 1–1. https://doi.org/10.1109/TSE.2022.3170122

17. Zhong Z, Tang Y, Zhou Y, et al (2021) A Survey on Scenario-Based Testing for Automated Driving Systems in High-Fidelity Simulation

18. Hoss M, Scholtes M, Eckstein L (2022) A Review of Testing Object-Based Environment Perception for Safe Automated Driving. Automot Innov 5:223–250. https://doi.org/10.1007/s42154-021-00172-y

19. King C, Ries L, Langner J, Sax E (2020) A Taxonomy and Survey on Validation Approaches for Automated Driving Systems. In: 2020 IEEE International Symposium on Systems Engineering (ISSE). pp 1–8

20. Schnelle S, Salaani MK, Rao S, et al (2019) Review of simulation frameworks and standards related to driving scenarios

21. Westhofen L, Neurohr C, Koopmann T, et al (2022) Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art. Arch Computat Methods Eng. https://doi.org/10.1007/s11831-022-09788-7

22. Zhao T, Yurtsever E, Paulson J, Rizzoni G (2022) Formal Certification Methods for Automated Vehicle Safety Assessment. IEEE Transactions on Intelligent Vehicles 1–1. https://doi.org/10.1109/TIV.2022.3170517

23. Gyllenhammar M, Rodrigues de Campos G, Törngren M (2022) Holistic Perspectives on Safety of Automated Driving Systems - Methods for Provision of Evidence

24. ISO (2023) ISO 34503:2023 “Road Vehicles – Test scenarios for automated driving systems – Specification for operational design domain”

25. ISO (2023) ISO/DIS 34504 “Road vehicles – Test scenarios for automated driving systems – Scenario categorization”

26. Menzel T, Bagschik G, Maurer M (2018) Scenarios for Development, Test and Validation of Automated Vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1821–1827

27. Neurohr C, Westhofen L, Butz M, et al (2021) Criticality Analysis for the Verification and Validation of Automated Vehicles. IEEE Access 9:18016–18041. https://doi.org/10.1109/ACCESS.2021.3053159

28. Rohloff M, Forschungsgesellschaft für Straßen- und Verkehrswesen (2008) Richtlinien für die Anlage von Autobahnen: RAA, Ausgabe 2008. FGSV-Verl, Köln

29. Guo J, Kurup U, Shah M (2020) Is it Safe to Drive? An Overview of Factors, Metrics, and Datasets for Driveability Assessment in Autonomous Driving. IEEE Transactions on Intelligent Transportation Systems 21:3135–3151. https://doi.org/10.1109/TITS.2019.2926042

30. Kang Y, Yin H, Berger C (2019) Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments. IEEE Transactions on Intelligent Vehicles 4:171–185. https://doi.org/10.1109/TIV.2018.2886678

31. Zhu J, Wang W, Zhao D (2018) A Tempt to Unify Heterogeneous Driving Databases using Traffic Primitives. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 2052–2057

32. Krajewski R, Bock J, Kloeker L, Eckstein L (2018) The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 2118–2125

33. Geyer S, Baltzer M, Franz B, et al (2014) Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. IET Intelligent Transport Systems 8:183–189. https://doi.org/10.1049/iet-its.2012.0188

34. Bagschik G, Menzel T, Maurer M (2018) Ontology based Scene Creation for the Development of Automated Vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1813–1820

35. Chen W, Kloul L (2018) An Ontology-based Approach to Generate the Advanced Driver Assistance Use Cases of Highway Traffic. In: 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. Seville, Spain

36. Klueck F, Li Y, Nica M, et al (2018) Using Ontologies for Test Suites Generation for Automated and Autonomous Driving Functions. In: 2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW). pp 118–123

37. Wotawa F, Li Y (2018) From Ontologies to Input Models for Combinatorial Testing. In: Medina-Bulo I, Merayo MG, Hierons R (eds) Testing Software and Systems. Springer International Publishing, Cham, pp 155–170

38. Li Y, Tao J, Wotawa F (2020) Ontology-based test generation for automated and autonomous driving functions. Information and Software Technology 117:106200. https://doi.org/10.1016/j.infsof.2019.106200

39. Khastgir S, Brewerton S, Thomas J, Jennings P (2021) Systems Approach to Creating Test Scenarios for Automated Driving Systems. Reliability Engineering & System Safety 215:107610. https://doi.org/10.1016/j.ress.2021.107610

40. Langner J, Bach J, Ries L, et al (2018) Estimating the Uniqueness of Test Scenarios derived from Recorded Real-World-Driving-Data using Autoencoders. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1860–1866

41. Bolte J-A, Bar A, Lipinski D, Fingscheidt T (2019) Towards Corner Case Detection for Autonomous Driving. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 438–445

42. Krajewski R, Moers T, Nerger D, Eckstein L (2018) Data-Driven Maneuver Modeling using Generative Adversarial Networks and Variational Autoencoders for Safety Validation of Highly Automated Vehicles. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 2383–2390

43. Jenkins IR, Gee LO, Knauss A, et al (2018) Accident Scenario Generation with Recurrent Neural Networks. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 3340–3345

44. Jesenski S, Stellet JE, Schiegg F, Zöllner JM (2019) Generation of Scenes in Intersections for the Validation of Highly Automated Driving Functions. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 502–509

45. Kruber F, Wurst J, Botsch M (2018) An Unsupervised Random Forest Clustering Technique for Automatic Traffic Scenario Categorization. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 2811–2818

46. Kruber F, Wurst J, Morales ES, et al (2019) Unsupervised and Supervised Learning with the Random Forest Algorithm for Traffic Scenario Clustering and Classification. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 2463–2470

47. Watanabe H, Malý T, Wallner J, et al Methodology of Scenario Clustering for Predictive Safety Functions. 8

48. Wang W, Zhao D (2018) Extracting Traffic Primitives Directly From Naturalistically Logged Data for Self-Driving Applications. IEEE Robotics and Automation Letters 3:1223–1229. https://doi.org/10.1109/LRA.2018.2794604

49. Esenturk E, Wallace AG, Khastgir S, Jennings P (2022) Identification of Traffic Accident Patterns via Cluster Analysis and Test Scenario Development for Autonomous Vehicles. IEEE Access 10:6660–6675. https://doi.org/10.1109/ACCESS.2021.3140052

50. Weber H, Bock J, Klimke J, et al (2019) A framework for definition of logical scenarios for safety assurance of automated driving. Traffic Injury Prevention 20:S65–S70. https://doi.org/10.1080/15389588.2019.1630827

51. Erdogan A, Ugranli B, Adali E, et al (2019) Real- World Maneuver Extraction for Autonomous Vehicle Validation: A Comparative Study. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 267–272

52. Beglerovic H, Schloemicher T, Metzner S, Horn M (2018) Deep Learning Applied to Scenario Classification for Lane-Keep-Assist Systems. Applied Sciences 8:2590. https://doi.org/10.3390/app8122590

53. Gruner R, Henzler P, Hinz G, et al (2017) Spatiotemporal representation of driving scenarios and classification using neural networks. In: 2017 IEEE Intelligent Vehicles Symposium (IV). pp 1782–1788

54. Dávid B, Láncz G, Hunyady G (2019) Highway Situation Analysis with Scenario Classification and Neural Network based Risk Estimation for Autonomous Vehicles. In: 2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI). pp 375–380

55. Bach J, Langner J, Otten S, et al (2017) Test scenario selection for system-level verification and validation of geolocation-dependent automotive control systems. In: 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC). pp 203–210

56. Hartjen L, Philipp R, Schuldt F, et al (2019) Classification of driving maneuvers in urban traffic for parametrization of test scenarios. In: 9. Tagung automatisiertes fahren

57. Zhou J, del Re L (2017) Identification of critical cases of ADAS safety by FOT based parameterization of a catalogue. In: 2017 11th Asian Control Conference (ASCC). pp 453–458

58. Zofka MR, Kuhnt F, Kohlhaas R, et al (2015) Data-driven simulation and parametrization of traffic scenarios for the development of advanced driver assistance systems. In: 2015 18th International Conference on Information Fusion (Fusion). pp 1422–1428

59. de Gelder E, Paardekooper J-P (2017) Assessment of Automated Driving Systems using real-life scenarios. In: 2017 IEEE Intelligent Vehicles Symposium (IV). pp 589–594

60. de Gelder E, Paardekooper J-P, Op den Camp O, De Schutter B (2019) Safety assessment of automated vehicles: how to determine whether we have collected enough field data? Traffic Injury Prevention 20:S162–S170. https://doi.org/10.1080/15389588.2019.1602727

61. Ponn T, Fratzke D, Gnandt C, Lienkamp M (2019) Towards Certification of Autonomous Driving: Systematic Test Case Generation for a Comprehensive but Economically-Feasible Assessment of Lane Keeping Assist Algorithms: In: Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems. SCITEPRESS - Science and Technology Publications, Heraklion, Crete, Greece, pp 333–342

62. Beglerovic H, Ravi A, Wikström N, et al (2017) Model-based safety validation of the automated driving functio highway pilot. In: Pfeffer ProfDrPE (ed) 8th International Munich Chassis Symposium 2017. Springer Fachmedien, Wiesbaden, pp 309–329

63. Rocklage E, Kraft H, Karatas A, Seewig J (2017) Automated scenario generation for regression testing of autonomous vehicles. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). pp 476–483

64. Kim B, Jarandikar A, Shum J, et al (2016) The SMT-based automatic road network generation in vehicle simulation environment. In: Proceedings of the 13th International Conference on Embedded Software. Association for Computing Machinery, New York, NY, USA, pp 1–10

65. Kim B, Masuda T, Shiraishi S (2019) Test Specification and Generation for Connected and Autonomous Vehicle in Virtual Environments. ACM Trans Cyber-Phys Syst 4:8:1-8:26. https://doi.org/10.1145/3311954

66. Majzik I, Semeráth O, Hajdu C, et al (2019) Towards System-Level Testing with Coverage Guarantees for Autonomous Vehicles. In: 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS). pp 89–94

67. Khastgir S, Dhadyalla G, Birrell S, et al (2017) Test Scenario Generation for Driving Simulators Using Constrained Randomization Technique. pp 2017-01–1672

68. Huang L, Xia Q, Xie F, et al (2018) Study on the Test Scenarios of Level 2 Automated Vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 49–54

69. Xie F, Chen T, Xia Q, et al (2018) Study on the Controlled Field Test Scenarios of Automated Vehicles. pp 2018-01–1633

70. Zhou J, Re L del (2017) Reduced Complexity Safety Testing for ADAS & ADF. IFAC-PapersOnLine 50:5985–5990. https://doi.org/10.1016/j.ifacol.2017.08.1261

71. Althoff M, Dolan JM (2012) Reachability computation of low-order models for the safety verification of high-order road vehicle models. In: 2012 American Control Conference (ACC). pp 3559–3566

72. Tuncali CE, Fainekos G (2019) Rapidly-exploring Random Trees for Testing Automated Vehicles. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp 661–666

73. Åsljung D, Nilsson J, Fredriksson J (2016) Comparing Collision Threat Measures for Verification of Autonomous Vehicles using Extreme Value Theory. IFAC-PapersOnLine 49:57–62. https://doi.org/10.1016/j.ifacol.2016.07.709

74. Åsljung D, Nilsson J, Fredriksson J (2017) Using Extreme Value Theory for Vehicle Level Safety Validation and Implications for Autonomous Vehicles. IEEE Transactions on Intelligent Vehicles 2:288–297. https://doi.org/10.1109/TIV.2017.2768219

75. Åsljung D, Westlund M, Fredriksson J (2019) A Probabilistic Framework for Collision Probability Estimation and an Analysis of the Discretization Precision. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 52–57

76. Åsljung D (2022) On Statistical Methods for Safety Validation of Automated Vehicles. Chalmers University of Technology

77. Wang X, Peng H, Zhao D (2021) Combining Reachability Analysis and Importance Sampling for Accelerated Evaluation of Highway Automated Vehicles at Pedestrian Crossing. ASME Letters in Dynamic Systems and Control 1:011017. https://doi.org/10.1115/1.4046610

78. Zhao D, Lam H, Peng H, et al (2017) Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques. IEEE Transactions on Intelligent Transportation Systems 18:595–607. https://doi.org/10.1109/TITS.2016.2582208

79. Zhao D, Huang X, Peng H, et al (2018) Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers. IEEE Transactions on Intelligent Transportation Systems 19:733–744. https://doi.org/10.1109/TITS.2017.2701846

80. Zhao D (2016) Accelerated Evaluation of Automated Vehicles. Ph.D., The University of Michigan

81. Prialé Olivares S, Rebernik N, Eichberger A, Stadlober E (2016) Virtual Stochastic Testing of Advanced Driver Assistance Systems. In: Schulze T, Müller B, Meyer G (eds) Advanced Microsystems for Automotive Applications 2015. Springer International Publishing, Cham, pp 25–35

82. O’ Kelly M, Sinha A, Namkoong H, et al (2018) Scalable End-to-End Autonomous Vehicle Testing via Rare-event Simulation. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.

83. Arief M, Glynn P, Zhao D (2018) An Accelerated Approach to Safely and Efficiently Test Pre-Production Autonomous Vehicles on Public Streets. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 2006–2011

84. Zhang S, Peng H, Zhao D, Tseng HE (2018) Accelerated Evaluation of Autonomous Vehicles in the Lane Change Scenario Based on Subset Simulation Technique. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 3935–3940

85. Huang Z, Zhao D, Lam H, et al (2017) Evaluation of automated vehicles in the frontal cut-in scenario — An enhanced approach using piecewise mixture models. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp 197–202

86. Huang Z, Lam H, Zhao D (2017) Sequential experimentation to efficiently test automated vehicles. In: 2017 Winter Simulation Conference (WSC). pp 3078–3089

87. Huang Z, Lam H, Zhao D (2017) Towards affordable on-track testing for autonomous vehicle — A Kriging-based statistical approach. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). pp 1–6

88. Huang Z, Lam H, Zhao D (2017) An accelerated testing approach for automated vehicles with background traffic described by joint distributions. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). pp 933–938

89. Huang Z, Guo Y, Arief M, et al (2018) A Versatile Approach to Evaluating and Testing Automated Vehicles based on Kernel Methods. In: 2018 Annual American Control Conference (ACC). pp 4796–4802

90. Huang Z, Lam H, LeBlanc DJ, Zhao D (2018) Accelerated Evaluation of Automated Vehicles Using Piecewise Mixture Models. IEEE Transactions on Intelligent Transportation Systems 19:2845–2855. https://doi.org/10.1109/TITS.2017.2766172

91. Huang Z, Arief M, Lam H, Zhao D (2018) Synthesis of Different Autonomous Vehicles Test Approaches. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 2000–2005

92. Huang Z, Arief M, Lam H, Zhao D (2019) Evaluation Uncertainty in Data-Driven Self-Driving Testing. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp 1902–1907

93. Akagi Y, Kato R, Kitajima S, et al (2019) A Risk-index based Sampling Method to Generate Scenarios for the Evaluation of Automated Driving Vehicle Safety. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp 667–672

94. So JJ, Park I, Wee J, et al (2019) Generating Traffic Safety Test Scenarios for Automated Vehicles using a Big Data Technique. KSCE J Civ Eng 23:2702–2712. https://doi.org/10.1007/s12205-019-1287-4

95. Fahrenkrog F, Wang L, Platzer T, et al (2019) Prospective Safety Effectiveness Assessment of Automated Driving Functions – From the Methods to the Results

96. Feig P, Schatz J, Labenski V, Leonhardt T (2019) Assessment of Technical Requirements for Level 3 and Beyond Automated Driving Systems Based on Naturalistic Driving and Accident Data Analysis

97. Stark L, Obst S, Schoenawa S, Düring M (2019) Towards Vision Zero: Addressing White Spots by Accident Data based ADAS Design and Evaluation. In: 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES). pp 1–6

98. Stark L, Düring M, Schoenawa S, et al (2019) Quantifying Vision Zero: Crash avoidance in rural and motorway accident scenarios by combination of ACC, AEB, and LKS projected to German accident occurrence. Traffic Injury Prevention 20:S126–S132. https://doi.org/10.1080/15389588.2019.1605167

99. Esenturk E, Khastgir S, Wallace A, Jennings P (2021) Analyzing Real-world Accidents for Test Scenario Generation for Automated Vehicles. In: 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, Nagoya, Japan, pp 288–295

100. Pierson A, Schwarting W, Karaman S, Rus D (2019) Learning Risk Level Set Parameters from Data Sets for Safer Driving. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 273–280

101. Althoff M, Lutz S (2018) Automatic Generation of Safety-Critical Test Scenarios for Collision Avoidance of Road Vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1326–1333

102. Klischat M, Althoff M (2019) Generating Critical Test Scenarios for Automated Vehicles with Evolutionary Algorithms. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 2352–2358

103. Koopman P, Fratrik F (2019) How Many Operational Design Domains, Objects, and Events? In: Espinoza H, hÉigeartaigh SÓ, Huang X, et al (eds) Proceedings of the AAAI Workshop on Artificial Intelligence Safety 2019. CEUR, Honolulu, Hawaii

104. Ponn T, Schwab A, Diermeyer DF, et al (2019) A Method for the Selection of Challenging Driving Scenarios for Automated Vehicles Based on an Objective Characterization of the Driving Behavior

105. Ponn T, Diermeyer F, Gnandt C (2019) An Optimization-Based Method to Identify Relevant Scenarios for Type Approval of Automated Vehicles

106. Qi Y, Luo Y, Li K, et al (2019) A Trajectory-Based Method for Scenario Analysis and Test Effort Reduction for Highly Automated Vehicle. SAE International, Warrendale, PA

107. Zhang C, Liu Y, Zhang Q, Wang L (2018) A Graded Offline Evaluation Framework for Intelligent Vehicle’s Cognitive Ability. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 320–325

108. Wang J, Zhang C, Liu Y, Zhang Q (2018) Traffic Sensory Data Classification by Quantifying Scenario Complexity. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1543–1548

109. Xia Q, Duan J, Gao F, et al (2018) Test Scenario Design for Intelligent Driving System Ensuring Coverage and Effectiveness. IntJ Automot Technol 19:751–758. https://doi.org/10.1007/s12239-018-0072-6

110. Xia Q, Duan J, Gao F, et al (2017) Automatic Generation Method of Test Scenario for ADAS Based on Complexity. pp 2017-01–1992

111. Gao F, Duan J, He Y, Wang Z (2019) A Test Scenario Automatic Generation Strategy for Intelligent Driving Systems. Mathematical Problems in Engineering 2019:e3737486. https://doi.org/10.1155/2019/3737486

112. Koschi M, Pek C, Maierhofer S, Althoff M (2019) Computationally Efficient Safety Falsification of Adaptive Cruise Control Systems. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp 2879–2886

113. Felbinger H, Klück F, Li Y, et al (2019) Comparing two systematic approaches for testing automated driving functions. In: 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE). pp 1–6

114. Tuncali CE (2019) Search-based Test Generation for Automated Driving Systems: From Perception to Control Logic. Ph.D., Arizona State University

115. Tuncali CE, Fainekos G, Prokhorov D, et al (2020) Requirements-Driven Test Generation for Autonomous Vehicles With Machine Learning Components. IEEE Transactions on Intelligent Vehicles 5:265–280. https://doi.org/10.1109/TIV.2019.2955903

116. Tuncali CE, Fainekos G, Ito H, Kapinski J (2018) Simulation-based Adversarial Test Generation for Autonomous Vehicles with Machine Learning Components. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1555–1562

117. Tuncali CE, Yaghoubi S, Pavlic TP, Fainekos G (2017) Functional gradient descent optimization for automatic test case generation for vehicle controllers. In: 2017 13th IEEE Conference on Automation Science and Engineering (CASE). pp 1059–1064

118. Tuncali CE, Pavlic TP, Fainekos G (2016) Utilizing S-TaLiRo as an automatic test generation framework for autonomous vehicles. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). pp 1470–1475

119. Abbas H, O'Kelly M, Rodionova A, Mangharam R (2019) Safe At Any Speed: A Simulation-Based Test Harness for Autonomous Vehicles. In: Chamberlain R, Taha W, Törngren M (eds) Cyber Physical Systems. Design, Modeling, and Evaluation. Springer International Publishing, Cham, pp 94–106

120. Nabhan M, Schoenauer M, Tourbier Y, Hage H (2019) Optimizing coverage of simulated driving scenarios for the autonomous vehicle. In: 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE). pp 1–5

121. Gangopadhyay B, Khastgir S, Dey S, et al (2019) Identification of Test Cases for Automated Driving Systems Using Bayesian Optimization. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp 1961–1967

122. Mullins GE, Stankiewicz PG, Gupta SK (2017) Automated generation of diverse and challenging scenarios for test and evaluation of autonomous vehicles. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp 1443–1450

123. Mullins GE, Stankiewicz PG, Hawthorne RC, Gupta SK (2018) Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles. Journal of Systems and Software 137:197–215. https://doi.org/10.1016/j.jss.2017.10.031

124. Mullins GE (2018) Adaptive Sampling Methods for Testing Autonomous Systems. https://doi.org/10.13016/M2PZ51Q5N

125. Ben Abdessalem R, Nejati S, Briand LC, Stifter T (2016) Testing advanced driver assistance systems using multi-objective search and neural networks. In: Proceedings

of the 31st IEEE/ACM International Conference on Automated Software Engineering. Association for Computing Machinery, New York, NY, USA, pp 63–74

126. Beglerovic H, Stolz M, Horn M (2017) Testing of autonomous vehicles using surrogate models and stochastic optimization. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). pp 1–6

127. Corso A, Du P, Driggs-Campbell K, Kochenderfer MJ (2019) Adaptive Stress Testing with Reward Augmentation for Autonomous Vehicle Validatio. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp 163–168

128. Lee R, Kochenderfer MJ, Mengshoel OJ, et al (2015) Adaptive stress testing of airborne collision avoidance systems. In: 2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC). pp 6C2-1-6C2-13

129. Lee R, Mengshoel O, Saksena A, et al (2018) Differential Adaptive Stress Testing of Airborne Collision Avoidance Systems. In: 2018 AIAA Modeling and Simulation Technologies Conference. American Institute of Aeronautics and Astronautics, Kissimmee, Florida

130. Koren M, Alsaif S, Lee R, Kochenderfer MJ (2018) Adaptive Stress Testing for Autonomous Vehicles. In: 2018 IEEE Intelligent Vehicles Symposium (IV). pp 1–7

131. Hallerbach S, Xia Y, Eberle U, Koester F (2018) Simulation-Based Identification of Critical Scenarios for Cooperative and Automated Vehicles. SAE Intl J CAV 1:93–106. https://doi.org/10.4271/2018-01-1066

132. Kirchhof JC, Kusmenko E, Rumpe B, Zhang H (2019) Simulation as a Service for Cooperative Vehicles. In: 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C). pp 28–37

133. Riedmaier S, Nesensohn J, Gutenkunst C, et al (2018) Validation of X-in-the-Loop Approaches for Virtual Homologation of Automated Driving Functions. p 12

134. Nicolas Wagener, Jean-Baptiste Coget, Athanasios Ballis, et al (2021) HEADSTART D2.1 Common methodology for test, validation, certification and criteria to allocate scenarios

135. Sjef van Montfort, Esra van Dam, Jacco van de Sluis, et al (2021) HEADSTART D4.2 Evaluation results of application and demonstration

136. Bernhard Hillbrand, Patrick Weissensteiner, Joaquim Castella Triginer, et al (2020) HEADSTART D3.1 Guideline of a comprehensive validation and certification procedure to ensure safe CAD systems

137. Bernhard Hillbrand, Patrick Weissensteiner, Michael Hoß, et al (2020) HEADSTART D3.2 Toolchain for mixed validation – integration of simulation and physical testing

138. Sjef van Montfort, Jeroen Broos, Sytze Kalisvaart, et al (2022) HEADSTART D3.3 Assessment criteria of CAD functionalities

139. Marcos Nieto, Patrick Weißensteiner, Camille Plestan, et al (2021) HEADSTART D3.4 Harmonisation proposals of test results

140. Nicolas Wagener, Jean-Baptiste Coget, Oihana Otaegui, et al (2019) HEADSTART D1.1 State of innovation of existing initiatives and gap analysis

141. PEGASUS method - an overview

142. Pegasus Method - pegasus-EN. https://www.pegasusprojekt.de/en/pegasus-method. Accessed 30 Jun 2023

143. Elrofai H, Paardekooper JP, de Gelder E, et al (2018) StreetWise: Scenario-based safety validation of connected automated driving

144. UNECE (2021) UN Regulation No 157 – Uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping Systems [2021/389]

145. Pütz A, Zlocki A, Bock J, Eckstein L (2017) System validation of highly automated vehicles with a database of relevant traffic scenarios. p 8

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

147. de Gelder E, Paardekooper J-P, Saberi AK, et al (2022) Towards an Ontology for Scenario Definition for the Assessment of Automated Vehicles: An Object-Oriented Framework. IEEE Transactions on Intelligent Vehicles 7:300–314. https://doi.org/10.1109/TIV.2022.3144803

148. Erwin de Gelder, Olaf Op den Camp, Niels de Boer (2020) Scenario Categories for the Assessment of Automated Vehicles. CETRAN

149. de Gelder E (2020) Procedure for the Safety Assessment of an Autonomous Vehicle Using Real-World Scenarios. In: Proceedings of the FISITA 2020 World Congress, Prague, 14-18 September 2020

150. De Gelder E, Op Den Camp O (2023) How certain are we that our automated driving system is safe? Traffic Injury Prevention 24:S131–S140. https://doi.org/10.1080/15389588.2023.2186733

151. Safety Pool Scenario Database. https://www.safetypool.ai/database. Accessed 17 Mar 2023

152. British Standards Institution (2020) PAS 1883:2020 Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) - Specification

153. Menzel T, Bagschik G, Isensee L, et al (2019) From Functional to Logical Scenarios: Detailing a Keyword-Based Scenario Description for Execution in a Simulation Environment. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp 2383–2390

154. Ulbrich S, Menzel T, Reschka A, et al (2015) 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

155. Chen S, Khastgir S, Babaev I, Jennings P (2020) Identifying Accident Causes of Driver-Vehicle Interactions Using System Theoretic Process Analysis (STPA). In: 2020 IEEE

International Conference on Systems, Man, and Cybernetics (SMC). IEEE, Toronto, ON, Canada, pp 3247–3253

156. Euro NCAP (2022) Euro NCAP Vision 2030: a safer future for mobility

157. Zhang X, Khastgir S, Jennings P (2020) Scenario Description Language for Automated Driving Systems: A Two Level Abstraction Approach. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, Toronto, ON, Canada, pp 973–980

158. BSI Group (2022) FLEX 1889:2022 Natural language description for abstract scenarios for automated driving systems – Specification

159. Bruto Da Costa AA, Irvine P, Zhang X, et al (2022) Writing Accessible and Correct Test Scenarios for Automated Driving Systems. In: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, Prague, Czech Republic, pp 1356–1363

160. Bruto da Costa AA, Irvine P, Zhang X, et al (2022) Translating automated vehicle test scenario specifications between scenario languages : learnings and challenges. In: Proceedings of the Driving Simulation Conference Europe 2022 VR. Driving Simulation Association, Strasbourg, France, pp 65–72

161. ASAM e.V. (2021) ASAM OpenLABEL v1.0.0. https://www.asam.net/standards/detail/openlabel/. Accessed 22 Jun 2023

162. Zhang X, Tao J, Tan K, et al (2021) Finding Critical Scenarios for Automated Driving Systems: A Systematic Literature Review

163. Thiolon G (2018) Real-world driving scenario identification for AV functional safety. Stuttgart, Germany

164. Foyer B, Revert E, Tlig M (2021) Scenario library management tools for ADS design & validation, a new way to manage driving simulation. Munich, Germany

165. PFA Safety (2019) SAFETY ARGUMENTATION FOR AUTOMATED VEHICLE SAE LEVEL OF AUTOMATION 3 AND 4. Automated Driving Safety Working Group

166. Nakamura H, Muslim H, Kato R, et al (2022) Defining Reasonably Foreseeable Parameter Ranges Using Real-World Traffic Data for Scenario-Based Safety Assessment of Automated Vehicles. IEEE Access 10:37743–37760. https://doi.org/10.1109/ACCESS.2022.3162601

167. Thal S, Znamiec H, Henze R, et al (2020) Incorporating safety relevance and realistic parameter combinations in test-case generation for automated driving safety assessment. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). pp 1–6

168. Stepien L, Thal S, Henze R, et al (2021) Applying heuristics to generate test cases for automated driving safety evaluation. https://doi.org/10.3390/app112110166

169. Eric Thorn, Shawn Kimmel, Michelle Chaka (2018) A Framework for Automated Driving System Testable Cases and Scenarios. National Highway Traffic Safety Administration, Washington, DC

170. SAE (2021) SAE J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International

171. VOICES - Confluence. https://usdot-voices.atlassian.net/wiki/spaces/VP/overview. Accessed 27 Jun 2023

172. VOICES (2021) Virtual Open Innovation Collaborative Environment for Safety

173. VOICES (2022) VOICES Concept of Operations and Capabilities

174. Cieslik I, Jiménez VJE, Martin H, et al (2022) State of the Art Study of the Safety Argumentation Frameworks for Automated Driving System Safety. pp 178–191

175. ISO (2018) ISO 26262:2018 Road Vehicles : Functional Safety. ISO

176. ISO (2023) ISO/FDIS 34503 "Road Vehicles – Test scenarios for automated driving systems – Specification for operational design domain"

177. ISO (2022) ISO/AWI 34505 "Road vehicles – Test scenarios for automated driving systems – Scenario evaluation and test case generation"

178. ISO (2021) ISO/SAE PAS 22736:2021 Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles

179. ISO (2022) ISO 21448:2022 - Road vehicles -- Safety of the intended functionality

180. ASAM Standards. https://www.asam.net/standards/. Accessed 22 Mar 2023

181. ANSI/UL (2022) 4600:2022 Standard for Safety for the Evaluation of Autonomous Products

182. ISO (2022) ISO/DIS 34503 "Road Vehicles – Test scenarios for automated driving systems – Specification for operational design domain"

183. ECE/TRANS/WP.29/2021/61 (2021) (GRVA) New Assessment/Test Method for Automated Driving (NATM) - Master Document | UNECE

184. Sun J, Zhang H, Zhou H, et al (2022) Scenario-based test automation for highly automated vehicles: A review and paving the way for systematic safety assurance. IEEE Transactions on Intelligent Transportation Systems 23:14088–14103. https://doi.org/10.1109/TITS.2021.3136353

185. Krajewski R, Bock J, Kloeker L, Eckstein L (2018) The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In: 2018 21st international conference on intelligent transportation systems (ITSC). pp 2118–2125

186. Re F, Kriston A, Broggi D, Minarini F (2021) Testing the Robustness of Commercial Lane Departure Warning Systems. Transportation Research Record 2675:385–400. https://doi.org/10.1177/03611981211029646

187. Rajabli N, Flammini F, Nardone R, Vittorini V (2021) Software Verification and Validation of Safe Autonomous Cars: A Systematic Literature Review. IEEE Access 9:4797–4819. https://doi.org/10.1109/ACCESS.2020.3048047

188. Sun J, Zhang H, Zhou H, et al (2022) Scenario-Based Test Automation for Highly Automated Vehicles: A Review and Paving the Way for Systematic Safety Assurance. IEEE Transactions on Intelligent Transportation Systems 23:14088–14103. https://doi.org/10.1109/TITS.2021.3136353

189. Christian Berger, Lukas Birkemeyer (2023) ZEBRA: Z-order Curve-based Event Retrieval Approach to Efficiently Explore Automotive Data. In: Proceedings of the 35th IEEE Intelligent Vehicles Symposium. Anchorage, Alaska, USA

190. (2019) DIN SAE SPEC 91381 Terms and Definitions Related to Testing of Automated Vehicle Technologies

191. PEGASUS research project. https://www.pegasusprojekt.de/en/home. Accessed 1 Dec 2022

192. SAKURA, SIP-adus, HEADSTART -project (2021) White Paper: Towards the Harmonization of Safety As-sessment Methods

193. Olaf Op den Camp, Erwin de Gelder, Jeroen Uittenbogaard, Jeroen Brood (2022) StreetWise: Scenario-Based Approach to Describe Real- World Traffic. Society of Automotive Engineers of Japan, Inc., Yokohama, Japan

194. ISO (2021) ISO/TR 21934-1:2021 Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation