Dipl.-Ing. Víctor J. Expósito Jiménez

Validation of Safety of the Intended Functionality for Perception Systems in Autonomous Driving

DOCTORAL THESIS
to achieve the university degree of
Doktor der technischen Wissenschaften

submitted to
Graz University of Technology

Supervisor
Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eugen Brenner

Institute of Technical Informatics

Co-Supervisor
Dipl.-Ing. Dr.techn. Georg Macher, BSc. MBA

Graz, April 2025

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material that has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present thesis.

Date	Signature

Acknowledgements

This thesis is a journey of many years in which many people have been involved, directly and indirectly, but who have all contributed to reaching the goal of completing this research. I would like to thank Prof. Eugen Brenner and Dr. Georg Macher for their support and advice during these years, pointing me the right direction through their valuable comments. Furthermore, I would like to express my gratitude to Virtual Vehicle Research GmbH, particularly my colleagues from the Dependable Systems and Embedded Systems Group, who provided me with invaluable assistance, guidance, and advice. Special mention goes to Dr. Helmut Martin and Dr. Christian Schwarzl for encouraging me to start and carry on despite all the difficulties during this time.

My greatest and heartfelt gratitude to my wife, who has lovingly supported me without any hesitation, and to my girls, Eva and Elena, whose joy and cheerfulness have given me strength in the most difficult moments. Not to forget my deep appreciation to my parents and my brother, who have been a pillar of support throughout all my life. Thanks to all my friends, specially Dr. José Romero Lopera, who has patiently been listening to me about this thesis more than he probably wanted to. Finally, I want to thank everyone who was there for me during this time. This thesis would not have been possible without all their incredible support.

This thesis was written at Virtual Vehicle Research GmbH in Graz, Austria. The author would like to acknowledge the financial support within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Labour and Economy (BMAW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorised for the programme management.

Graz, April 2025
Víctor J. Expósito Jiménez

Abstract

Highly automated vehicles have increased the reliance on perception sensors, which model the reality that is perceived by the vehicle. Therefore, the safety validation of autonomous driving functions has to be evolved since the usage of classical strategies cannot be applied due to the large number of scenarios to be covered as well as the identification of unknown scenarios that may lead to a hazardous behaviour at the vehicle level. The main goal of this thesis is to provide a safety argumentation that could be used in this new paradigm in which the Safety Of The Intended Functionality (SOTIF) has a key role. First, the methodology reuses the existing hazard analysis and risk assessment from functional safety, adding the relevant parameters for SOTIF to identify the potential triggering conditions associated with each identified hazardous behaviour.

As part of the validation, this research work also provides a way to include the potential triggering conditions into a validation tool by their decomposition in one or more scenario constraints. The uncertainty of validating all possible scenarios, not only to cover the known hazardous scenarios but especially to discover unknown scenarios is one of the main issues that a SOTIF validation has to face. Therefore, we investigate a feasible way to cover the maximum number of scenarios by focusing on the impact on the system instead of the cause of the hazardous behaviour. In this context, a classification of performance insufficiencies is given based on the impact that they can include on the output of the system. Then, these defined performance insufficiencies are injected into the perception part of the system to determine if they may lead to a hazardous behaviour at the vehicle level. The results of these simulations are used to quantify the risk of the function against the injected insufficiency. The proposed argumentation is finally implemented in practice through a use case. The novel SOTIF validation proposed in this thesis can support engineers to quantify the system reliability, which provides not only safer systems but also helps reduce the costs of implementing autonomous driving.

Kurzfassung

Hochautomatisierte Fahrzeuge haben eine erhöhte Abhängigkeit von Wahrnehmungssensoren, die die Realität modellieren Basis für die durch Fahrzeug wahrgenommen wird. Die Sicherheitsvalidierung autonomer Fahrfunktionen muss weiterentwickelt werden, da die Verwendung klassischer Strategien aufgrund der großen Anzahl der abzudeckenden Szenarien, sowie der Identifizierung unbekannter Szenarien, welche zu einem gefährlichen Verhalten auf Fahrzeugebene führen können, nicht möglich ist. Das Hauptziel der vorliegenden Arbeit ist es, eine Sicherheitsargumentation zu liefern, die in diesem neuen Paradigma verwendet werden kann, in dem der Begriff Safety Of the Intended Functionality (SOTIF) eine Schlüsselrolle spielt. Die Methodik greift zunächst auf die bestehende Gefahrenanalyse und Risikobewertung aus der funktionalen Sicherheit zurück und fügt die relevanten Parameter für SOTIF hinzu, um die potenziellen Auslösebedingungen zu ermitteln, die mit jeder identifizierten, gefährlichen Verhaltensweise verbunden sind.

Als Teil der Validierung bietet diese Forschungsarbeit auch eine Möglichkeit, die potenziell auslösenden Bedingungen in ein Validierungswerkzeug einzubeziehen, indem sie in eine oder mehrere eingeschränkte Szenarien zerlegt werden. Die Ungewissheit, alle möglichen Szenarien zu validieren, nicht nur um die bekannten gefährlichen Szenarien abzudecken, sondern vor allem um unbekannte Szenarien zu entdecken, ist eines der Hauptprobleme, mit denen eine SOTIF-Validierung konfrontiert ist. Daher untersuchen wir einen praktikablen Weg, um eine maximale Anzahl von Szenarien abzudecken, indem wir uns auf die Auswirkungen im System konzentrieren und nicht auf die Ursache des gefährlichen Verhaltens. In diesem Kontext wird eine Klassifizierung von Leistungsmängeln auf der Grundlage der Auswirkungen vorgenommen, die sie auf das System haben können. Anschließend werden diese definierten Leistungsmängel injiziert, um festzustellen, ob sie zu einem gefährlichen Verhalten auf Fahrzeugebene führen können. Die Ergebnisse dieser Simulationen werden verwendet, um das Risiko der Funktion gegenüber der eingebrachten Unzulänglichkeit zu quantifizieren. Diese vorgeschlagene Argumentation wird schließlich anhand eines Anwendungsfalls in die Praxis umgesetzt. Die neuartige SOTIF-Validierung, die in dieser Arbeit vorgeschlagen wird, kann Ingenieure dabei unterstützen, die Systemzuverlässigkeit zu quantifizieren, was nicht nur zu sichereren Systemen führt, sondern auch dazu beiträgt, die Kosten für die Implementierung des autonomen Fahrens zu senken.

Contents

List of Figures

List of Tables

List of Abbreviations

AD	Automated Driving
ADAS	Advanced Driver Assistance Systems
ADS	Automated Driving System
AEB	Automated Emergency Brake
AI	Artificial Intelligence
AIS	Abbreviated Injury Scale
ALARP	As low as reasonably Practicable
ASAM	Association for Standardization of Automation and Measuring Systems
ASIL	Automotive Safety Integrity Level
BMAW	Austrian Federal Ministry for Labour and Economy
BMK	Austrian Federal Ministry for Climate Action
BSI	British Standards Institution
CPI	Crash Potential Index
DDT	Dynamic Driving Task
DST	Deceleration to Safety Time
FFG	Österreichische Forschungsförderungsgesellschaft (engl.: Austrian Research Promotion Agency)
FoV	Field of View
GmbH	Gesellschaft mit beschränkter Haftung
GPI	Generic Performance Insufficiency
HARA	Hazard Analysis and Risk Assessment
HIL	Hardware-in-the-Loop
ISO	International Organization for Standardization
KPI	Key Performance Indicator
MAIS	Maximum Abbreviated Injury Scale
ODD	Operational Design Domain
PAS	Publicly Available Specification

ROS Robotics Operating System
RSS Responsibility-Sensitive Safety

SAE Society of Automotive Engineers
SFG Styrian Business Promotion Agency
SOTA State of the Art
SOTIF Safety of the Intented Functionality
SPI Safety Perfomance Indicator
STPA System-Theoretic Process Analysis

TCPI Triggering Condition Performance Insufficiency
TPI Technology Performance Insufficiency
TTC Time To Collision
TTM Time To Maneuver

CHAPTER 1

Introduction

This chapter settles the basis of the dissertation providing a motivation for the topic and the research questions that arise from the outline problem. Moreover, the structure of the thesis is described with a brief description of each chapter and the involved publications.

1.1 Motivation

The increasing complexity of modern Advanced Driver Assistance Systems (ADAS) and Automated Driving (AD) functions has driven the automotive industry to shift from traditional component-based validation to scenario-based validation. This shift arises from the growing challenges in safety assurance, as validation now requires addressing not only malfunctions and external attacks but also scenario conditions and complex algorithms. As a result, the effort needed to generate quantitative evidence for system safety has increased significantly. While component-based validation focuses on ensuring that individual components function correctly, scenario-based validation expands the scope to include situations where all components operate as intended, but interactions within the scenario could still lead to hazardous behaviour. In this context, Safety of the Intented Functionality (SOTIF) plays a crucial role by addressing hazards caused by misuse or technical limitations (e.g., performance insufficiencies) that may be triggered under specific scenario conditions.

In higher levels of driving automation (Society of Automotive Engineers (SAE) L3+), the system increasingly relies on perception sensors to collect and process environmental data, which is used to make and take decisions. It is crucial that the systems are validated for most wide scenario conditions to ensure that it is safer in the highest number of scenarios. Therefore, the major benefit of providing SOTIF argumentation is producing more reliable systems that lead to safer systems, i.e. less accidents and fatalities. An argumentation like this also speeds up the implementation of automated or autonomous vehicles since the process helps to discover unknown and hazardous scenarios that could be validated during the development phase and could be reused along different developments. Moreover, it can decrease the cost of the implementation of an autonomous system because the quantified risk could be compared for different sensor technologies and configurations to adopt the safest and most cost-effective setup.

1.2 Problem Statement and Research Questions

One of the main challenges in ADAS/AD is the validation of the safety of autonomous driving functions, due to the infinite number of possible driving scenarios a vehicle can get into. Since not all possible driving scenarios are known during the development phase of an ADAS/AD function, these remain unknown until they are discovered during operation in the field, which may cause hazardous situations. The usage of classical test strategies cannot be applied to these unknown scenarios because its expected outcome is a prerequisite for testing. In addition, the large number of scenarios and conditions requires consideration of new validation strategies that could maximise scenario coverage in a manageable way. Another point to notice is the necessity of the usage of metrics that can provide an objective quantification. Although qualitative metrics provide valuable insights, especially in understanding the context, quantitative metrics offer precision, objectivity, and scalability, making them indispensable in this kind of domain.

Based on the research gaps highlighted, the following research questions for this thesis are formulated:

RQ1 How could the validation and verification of ADAS/AD functions be improved by using the injection of triggering conditions?

- **RQ1.1** How to include triggering conditions to be validated into software tools?
- **RQ1.2** How to find a way to maximise the coverage of the validation of potential triggering conditions?
- **RQ1.3** How to provide a quantitative metric that can measure the risk of an ADAS/AD system?

Therefore, the main goal of this thesis is the research of an argumentation that provides quantitative information about the reliability of the function and the usefulness of the introduced safety measures to achieve an acceptable risk to humans. The scope of the thesis is focused on the perception side to enclose the topic and make it feasible.

1.3 Structure of the Thesis

The content of this dissertation is taken from the contributions of the author in the following published publications:

- C. Schwarzl, N. Marko, H. Martin, V. Expósito Jiménez, J. Castella Triginer, B. Winkler, and R. Bramberger, "Safety and Security Co-engineering for Highly Automated Vehicles," *e & i Elektrotechnik und Informationstechnik,* vol. 138, pp. 469–479, nov 2021. [4] – Journal
- I. Cieslik, V. J. Expósito Jiménez, H. Martin, H. Scharke, and H. Schneider, "State of the Art Study of the Safety Argumentation Frameworks for Automated Driving System," in *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops* (M. Trapp, E.

Schoitsch, J. Guiochet, and F. Bitsch, eds.), (Cham), pp. 178–191, Springer International Publishing, 2022. [5] – Conference

- V. J. Expósito Jiménez, H. Martin, C. Schwarzl, G. Macher, and E. Brenner, "Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions," in *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops* (M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, eds.), (Cham), pp. 11–22, Springer International Publishing, 2022. [6] – Conference

- V. J. Expósito Jiménez, B. Winkler, J. M. Castella Triginer, H. Scharke, H. Schneider, E. Brenner, and G. Macher, "Safety of the Intended Functionality Concept Integration into a Validation Tool Suite," *ACM SIGAda Ada Letters*, vol. 43, pp. 69––72, jun 2024. [7] – Journal

- V. J. Expósito Jiménez, G. Macher, D. Watzenig, and E. Brenner, "Safety of the Intended Functionality Validation for Automated Driving Systems by using Perception Performance Insufficiencies Injection," *Vehicles*, vol. 6, no. 3, pp. 1164–1184, 2024. [8] – Journal

- H. Scharke, H. Goossens, S. Kalisvaart, V. J. Expósito Jiménez, "Systematic SOTIF-aligned approach to explore residual risk in autonomous driving systems," in *Book Of Proceedings - International Congress: SIA VISION 2024 - 16 and 17 OCTOBER 2024*, (Paris, France), pp. 185–192, Société des Ingénieurs de l'Automobile, 2024. [9] – Conference

While many text passages and figures are taken directly from the original publications, the objective of this dissertation is to present a unified document that encompasses all research conducted in the previous publications, thereby providing a comprehensive overview of the proposed argumentation. The structure of the thesis is as follows. The next chapter provides the reader with the technical background knowledge to understand the main topics that are covered during this dissertation. Then, the State of the Art (SOTA) of the main topics involved in the thesis can be found in Chapter 3. The proposed argumentation is explained in Chapter 4, where the main concepts are described in detail. Subsequently, the approach is evaluated in the next Chapter 5. In this chapter, the implementation and its results are shown through a use case. Conclusions and outlook are given in Chapter 6. Finally, the full-text publications are collected in the appendix together with the contribution that each publication has brought to the thesis.

CHAPTER 2

Background and Terminology

This chapter sets out the principal background knowledge that is required for the dissertation, with a particular focus on the topics that are related to the work described in this thesis. The field of automotive safety is a vast and multifaceted domain, encompassing a diverse list of topics. A general definition of safety is the absence of unacceptable risk in which a risk is considered as a physical injury or damage to the health of people or to the environment. In the context of automotive safety, the dependability of a system is based on three main areas: functional safety, cybersecurity, and Safety Of The Intended Functionality (SOTIF) as it is depicted in Figure 2.1. Functional safety is focused on failures and malfunctions as described in International Organization for Standardization (ISO) 26262:2018 [10]. External attacks are covered on cybersecurity in the standard ISO/SAE 21434:2021 [11]. Finally, SOTIF is focused on technical shortcomings and human misuses, which is described in ISO21448:2022 [1].

2.1 Functional Safety

The ISO26262:2018 [10] is an essential standard in the automotive domain. The first version was published in 2011. A second version was published in 2018, which is the current version. The main change in this second version was the addition of a guideline for semiconductors, whose implementation has increased significantly in recent years. Based on the standard, functional safety is the absence of risk due to hazards caused by failure of electrical and electronic systems or unintended behaviour, focusing on avoiding systematic faults, controlling systematic faults and random hardware failures as well as including a safety culture that provides management commitment and responsibility. The standard provides an automotive safety lifecycle, including an automotive-specific risk-based approach to determine safety integrity levels. The determination of these integrity levels is intended to specify applicable requirements to avoid unacceptable residual risks, which, in turn, are used as confirmation measures to ensure a sufficient and acceptable level of safety is achieved.

The Hazard Analysis and Risk Assessment (HARA) methodology is always performed in functional safety processes in which associated system hazards are analysed and classified according to their hazardousness. The HARA methodology assigns a safety level to each defined hazard called the Automotive Safety Integrity Level (ASIL). The calculated ASIL of

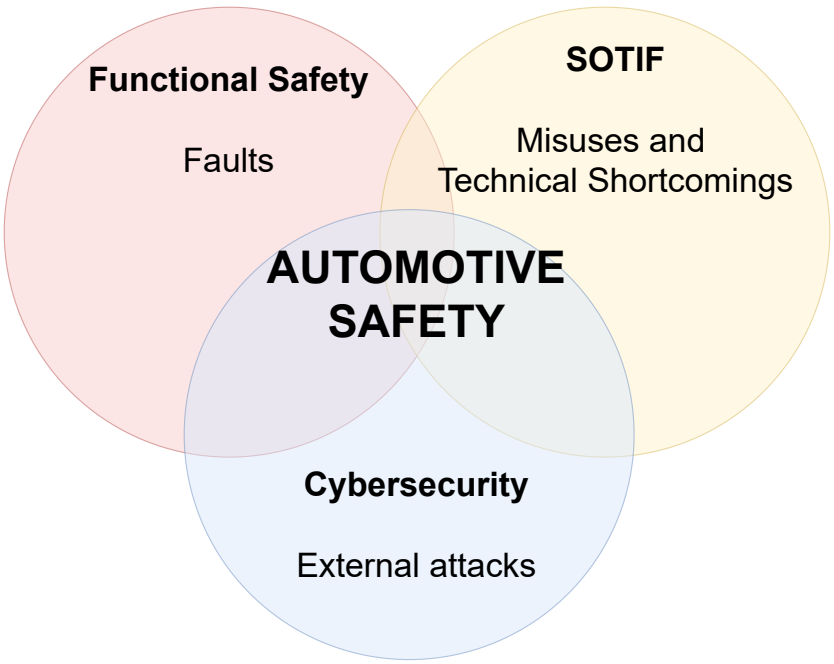


Figure 2.1: Dependability of a system.

a hazard is based on three main variables:

- Severity (S): the level of injury to the driver and passengers. The definition of the injuries is based on the Abbreviated Injury Scale (AIS). [12]
- Exposure (E): how often the hazard occurred during the driving time. There are two types of exposures, one based on the duration and the other one regarding the frequency in operational situations.
- Controllability (C): if the hazard could be controlled by the driver.

These variables are defined in different classes according to the influence of the hazardous events, ranging from 0 (mildest) to 4 (worst). The different classes are defined in the standard (see ISO262626:2018 [10] part 3) and are described in the tables below (Table 2.1, Table 2.2, Table 2.3, Table 2.4). The sum of these three variables (e.g., $S0+C2+E1$) is used to determine the ASIL level. ASIL A is assigned to the lowest risk, while ASIL D is assigned to the highest risk. An example of a HARA is given in Figure 2.2, whereas Figure 2.3 shows the ASIL determination according to all possible severity, exposure, and controllability classes.

Once a HARA is performed and the hazards have been assigned to an ASIL level, a safety goal is defined for each hazard to reach the acceptance criteria to consider the item safe. They are top-level requirements with a usually abstract description. The functional safety requirements are derived from the safety goals and are on the level of the system function. Following the requirement hierarchy, technical safety requirements are also derived from these

Hazard	Malfunction	Situation	S	Argumentation for S	E	Argumentation for E	C	Argumentation for C	ASIL
Unintendend omnidirectional movement lead to collison with objects or persons	AGV accelerate inverse	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: No special Maneuver	S2	Serious injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL C
Unintendend omnidirectional movement lead to collison with objects or persons	AGV accelerate inverse	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Following and Curve	S2	Serious injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL C
Unintendend omnidirectional movement lead to collison with objects or persons	AGV accelerate inverse	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Overtaking of other vehicle	S2	Serious injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL C
Unintendend stop leads to collision with moving object behind	False object dectection	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: No special Maneuver	S1	Light and moderated injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL B
Unintendend stop leads to collision with moving object behind	False object dectection	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Following and Curve	S1	Light and moderated injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL B

Figure 2.2: HARA example.

		C1	C2	C3
S1	E1	QM	QM	QM
	E2	QM	QM	QM
	E3	QM	QM	ASIL A
	E4	QM	ASIL A	ASIL B
S2	E1	QM	QM	QM
	E2	QM	QM	ASIL A
	E3	QM	ASIL A	ASIL B
	E4	ASIL A	ASIL B	ASIL C
S3	E1	QM	QM	ASIL A
	E2	QM	ASIL A	ASIL B
	E3	ASIL A	ASIL B	ASIL C
	E4	ASIL B	ASIL C	ASIL D

Figure 2.3: ASIL determination.

Table 2.1: Severity classification.

S0	**S1**	**S2**	**S3**
No injuries	Light and moderate injuries	Severe and life-threatening injuries, survival probable	Life-threatening injuries (survival uncertain), fatal injuries
AIS 0 and less than 10% probability of AIS 1-6. Damage that cannot be classified safety-related	More than 10% probability of AIS 1-6 (and not S2 or S3)	More than 10% probability of AIS 3-6 (and not S3)	More than 10% probability of AIS 5-6

Table 2.2: Classes of probability of exposure regarding duration in operational situations.

E1	**E2**	**E3**	**E4**
Very low probability	Low probability	Medium probability	High probability
Duration not specified	<1% of average operating time	1 % to 10 % of average operating time	>10 % of average operating time

Table 2.3: Classes of probability of exposure regarding frequency in operational situations.

E1	**E2**	**E3**	**E4**
Very low probability	Low probability	Medium probability	High probability
Occurs less often than once a year for the great majority of drivers	Occurs a few times a year for the great majority of drivers	Occurs once a month or more often for an average driver	Occurs during almost every drive on average

safety goals. The standard also introduces the ASIL decomposition to help to achieve the safety goals. This method allows the engineers to split a high ASIL level safety goal into two lower ASIL goals. For example, an $ASIL-D$ hazard can be split into an $ASIL-A(D)$ and $ASIL-C(D)$ hazards. Be noticed that each decomposed ASIL shall be marked by giving the ASIL of the parent safety goal in parenthesis. In this context, the E-Gas three layers concept [13] represents a well-established methodology for achieving successful an ASIL decomposition.

Table 2.4: Controllability classification.

C0	C1	C2	C3
Controllable in general	Simply controllable	Normally controllable	Difficult to control or uncontrollable
Controllable in general	More than 99 % of the average drivers or other traffic participants are able to avoid harm	Between 90 % and 99 % of the average drivers or other traffic participants are able to avoid harm	Less than 90 % of the average drivers or other traffic participants are able to avoid harm

2.2 Safety Of The Intended Functionality

In the context of the new safety paradigm for highly automated driving (L3+) systems, SOTIF has been developed to address gaps that neither cybersecurity nor functional safety fully cover. While functional safety focuses on failures and malfunctions, and cybersecurity addresses external attacks, SOTIF concentrates on technical shortcomings and human misuse, as well as the hazardous behaviors these may cause. The major distinction between functional safety and the SOTIF approach lies in the scope. SOTIF extends beyond the traditional considerations of electrical and electronic automotive systems to include challenges arising from the vehicle's perception of and interaction with its complex environment. This expansion aims to identify and address functional insufficiencies. For example, lidar sensors may generate ghost reflections in heavy snow, making such weather conditions a potential triggering condition. The main goal of SOTIF is to improve the identification of hazardous scenarios for validation and to minimize the likelihood of unknown hazardous scenarios occurring. Following the ISO 21448 standard, it is essential to clarify certain terms and definitions to ensure a comprehensive understanding of this thesis's content:

- A *hazardous behaviour* is defined as the system behaviour of a system outside a specified acceptance criteria. Definition of the acceptance criteria could be established on the basis of a Key Performance Indicator (KPI) or, in the case of safety, a Safety Perfomance Indicator (SPI).
- A *triggering condition* is defined as a specific condition of a scenario that starts a reaction in the system contributing to hazardous behaviour. Potential could be included as a prefix when it is not yet validated, but experts see evidence that it could turn out to be a triggering condition in the end. Another definition is given in [14], where a triggering condition is defined as: *"an external condition (relative to ego-vehicle) in a scenario that triggers one or multiple functional insufficiencies and further results in hazardous behaviour. They are system-dependent as well"*.

- A *performance insufficiency* is defined as *"a limitation of technical capability contributing to hazardous behaviour when activated by one or more triggering conditions"*.
- An *output insufficiency* is an insufficiency on a functional level and, like the other insufficiency, can be activated by one or more functional insufficiency or triggering conditions.
- A *functional insufficiency* is defined as an insufficiency of specification or performance insufficiency. The scope of this dissertation is only focused on performance insufficiencies, therefore the insufficiencies of specification are not considered in this document.
- A *hazard* is defined as a potential source of harm caused by hazardous behaviour at the vehicle level.

The relationship between these concepts is illustrated in Figure 2.4. This figure shows how a scenario condition, such as heavy fog, can start as a triggering condition, spreading across the system until a hazardous behaviour occurs in the form of unintended braking. A vehicle exiting a tunnel is another example. If an Automated Driving System (ADS) relies solely on a camera as its perception sensor, the sudden exposure to extremely high-contrast images can disrupt the system's performance and potentially impact the ADS's behaviour.

From a scenario perspective, one of the primary objectives of SOTIF is to minimize potentially hazardous scenarios. Unlike ASIL, SOTIF validation does not classify scenarios based on a specific metric. Instead, the validation process focuses on iterative improvement by reducing known hazardous scenarios and identifying new ones, which may emerge with each iteration. Figure 2.5 illustrates how scenarios are categorized into four main areas: safe and known scenarios, and the worst-case category, unsafe and unknown scenarios. As shown, each validation cycle aims to shrink the number of unsafe unknown scenarios. This involves first validating known safe and unsafe scenarios, and then identifying and assessing new scenarios that may fall into the unsafe and unknown category. However, it is impossible to entirely eliminate hazardous unknown scenarios, as such situations can always arise unexpectedly. These remaining scenarios must be addressed as residual risks of the ADS.

An automated driving system is typically composed of three primary components: sense, plan, and act. Figure 2.6 illustrates these blocks and their interconnections. The sense block is responsible for perceiving the environment using sensors such as cameras, lidar, and radar. In our approach, this block is further divided into two sub-blocks: the perception block and the algorithm block. The sense-perception block creates a model of the observed environment based on sensor inputs. For example, a lidar sensor generates a point cloud that reflects the surrounding area. The sense-algo block processes the perception block's output to extract actionable information, such as generating an object list from the point cloud. The plan block handles decision-making by interpreting the information provided by the sense block and determining the appropriate actions based on the system's functionality and the current situation. Finally, the act block translates the plan block's control commands into tangible actions, such as applying the brakes or steering the vehicle. This model is also referred to by different names such as Sense–Decide–Act [15], but the meaning of each block remains the same.

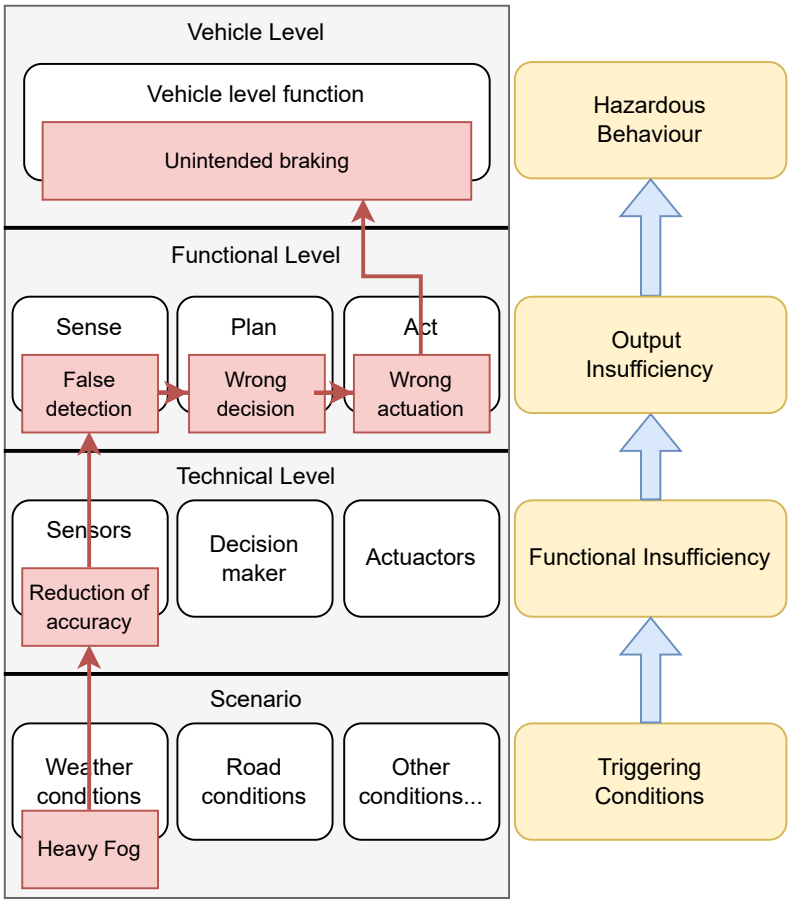


Figure 2.4: Cause and effect model between potential functional insufficiencies and triggering conditions.

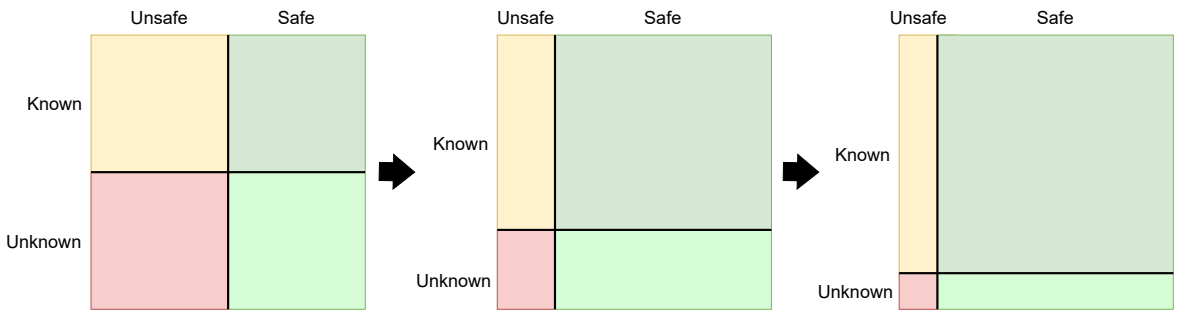


Figure 2.5: Evolution of the scenario categories through the SOTIF validation iterations.

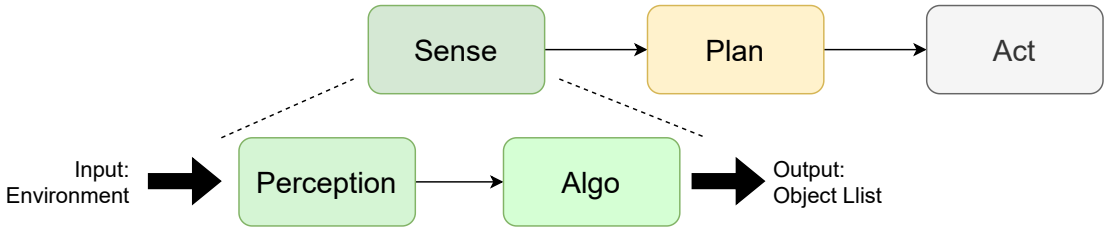


Figure 2.6: Sense - Plan - Act Model.

Lastly, Figure 2.7 shows the complete workflow of the standard showing the main activities and the relationship between them. The SOTIF standard is structured into numbered clauses, which are represented in the workflow. For instance, point 6 of the workflow pertains to clause 6, and so forth. The objectives of the main relevant clauses are:

- Clause 5: Design and specification that shall contain the information to conduct the SOTIF-related activities.
- Clause 6: The identification of the hazards arising from the intended functionality defined at the vehicle level, including the parameters that defined the situations in which the behaviour is considered hazardous. The acceptance criteria for the residual risk is also specified in this clause.
- Clause 7: Identification and evaluation of potential functional insufficiencies and potential triggering conditions, including reasonably foreseeable. Direct misuse shall be identified and those leading to hazardous behaviour shall be determined including the associated system responses.
- Clause 8: Specification and application of functional modifications addressing the SOTIF-related risks and updated from the specification and design.
- Clause 9: Definition of the verification and validation strategy, including the necessary evaluation of potentially hazardous scenarios, the relevant scenario coverage, evidences and procedures.
- Clause 10: Evaluation of known hazardous scenarios, the potentially hazardous behaviour due to the specified behaviour at the vehicle level. Known scenarios shall be sufficiently covered according to the verification and validation, providing evidences that demonstrate the validation targets are met.
- Clause 11: Evaluation of unknown scenarios by using the validation results provided with sufficient confidence that the residual risk from unknown hazardous scenarios meets the acceptance criteria.
- Clause 12: Evaluation of the achievement of the SOTIF, reviewing the work products and providing the fulfilment of the objectives of the clauses of this document and the corresponding work products.

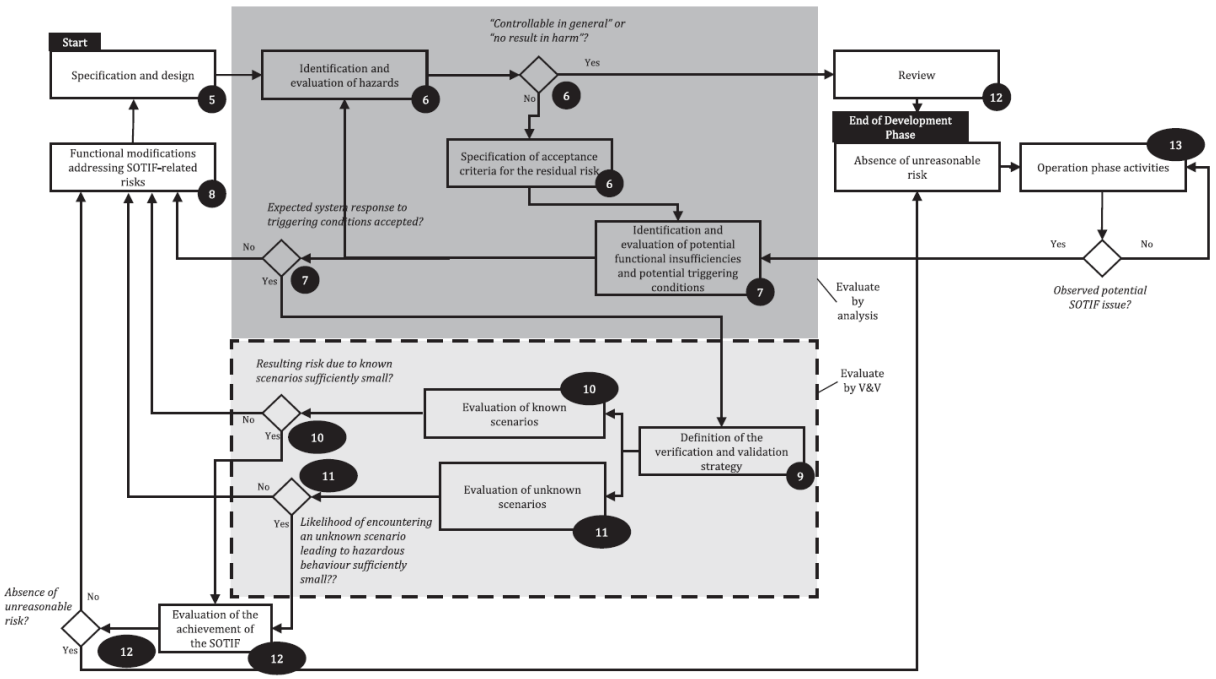


Figure 2.7: Dependencies between the ISO 21448 activities [1].

2.3 Scenario-Based Validation

As mentioned previously, scenarios have become a key element in automotive safety. Scenario classification extends beyond the definitions provided by the SOTIF standard. An additional classification [16] categorizes scenarios based on their frequency of occurrence or level of hazard. Corner scenarios occur under rare conditions while maintaining normal operational parameters, such as a low sun angle or an ice-covered road. Edge cases, in contrast, also occur rarely but involve extreme values or conditions. While some scenarios may be both corner and edge cases, not all corner cases qualify as edge cases, and vice versa. Nominal scenarios represent typical traffic situations involving regular, non-critical driving manoeuvrers, as defined by [17]. It also defines the critical scenarios are those requiring emergency manoeuvrers to prevent harm or respond to a system failure. These classifications provide a nuanced framework for understanding and analysing driving scenarios. A well-known methodology for the description of a scenario is given in the PEGASUS project [18, 19] in which the definition of the scenario is split into six layers, each concentrating on the context of the scenario. The layers are depicted in Figure 2.8 and described in the following list:

Layer 1 *Road network and traffic guidance objects*: e.g. road markings, and traffic signs and traffic lights.

Layer 2 *Roadside structures*: e.g. buildings, vegetation, street lamps, and advertising boards.

Layer 3 *Temporary modifications of L1 and L2*: e.g. roadwork signs, temporary markings, and covered markings.

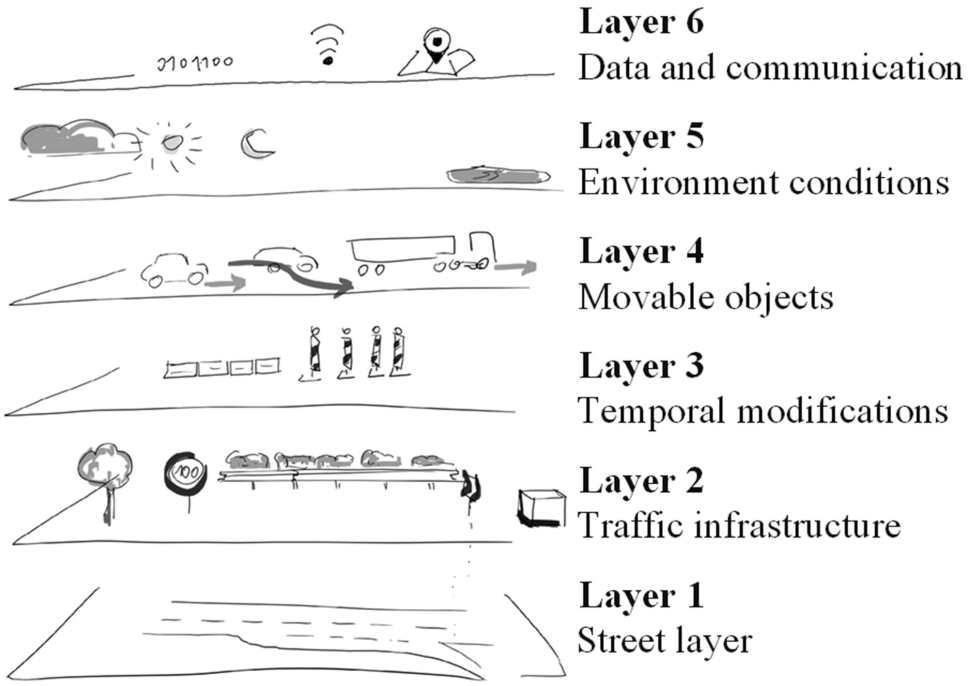


Figure 2.8: PEGASUS 6-layer scenario model [2].

Layer 4 *Dynamic Objects*: e.g. vehicles (moving and non-moving), pedestrians (moving and non-moving), trailers, and animals.

Layer 5 *Environmental Conditions*: e.g. illumination, precipitation, and road weather.

Layer 6 *Digital information*: e.g. state of traffic lights, switchable traffic signs, and V2X messages.

The Operational Design Domain (ODD) is a fundamental concept in the scenario validation process. Although it is defined in the SOTIF standard, the UL4600 standard [20] provides a more precise perspective on what an ODD encompasses. According to UL4600, an ODD is defined as: *"The set of environments and situations the item is intended to operate within. This includes not only direct environmental conditions and geographic restrictions, but also a characterization of the set of objects, events, and other conditions that will occur within that environment"*. In this context, the ODD specifies all the scenarios in which an ADS is designed to function safely. Numerous definitions of ODD exist, as highlighted in [21], ranging from the SAKURA project's interpretation [22] to the framework provided by Association for Standardization of Automation and Measuring Systems (ASAM) in the ASAM OpenODD standard [23]. ASAM OpenODD offers a standardized syntax for integrating a defined ODD into software simulations, enabling effective validation. This syntax also supports the reuse, sharing, and combination of different ODD definitions, fostering greater flexibility and collaboration among partners in ADS development. ISO34503 [24] outlines the principles for defining an ODD, emphasizing the importance of a detailed scenario description in this process. Numerous

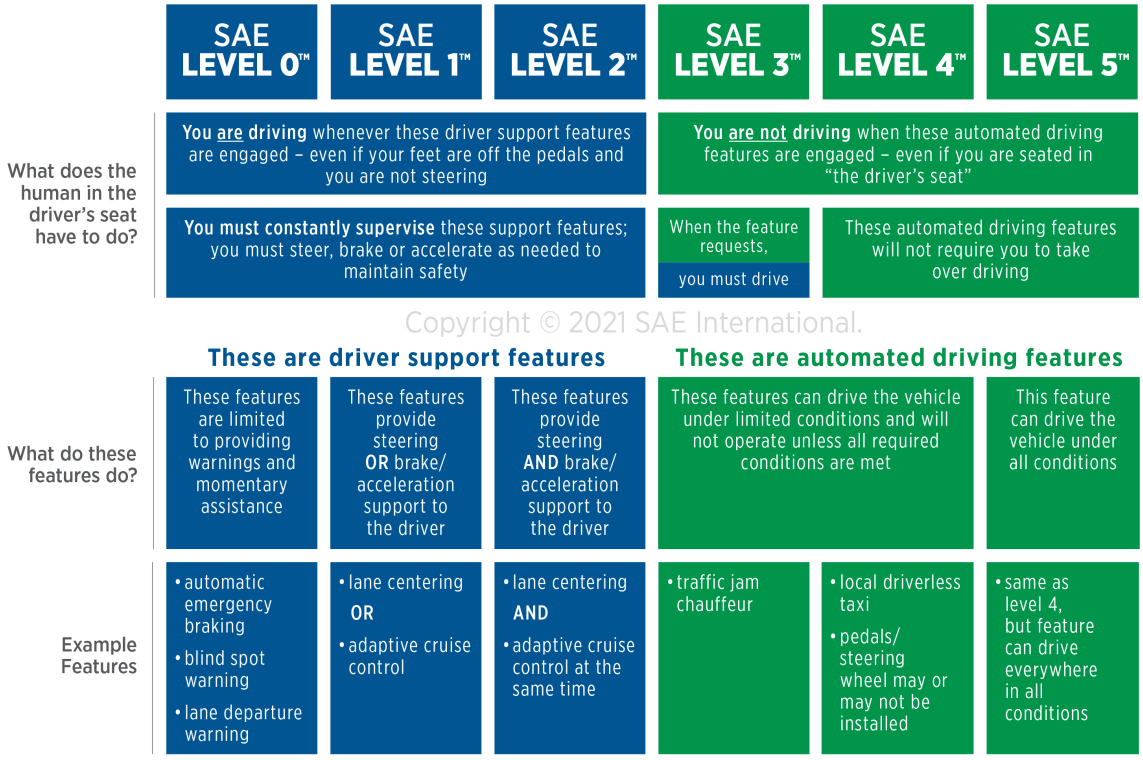


Figure 2.9: Levels of driving automation [3].

taxonomies [25, 26], including those previously mentioned, have attempted to address this challenge by categorizing various factors, ranging from different weather conditions (e.g., wind, snow) to road topologies.

Scenarios become more relevant in the validation process as the driving automation level increases. Driving automation is categorized based on the division of responsibilities between the driver and the system, specifically regarding control of the Dynamic Driving Task (DDT). The DDT encompasses the real-time tasks necessary for operating a vehicle in traffic, such as lateral and longitudinal motion control. The SAE classification defines six primary levels of automation, as illustrated in Figure 2.9.

CHAPTER 3

Related Work

Extensive related work has been carried out during this thesis. Based on this literature review, this chapter presents the most relevant related work according to the main topics of the thesis: safety validation, sensor models and risk evaluation in the field of autonomous driving.

3.1 Safety Validation

New regulations, such as the European Regulation 2022/1426 [27, 28] and UN Regulation No.157 [29], required the provision of validation evidence to gain authorization for operation on public roads and, more importantly, to prevent accidents like those involving Tesla, Uber, and other autonomous vehicles in the past [30, 31, 32, 33]. It is noteworthy that one of the publications [5] included in the thesis shows an extensive research work of the state-of-the-art on the automotive safety validation domain. This publication goes through many aspects of safety including standardisation, existing regulations, and current and past projects covering this topic. In a similar way, researchers in [34, 35] also provide an overview of this domain, highlighting key requirements and concepts. Meanwhile, new standards such as ISO34502 [36] are being developed to support the shifting safety paradigm, emphasizing the growing importance of scenario-based validation.

Many research works [37, 38] have investigated the ways to reuse and integrate the processes outlined in the ISO26262 standard into the SOTIF, including methodologies like System-Theoretic Process Analysis (STPA). In this context, the first publication [4] included in this thesis presents a co-engineering methodology for safety and security to reduce efforts by applying integrated approaches, where also SOTIF started to be relevant in this kind of analysis. SOTIF is a topic that has been covered in many research work. For instance, the publication [39] gives an overview of triggering conditions in automated driving systems. Another critical aspect of SOTIF is identifying the limitations of each technology utilized in the automated driving domain [40], which provides a methodology to recognise and understand the causes of unknown scenarios. Additionally, the research work [41] provides a comprehensive survey of current safety standards in the automated driving domain, highlighting common perception failures and key metrics for evaluating perception systems. An additional overview of safety standards, emphasizing object-based environment perception, is presented in [42].

Similarly, other researchers [43] explore the relationship of SOTIF with other standards and its application in verification and validation processes.

There are many KPIs or SPIs defined that are used in the context of the safety validation. An extensive overview of available metrics is given in [44]. One of the most known metrics is the Time To Collision (TTC) [45], which returns the minimal time until two actors collide. There are also multiple variables of the TTC, for example, the Inverse TTC [46] or the called Modified TTC [47] that includes the velocity of the collision in the metric estimation. Another time-based metric is the Time To Maneuver (TTM) [48]. This metric provides the latest possible time in which the vehicle is able to perform a manoeuvrer such as braking or steering to avoid the collision. Metrics can also be defined based on other aspects beyond time. For instance, the Deceleration to Safety Time (DST) [49] is an acceleration-based metric or the Crash Potential Index (CPI) [50] that calculates the probability an actor can not avoid a collision by deceleration.

As previously described, one of the main concepts the SOTIF standard introduces is triggering conditions. There are already research studies focused on this topic. For instance, the authors in [14] describe a methodology to provide a systematization and categorization of triggering conditions, which offers a structured manner to manage them. Researchers explore machine learning techniques in this context in [51]. They finally conclude the absence of sufficiently meaningful training data prevents this approach from fully supporting the complete SOTIF argumentation at this stage. In the context of ODD, the approaches detailed in [52, 53] focus on strategies to maximize ODD coverage during the scenario validation process. Validating ADAS/AD functions is a challenging task due to the complexity of covering all possible variations and scenarios. This issue has been faced by using different approaches. Relying solely on real-world driven miles for validation is impractical due to the exorbitant cost and time requirements as [54] describes. An alternative way to handle it is by using a simulation-based validation. This kind of validation has gained traction in recent years with the development of specialized environment simulators like CARLA [55] and SVL Simulator [56]. Research studies [57, 58] provide an overview of the most relevant simulators in this field and their key features. The most significant challenge with simulation-based validation lies in creating realistic, high-fidelity sensor models. These sensor models are not only complex to develop but also demand substantial computational resources and high-performance hardware [59]. Moreover, the generation of scenarios that cover all possible scenario conditions and variations remains a critical hurdle as publication shows [60].

3.2 Sensor Models

Although there are numerous Hardware-in-the-Loop (HIL) platforms [61, 62] available for collecting data to validate automated driving systems, two significant challenges arise with this approach. The first is determining how much real-world data is enough to validate an ADS [63]. Secondly, it is the inherent difficulty in collecting data for all possible scenarios, including different types and intensities of weather conditions [64, 65]. Therefore, virtual validation emerges as a more practical solution for ADS validation. A hybrid approach combining virtual validation and HIL platforms offers an intermediate step, as demonstrated

by the methodologies in [66] and [67]. Extensive research works have been focused on understanding the behaviour of different sensor technologies in harsh environments. For example, [59] reviews the state-of-the-art in sensor models for virtual validation, highlighting different fidelity levels. In this context, numerous studies [68, 69, 70, 71, 72, 73, 74] explore sensor performance in harsh conditions to model the sensor error behaviour in such harsh conditions. More specific research works such as [75, 76, 77] analyze the performance of camera and lidar systems in adverse weather like fog. A detailed fog error model for generated point cloud messages from lidar sensors was introduced in [78]. The same authors have expanded their work to model lidar performance insufficiencies in snowfall conditions [79]. Additionally, [80] presents a library for editing lidar point clouds, incorporating effects like cropping and added reflections. This research also extends into security, as these modifications can serve as spoofing attacks by external actors.

Machine learning has also been employed to develop sensor error models, as demonstrated in [81]. In this research work, the authors detail a process for simulating rain effects on image frames generated by a camera. Such research and tools may be integrated into our approach to incorporate them as perception performance insufficiencies in the injection process. Unfortunately, high-fidelity models often demand significant computational resources for simulations. The authors in [82] describe an approach that may solve this issue. They propose low-fidelity models that are able to replicate similar behaviours than high-fidelity models with less computational costs. Like this thesis, Andrea Piazzoni et al. [83, 84, 85] remark on the importance of considering perception errors in the virtual validation process and its lack in current approaches. Their work extensively explores perception error models and integration into simulation pipelines. Moreover, the approach is implemented across various sensor configurations to illustrate the relationship between system safety and sensor setups. Perception errors have also been utilised for virtual validation through adversarial attacks [86] as proposed in [87]. The authors conclude that, although these attacks may be seemingly harmless, can influence the system's final behaviour. Another approach is presented in [88], where authors develop RGB-camera perception error models to estimate rare failure probabilities and learn high-likelihood failure trajectory distributions.

3.3 Risk Evaluation for Autonomous Driving

Recent scientific works have introduced various approaches in the domain of quantitative risk assessment. The publication [89] describes a comprehensive list of deficiencies in the existing SOTIF standard and suggests potential corrections. It also outlines the possibility of employing a statistical approach for SOTIF validation. The studies in [90, 91] propose a method for assigning quantitative values to the key categories of HARA analysis (i.e. exposure, controllability, and severity) outlined in ISO26262. This enables the calculation of risk for an ADS based on these values and statistical approaches. Using these values, the probability of risk for specific scenarios, both with and without triggering conditions, can be estimated. Unlike our approach, which is centred on fully virtual validation, their methodology emphasizes real-world data for validation purposes. Another risk quantification method is introduced in [92], where the authors apply fault tree analysis to derive a quantitative metric

for assessing ADS safety.

A purely statistical approach to provide quantitative metrics validation on ADAS is given in [93]. The authors employ one-sided binomial and Poisson distributions; however, the authors acknowledge that their analysis is heavily simplified and tailored to a specific ADS model. The approach demonstrates that safety conditions are met at the component level by using reduced-size datasets compared to those required for vehicle-level validation. As a main disadvantage, its application is constrained by its specificity to the described scenario. Another statistical validation methodology is proposed in [94]. This approach uses reinforcement learning to identify scenarios that result in outcomes outside the acceptance criteria. This method reduces the number of scenario simulations needed to validate a collision avoidance system. The authors in [95] describe a perception validation methodology based on failure rate probabilities. Similar to our approach, the publication focuses on the Responsibility-Sensitive Safety (RSS) [96] as the safety-relevant area during the validation. However, this approach only addresses perception components, omitting the broader impact on the entire system. Additional research is done in [97], which introduces a system for monitoring, quantifying, and mitigating SOTIF risks. The methodology is validated by using a HIL platform, which also includes Artificial Intelligence (AI) algorithms in the perception system.

CHAPTER 4

Proposed SOTIF Argumentation

This is the main section of this dissertation in which the steps of the proposed approach of the SOTIF validation are described. Figure 4.1 depicts the workflow of the proposed argumentation, splitting it into blocks, and showing the relationship between them. Each numbered square represents the associated clause from the standard. The flowchart was first introduced in the publication [6] and has been refined during this thesis. The main goal of the displayed workflow is to adapt the guidelines provided by the standard and to put into practice the methodology described in this thesis. For the sake of clarity, the number shown in each box referees the corresponding clause from the standard. As shown in this figure, the argumentation is divided into four main blocks that build the whole approach. Each section of this chapter details one of the main parts of the validation process. Section 4.1 explains the qualitative analysis performed in the argumentation. The next section, Section 4.2, details how triggering conditions are parametrised to be able to implement them into an existing validation tool. The injection of performance insufficiencies is described in Section 4.3, where their categorisation and formal definition to be integrated into the validation process of an ADS are given. Finally, Section 4.4 explains the calculation of a quantitative metric of the risk based on the results of the previous performance insufficiency injections. The obtained risk results provide the stakeholders with comparative results that can be used to improve the system safety in further validation iterations.

4.1 Qualitative Analysis

The first step of our approach is to select the relevant scenarios for the system based on the defined ODD. The main point of this step is to filter out the scenarios of a database that are outside of the parameters of the ODD and do not have any impact on the validation process for the ADS validation process. For example, in the case of a defined ODD that only includes urban and local roads. Legislators define the maximum allowed speed on these roads as 50km/hour [98], therefore, scenarios where the vehicle speed is higher than this should be dismissed and not included in the validation process. Industry initiatives, such as ASAM OpenLabel [99], have been used to standardise this process, creating interoperable scenario labels that could be used between scenario databases from different providers.

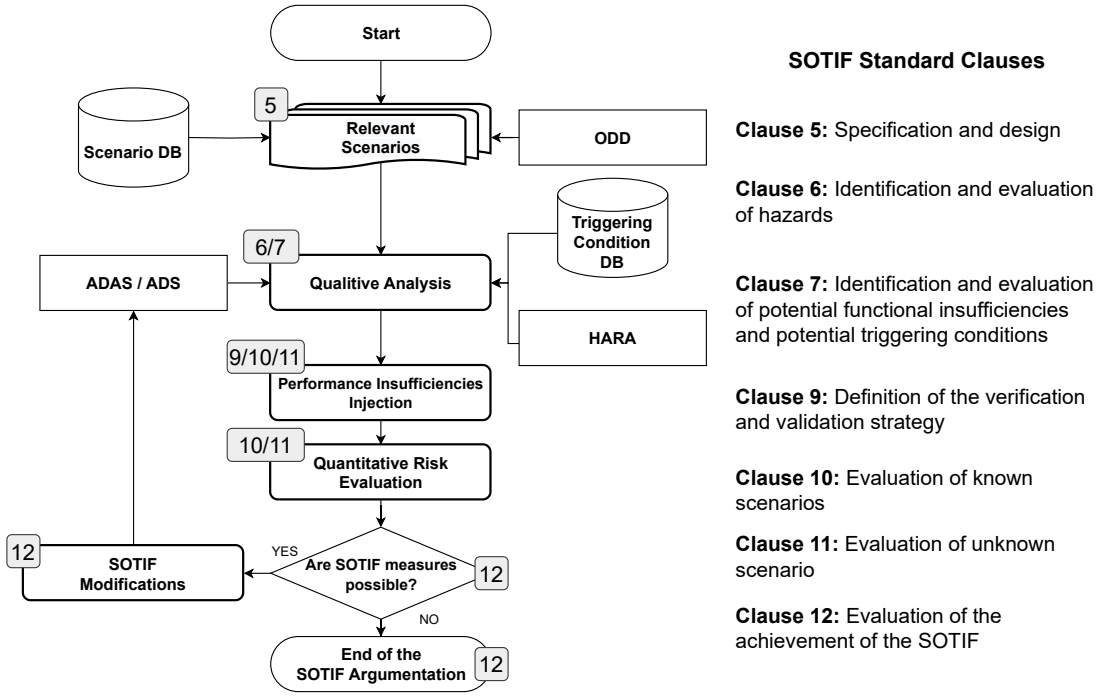


Figure 4.1: SOTIF validation process flowchart.

As described in Section 2.1, the HARA method is mandatory in the ISO26262 to identify and categorise hazards, classify them with an ASIL, and specify a safety goal for each one. Unlike the process implemented in functional safety standard, the ASIL level is not used in the SOTIF analysis process and only the identification of the hazards is used. The identification process of relevant hazardous scenarios and parameters will be conducted, in the first instance, by a qualitative analysis to determine the potential triggering conditions and their impact on the scenario category.

In our approach, the filled HARA is reused by including some columns that make it suitable for a SOTIF analysis. The addition of the columns "Potential Triggering Conditions", "Description of Possible Functional Insufficiency", and "Hazardous Behaviour" were added to follow the "Cause-Effect-Model" described in the SOTIF standard (Figure 4.2). Moreover, the probability of occurrence of each hazardous situation is also added to evaluate the risk based on the severity and its probability of occurrence. Columns related to exposure and controllability are omitted since ASILs are not relevant in the SOTIF scope. The occurrence frequency classification follows the approach given in [14], where the authors define three qualitative levels:

- Common Cases: denote the occurrence in a daily basis (e.g., a crowd of pedestrians).
- Reported Cases: are validated by media (e.g., news) or mentioned in safety reports by automotive manufacturers, which exhibit lower exposure by normal drives.
- Hypothesized Cases: are theoretically possible but are rarely encountered in reality.

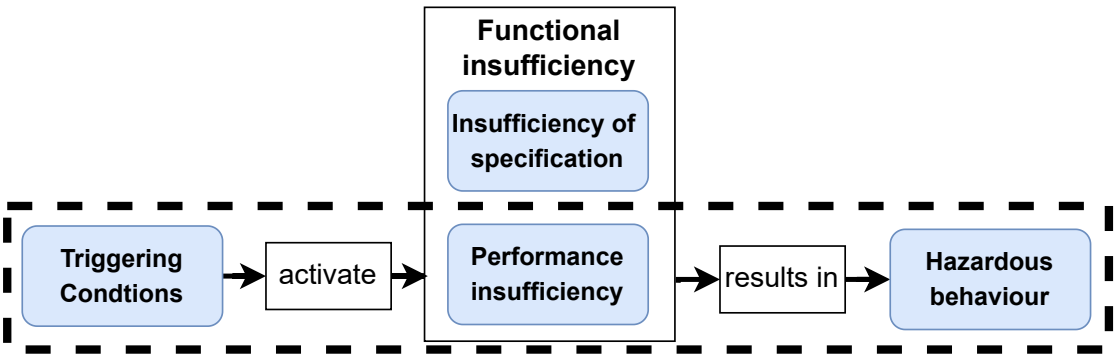


Figure 4.2: Cause-and-effect model from ISO21448.

A triggering conditions' database (Figure 4.3) has been developed based on definitions of different standards (Annex C of the SOTIF Standard [1], the British Standards Institution (BSI) Publicly Available Specification (PAS) 1883:2020 [26], SAE [25]). This database is used to identify the potential triggering conditions associated with each hazardous situation. The database classifies triggering conditions into different categories and subcategories, creating a hierarchy, and including the parameters that define each triggering condition, from the unit used to different levels of the triggering conditions, with examples of the parametrisation.

This database is organized into categories and progressively detailed subcategories, providing a high level of granularity. For example, the main categories are broad topics such as environmental conditions or road conditions. These main categories branch into subcategories like weather conditions or illumination, which are further divided into more specific conditions. In the case of weather conditions, it is broken down into types like rain, snow, fog, and other climate phenomena, each of which is further classified by intensity levels (e.g., light, medium, heavy). The primary goal of this structured approach is to achieve fine granularity, enabling the representation of a wide variety of scenarios. This, in turn, supports the identification of additional functional shortcomings during the analysis process.

4.2 Triggering Conditions Parametrisation

Once the qualitative risk analysis of the ADS is carried out and the potential triggering conditions have been identified, these conditions are tested in the validation toolchain to obtain a risk evaluation of the system. Therefore, it is necessary to have a way to integrate triggering conditions to achieve this evaluation, where they have to be parametrised and included in the validation tools. The parametrisation is done based on scenario constraints that each triggering condition could include, for example, perception distance limitation or friction limitations. The next integration step is to link the triggering conditions with the

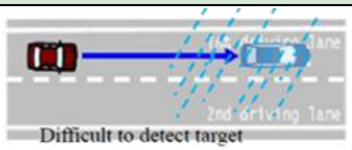
Subcategory of Triggering Condition	**Variation acc. to taxonomy**	**Scale**	**Example of Parameters / Boundry values from existing standards or scale**
Rain (Precipitation Types)	Scale for rain: 1. Light rain (low visibility) 2. moderate rain 3. Heavy rain (poor visibility) Other ways to describe the rain: size of rain droplets, pollution level (transparent, dark colour..), viscosity (rain with mud)	BSI PAS 1883 SAE AVSC00002202004	NOTE: BSI Stakeholders may classify rainfall intensity as: i) light rain: when the precipitation rate is < 2.5 mm/h; ii) moderate rain: when the precipitation rate is between 2.5 mm/h and 7.6 mm/h; iii) heavy rain: when the precipitation rate is between 7.6 mm/h and 50 mm/h; iv) violent rain: when the precipitation rate is between 50 mm/h and 100 mm/h; v) cloudburst: when the precipitation rate is > 100 mm/h. From American Meteorological Society: Precipitation in the form of liquid water drops that have diameters greater than 0.5 mm, or, if widely scattered, the drops may be smaller. Rainfall Rates acc. to the SAE Lexicon: Description - Inches per Hour - Centimeters per Hour a) Light rain 0.01 (trace)-0.10 < 0.25 b) Moderate rain 0.11 - 0.30 0.26 - 0.76 c) Heavy rain >0.3 > 0.76
Category / Factor	**Example / Description**	**Unit**	
Environment (Weather state / natural phenomena)	Difficult to detect a target, lane, other actors In air below 0° F, raindrops may start as snow or ice crystals but melt when they fall into warmer air EXAMPLE 1 Rain and snow can affect radar performance.	mm/hr, ø- size of rain droplets, ...	
Additional reference		**Source of potential statistics**	**Image**
American Meteorological Society, "Rain," American Meteorological Society, 25 April 2012. [Online]. Available: http://glossary.ametsoc.org/wiki/Rain		Historical weather forecast Probability Forecast	

Figure 4.3: Excerpt from the triggering conditions database for the rain scenario condition.

existing defined taxonomy to be able to parametrise as many as possible triggering conditions and include them in the test scenario. Therefore, an intermediate block is added to the cause-and-effect model defined in the SOTIF standard. Figure 4.4 shows this new approach to the model. In this picture, a scenario constraint block has been added to connect the triggering conditions with the functional insufficiencies.

This approach allows the definition of each triggering condition as a combination of one or many scenario constraints. A scenario constraint is a new entity, which is defined by a type of constraint and an ontology value linked to the ontology from a validation tool. The type of constraint defines its impact in the scenario, like a maximum or minimum value. A scenario constraint could be related to one or many triggering conditions and vice versa. Therefore, scenario constraints and triggering conditions have a many-to-many relationship. A triggering condition can be defined not only by scenario constraints but also for *Values*. A value is a phenomenon that is not yet modelled, for example, a snowfall or rainy condition. Values are defined by one or more scenario constraints as shown in Figure 4.5. In the same way, each scenario constraint can also be related to one or many value entities. Consequently, triggering conditions are defined as one or more values and scenario constraints as Figure 4.6

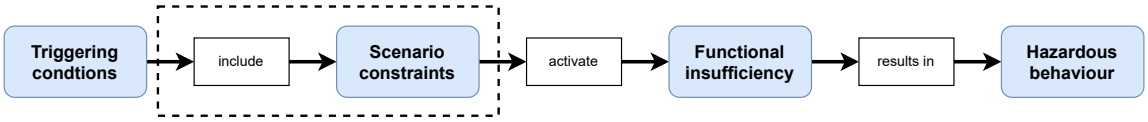


Figure 4.4: Updated cause-and-effect model including the scenarios constraints.

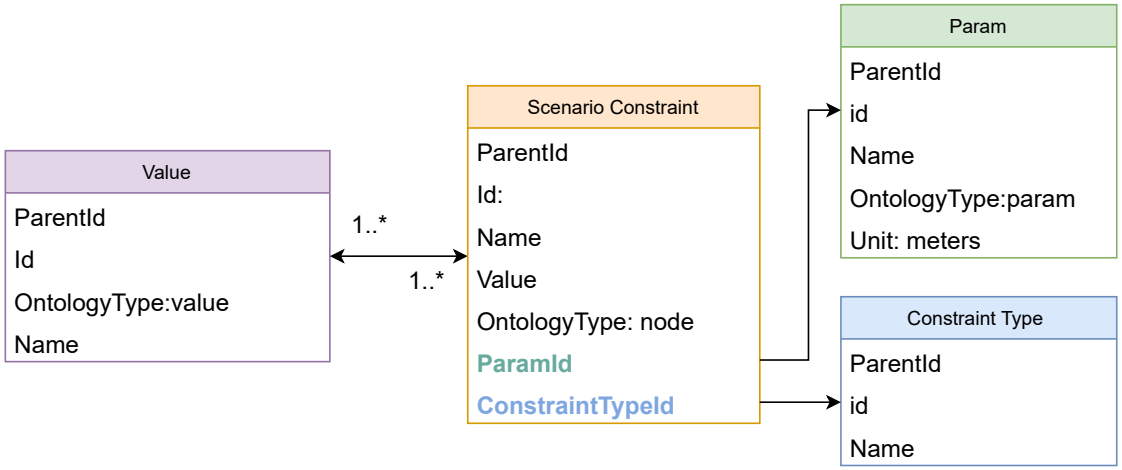


Figure 4.5: Value definition.

depicts.

The SOTIF concept is designed with scalability in mind. The complexity of the triggering conditions and their parametrization shall increase due to the test scenario. In case complex triggering conditions are included in the system, a decomposition is applied to define the triggering condition as a combination of triggering conditions. For example, for the triggering condition *Heavy Snow during Night-time*, a decomposition of two triggering conditions can be applied: *Heavy Snow + Night-time*, which is shown in Figure 4.7. A thing to remark in this case is that two reduced illuminances are defined. In this particular triggering condition, the illuminance is restricted due to night-time, therefore, the illuminance is limited to 1 lux according to the standards. In this context, the most restrictive constraint is applied. It means the illuminance constraint defined by the heavy snow condition is overridden due to the most restrictive defined in the night-time condition. The restriction is applied based on the constraint type and parameter associated with the scenario constraint. For example, the restriction applied in the illuminance constraint is because the *ParamId*(illumination_id) and *ConstrainTypeId*(ct_max_id) have the same values. The behaviour to determine the most restrictive values is given by each *ParamId* because each parameter value can have a different effect on the system. In this case, the lowest illuminance value is more restrictive than a higher one. On the other hand, for other parameters such as temperature or speed, the highest values are detrimental to the system. When one or more potential triggering conditions are selected for testing, all associated scenario constraints (e.g., limited visibility and reduced friction) are incorporated into the generated test cases. The metrics derived from these test cases illustrate

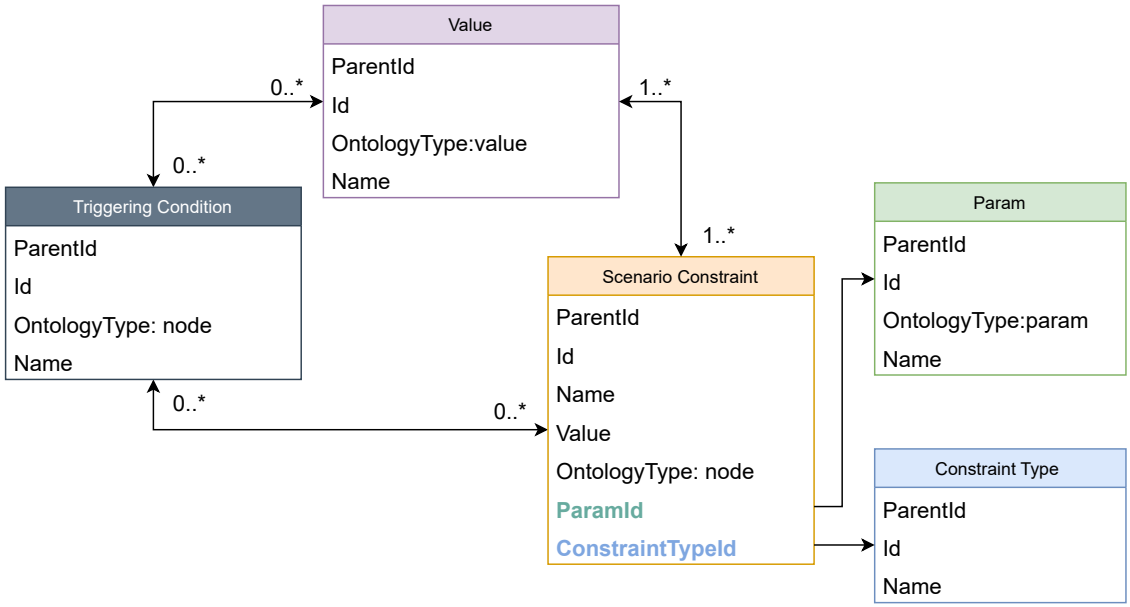


Figure 4.6: Triggering condition definition.

the impact of the selected potential triggering conditions on the function. These results are then compared to the nominal performance of the function (i.e. without potential triggering conditions) to assess whether these scenario constraints affect the system. As a final remark of this section, this integration process was presented in [100] and published in [7, 9].

4.3 Perception Performance Insufficiencies Injection

The validation of all triggering conditions that could occur in a scenario is an unmanageable task. Unlike the most common approaches validating all possible potential triggering conditions, our approach is based on validating the impact that the scenario conditions or triggering conditions may have in the system. An example of entering or exiting a tunnel can be used to illustrate this approach. Validating all possible tunnels should include all definitions of all parameters that define the tunnel, such as the number of lanes or the sizes. This validation approach should lead to an unwieldy number of test cases, making the validation unfeasible. On the other hand, the impact may have in the ADS is validated in our approach. In the case the ADS relies on a camera as a perception sensor, the validation of extremely high contrast images for a short period of time is carried out. By using this strategy, not only the entry or the exit of a tunnel is validated but also all other scenario conditions that may have the same impact on the system such as dazzle during sunrise or sunshine. This validation approach also contributes to identifying unknown scenarios which is one of the main goals of SOTIF. The description of this approach follows the publication [8] included in the thesis.

Perception performance insufficiencies injection are addressed at a high level of abstraction in our approach. Thus, the injection is applied to the raw data provided by the perception

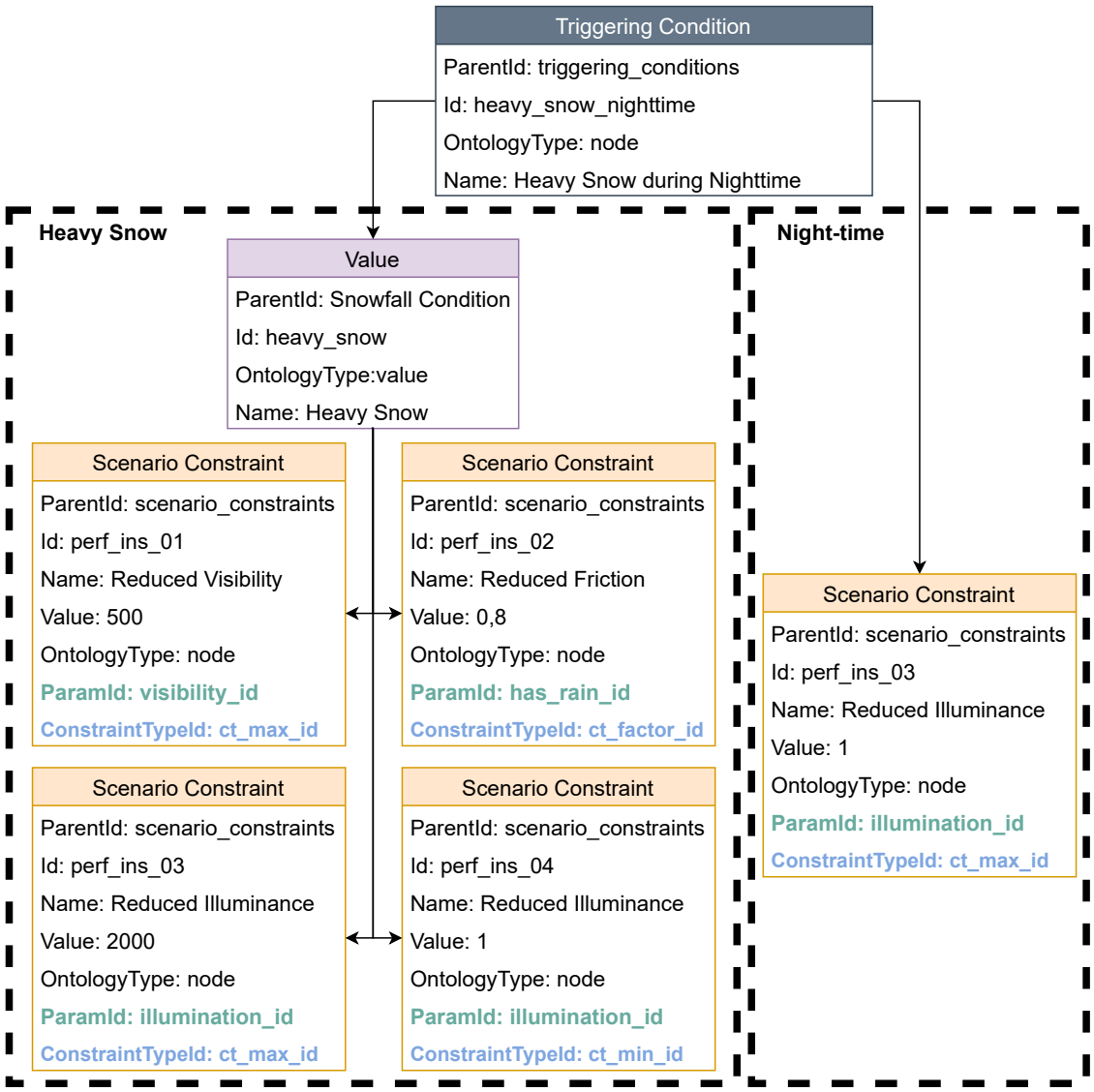


Figure 4.7: Definition of a triggering condition by the combination of scenario constraints.

sensor. For example, the performance insufficiency injection in a lidar sensor is carried out by modifying the point cloud that the sensor delivers. Similarly, each image frame of a camera sensor is altered to include the injection. The architecture of our approach is depicted in Figure 4.8 where the main blocks of the injection are depicted. While our focus is on perception-related performance insufficiencies, the impact of these injections can be manifested in any part (i.e. sense, plan, and act blocks) of the ADS.

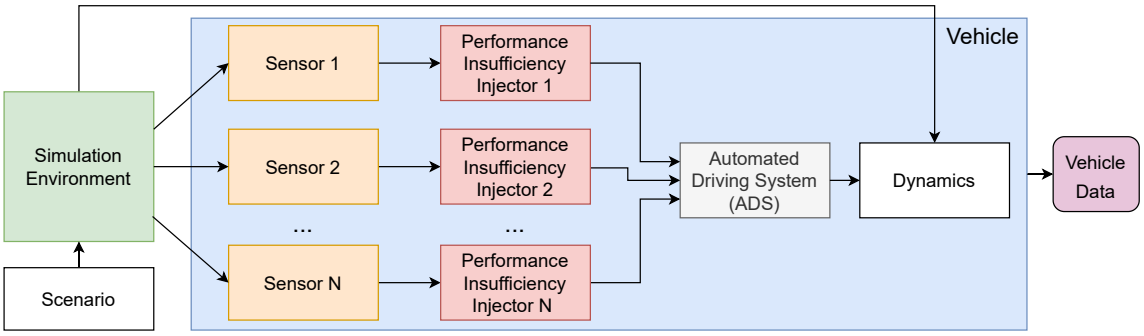


Figure 4.8: Architecture of the injection approach.

A classification of performance insufficiencies is defined for an easy integration of the injection. The main category defines the general impact of the insufficiency while the other categories define the performance insufficiencies for specific technologies and specific triggering conditions. The classification of the performance insufficiencies is as follows:

Generic Performance Insufficiency (GPI) This classification defines the list of performance insufficiencies based on the behaviour that can occur on the sensor but is not associated with any specific sensor technology. Table 4.1 shows an excerpt of some identified performance insufficiencies, impact their impact on the sensor.

Technology Performance Insufficiency (TPI) These insufficiencies derive from the GPI but, in this case, they are modelled for a specific technology. For example, the defined reduction field of view insufficiency is defined for the lidar technology cropping in the point cloud message provided by the lidar sensor function. The injection removes the points from the point cloud messages that are not within the defined sensor field of view. Table 4.2 shows an excerpt of the performance insufficiencies for the lidar technology and how they are modelled to be included in the injections.

Triggering Condition Performance Insufficiency (TCPI) This refers to a performance insufficiency modelled for a specific triggering condition and technology, such as the lidar snowfall model from [79] or the camera rain model from [81]. This category also encompasses the taxonomies defined in standards (e.g., SAE [3], BSI [26], SOTIF [1]) that can be applied as triggering conditions in the validation process. For instance, the SAE standard [3] limits visibility to 500 meters for heavy snow scenarios. It is important to note that these performance

insufficiencies are not system-independent and must therefore be applied simultaneously to all relevant sensors. Therefore, if a triggering condition is validated for an ADS with radar, camera, and lidar sensors, the performance insufficiency injections must be applied consistently across all sensors at the same fidelity level to ensure accurate results.

A test case function (f_{TC}) is defined in which the result is within the acceptance criteria ($f_{TC}() \in \varepsilon_{acceptance-criteria}$) and no performance insufficiency injections are included. Then, the same test case with performance insufficiency injections for all injection levels (S) is defined as follows:

$$f_{TC}(PI_i) \quad \forall \quad 1 \leq i \leq S \tag{4.1}$$

In case more than one performance insufficiencies are validated, Equation 4.2 shows how it can be defined for all different performance insufficiencies (N) and their levels of injection (S).

$$f_{TC}(PI_{ji}) \quad \forall \quad 1 \leq i \leq S \quad \text{and} \quad 1 \leq j \leq N \tag{4.2}$$

Triggering condition performance insufficiencies are defined as the combination of one or many performance insufficiencies. For instance, a heavy fog can be defined as the injection of different performance insufficiencies simultaneously such as visibility, illuminance and accuracy insufficiency. Therefore, a test case that includes a TCPI is defined as expressed in Equation 4.3.

$$f_{TC}(TCPI) = f_{TC}(PI_{1i}, PI_{2i}, ..., PI_{ji}) \quad \forall \quad 1 \leq i \leq S \quad \text{and} \quad 1 \leq j \leq N \tag{4.3}$$

Following the same argumentation, Equation 4.4 is defined in case more than one TCPIs (M) is included in the test case.

$$f_{TC}(TCPI_k) \quad \forall \quad 1 \leq k \leq M \tag{4.4}$$

4.4 Risk Quantification

Conducting a quantitative evaluation is crucial to ensure objective validation of the function. Therefore, the next step in our methodology focuses on quantifying the risk to validate the ADS, facilitating comparisons with newer iterations and enabling continuous improvements. Similar to the previous section, this section is taken from [8]. Unlike functional safety, the risk in SOTIF is defined as the product of controllability and severity. The primary objective is to increase the controllability and to reduce the severity to achieve the acceptable risk of the system. Ideally, controllability and severity should be zero ($C = 0$ and $S = 0$). The proposed approach to calculate the risk is illustrated in Figure 4.9. This figure shows the risk

Table 4.1: Generic performance insufficiencies excerpt list

GPI ID	Generic Performance Insufficiency (GPI)	Impact
PI-01	Reduction of the Field of View (FoV)	The visual range of the sensor is reduced from the nominal sensor performance.
PI-02	Light disturbance	An external light source affects the sensor perception.
PI-03	Misalignment	The position of the sensor was changed from the calibrated sensor position.
PI-04	Reduction of resolution	Sensor resolution is reduced according to the nominal performance provided by the manufacturer.
PI-05	Reduction of accuracy	Sensor accuracy decreases according to the nominal performance.
PI-06	Reduction of luminous intensity	The luminous intensity of the sensor is reduced according to the technical specifications.
PI-07	Slower processing time	Sensor processing time is slower than the maximum processing time in nominal conditions.

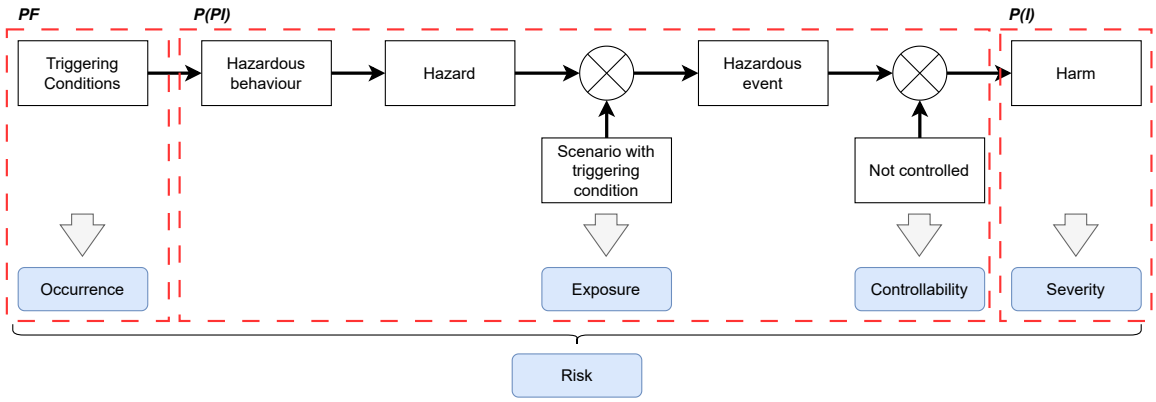


Figure 4.9: Risk quantification from our approach and compared with ISO21448.

displayed in the SOTIF standard. The red blocks have been added from our approach and symbolise the three parameters defined in our approach to calculate the risk of the system.

Table 4.2: Lidar technology performance insufficiency excerpt list

Technology Performance Insufficiency (TPI)	Parent Generic Performance Insufficiency (GPI)	Potential Triggering Conditions	Performance Insufficiency Injection
Reduction of Field of View (FoV)	PI-01	Snowfall, fog conditions,...	Crop the raw point cloud (vertical and horizontal cropping) generated by the lidar sensor.
Light Disturbance	PI-02	Mirrors, water on the street,...	Added random points into the point cloud message.
Misalignment	PI-03	Wrong calibration, earthen or gravel roads, potholes,...	Change the position of the sensor.
Reduction of accuracy	PI-05	Sensor cover, housing dirtiness, occlusion,...	Include noise into the point cloud message.
Slower Processing Time	PI-07	Driving in urban areas,...	Include random objects into the point cloud message.

$$Risk = PF * P(PI) * P(I) \tag{4.5}$$

Equation 4.5 describes how the risk is calculated in the proposed approach. This calculation is based on two different probabilities, $P(PI)$ and $P(I)$, and a factor, PF. The probability of performance insufficiency ($P(PI)$) shows the likelihood that the performance insufficiency may have an impact on the expected performance of the system, called nominal performance. The nominal performance is calculated based on the values given from the Montecarlo simulations without any performance insufficiency being injected. An acceptable window based on the nominal performance is set to determine whether the results have been affected by the injected performance insufficiency. The acceptable window is calculated based on the values of the nominal performance and their standard deviation. Following the approach depicted in Figure 4.9, if the $P(PI)$ is bigger than zero, the ADS is vulnerable to the injected insufficiency and uncontrolled by the ADS. The other probability, the probability of injury ($P(PI)$), shows if the injection could cause harm and its magnitude. For simplicity, it is assumed that the probabilities are independent; however, future research will account for dependent probabilities to achieve more precise quantification results. The calculation of the

$P(PI)$ uses the methodology used in [90, 101], which in turn used the model developed by Kusano and Gabler [102, 103] to calculate the likelihood of injury. The model calculates the probability of injury for any injury level equal to or greater than level 2 according to the Maximum Abbreviated Injury Scale (MAIS) [104]. MAIS level 2 is characterized by moderate injuries with a low probability of fatality (1–2%). In our approach, this is treated as a severity level greater than zero ($S > 0$), and thus, the associated risk is considered. Lower MAIS levels are not considered and, therefore, the risk is set to zero. In this case, while the performance insufficiency continues to impact the system, SOTIF modifications can be implemented in the ADS to improve system reliability. The model [102, 103] utilised to calculate the probability of injury is defined in the equation as follows:

$$P(I) = \begin{cases} \frac{1}{1+e^{-(\beta_0+\beta_1\Delta v+\beta_2 b)}} & \text{MAIS} \geq 2 \\ 0 & \text{MAIS} < 2 \end{cases} \tag{4.6}$$

The values β values are given as constants by the model, which are $\beta_0 = -6.068$, $\beta_1 = 0.1000 s/m$, and $\beta_2 = 0.6234$. b is related to the usage of the seat belt. $b = -1$ is used if the seat belt is not used, otherwise, $b = 1$. We have assumed that the seat belt is always used as mandatory by the law, therefore, $b = 1$ was set in all our calculations. Δv was calculated based on the maximum and minimum speed values at the time next to the collision.

The third parameter of the risk quantification is a Plausibility Factor (PF) that is added to adjust the different levels of the injected performance insufficiency. This factor compensates for the impact of the performance insufficiency as long as it becomes more severe. For example, it is less probable there is an extreme limitation of visibility that makes any object not visible a few meters away from the ego vehicle than a less severe visibility reduction in which the perception system still can perceive objects hundreds of meters away. Currently, the PF values are given by expert judgment, but using statistics to set this factor is an open question for further research. The value range of the plausibility factor is from zero to one. Finally, Equation 4.7 shows the calculation of the risk for a performance insufficiency for each injection level denoted as S.

$$Risk_{PI} = \sum_{i=0}^{S} PF_i * P(PI_i) * P(I_i) \quad \forall \quad 1 \leq i \leq S \tag{4.7}$$

Similarly, if more than one performance insufficiency is included in the risk calculation, the risk of the ADS is the sum of all the risks calculated for each performance insufficiencies (N) as Equation 4.8 shows.

$$Risk_{ADS} = Risk_{PI_1} + Risk_{PI_2} + \cdots + Risk_{PI_N} = \sum_{i=1}^{N} Risk_{PI_j} \quad \forall \quad 1 \leq j \leq N \tag{4.8}$$

This methodology provides a quantitative risk metric, but it is up to the stakeholder to assess whether the risk obtained is within the acceptance criteria of the SOTIF validation process. It

is the responsibility of the stakeholders to minimise the obtained risk to achieve the lowest risk level using principles such as As low as reasonably Practicable (ALARP) [105, 106]. This is also in line with improving the safety of the system in each iteration, which is one of the main objectives of the SOTIF standard.

CHAPTER 5

Approach Implementation and Results

This chapter presents a use case illustrating the potential implementation of the proposed approach. The use case focuses on the validation of an ADS in the event of field of view limitation. As in the preceding chapter, the section goes through each main argumentation part explaining the process of the implementation. The qualitative analysis described how it is possible to extend the carried-out risk analysis to be relevant for SOTIF. Then, the integration of triggering conditions necessary for the validation process is shown through an existing validation tool. The perception performance insufficiencies injection and the quantitative risk evaluation for this use case are in the following sections. Moreover, these final sections also show another type of perception insufficiency to reflect how different kinds of insufficiencies could impact the risk evaluation.

5.1 Deceleration Scenario Definition

The scenario used in the simulations is a deceleration scenario as is depicted in Figure 5.1. The ego vehicle, which is shown as green in the figure, reduces or stops its velocity to avoid the target vehicle (red) located at the front through of an Automated Emergency Brake (AEB) function implemented in the system. The integrated function is labelled as a SAE Level 4 because the dynamic driving task has no human interaction. The waypoint followed by the ego vehicle is straight without turning in any direction. The starting speed of both vehicles is zero where the maximum speed by the ego vehicle in the scenario is 80kph. On the other hand, the target vehicle remains still during the test execution.

5.2 SOTIF Extended Analysis

Figure 5.2 shows an excerpt of a HARA analysis for an AEB function. Severity, exposure, and controllability values are given for each analysed hazard to assign an ASIL level consequently. As a part of the given scenario, the HARA shows the identified hazardous event *"a collision with the front vehicle"* based on different operational situations. Both hazards are classified as a *ASIL D* as both hazards are set to the maximum values for severity, exposure and controllability. Based on the given analysis, Figure 5.3 depicts the adaptation of the given

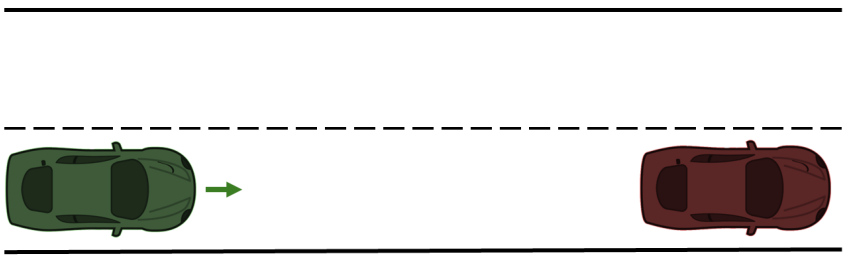


Figure 5.1: Deceleration scenario.

Hazard	Operational situation	Hazardous Event	Severity 'S'	Justification for 'S'	Exposure/ Frequency 'E'	Justification for 'E'	Controllability 'C'	Justification for 'C'	Resulting ASIL
Delayed emergency brake (HZ_01)	Vehicle operation driving at high speed on the highway	**Collision with the front vehicle**	3	Life-threatening injuries (survival uncertain), fatal injuries	4	Highway driving. High probability	3	SAE Level 4 Function - Not controllable by the driver	**ASIL D**
Emergency brake is not activated (HZ_02)	Vehicle operation driving at high speed on the highway	**Collision with the front vehicle**	3	Life-threatening injuries (survival uncertain), fatal injuries	4	Highway driving. High probability	3	SAE Level 4 Function - Not controllable by the driver	**ASIL D**

Figure 5.2: HARA excerpt of the analysed function.

analysis to be reused in the SOTIF context to discover potential triggering conditions of the system. As explained in Section 4.1, three new columns are added: *"Potential Triggering Condition"*, *"Description of Possible Functional Insufficiency"*, and *"Hazardous Behaviour"*, whose impact on the identified potential triggering conditions may lead to the system following the cause-and-effect model. Occurrence values were also added, which is relevant to know when the triggering conditions may appear in the scenario. As can be observed, a hazard could be related to one or more potential triggering conditions. In the first row, *"Darkness condition"* has been identified as a potential triggering condition since it may lead to a visibility reduction of the system. This triggering condition applies for systems in which the perception system is camera-based, although not being especially relevant for perception systems based on lidar or radar technology. On the other hand, the identified *"Heavy Snow condition"* may have an impact on more varied perception systems. Ultimately, the analysis will provide a list of the most relevant potential triggering conditions for the specified system and the scenario.

5.3 Triggering Conditions Integration into a Validation Tool Suite

The integration concept of the triggering conditions has been implemented within the AVL SCENIUS^{TM} [107] tool suite. This suite was designed specifically for scenario-based validation and covers the entire process, including scenario design, management, test case generation, test allocation, and result reporting. Users can manage all aspects of a scenario, such as

Potential Triggering Condition	Description of Possible Functional Insufficiency	Hazardous Behaviour	Hazard	Operational situation	Hazardous Event	Severity 'S'	Justification for 'S'	Occurrence frequency 'O'	Justification for 'O'
Darkness condition	Reduced visibility (PI_02)	Delay of the object detection (HB_06)	Delayed emergency brake (HZ_01)	Vehicle operation driving at high speed on the highway	**Collision with the front vehicle**	3	Life-threatening injuries (survival uncertain), fatal injuries	Common Case	Common cases denote the occurrence on a daily basis
Heavy snow condition	Reduced visibility (PI_02)	Delay of the object detection (HB_06)	Emergency brake is not activated (HZ_02)	Vehicle operation driving at high speed on the highway	**Collision with the front vehicle**	3	Life-threatening injuries (survival uncertain), fatal injuries	Common Case	Common cases denote the occurrence on a daily basis
Bridge on the road next to another vehicle that is in front of the ego vehicle	Reduced sensor perception accuracy (PI_03)	Object detection False-Negative (HB_01)	Emergency brake is not activated (HZ_02)	Vehicle operation driving at high speed on the highway	**Collision with the front vehicle**	3	Life-threatening injuries (survival uncertain), fatal injuries	Reported Case	Reported cases by media or mentioned in safety reports

Figure 5.3: HARA excerpt including SOTIF analysis.

parametrization, through the scenario designer. It offers full support for ASAM OpenScenario [108] and OpenDrive [109], ensuring immediate verification of scenarios for standard compliance as well as enhanced data and logic checks. Then, the scenario data manager module organizes and stores all scenarios in a central database, managing essential scenario elements such as road content, traffic content, and environmental data. Finally, the test case generator module allows users to define test orders for simulation or transfer to alternative execution environments. Its smart testing algorithms automatically reduce the number of test cases and parameter variations, optimizing the testing process. In addition to benefits like time and cost savings, efficiency, rapid integration, and enhanced traceability, the tool suite also incorporates the SOTIF concept. The enhancement in the identification of both known and unknown hazardous scenarios provides a more robust safety argumentation.

An ontology is employed to define the scenarios used for testing. The ASAM OpenXOntology [110] has been adapted with modifications to better align with the requirements of the toolchain. Within the ontology, the entities and their relationships are categorized into four main types:

- *Node:* A node represents the fundamental unit of the ontology hierarchy. It can be either a child of another node or a parent for enumerations or parameters. For instance, the ambient or weather conditions of a given scenario may be the parent of scenario parameters such as rain or illumination.
- *Enum:* Defines a list of values that are related to each other. For example, an enum is the snowfall condition, which is defined by three different levels of severity: heavy snow, light snow, and moderate snow.
- *Value:* A value is used to define a phenomenon that has not yet been fully modelled. In this context, the phenomenon in question already exists within the system but has not yet been fully characterised. For example, this type of entity are the ones previously mentioned: heavy snow, light snow, and moderate snow.

- *Param:* Defines an entity that can be quantified. Each one is associated with a unit to be measured. For example, scenario illuminance is associated with lux units.

The next integration step is to link the triggering conditions with the existing defined ontology to be able to parametrise each triggering condition and include them in the test scenario. By using this approach, we are able to define each triggering condition as a combination of one or many scenario constraints as explained in Section 4.2. The integration follows the identified triggering condition from the previous version, a heavy snow triggering condition. The BSI PAS 1883:2020 [26] standard defines a heavy snow condition as a visibility limitation up to 500 meters, where a *Node* entity defined as a scenario constraint is associated with the *Param* from the inner ontology (visibility) and the type of constraint (MAX). As already explained, a more precise definition could be provided by the combination of more than one scenario constraint but this definition is used for simplicity.

The complete relationship diagram of the explained triggering condition is shown in Figure 5.4. This diagram shows the two root concepts, Dynamic (1) and SOTIF (2), and how they are related to each other in an illustrative way. As described previously, the visibility parameter (9) inherits from the ambient conditions (5), which in turn inherits from the domain concept (3). Similarly and following the hierarchy, the snowfall condition (6) also derives from the ambient conditions (5). The heavy snow triggering condition (5) is linked with the ontology in which the snowfall conditions are defined in three levels, light, moderate, and heavy. Finally, the heavy snow condition is defined as a visibility scenario constraint (8) up to a maximum of 500 meters. Figure 5.5 illustrates the integration in a more general manner that can be employed to describe different triggering conditions. In addition to the given schema, Figure 5.6 depicts how this integration is shown from the user perspective in the validation tool chain.

5.4 Visibility Insufficiency

5.4.1 Simulation Considerations

The simulations were carried out through the open-source environment simulator CARLA [55]. All scenario elements were removed to avoid interferences in the perception system. Only both vehicles, ego and target, and the road remain. The system relies on a lidar sensor located at the top of the vehicle as a unique perception sensor. The ADS is developed in Robotics Operating System (ROS) [111] in which object detection and clustering use the implementation provided by the Autoware [112] software stack. Figure 5.7 depicts the architecture for this use case. It is assumed that the performance insufficiency injection does not include a significant delay in the system and does not affect the results. Figure 5.8 shows the visualisation of the scenario from both the environment simulator and vehicle perspective.

The performance insufficiency has to be modelled to be included in the injections. The reduction of the field of view is modelled as the cropping of the point cloud provided by the lidar sensor in which the ADS relies. Only the RSS [96, 113] area (A_{RSS}) is considered in the tests to increase the performance of the simulation and reduce the test cases. This decision

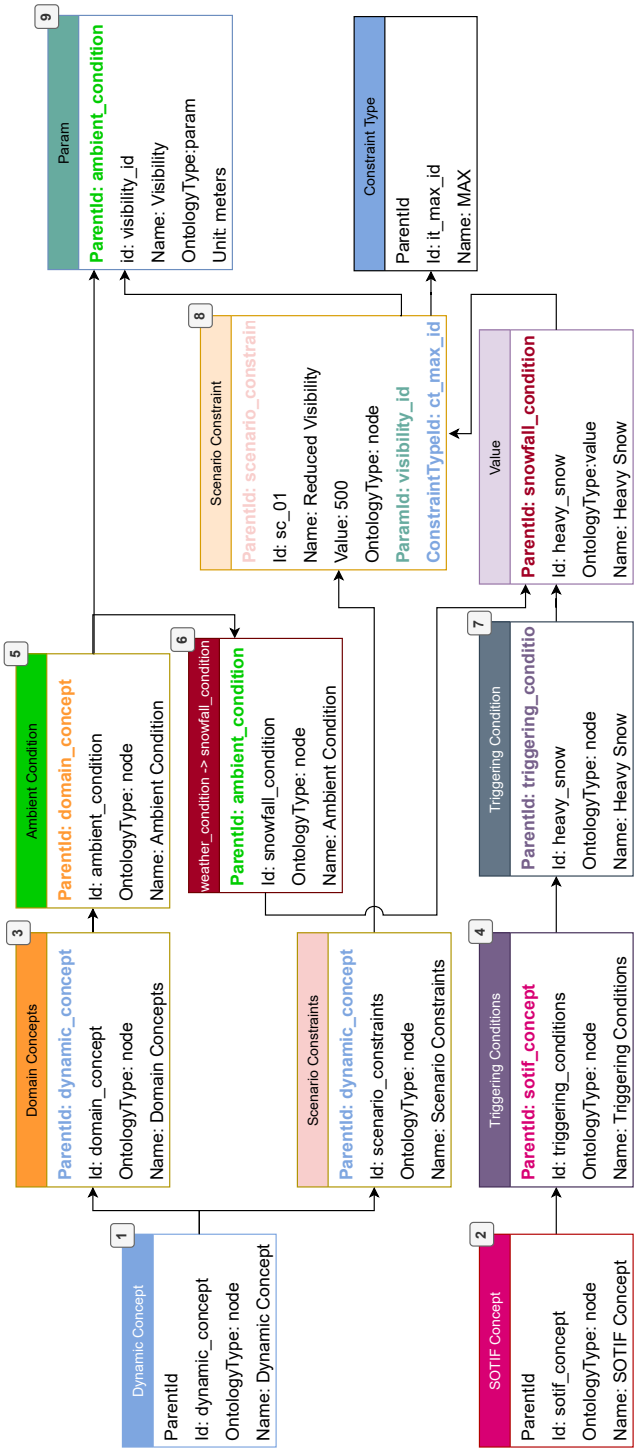


Figure 5.4: Heavy snow triggering condition entities' relationships within the validation tool suite.

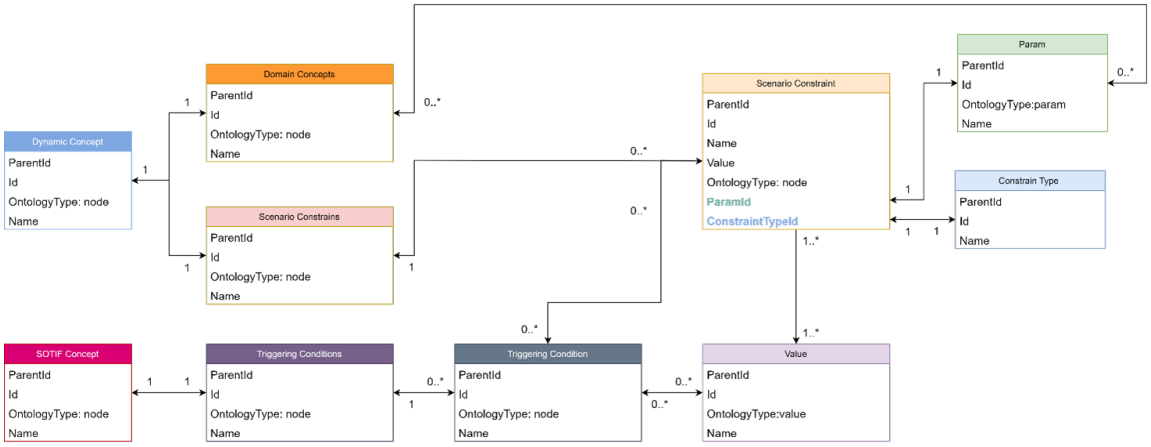


Figure 5.5: SOTIF concept implementation diagram.

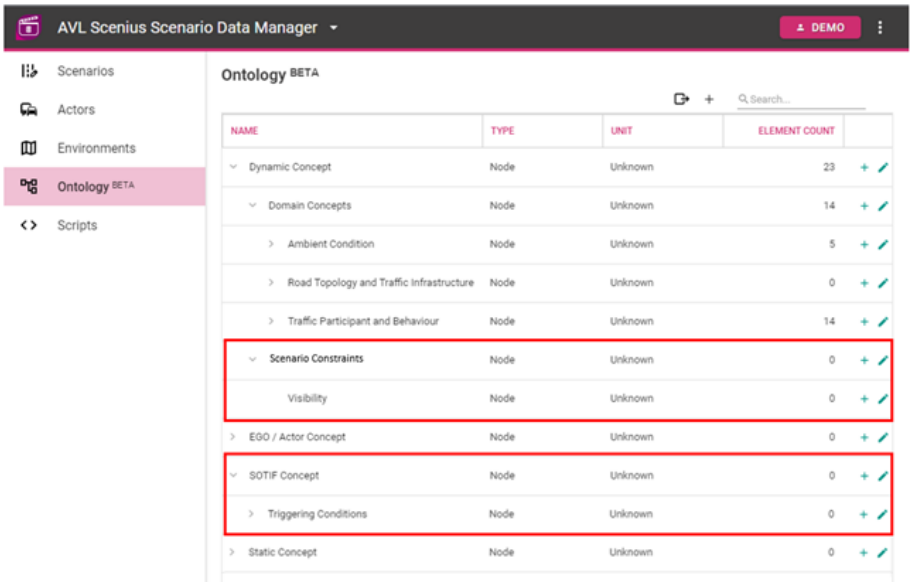


Figure 5.6: Integration of the SOTIF concept within AVL SCENIUS^{TM}.

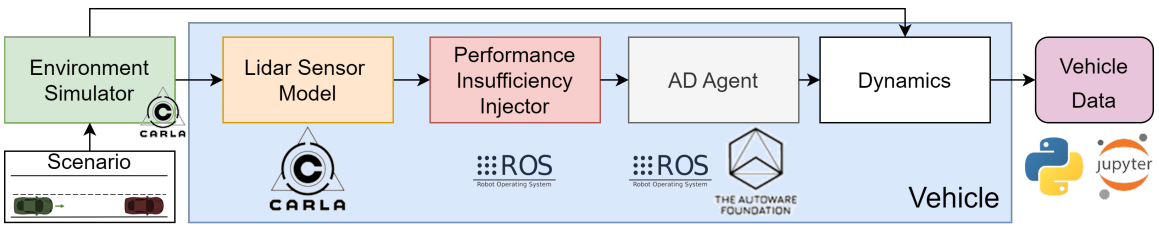


Figure 5.7: Use cases' architecture.

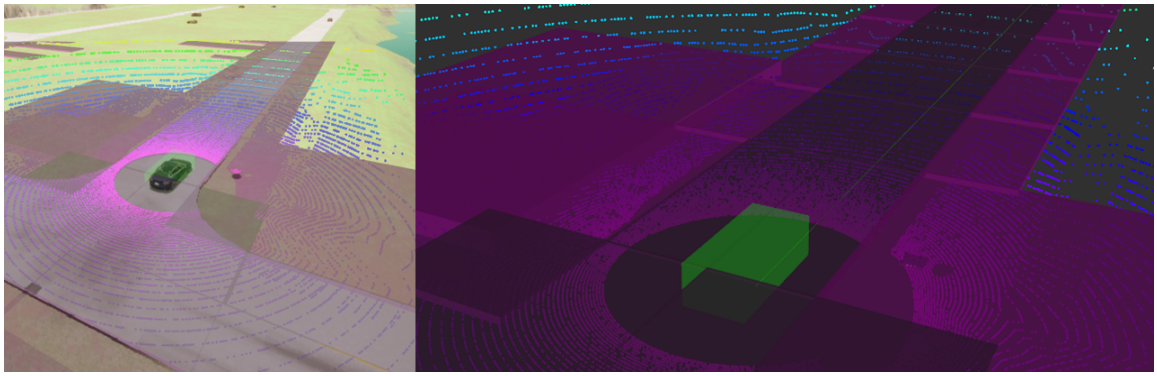


Figure 5.8: Simulation visualisation of the deceleration scenario. Left: CARLA visualisation. Right: vehicle visualisation.

was made because, based on the model, only the A_{RSS} is relevant for the safety of the vehicle. Moreover, to be more conservative in our approach, the velocity of the target vehicle was set to zero in the RSS model.

$$D_{rss} = \left[v_r \rho + \frac{1}{2} a_{max,accel} \rho^2 + \frac{(v_r + \rho a_{max,accel})^2}{2 a_{min,brake}} \right]_+ [x]_+ := max\{x, 0\} \tag{5.1}$$

The meaning and the given value for each parameter of the formula are:

- D_{RSS}: Minimum distance to ensure that there is no crash with the obstacle.
- v_r: Max ego vehicle velocity (m/s) in the test scenario. Value: 22.22m/s (80 km/h).
- ρ: Response time in seconds. 0.5s.
- $a_{max,accel}$: Maximum acceleration of the robot (m/s^2). Value: 5.5m/s^2.
- $a_{min,brake}$: Minimum braking acceleration of the ego vehicle (m/s^2). Value: 4.5m/s^2.

The calculated minimum distance by using the model is only used for the longitudinal value. The given distance, calculated using the above values, is 81.09 meters (D_{RSS}). On the other hand, the standard width for a highway is applied for the latitude value, 3.75 meters [114] (R_{width}). The vehicle length (V_{length}) is assumed as five meters. Figure 5.9 illustrated the calculated area (A_{RSS}) used in the simulations.

$$V_{length}/2 < X < V_{length}/2 + D_{rss} \qquad -R_{width}/2 < Y < R_{width}/2 \tag{5.2}$$

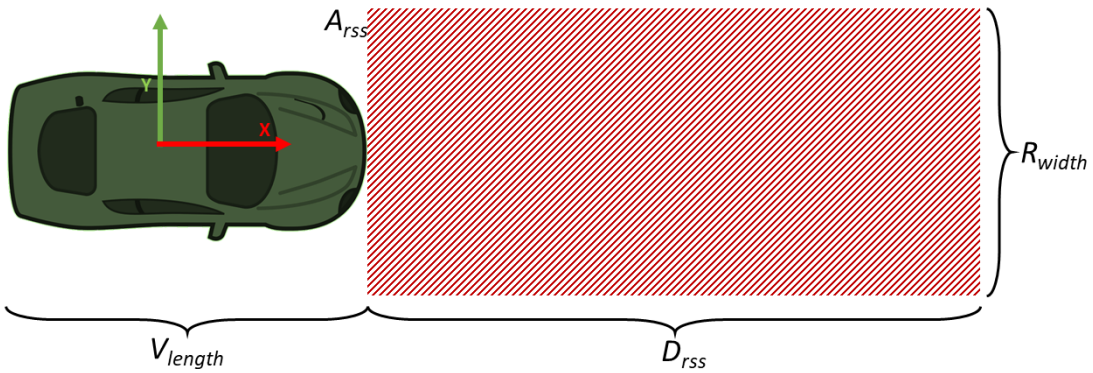


Figure 5.9: Considered RSS area in the use case.

5.4.2 Perception Performance Insufficiencies Injection

The nominal performance of the ADS is shown in Figure 5.10. The nominal performance is the performance of the system without the injection of the performance insufficiencies. In this use case, one hundred simulations ($B = 100$) were performed to obtain the probabilistic values of the nominal performance. The graph shows the distance travelled (y-axis) over the time (x-axis). Black lines show the travelled distance for each simulation, where the cyan line represents the mean value of all simulations carried out. The execution time tolerable window is shown with vertical blue lines. Similarly, green lines are used to limit the tolerance window of the ADS distance travelled. The tolerable windows are calculated based on the mean value from the nominal performance simulations in which the standard deviation is multiplied by a factor set to the upper and lower limits. A test case is considered hazardous behaviour if the execution time or the distance travelled is outside the defined tolerance window.

The injection levels introduced into the system for the limited visibility performance insufficiency are 80, 60, 45, 30, 20, and 15 meters. The results of these injections are illustrated in Figure 5.11 and Figure 5.12. Figure 5.11 shows the level 0 (80 meters), level 1 (60 meters), and level 2 (45 meters). The graphics show that these injections have no relevant impact on the system output. The black lines show the simulation results for each test execution, whereas the cyan line depicts the nominal performance of the system. The behaviour of the system with injections remains very similar to the nominal performance. In these three first injection levels, all simulations are within the defined boundaries, both in time, represented in blue lines, and travelled distance shown with green lines. On the other hand, Figure 5.12 shows the results of the three next injection levels, defined as level 3 (30 meters), level 4 (20 meters), and level 5 (15 meters). Unlike the less restrictive levels, these injection levels do include an impact on the system output. The level 3 simulation results show the differences between the nominal performance and the performance with injections. The behaviour starts to differentiate from the second ten due to the delay in the detection of the target vehicle, which also delays the triggering of the braking. The impact of the injection leads to a collision with the target vehicle, which is shown with red lines. Consequently, the following injection levels have similar behaviour but with a bigger impact due to the increasing delay in the detection of the target vehicle. Therefore, the collisions from level 4 occur later than the collisions derived

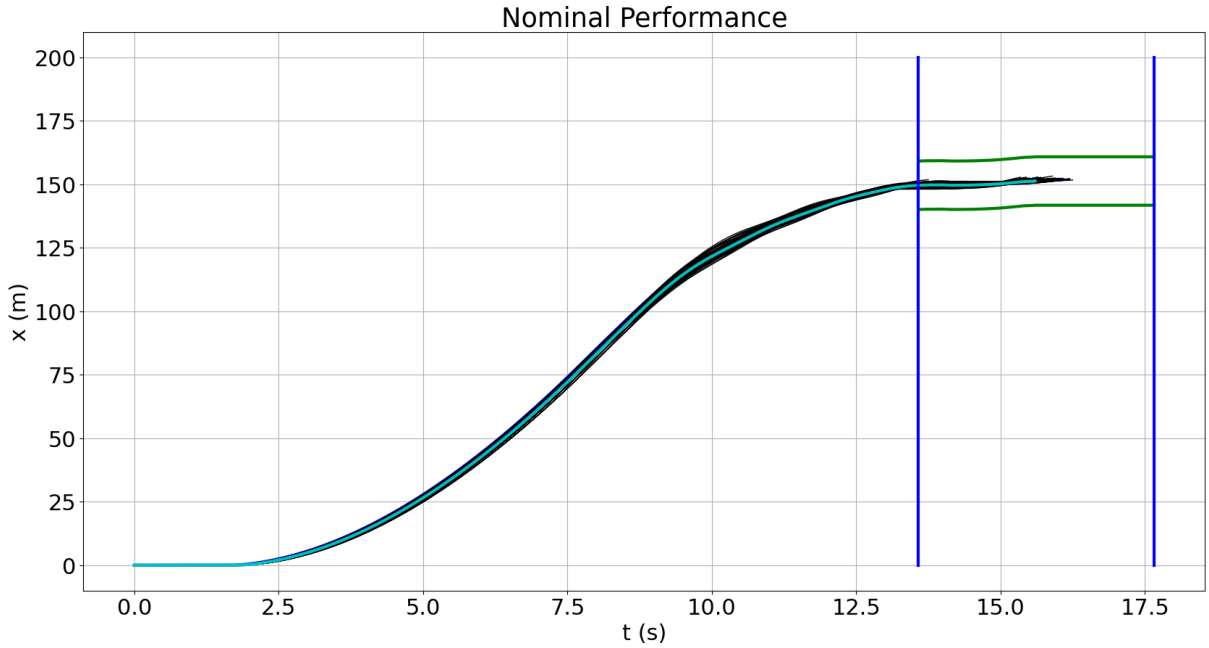


Figure 5.10: Nominal performance results.

from the injection of the most restrictive level. The delay induced by the injection not only has an impact on the target vehicle detection but also on the speed of collision. The longer the detection delay, the higher the collision speed between the ego vehicle and the target vehicle. The vehicle speed is relevant because it is one of the parameters to calculate the injury in a collision. The severity of injuries in a collision is directly proportional to the speed at which the collision occurs.

The results of performance insufficiency are summarised in Table 5.1. This table shows the probabilities of hazardous behaviour ($P(PI)$) and collision ($P(C)$) of each injection level. As shown in the graphics, there is no impact on the output for the three first injection levels. That means, neither $P(PI)$ nor $P(C)$. In level 3, two-thirds of all simulations lead to a hazardous behaviour, leading to a collision in all situations where there is a hazardous behaviour. According to the results, there is a visibility insufficiency between level 2 and level 3 in which the insufficiency starts to have an impact on the system. Unfortunately, the exact value when that occurred was not found by using the defined injection levels in this use case. The last two levels have a full impact both in terms of hazardous behaviour and collision. All simulations in these levels lead to a hazardous behaviour and end in a collision. These results follow the cause-and-effect model where the visibility reduction injected at the technical level leads to an output insufficiency at the functional level, which ends in hazardous behaviour.

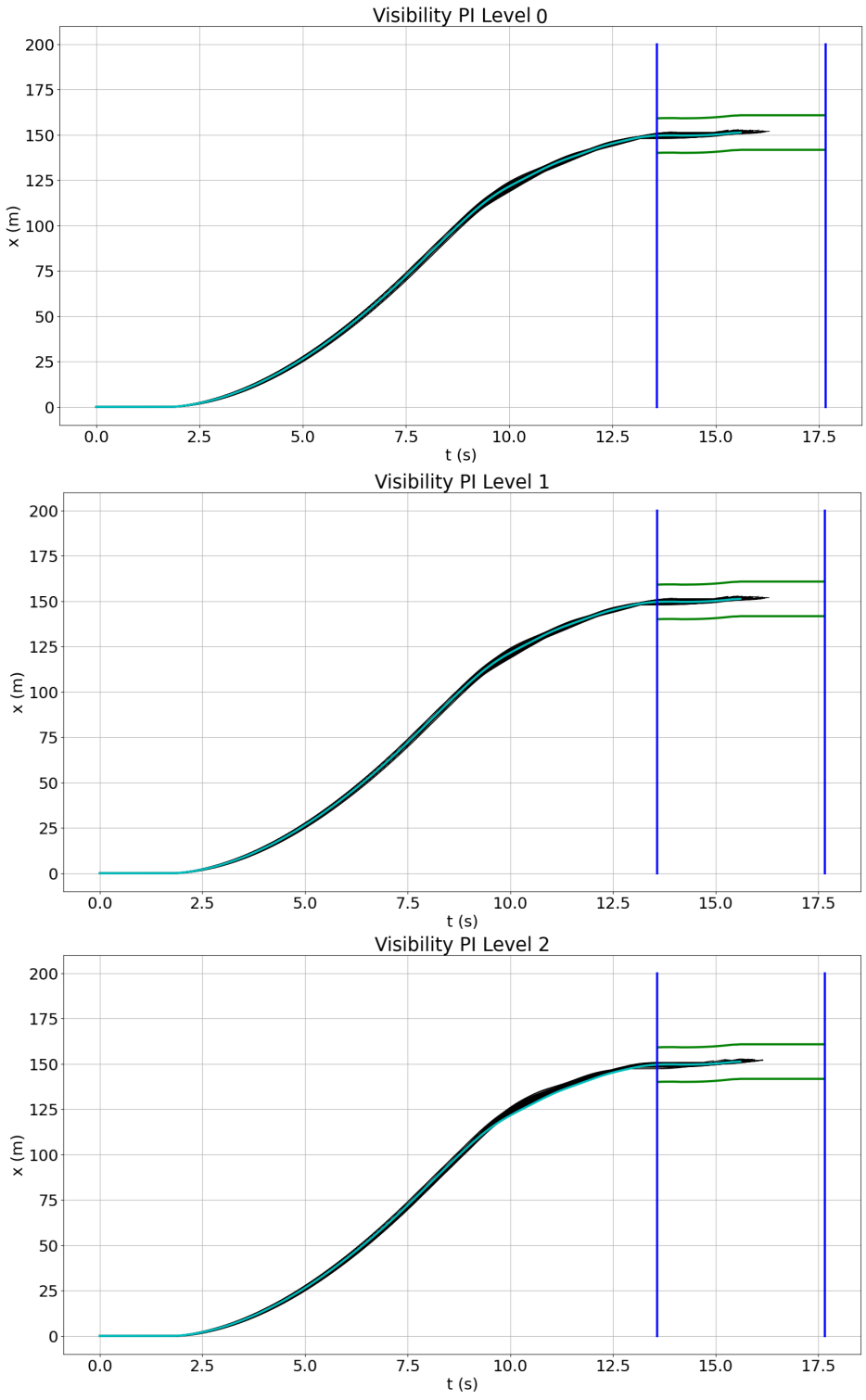


Figure 5.11: Simulation results of the visibility performance insufficiency injections from injection levels 0 to 2.

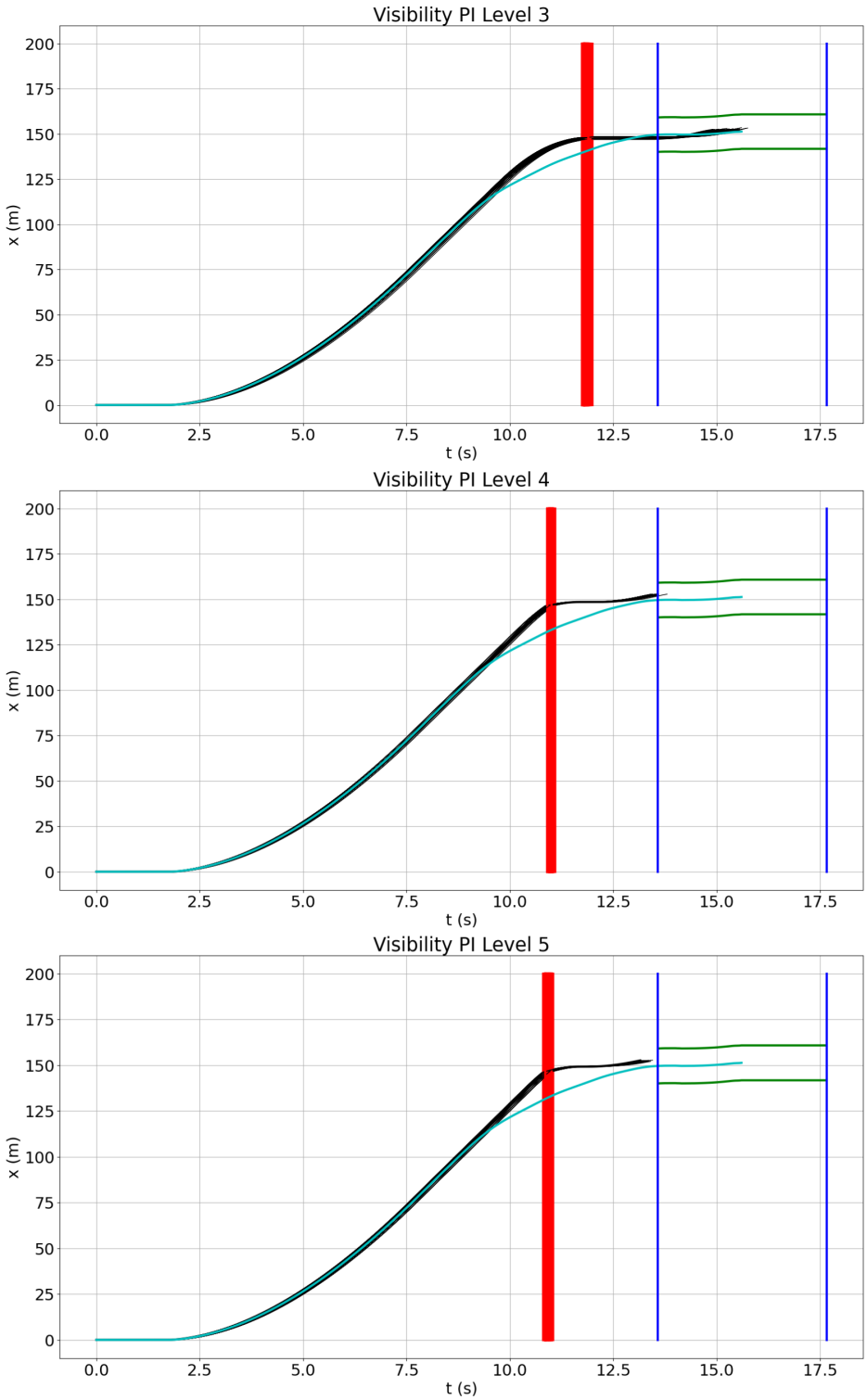


Figure 5.12: Simulation results of the visibility performance insufficiency injections from injection levels 3 to 5.

Table 5.1: Simulation results for each visibility performance insufficiency level.

PI_{vis} **Level (Meters)**	**Hazardous Behaviour *P(PI)***	**Collision *P(C)***
Level 0 (80 meters)	0.00	0.00
Level 1 (60 meters)	0.00	0.00
Level 2 (45 meters)	0.00	0.00
Level 3 (30 meters)	0.66	0.66
Level 4 (20 meters)	1.00	1.00
Level 5 (15 meters)	1.00	1.00

5.4.3 Risk Quantification from Simulation Results

As described in Section 4.4, the quantification of the risk in our approach depends on three main parameters, the plausibility factor (PF), the probability of performance insufficiency ($P(PI)$), and the probability of injury ($P(I)$). In this use case, the values of the PF are set by expertise judgment. The plausibility factor was calculated based on the values from an exponential distribution with a given lambda value (λ = 1) and a random variable (X), which is defined by each injection level. All values are shown in Table 5.2.

Table 5.2: Plausibility factor for the visibility performance insufficiency.

PI_{vis} **Level**	**Visibility Limitation**	**X**	**Given *PF***
Level 0	80 meters	P(X>=0)	$PF_{vis_{80}}$ = 1.00000
Level 1	60 meters	P(X>=1)	$PF_{vis_{60}}$ = 0.36788
Level 2	45 meters	P(X>=2)	$PF_{vis_{45}}$ = 0.13534
Level 3	30 meters	P(X>=3)	$PF_{vis_{30}}$ = 4.979x10^{-2}
Level 4	20 meters	P(X>=4)	$PF_{vis_{20}}$ = 1.832x10^{-2}
Level 5	15 meters	P(X>=5)	$PF_{vis_{15}}$ = 6.74x10^{-3}

Equation 5.3 shows the evaluated risk for the lowest level of the reduced visibility performance level. In this case, the plausibility factor given for this injection is the maximum, $PF_{vis_{80}} = 1.0000$. The plausibility factor given is maximum because it is assumed this performance insufficiency is highly probable. Therefore, its impact on the behaviour of the system is not minimised. The probability of performance insufficiency ($P(PI_{vis_{80}})$) is zero, which is expected since the reduced visibility is close to the nominal field of view of the sensor.

Table 5.3: Risk evaluation for each visibility performance insufficiency level.

PI_{vis} **Level (Meters)**	***PF***	***P(PI)***	***P(I)***	**Risk**
Level 0 (80 meters)	1.00000	0.00	0.00	0.00
Level 1 (60 meters)	0.36788	0.00	0.00	0.00
Level 2 (45 meters)	0.13534	0.00	0.00	0.00
Level 3 (30 meters)	0.04979	0.66	$1.22966\text{x}10^{-2}$	$4.04083\text{x}10^{-4}$
Level 4 (20 meters)	0.01832	1.00	$3.83674\text{x}10^{-2}$	$7.02891\text{x}10^{-4}$
Level 5 (15 meters)	0.00674	1.00	$3.83675\text{x}10^{-2}$	$2.58597\text{x}10^{-4}$

Consequently, the probability of injury ($P(I_{vis_{80}})$) is zero due to the absence of collisions. Following the given risk quantification, the calculated risk is zero for this performance insufficiency level.

$$Risk_{vis_{80}} = PF_{vis_{80}} * P(PI_{vis_{80}}) * P(I_{vis_{80}}) = 1.00 * 0.00 * 0.00 = 0.00 \tag{5.3}$$

On the other hand, the quantified risk for level 3 of the performance insufficiency injection ($Risk_{vis_{30}}$) is not zero, unlike the previously quantified risk. The given plausibility factor is not one in this case because this kind of performance insufficiency level does not occur as regularly. Moreover, the performance insufficiency injection at this level has an impact on two-thirds of the simulations. Subsequently, the probability of injury is not zero because some injections lead to collision at this injection level.

$$Risk_{vis_{30}} = 0.04979 * 0.66 * 1.229\,66 \times 10^{-2} = 4.040\,83 \times 10^{-4} \tag{5.4}$$

Finally, the quantitative risk evaluation for this reduced visibility performance insufficiency is shown in Equation 5.5, which sums up all quantified risks for all injection levels. It should be noted that the values of the probability of injury increase when more strict levels are injected because there is less time to brake for the vehicle and therefore the crash speed is greater for each level. The overall risk for each level is not always higher than the previous level because of the given plausibility factor.

$$Risk_{PI_{vis}} = Risk_{vis_{80}} + Risk_{vis_{60}} + Risk_{vis_{45}} + Risk_{vis_{30}} + Risk_{vis_{20}} + Risk_{vis_{15}} = 1.365\,57 \times 10^{-3} \tag{5.5}$$

This quantitative risk evaluation provides a reference point for the minimisation of risk in subsequent iterations of the SOTIF validation. It is noted that these results could be used to

validate the function for specific triggering conditions defined in the standards. For example, the SAE standard [25] classifies fog into six levels based on system visibility.

- Level 5: 0 meters $\leq$ visibility < 61 meters
- Level 4: 61 meters $\leq$ visibility < 244 meters
- Level 3: 244 meters $\leq$ visibility < 805 meters
- Level 2: 805 meters $\leq$ visibility < 1609 meters
- Level 1: visibility $\geq$ 1609 meters

Therefore, the system is validated for the SAE fog scale up to level 4 ($visibility > 60meters$), ensuring zero risk at that level in the system:

$$Risk_{PI_{visSAELevel4}} = Risk_{vis_{60}} = 0.00 \tag{5.6}$$

Consequently, the quantified risk for the most severe fog level, level 5, in by SAE can quantified by the sum of all risks below 60 meters.

$$Risk_{PI_{visSAELevel5}} = Risk_{vis_{45}} + Risk_{vis_{30}} + Risk_{vis_{20}} + Risk_{vis_{15}} = 1.365\,57 \times 10^{-3} \tag{5.7}$$

This gives a link to adapt the obtained metrics by the performance insufficiencies injections to be used to provide specific metrics for the defined scenario conditions defined in the standards.

5.5 Accuracy Insufficiency

A reduction of accuracy of the perception part is included in the system based on the classification from Table 4.1 in this use case. The primary objective of this use case is to demonstrate how reflections in different density levels can affect the system's object detection and resulting behaviour.

5.5.1 Perception Performance Insufficiencies Injection

The accuracy insufficiency was modelled by including random points with different intensity levels in the point cloud message generated from the lidar perception sensor. Only the RSS area (A_{RSS}) is taken into consideration since it is the relevant safety area. Moreover, it increases the performance of the simulation because the injection is limited only to the safety-relevant area. Different levels of reflection density are injected into the system. The density is calculated based on the number of points in the A_{RSS}, which depends on the sensor resolution and the number of injected points for this area ($injection\_density = number\_injected\_points/number\_A_{RSS}\_points$).

Figure 5.13 shows how the injections are perceived from the vehicle perspective. The left picture shows the vehicle perception where a low level of injection is utilised. To be noticed that many false positives appear and make the vehicle constantly brake and accelerate. On the other hand, the right picture shows the perception with a more severe injection where the amount of false positives causes the ego vehicle to stop completely. The results of all injection levels are depicted in Figure 5.14 and Figure 5.15. The first level of injection has no visible impact on the output of the function. Level 1, however, has a visible impact on the behaviour of the function. In this case, the ego drives slower than the nominal performance. Although it does affect the behaviour of the system, all results of the simulation tests are within the acceptance boundaries, therefore, no hazardous behaviours are considered in this case. By contrast, the impact produced by the injection of level 3 makes the ego vehicle not reach the expected behaviour. The addition of reflections in this level produces a no-uniform behaviour due to constant random false positives that make each simulation produce a different result. The increasing number of injected reflections leads the vehicle completely stop in the next two injection levels. In these levels, the ego vehicle remains nearly at the start point until the test timeout is reached. The last level, level 5, presents a behaviour worth mentioning. In this level, the amount of reflections is so high that they are omitted by the ADS, producing the target vehicle is not detected. Therefore, all test executions lead to a collision with the target vehicle. Table 5.4 summarises the results of all accuracy insufficiency levels.

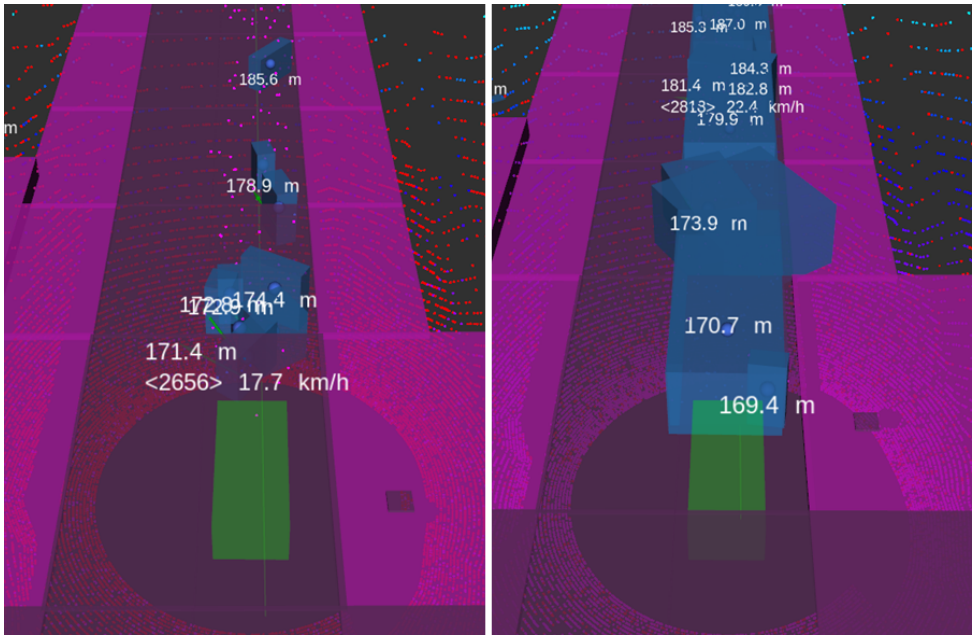


Figure 5.13: Accuracy performance insufficiency injection visualised from ego vehicle.

5.5.2 Risk Quantification from Simulation Results

Once the probability of performance insufficiency is calculated, the risk can be evaluated. Similar to the previous insufficiency, the plausibility factor is determined based on expert

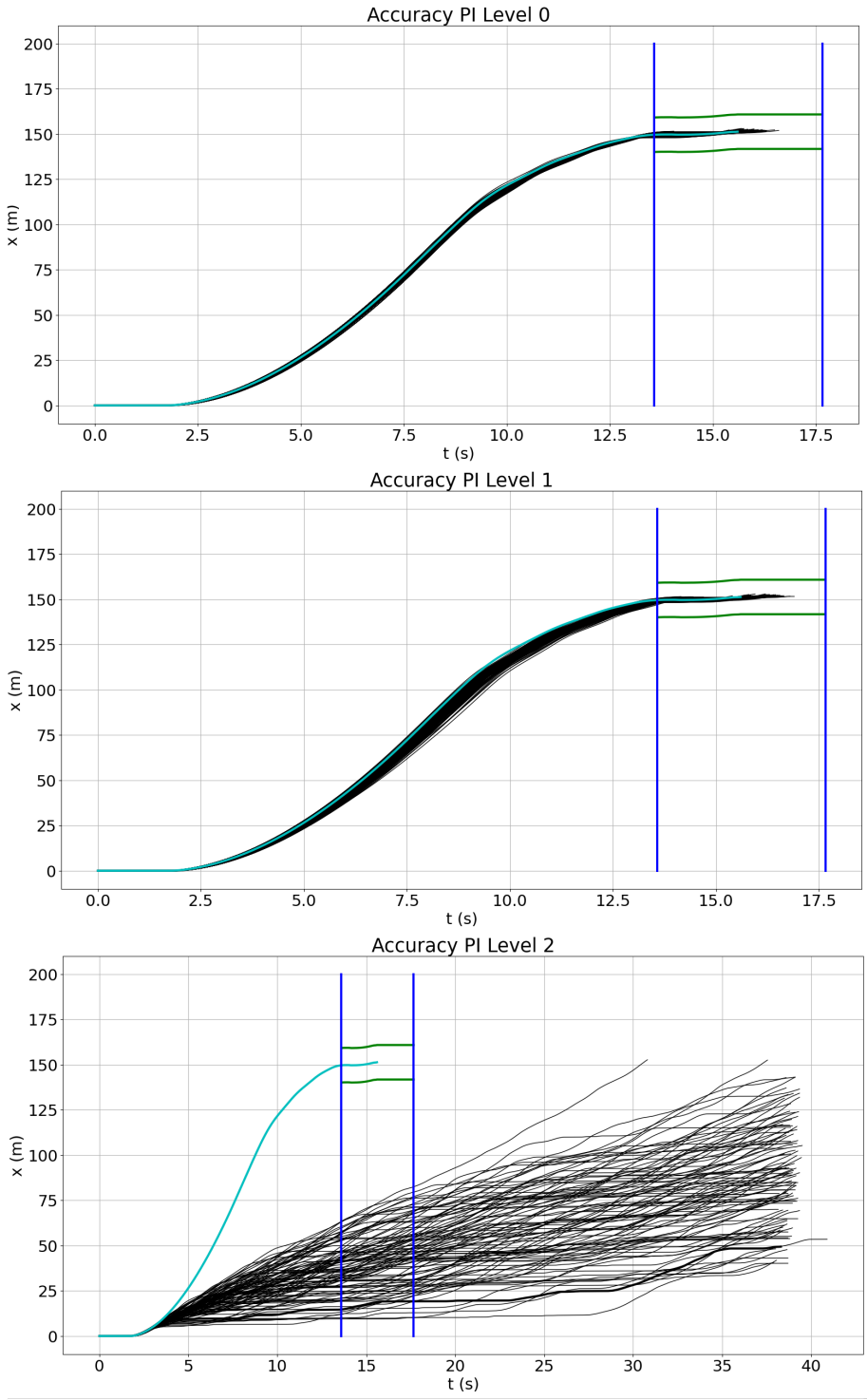


Figure 5.14: Simulation results of the accuracy performance insufficiency injections from injection levels 0 to 2.

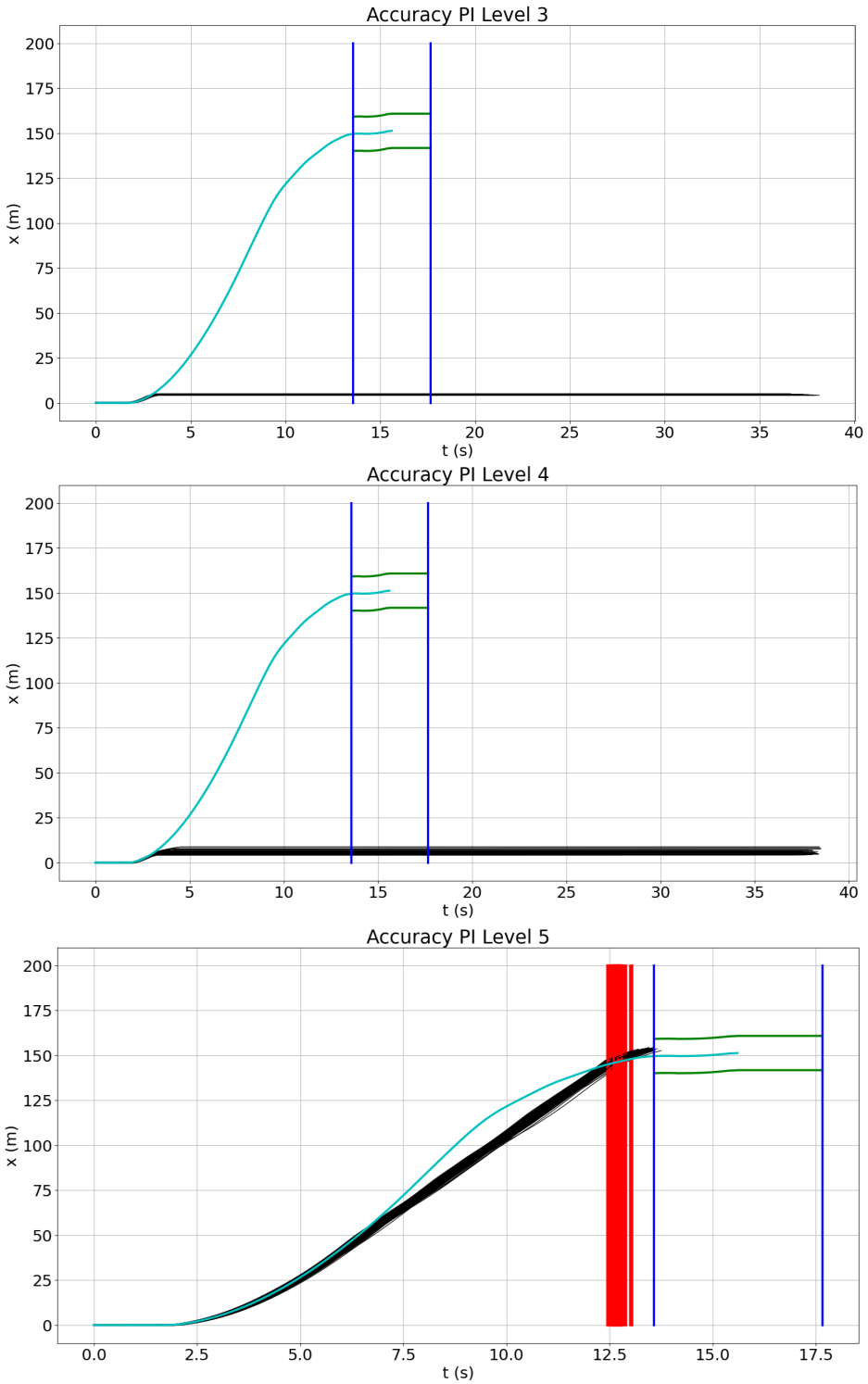


Figure 5.15: Simulation results of the accuracy performance insufficiency injections from injection levels 3 to 5.

Table 5.4: Simulation results for each accuracy performance insufficiency level.

PI_{acc} **Level (Injection Density in %)**	**Hazardous Behaviour *P(PI)***	**Collision *P(C)***
Level 0 (0.15 %)	0.00	0.00
Level 1 (0.30 %)	0.00	0.00
Level 2 (0.75 %)	1.00	0.00
Level 3 (1.49 %)	1.00	0.00
Level 4 (2.99 %)	1.00	0.00
Level 5 (5.97 %)	1.00	1.00

judgment. Table 5.5 shows the risk values for each injection level. In this performance insufficiency, only the more severe injection level leads to a risk value bigger than zero due to the occurrence of collisions. It should be noted that there are many performance insufficiency levels which, although they affect the ADS, do not lead to collisions. Therefore, SOTIF measures should be implemented to minimise the probability of hazardous behaviours. One example of a SOTIF measure could be the inclusion of a diversity of perception sensor technologies. The sensor diversity is beneficial as each sensor technology has its advantages and disadvantages in specific environmental situations, which could help mitigate the impact of certain scenarios.

Table 5.5: Risk evaluation results for each accuracy performance insufficiency level.

PI_{acc} **Level (Injection Density in %)**	***PF***	***P(PI)***	***P(I)***	**Risk**
Level 0 (0.15 %)	1.00000	0.00	0.00	0.00
Level 1 (0.30 %)	0.36788	0.00	0.00	0.00
Level 2 (0.75 %)	0.13534	1.00	0.00	0.00
Level 3 (1.49 %)	0.04979	1.00	0.00	0.00
Level 4 (2.99 %)	0.01832	1.00	0.00	0.00
Level 5 (5.97 %)	0.00674	1.00	$1.26752 \text{x} 10^{-2}$	$8.54309 \text{x} 10^{-5}$

Equation 5.8 the calculated risk for the reduction of accuracy. Although most of the performance insufficiency injections have an impact on the system output, only the last injection has an impact on the risk quantification. The aim of showing this performance insufficiency is to remark both risk and SOTIF modifications are relevant in the validation

process.

$$Risk_{PI_{acc}} = \sum_{i=0}^{5} Risk_{PI_j} = 8.54309 \times 10^{-5} \quad \forall \quad 0 \leq j \leq 5 \tag{5.8}$$

The risk quantification for the validated ADS is the sum of all performance insufficiencies validated along this use case as Equation 5.9 shows. As mentioned in the document, one of the main goals of SOTIF is to improve the system's safety through each iteration. Therefore, the calculation of this risk provides a metric that is used as a reference metric to minimise the risk in further validation iterations.

$$Risk_{ADS} = Risk_{PI_{vis}} + Risk_{PI_{acc}} = 1.36557 \times 10^{-3} + 8.54309 \times 10^{-5} = 1.45100 \times 10^{-3} \tag{5.9}$$

The main goal of this section is to show how the proposed argumentation can be put into practice through the different steps explained in Chapter 4. First, the performed HARA analysis for the deceleration scenario is adapted to identified hazardous scenarios. Then, the parametrisation of the triggering conditions is described and integrated into an existing validation tool, demonstrating that the proposed approach is prepared for deployment in the industry. Risk quantification is also shown by the injection of two performance insufficiencies. For simplicity, only two performance insufficiencies are included in this use case, but it is expected a more wide of insufficiencies to cover the maximum number of scenario conditions in the validation. The selection of the chosen two insufficiencies is to show different behaviours that insufficiencies can included in the system. The visibility insufficiency leads to collision each time it has an influence, while most of the accuracy insufficiency injections have an impact on the system but don't lead to collisions, so there is no risk. In this case, the goal is to show that, although there is no risk, SOTIF modifications of the system are necessary to avoid hazardous behaviour due to performance insufficiency. Finally, the obtained risk metrics of the use case are used as a reference to compare the ADS with other systems as well as to improve the safety system by minimizing this metric in further iterations.

CHAPTER 6

Conclusions and Outlook

In this concluding chapter, a brief overview of the chapters covered in this work is presented, providing a summary of the motivation behind the research activities during the PhD studies for each chapter, highlighting the key findings and outlining potential avenues for further research. The thesis starts by introducing the main motivation and the research questions in the first chapter. Chapter 2 introduces the main thesis' knowledge necessary to understand the proposed work. This chapter is focused on automotive safety, discussing the main functional safety concepts, SOTIF and scenario-based validation. The next chapter, Chapter 3, covers the state of the art on the main fields required to achieve the argumentation proposed in this thesis. Existing safety works are covered in the first section of chapters such as regulations, tools, and standards, as well as different approaches from other research works. The following sections cover the existing research on sensor models, necessary for the perception performance insufficiencies injection, and risk evaluation for automated driving.

The main work of this thesis is described in Chapter 4, which starts with showing the workflow of the argumentation with its main steps. The quantitative analysis shows how existing analysis from functional safety can be adapted to be used for a SOTIF argumentation. The methodology also describes how triggering conditions can be parametrised to convert it from a textual description to a formal definition, making them feasible to be used in a software tool. Due to the impossibility of validating all possible potential triggering conditions, this thesis is focused on the validation of the impact the triggering conditions may include in the system. Perception performance insufficiencies injections are utilised to face this issue. The final part of the argumentation is covered in the last section, in which a quantitative metric is calculated from the results of the injections. Chapter 5 describes how the proposed argumentation can be carried out. First, the chapter illustrates how the triggering conditions are integrated into an existing validation tool. This description demonstrates that the proposed approach is not only a conceptualisation but rather an operational proposition that is ready to be deployed in the industry. Moreover, the perception performance insufficiencies and risk quantification are evaluated by using two different use cases, visibility and an accuracy insufficiency. The injection of these different insufficiencies demonstrates that, in certain circumstances, despite the absence of risk evidence, SOTIF modifications to the system are required. Finally, all publications are collected in the Annex, including the main contribution to this thesis for each publication.

6.1 Research Contributions

An overview of the results achieved by this thesis is discussed in this section. Following back to the problem statements described at the beginning of this thesis, the main research question this thesis faces is to improve the validation and verification of ADAS/AD functions by using triggering condition injections. This question has been addressed by the achieved contributions during the thesis:

C1 *Qualitative analysis by adapting existing methodology*: The qualitative identification of potential triggering conditions is shown by adapting the HARA methodology from the functional safety standard. SOTIF relevant parameters are added to each identified hazard from the analysis, which are used to identify known scenarios, both non-hazardous and hazardous.

C2 *Integration of triggering conditions into software tools*: Only the identification of potential triggering conditions of a system is not enough. Therefore, they have to be translated from a textual description to a representation that can be used in software and be able to discern if a scenario is hazardous or not. This thesis also provides a way to face this issue. It is solved by decomposing the triggering conditions into scenario constraints that can be parametrised and integrated into existing validation tools. The decomposition is carried out by the taxonomy definitions given in the most common standards to ensure interoperability between different systems.

C3 *Usage of performance insufficiencies injection to ensure the system is reliable and to detect unknown hazardous scenarios*: The validation process is done through perception performance insufficiencies injection. The injection process described in this document faced two main problems. The first problem is the impossibility of validating all possible triggering conditions since the number of combinations tends to be unmanageable. Therefore, the argumentation covers the impact that the triggering condition could have on the system. The second problem is the identification of edge and corner scenarios by using the injection of performance insufficiencies in different levels of intensity to check when the system's output differs from the nominal performance without any insufficiency and thus define when a performance insufficiency leads to a hazardous behaviour.

C4 *Providing a quantitative metric as a reference for improvement in further validation iterations of the system*: Risk assessment is also addressed in this research. The metric is obtained from the calculated probability of hazardous behaviour and the probability of injury from the injection results. A third factor, which models the occurrence of performance insufficiency, is also taken into account for risk quantification. The quantitative risk metric defined in the thesis provides a reference metric for improvement in further validation iterations of the system.

C5 *Validation process that covers main clauses of the standard*: As a final contribution, the argumentation itself, integrates all the contributions during the thesis. It provides a

complete workflow that covers all the main objectives of the SOTIF standards. These objectives include the identification of unknown scenarios and the minimisation of the risk associated with the system.

The aforementioned contributions collectively address the primary objectives set forth in the thesis.

6.2 Future Work

This research has raised several questions that require further investigation. The parametrisation of the triggering conditions needs to be extended to cover more complex conditions, the impact of which is not easy to quantify. Although the risk evaluation provides a quantitative metric, it is not related to any specific measure such as the number of hours driven or kilometres travelled. Future research should address this issue to better link the obtained metric with real-world measurements. Moreover, the current usage of the plausibility factor should be calculated through statistics instead of current expertise judgment. Another important question that needs to be addressed is how to determine when a performance insufficiency has been fully validated, including more accurate models for each performance insufficiency. Additionally, a more effective approach to identifying edge cases in performance insufficiency injection testing should be implemented in the future. Moreover, in the following research, we would like to extend the use case to different sensor setups to see how different sensor configurations can affect the function validation. As a final open research question remains the extension of this argumentation to cover not only the sense block but also the other blocks of the sense-plan-act model.

CHAPTER A

Publications

This chapter provides an overview of the papers included in the thesis. Initially, a list of all references is presented, accompanied by the author's contribution to each paper. Subsequently, an overview of the publications and their relationship with the contributions achieved during the course of the thesis is provided. Finally, the full versions of all individual publications are included in the chapter.

The list of the included publication and the given contribution for each publication is as follows:

P1 C. Schwarzl, N. Marko, H. Martin, V. Expósito Jiménez, J. Castella Triginer, B. Winkler, and R. Bramberger, "Safety and Security Co-engineering for Highly Automated Vehicles," *e & i Elektrotechnik und Informationstechnik,* vol. 138, pp. 469–479, nov 2021. [4] – Journal. **My Contribution**: Use case description and HARA analysis of the collision avoidance function of the robot.

P2 I. Cieslik, V. J. Expósito Jiménez, H. Martin, H. Scharke, and H. Schneider, "State of the Art Study of the Safety Argumentation Frameworks for Automated Driving System," in *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops* (M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, eds.), (Cham), pp. 178–191, Springer International Publishing, 2022. [5] – Conference. **My Contribution**: The main contribution in this paper is related to SOTIF safety argumentation, which includes the definition of the scope, safety goals, acceptance criteria, and the quantitative Hazard and Risk Analysis as a part of the SOTIF Processes.

P3 V. J. Expósito Jiménez, H. Martin, C. Schwarzl, G. Macher, and E. Brenner, "Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions," in *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops* (M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, eds.), (Cham), pp. 11–22, Springer International Publishing, 2022. [6] – Conference. **My Contribution**: This paper was completely written and conceived by me. The contribution of the other authors was in the form of discussions, feedback for improvements, and corrections.

P4 V. J. Expósito Jiménez, B. Winkler, J. M. Castella Triginer, H. Scharke, H. Schneider, E. Brenner, and G. Macher, "Safety of the Intended Functionality Concept Integration into

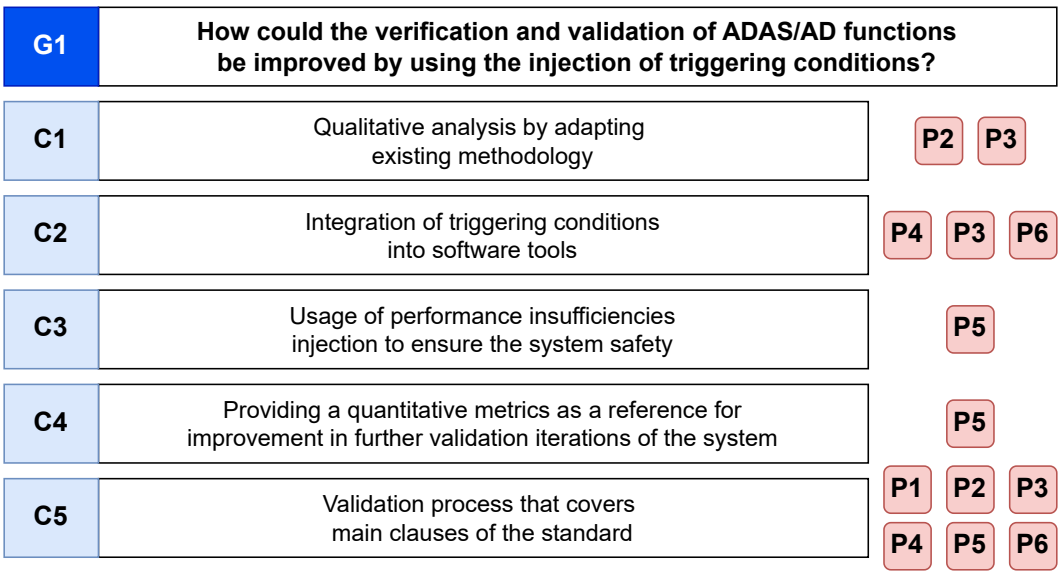


Figure A.1: Relationship between the publications and their contribution in the thesis.

a Validation Tool Suite," *ACM SIGAda Ada Letters*, vol. 43, pp. 69––72, jun 2024. [7] – Journal. **My Contribution**: This paper was completely written and conceived by me. The contribution of the other authors was in the form of discussions, feedback for improvements, and corrections.

P5 V. J. Expósito Jiménez, G. Macher, D. Watzenig, and E. Brenner, "Safety of the Intended Functionality Validation for Automated Driving Systems by using Perception Performance Insufficiencies Injection," *Vehicles*, vol. 6, no. 3, pp. 1164–1184, 2024. [8] – Journal. **My Contribution**: This paper was completely written and conceived by me. The contribution of the other authors was in the form of discussions, feedback for improvements, and corrections.

P6 H. Scharke, H. Goossens, S. Kalisvaart, V. J. Expósito Jiménez, "Systematic SOTIF-aligned approach to explore residual risk in autonomous driving systems," in *Book Of Proceedings - International Congress: SIA VISION 2024 - 16 and 17 OCTOBER 2024*, (Paris, France), pp. 185–192, Société des Ingénieurs de l'Automobile, 2024. [9] – Conference. **My Contribution**: The main contribution in this paper is on the safety validation toolchain section in which the integration of the SOTIF concept into the validation tool is described. Moreover, I have also contributed in the form of discussions, feedback for improvements, and corrections during the publication process.

Figure A.1 shows the contribution of each included publication in the thesis in the final research contributions of the theses. Publication 1 and Publication 2 set up the basis in the form of a state-of-the-art of the topic that has been used to develop the research work. Publication 3 defines the first version of the proposed workflow used in the proposed argumentation. Publication 4 described the process to parametrise the triggering conditions and integrate into

a validation tool. This process is also shown in Publication 6, where the paper is not only focused on the triggering conditions but in showing the complete validation suite. Finally, performance insufficiency injections and risk quantification are detailed in Publication 5.

A.1 Safety and Security Co-engineering for Highly Automated Vehicles

Elektrotechnik & Informationstechnik https://doi.org/10.1007/s00502-021-00934-w

ORIGINALARBEIT

Safety and security co-engineering for highly automated vehicles

C. Schwarzl, N. Marko, H. Martin, V. Expósito Jiménez, J. Castella Triginer, B. Winkler, R. Bramberger

Highly automated driving will have a great impact on people's social life, changing the way we perceive mobility and its actual meaning and how vehicle occupants act while traveling to their desired destinations. Future highly automated vehicles (HAVs) will have to be updated periodically to continuously improve them and to keep up with the enormous development speed of the entire automated driving (AD) ecosystem. The updating process as well as the high connectivity of HAVs lead to a high risk of cybersecurity attacks through all kinds of internal and external electrical interfaces. Through such attacks, information could be stolen or, even worse, the control over vehicles could be assumed. Hence, security directly influences safety of vehicles. Attacks must be mitigated during all stages of the vehicle's life cycle, including development, operation, maintenance, and disposal, to reduce security risks and, consequently, safety risks. Currently, there is no well-defined and officially accepted approach to combine safety and cybersecurity activities. Both the standards for functional safety and cybersecurity have to be met and taken into account accordingly during the (development) processes. In this paper, well-known safety and security methods in the automotive sector are summarized. Safety and cybersecurity co-analysis and co-design methods are outlined for the automotive sector with a focus on HAVs. Furthermore, these safety, cybersecurity, and co-engineering methods are evaluated in practice using a real vehicle and the first results are shown. The examined vehicle is the mobile test platform SPIDER. This platform enables the testing of components and vehicle functions in real-world situations and under harsh environmental conditions, which is a prerequisite to ensure safety.

Keywords: safety; security; co-engineering; systems engineering; highly automated driving

Integrierte Entwicklungsmethodik für funktionale Sicherheit und IT-Sicherheit für automatisierte Fahrfunktionen.

Hochautomatisiertes Fahren wird einen großen Einfluss auf das gesellschaftliche Leben des Menschen haben und die Art und Weise verändern, wie wir Mobilität und ihre tatsächliche Bedeutung wahrnehmen und wie sich die Fahrzeuginsassen während der Fahrt zu den gewünschten Zielen verhalten werden. Zukünftige hochautomatisierte Fahrzeuge (HAVs) müssen regelmäßig aktualisiert werden, um sie kontinuierlich zu verbessern und um mit der enormen Entwicklungsgeschwindigkeit des gesamten Automated Driving (AD)-Ökosystems Schritt zu halten. Der Aktualisierungsprozess sowie die hohe Konnektivität von HAVs führen zu einem hohen Risiko an Angriffen auf die Cybersicherheit über alle Arten von internen und externen elektrischen Schnittstellen. Durch solche Angriffe könnten Informationen gestohlen oder, noch schlimmer, die Kontrolle über Fahrzeuge übernommen werden. Die Cybersicherheit wirkt sich daher direkt auf die funktionale Sicherheit von Fahrzeugen aus. Angriffe müssen in allen Phasen des Fahrzeuglebenszyklus, einschließlich Entwicklung, Betrieb, Wartung und Entsorgung, abgeschwächt werden, um die Cybersicherheits- und damit die funktionalen Sicherheitsrisiken zu reduzieren. Derzeit fehlt ein klar definierter und offiziell akzeptierter Ansatz zur Kombination von funktionalen Sicherheits- und Cybersicherheitsaktivitäten. Sowohl die Standards für funktionale Sicherheit als auch für Cybersicherheit müssen erfüllt und entsprechend in den (Entwicklungs-) Prozessen berücksichtigt werden. In diesem Beitrag werden die im Automobilbereich bekannten Sicherheitsmethoden zusammengefasst. Co-Analyse- und Co-Design-Methoden für funktionale Sicherheit und Cybersicherheit werden für den Automobilbereich mit einem Schwerpunkt auf HAVs erläutert. Des Weiteren werden diese Methoden und implementierte Sicherheitsmaßnahmen praxisnah an einem realen Fahrzeug evaluiert und erste experimentelle Ergebnisse gezeigt. Das untersuchte Fahrzeug ist die mobile Testplattform SPIDER. Diese Plattform ermöglicht es, Komponenten und Fahrzeugfunktionen in realen Situationen und unter rauen Umgebungsbedingungen zu testen, was eine Voraussetzung ist, um Sicherheit zu gewährleisten.

Schlüsselwörter: funktionale Sicherheit; Cybersicherheit; integrierte Entwicklung; Entwicklungsmethodik; hochautomatisiertes Fahren

Received July 9, 2021, accepted September 8, 2021

1. Introduction

To make Automated Driving (AD) a success, economical aspects and customer needs have to be considered. Safety and quality of automated vehicles are of highest importance for buying decisions. Based on this information it is self-evident that Automated Driving functions shall be thoroughly tested during development and type approval, to achieve the required quality level and to increase customer´s trust. Testing of a complex AD function requires many test drives for achieving a similar accident probability for Highly Automated Vehicles (HAVs) in comparison to manual driving. Since this large amount of test driving is economically infeasible and practically nearly impossible, new virtual testing and type approval techniques are inevitable [1].

Schwarzl, Christian, Virtual Vehicle Research Center, Inffeldgasse 21a, 8010 Graz, Austria (E-mail: christian.schwarzl@v2c2.at); **Marko, Nadja,** Virtual Vehicle Research Center, Graz, Austria; **Martin, Helmut,** Virtual Vehicle Research Center, Graz, Austria; **Expósito Jiménez, Víctor,** Virtual Vehicle Research Center, Graz, Austria; **Castella Triginer, Joaquim,** Virtual Vehicle Research Center, Graz, Austria; **Winkler, Bernhard,** Virtual Vehicle Research Center, Graz, Austria; **Bramberger, Robert,** Virtual Vehicle Research Center, Graz, Austria

Fig. 1. Harsh environment conditions like rain and water on streets can reduce performance of perception system [© Schwarzl]

In addition to virtual testing, the rising vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) connectivity increases the importance of cybersecurity aspects within the development process. Automotive systems are developing from stand-alone systems towards systems of systems, interacting and coordinating with each other and influencing vehicle actions. While safety is already established in companies, security engineering (cybersecurity and security is synonymous in this paper) is a rather novel aspect in automotive. Know-how and best practices from other application domains, such as IT, must be adopted to fulfill the special requirements coming along with automated vehicles. Whereas in IT, security risks often cover the impact on privacy or confidentiality of data, the main important aspect in context of automated driving is the impact of security risks on safety. This makes security very important for the automotive domain as security risks can have high impact on road safety and vehicle operations.

Safety is one of the key factors to become automated driving reality. There are several challenges that have to be solved in the near future to provide the needed safety and reliability of HAVs.

The responsibility handover from the driver to the HAV increases the importance of cybersecurity aspects within the development process. One of the main reasons is that security threats can directly impact the functional safety if a vehicle vulnerability is used to negatively impact implemented safety features. In addition, data privacy must be ensured, which requires secure technical solutions not only within the vehicle network but also for all other connections (e.g., back-end systems operated by OEMs or consumer electronics like smartphones).

A technical solution which is safe and secure at the same time, must fulfil mutual exclusive requirements regarding maintenance. A functional safe system is designed in such a way, that it can deal with internal system faults, like hardware or software errors, and incorrect stimuli from the environment without putting humans at unreasonable risk. For this reason, a safe system should not be changed over lifetime, ensuring that the safety measures work as expected. To maintain security quite the contrary is the case. Regular software updates are needed to remove newly found system vulnerabilities and to react to latest attacks. Consequently, a secure system must be constantly adapted and changed (software updates and upgrades), which conflicts with the safety requirements regarding continuity [1].

One of the main challenges in testing HAVs is sensor testing as sensors provide the input for decisions of automated driving functions. Hence, a safe vehicle operation is dependent on the correct environmental perception of sensors. A difficult testing task is the coherent sensor stimulation, ensuring that all sensors detect the same environment. Especially in a simulated environment, in which the environment objects have to be transformed into coherent sensor inputs, this is challenging. To achieve coherency, the different sensors need to perceive the same objects with comparable precision, ensuring the plausibility of measurements produced by different sensors. The coherent sensor stimulation must also consider different weather conditions, which can heavily influence the performance of real sensors (cp. Fig. 1). A second big challenge in testing AD functions is the definition of the needed test scenarios, ensuring that all critical scenarios are sufficiently covered with a minimal number of tests. Critical scenarios have to cover any tests for safety functions and include security attacks.

The complexity of AD and its functions also pose severe challenges on the type-approval/homologation of such functions, which must be done before a car is permitted to be brought onto the market. The type-approval of a vehicle is performed by an independent party (e.g., TÜV), which checks the functional correctness of the vehicle by testing. Consequently, the type-approval/homologation procedure has to deal with the same challenges as stated above for testing and has to find general solutions applicable for all function variations coming from different OEMs (Original Equipment Manufacturer). The criticality of these testing activities is much higher than

during function development. The reason is that if type approval/homologation is passed, the vehicle is deemed safe by the public authority and therefore the vehicle must perform according to public regulations and should reflect public expectations [1].

Therefore, we need to integrate continuous development including safety and security co-engineering into homologation and certification processes that are (a) fit to their purpose; a system passing homologation should have the required safety, security, reliability, and other required qualities, (b) accepted by regulation authorities, i.e. homologation should be performed in a way that all safety arguments are understandable and verifiable by a third party, and (c) accepted by the end user of these systems, i.e., the public must be able to trust that these systems have the desired qualities [2].

Therefore, we need to change the way HAVs are developed and tested. Existing design and testing processes need to be extended to processes covering the whole life cycle of an HAV. These extended processes use data collected in the field for continuous engineering of HAV. Together with corresponding monitoring and update methods as well as in conjunction with systems properties like fail-awareness and fail-operationality, these processes allow for validation of safety cases that are much more appropriate to the dynamic nature of HAVs [2].

Currently, a well-defined and officially accepted approach to combine safety and security activities is missing. However, the dependencies and synergies of the different analysis methods are currently investigated in the scientific community and automotive industry to find an efficient task- and review-alignment minimizing the overall effort.

For the development of Automated Driving functions and vehicles, especially the international standards ISO 26262 [3], ISO PAS 21448 [4] and ISO/SAE 21434 [5] shall be considered, where ISO 26262 and ISO PAS 21448 focus on functional safety and ISO/SAE 21434 focuses on cybersecurity for road vehicles. An integrated safety and security co-engineering approach, which fulfils the requirements of all relevant standards and precisely defines the interactions between the tasks and the content in the produced work products is needed.

We call this integrated approach safety and security co-engineering. A safety and security co-engineering approach should include all engineering activities in the automotive system development life cycle. However, important activities are co-analysis of safety and security risks starting in the concept phase, co-design of safety and security measures by considering interdependencies and the resulting impact on the system as well as efficient test methods to reduce the number of test cases.

In this paper an overview of safety and security analysis and development methods are presented, which need to be harmonized and aligned to establish an effective and efficient co-engineering method. An integrated workflow for co-analysis and co-design methods is described and their application in a practical Automotive use case is shown.

2. Security/safety methods

The rising connectivity of vehicles makes security more important for the automotive domain as security risks can have high impact on road safety and vehicle operations. A well-known demonstration of these security risks was the hack of a Jeep Cherokee [6] where the attackers were able to influence braking, steering and acceleration. Hence, security can have high impact on safety. In the upcoming automotive security standard ISO/SAE FDIS 21434 [5] the financial, operational, privacy and safety impact has to be distinguished and analyzed in development. For the safety security co-engineering approach, mainly the impact on safety is relevant as a safe operation of HAVs is the premise for user acceptance.

Safety and cybersecurity should be built into the system rather than added on at the end of development. Hence, analysis of security and safety risks is of great importance already at the beginning of the development process (concept phase). Therefore, several analysis methods exist. For safety, well known methods defined in the ISO 26262 standard [3] are: HARA (Hazard Analysis and Risk Assessment), FMEA (Failure Mode and Effect Analysis), FTA (Fault Tree Analysis), and HAZOP (Hazard and Operability study). A further notable analysis method is STPA (Systems-Theoretic Process Analysis) [7, 8] which represents a safety hazard analysis technique based on a control model.

General safety design solutions are of course not suggested in the ISO standard as systems are diverse and solutions are variable and intellectual property of companies. However, safety design patterns can be used to apply already proven solutions [9], e.g., hardware redundancy, and model-based methods are helpful for deriving architectures and design by supporting traceability aspects and hence impact analyses.

For security, a major activity is the analysis of potential attacks. According to the security ISO standard, Threat Analysis and Risk Assessment (TARA) includes asset identification, threat scenario identification, impact rating, attack path analysis, attack feasibility rating, risk value determination, and risk treatment decision [5].

Security methods mentioned in the cybersecurity standard [5] are threat modelling approaches based on frameworks and projects such as EVITA[1] (E-safety vehicle intrusion protected applications), TVRA (Threat, Vulnerability and Risk Assessment), PASTA (Process for Attack Simulation and Threat Analysis) or STRIDE. EVITA classification separates different aspects of the consequences of security threats and is an outcome of a European research project. TVRA models the likelihood and impact of attacks. PASTA threat modelling combines an attacker perspective of a business with risk and impact analysis and their vulnerability to attack. STRIDE is a threat modelling approach and an acronym for spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privileges [10].

For attack path analysis, the ATA (attack tree analysis) identifies threats in a hierarchical manner and is adequate for exploiting combinations of threats (attack patterns). With an attack feasibility rating the probability of attacks is estimated. The rating is defined in the cybersecurity standard [5]. Finally, the risk is determined based on impact of attacks and the attack feasibility rating. From the data generated by TARA a cybersecurity concept with mitigating measures can be developed. In order to support the development of security measures, security patterns can be used to reuse already existing solutions. Security patterns can be seen as best practices to solve security problems in new scenarios. In the European project SECREDAS, security patterns are described and collected in the context of connected vehicles. The SECREDAS approach and some further existing security pattern catalogues/best practice collections are summarized in [11].

For safety and security co-engineering a combination of safety and security methods is needed. Currently, not many approaches are established. STPA-sec (STPA for security), which was originally developed for safety (STPA), has later been extended for security. According to STPA-sec, safety prevents losses due to unintentional

[1] https://www.evita-project.org/.

actions by benevolent actors, while security prevents losses due to intentional actions by malevolent actors. STPA-sec focuses on preventing losses regardless of actor intent and can be seen as an approach combining safety and security analyses [12].

SAHARA combines the automotive hazard analysis and risk assessment (HARA) with the security domain STRIDE approach to quantify impacts of security threats and safety hazards on system concepts at initial concept phase [10].

However, the integration of safety and security requires analysis of impact and tradeoff between safety and security requirements/measures. This tradeoff analysis has to be handled in an integrated approach which has to be developed for the automotive domain.

3. Safety and security co-engineering

Currently, automotive safety and security disciplines are not similarly mature (security is less mature than safety). While the ISO/SAE 21434 standard regarding automotive cybersecurity is now available in the first edition, for the established automotive functional safety standard ISO 26262, already edition 2 is available. Both documents note interaction points of functional safety and cybersecurity, but only in an informative way.

Safety and security co-engineering refers to engineering activities that consider both safety and security and their interactions in the development life cycle. In order to improve safety and security co-engineering methods, the safety and security aspect must be considered in a holistic way. This includes safety & security co-analysis, which refers to methods and techniques that can be used to identify safety hazards and security threats in a joint approach. Safety and security analyses are applied in the early stage of the development life cycle such as requirements and concept phase. However, the safety and security analyses and co-analysis do not end at this stage of development but must be part of the whole life cycle of HAVs. For example, to provide appropriate security measures, updates are necessary. These updates make further actions necessary. Impact analyses and if changes are required, further hazard, threat and risk analyses. Hence, co-engineering considers all phases of the life cycle, in which co-analysis is an integral part. Safety analysis does not end after development but has to be extended until the vehicle is taken off the market.

Security can have impact on several aspects. For the safety and security co-engineering the impact of security on safety is relevant. Hence, in addition to safety analysis and security analysis the correlation between both has to be analyzed. Moreover, co-engineering must cover design methods by considering trade-off of safety and security, shared testing, verification, and validation. In this paper, we summarize selected co-engineering methods to show examples for how to integrate security with safety.

Safety and security co-analysis In context of the automotive domain, existing safety analysis methods, standardized in ISO 26262 can be integrated with security analysis. Security/Safety methods such as STPA-Sec extends the analyses by including unintentional system faults with intentional system faults. In fact, we already proposed the integration of the STPA-sec method into the methods defined in the standards [13]. However, automotive standards take as a reference HARA for safety and TARA for security. Therefore, the automotive industry would take benefit from the integration of these two analysis methods.

The safety standard uses the HARA method to identify and categorize several types of hazardous events. The identified hazardous events determine the ASILs (Automotive Safety Integrity Levels), which is the basis for specifying high level safety requirements (safety goals) that prevent or mitigate the associated hazards. The security standard proposes TARA to identify and assess security threats. Based on the assessment of security threats, the risk level is determined, and counter measures are selected to mitigate the associated vulnerabilities. The analysis methods have parallel activities but with several differences. Hazardous events are a selection of relevant hazardous scenarios resulting from combining effects of the situation analysis and the hazard identification. Threat assessment is the result of combining vulnerabilities with damage scenarios and threat scenarios. In both cases, we evaluate scenarios, however, safety analyses malfunctions of the item (unintentional system faults) whereas security analyses vulnerabilities of the item which can be exploited by a hacker for example (intentional system faults). Further, the risk evaluation and classification differ for safety and security. However, the security risk assessment evaluates the safety impact level depending on the results of the safety analyses.

The proposed safety and security co-analysis method integrates the HARA and TARA at two points: a) assigning the relations between item malfunctions and item vulnerabilities and b) use the safety risk assessment for the impact rating in the security risk assessment. The flow diagram shown in Fig. 2 shows the integrated activities S1 and S2:

(S1.) *Assign relations.* This step consists of two parts. The first part assigns relations between functions of the item and the defined security assets. The analysis of the safety attributes and the security properties supports finding the relations. The second part uses the hazardous scenarios from the HARA and integrates them into the damage scenarios from the TARA to take advantage of the situation analysis already carried out by safety analyses.

(S2.) *Integrate Safety Classification.* This step relates the severity classification of hazardous events from the safety analyses to the safety impact category of the damage scenarios from the security analyses. In this activity, safety analysis is used to improve the impact rating of security threats regarding safety.

Safety and security co-design The safety domain has collected valuable experience in using model-based approaches for safety analysis and integration of safety engineering with systems engineering by using a shared model as common viewpoint. For co-engineering this is extended towards security so that model-based engineering can support system, safety, and security engineering. For safety and security co-design in the automotive domain, we propose a systematic pattern-based and ISO 26262-oriented approach. Patterns, inspired by software design patterns [14], are used to collect and organize solutions for similar problems with a general and universal solution. A well-known and proven solution for a specific problem is generalized so that it can be reused for similar recurring problems in other projects. The benefits of patterns are conservation and reuse of design knowledge, best practices, and tested solutions.

Security patterns can be seen as the essence of sound security designs and best practices that can be used to solve security problems in new scenarios. During the security engineering process, security patterns can be used in requirements analysis and design to eliminate security flaws and provide additional information for security validation.

For safety and security co-engineering, we design the security pattern to follow the same approach as the safety pattern, adopt the generic template and describe the security content as well as additional information regarding potential safety implications.

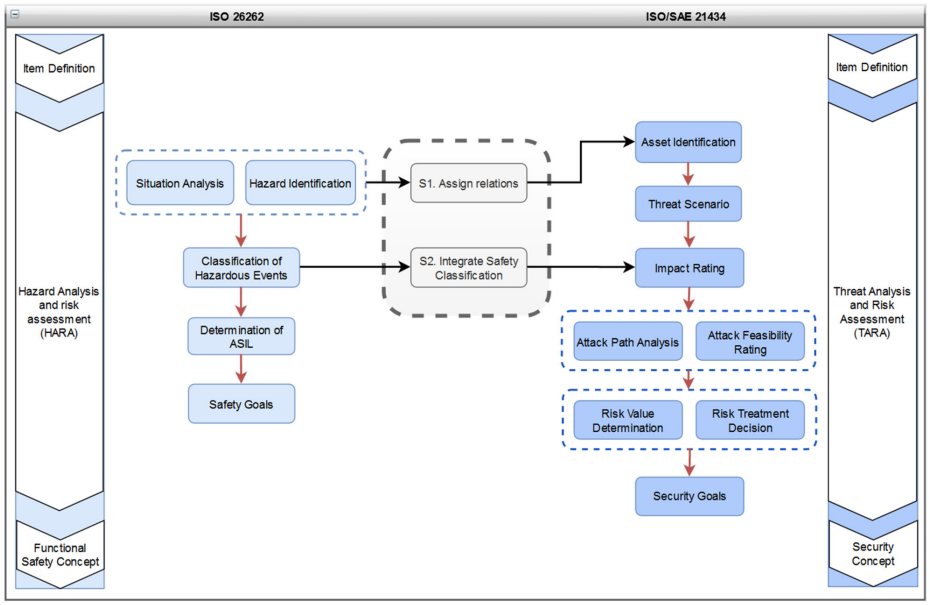


Fig. 2. Safety and security co-analysis activities

A generic combined pattern-based engineering workflow is proposed to guide engineers selecting and applying safety and security patterns to develop safe and secure systems/products. This workflow is meant to be used in an integrated way with the usual safety and security engineering approaches. It therefore does not substitute established approaches but enhances them with further tasks.

The approach explicitly focuses on security for safety (i.e., safety concerns are the main engineering drivers), i.e., security should start working when the first safety architecture is almost defined.

The workflow is divided into three main phases happening one after the other in a waterfall fashion: Safety Pattern Engineering, Security Pattern Engineering, and Safety and Security Co-Engineering Loop. In the first two phases specific teams are involved: Safety Pattern Engineering is performed by systems and safety engineers, and Security Pattern Engineering is performed by systems and security engineers. We introduce the Safety and Security Co- Engineering Loop, the third phase of the workflow, as security measures can influence system properties that can alter safety. In this phase all involved disciplines (systems, safety, and security engineers) work together in a joint approach.

More details regarding this approach can be found in [15]. In this paper, we just describe the safety and security co-engineering loop. In the co-engineering loop iterative activities have been defined, which are summarized in Fig. 3. The flow diagram shown in this figure includes the following tasks (Sx... Step x, D... Decision):

(S1.) *Safety Impact Analysis.* After successful instantiation of the security pattern in the architecture, a safety impact analysis must be performed. This analysis should check what happens if the introduced security measures may have a safety impact (a complete safety analysis of failure modes of new/modified elements as well as an impact analysis of existing elements must be done).

(S1.D) *Safety Impact Decision.* Is there an impact of security on safety?

(S1.D.a) *NO.* Safe and Secure Concept available

(S1.D.b) *YES.* Continue with Co-Analysis

(S2.) *Design Co-Analysis.* Perform co-analysis for safety and security aspects by methods, such as SAHARA or FMVEA to identify new failure modes (safety) or new vulnerabilities (security).

(S3.) *Harmonization.* Joint engineering team elaborates and defines safety and security measures based on available patterns if possible.

(S3.D) *Harmonized Decision.* Is harmonization successful without any side-effect between safety and security measures?

(S3.D.a) *YES.* A safe and secure concept has been elaborated, in which all safety and security requirements are covered without any contradiction ("freedom of interference"). Continue with Argumentation.

(S3.D.b) *NO.* Proposals for different solutions are available. Continue with Trade-Off.

(S4.) *Trade-Off.* A joint engineering team must prioritize which aspect is more relevant and a trade-off analysis between safety and security measures has to be done. The trade-off solution can again be supported by safety and security patterns, where the patterns can differ from the previously applied ones to fulfill the safety and security goals.

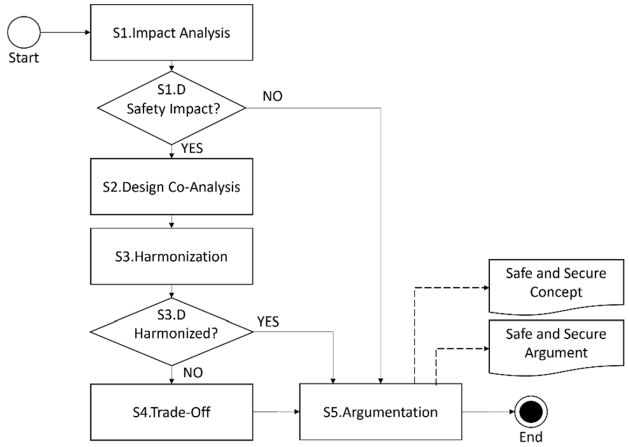


Fig. 3. Flow chart of co-engineering loop [15]

(S5.) *Argumentation*. The safe and secure concept is available and in this task all rationales are collected to compile the final safe and secure argument of the elaborated concept, that provides a piece to the Assurance Case.

The argumentation demonstrates that the system under consideration contains no unreasonable risk and consolidates functional safety and security. All safety and security requirements have to be fulfilled. The generated documents, argumentation, and safety and security case support certification which can be helpful for homologation.

4. Use case

The described safety and security co-engineering methods are applied for developing safety and security measures for a mobile Hardware-in-the-Loop (HiL) platform. This mobile platform is called Smart Physlcal Demonstration and Evaluation Robot (SPIDER) and has been built by Virtual Vehicle (cp. Fig. 4). The SPIDER allows the reproducible execution of test scenarios on proving grounds and with a limited speed of 10 km/h on public roads. Due to its modular system design, it is possible to test individual components, electronic control units and even whole vehicle functions in real world situations and harsh environment conditions.

The constructed robot chassis consists of a robust aluminum frame with a wheelbase of 1.2 m. Each of the four wheels comes with an electrical drive as well as a steering engine. Thus, the platform can change its heading while remaining in the same location. The SPIDER is equipped with multiple electronic components. The chassis holds enough space for the battery, drive-controllers, and the processing unit in the waterproof interior of the robot. The robot is equipped with an on-board-unit (OBU, Cohda Wireless) for V2X communication, a WiFi antenna to communicate with the on-board processing unit and a GPS module for coarse location estimation. Additionally, the robot has a mounting point for an environmental perception platform on top of the robot. For safety reasons, the robot is also equipped with an emergency-stop button on its rear end, which can stop all operations and apply the brakes. Additionally, the platform can be stopped manually with a

Fig. 4. SPIDER on a proving ground equipped with a sensor-box, which is an optional module used for integration of additional sensors [© Virtual Vehicle]

remote emergency break connected via an active-on connection. The platform provides two different control functions: automatic control and manual control. In manual mode the platform is controlled using an operator panel (e.g., a video game controller). In automatic mode, the platform can drive a pre-defined path. The robot calculates the resulting controls and sends them to the actuators. For the high-level control, the robot is equipped with an x86-based processing unit running Linux and ROS. The computer is located in the inside of the waterproof chassis and is powered by the robot's battery. The computer has to handle requests from the operator panel, from the sensors and to perform the respective processing. The processing unit is capable of logging the sensor data of the robot via the ROS tool rosbag. For low-level control, the robot is equipped with an automotive state-of-the-art, safety-focused microcontroller (Infineon's AURIX). The microcontroller communicates with the motor/steering controller of the four wheels and handles the battery/charging management. Further information to SPIDER can be found in [16].

Table 1. Extract of the HARA analysis for the collision avoidance function of SPIDER

Hazard	Malfunction	Situation	S	Argumentation for S	E	Argumentation for E	C	Argumentation for C	ASIL
Unintendend omnidirectional movement lead to collison with objects or persons	AGV accelerate inverse	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: No special Maneuver	S2	Serious injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL C
Unintendend omnidirectional movement lead to collison with objects or persons	AGV accelerate inverse	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Following and Curve	S2	Serious injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL C
Unintendend omnidirectional movement lead to collison with objects or persons	AGV accelerate inverse	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Overtaking of other vehicle	S2	Serious injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL C
Unintendend stop leads to collision with moving object behind	False object dectection	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: No special Maneuver	S1	Light and moderated injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL B
Unintendend stop leads to collision with moving object behind	False object dectection	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Following and Curve	S1	Light and moderated injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL B
Unintendend stop leads to collision with moving object behind	False object dectection	- DM: Drive (Fully automated) - LOC: City - EC: Normal environment - WC: No special weather - MA: Overtaking of other vehicle	S1	Light and moderated injuries	E4	> 10% of average operation time	C3	Less than 90 % of all drivers or other traffic participants are usually able, or barely able, to avoid harm	ASIL B

The SPIDER uses several interfaces to communicate with its environment and the user. Since most of these interfaces rely on wireless communication, the implementation of security measures is necessary to ensure safety. Given its weight of 380 kg and its maximum speed of 50 km/h, the SPIDER can endanger the testing crew and other test equipment if control is lost.

Safety and security co-engineering for collision avoidance function The collision avoidance function (CAF) is a safety critical function. Malfunctioning or intentional disruption may have hazardous consequences. Moreover, the function has to fulfill mutual dependent requirements:

The collision avoidance function needs to be configurable via an operator panel. The operator panel is a remote control for SPIDER that is using a WLAN connection for communication. This connection is essential and could be vulnerable to attacks. If an attack is successful, the attacker can control SPIDER and cause harm. Incorrect configuration or tampering with the function itself can lead to hazardous, and in the worst-case, to fatal events. Further, the operator panel function is obligatory for operation. If it is not working, the SPIDER cannot be operated.

Consequentially, the following high-level safety and security requirements should be fulfilled:

- All publicly accessible interfaces of the SPIDER should be secured.
- The configuration of the collision avoidance function should enable a safe operation of SPIDER.

Safety and security co-analysis for collision avoidance For the complete analysis of the collision avoidance function, first safety analyses, second security analyses, and third co-analysis have to be performed, which is described in the following paragraphs.

Safety analyses are comprehensive and it's not the focus of this paper to describe a complete safety analysis. In this section an example hazard from the HARA is described to show how critical a malfunction in the collision avoidance could be. The typical driving scenario for SPIDER is on a proving ground where different kinds of tests for driving functions or sensors are executed. Before executing the tests, operators are preparing the surrounding objects needed for the planned test. In this situation, the SPIDER is stopped, but the collision avoidance function is always activated even if the velocity of the robot is zero. This is a safety feature to avoid the possibility of crashing into an object or hitting a person. By design, the collision avoidance function sends a signal to activate the emergency brake in case an object is detected. However, the function cannot activate the emergency brake on its own. A malfunction in the low-level controller could unintentionally start the SPIDER and if at the same time the emergency brake does not work correctly, a collision could be the result. A malfunction of the emergency brake is assumed in the analyzed scenario. Therefore, a hazard has to be considered even if the collision avoidance function detects all obstacles correctly and works properly. In the HARA, the described situation could lead to the specified hazard "unintended omnidirectional movement lead to a collision with objects or persons". According to our analysis and the specifications given by the standard ISO 26262 Part 9,[2] the hazard was defined as ASIL C, which is the second most critical safety level. Following the standard, ASIL C was defined by the combination of severity (S), controllability (C), and exposure (E). The severity was defined as S2, which means more than 10% probability of Abbreviated Injury Scale (AIS) 3-6.[3] It applies for front or side collisions with low speed (defined as less than 15 kph). Controllability was defined as the maximum value C4 since, in this situation, most of the participants are not able to avoid harm. Finally, this hazard can occur at all the time on the proving ground, therefore, it was set to E4. The same methodology was followed for the other analyzed

[2]ISO 26262-9, Road vehicles—Functional safety—Part 9: Automotive safety integrity level (ASIL)-oriented and safety-oriented analyses https://www.iso.org/standard/68391.html.

[3]Abbreviated Injury Scale (AIS) https://www.aaam.org/abbreviated-injury-scale-ais/.

Table 2. Extract of TARA analysis for collision avoidance function of SPIDER

Asset	Damage Scenario	CIA affected	Safety Impact	Financial Impact	Operational Impact	Privacy Damage	Impact Rating	Threat Scenario
Collision avoidance function	Unintended stop due to functionality loss.	Availability	Moderate	Moderate	Major	Negligible	Major	collission avoidance function Denial of Service: collision avoidance function is stopping due to unintended input
	Unintended crash due to manipulated functionality.	Integrity	Major	Major	Major	Negligible	Major	Tampering: changing code of collision avoidance function
Collision avoidance function configuration	Unintended stop due to manipulated configuration.	Integrity	Moderate	Moderate	Major	Negligible	Major	Denial of Service: CAF is stopping due to unintended/missing configuration file
	Unintended crash due to manipulated configuration.	Integrity	Major	Major	Major	Negligible	Major	Tampering: changing configuration of collision avoidance function

hazards such as the hazard "unintended stop due to functionality loss". This may occur when the function perceives an object that does not exist in reality, for example, due to light reflections of the lidar sensor. The examples above only show some of the analysed hazards that have been covered to make the function passing the acceptable safety criteria.

Table 1 shows an extract of the performed safety analysis for the collision avoidance function, in which all potential hazards and situations are analyzed applying the described methodology.

Security Analysis—Automotive threats range from reputation and brand damage to hazardous safety consequences. For the co-engineering the security threats and vulnerabilities which have impact on safety are relevant.

Security analyses for an item start at the concept phase and should result in a work product providing the threat analysis and risk assessment (TARA) (cp. ISO SAE 21434). The main outcome is the risk determination which is the basis for developing security goals and the security concept. For the risk assessment, the following activities have to be done: asset identification, threat scenario identification, impact rating (including safety impact), attack path analysis, attack feasibility rating, risk determination and risk treatment decisions.

The collision avoidance function itself has two main damage scenarios (cp. Table 2): 1) unintended stop, which affects the availability of SPIDER, and with regard to safety the more severe, 2) unintended crash, which affects the integrity of SPIDER. For the unintended stop, the safety impact is moderate as on the proving ground no serious injuries will be expected. However, there is an operational impact as SPIDER cannot be used, at least for a while. For the unintended crash, the safety impact is major which means that severe and life-threatening injuries can happen. Depending on the attack feasibility, the risk is determined, and the risk treatment has to be chosen.

For SPIDER, the security measures firewall, encryption of configuration on hard disc, and encryption of the communication between operator panel and SPIDER including authentication have been chosen. With the firewall unauthorized access to SPIDER is prohibited and with the encryption of the communication between operator panel and SPIDER, tampering the data is hindered, at least via the network which seems the easiest attack path.

Safety and Security Co-Analysis optimizes the individual methods and provides a more integrated view. As presented in Sect. 3, the first step (S1.) of the co-analysis proposes two main activities. The first one interconnects the safety function collision avoidance with the security assets. The second one integrates the hazardous malfunctions such as "unintended omnidirectional movement" together with multiple scenarios into the damage scenario. The second step (S2.) integrates the severity classification of the safety analysis into the security safety impact as it is summarized in Table 3. As a result, we can see that the consequences of malfunctions (unintentional) and damage scenarios (intentional) are the same, but the causes differ.

Safety and security co-design *Design Patterns*—Based on co-analysis results obtained in the concept phase of SPIDER development, a set of measures have been implemented. These measures follow the best practices defined in the safety patterns and can for example consist of redundancies such as an on-vehicle and remote controlled emergency button.

An important safety function is the emergency brake, which is triggered in case an obstacle comes too close to the SPIDER. The emergency brake is activated by the signal sent from the collision avoidance function if an object is detected in the danger zone of SPIDER (cp. Fig. 5). Having different safety zones is an applied safety pattern of SPIDER as well. In case an object is detected in the safety zone, the distance to this object is calculated and sent to the speed limiter function to reduce the velocity and avoid entering inside the danger zone.

Additionally, some first security patterns have been implemented to secure the SPIDER architecture shown in Fig. 6. These security patterns are:

- *Firewall* at high level control for all public interfaces restricting unauthorized access.
- *Encryption* between user and high-level control. The communication between these two participants is encrypted via a session key, exchanged using public key cryptography.
- *Plausibility checks* on low-level control mitigating attacks on vehicle control.

For the co-design, the included security patterns firewall and encryption are analyzed in this paper. The main goal of the analysis is to find out whether the introduced security measures (patterns) influence the safety of SPIDER. Therefore, the activities of the co-engineering loop (cp. Fig. 3) have been applied. The security measures for the CAF have no direct impact on SPIDER´s safety. Hence, not the whole workflow must be applied.

Step 1 (S1): Impact analysis. The impact of the security patterns firewall and encryption on safety are analyzed. For this the architecture

Table 3. Extract of safety and security co-analysis for collision avoidance function of SPIDER

Funtions / Assets	Hazardous Malfuntions	Scenario Situation	Damage Scenario	Severity / Safety impact
Collision avoidance function	Unintendend stop leads to a collision with moving objects behind	Scenarios considered: - No special maneuver - Following and curve - Overtaking	Unintended stop due to manipulated functionality leads to a collision with objects behind	S1 / Moderate
Collision avoidance function	Unintended omnidirectional movement lead to a collision with objects or persons due to acceleration inverse	Scenarios considered: - No special maneuver - Following and curve - Overtaking	Unintended omnidirectional movement due to manipulated functionality leads to a collision with objects or persons	S2 / Major
Collision avoidance function configuration	Unintendend stop leads to a collision with moving objects behind	Scenarios considered: - No special maneuver - Following and curve - Overtaking	Unintended stop due to manipulated configuration leads to a collision with objects behind	S1 / Moderate
Collision avoidance function configuration	Unintended omnidirectional movement lead to a collision with objects or persons due to acceleration inverse	Scenarios considered: - No special maneuver - Following and curve - Overtaking	Unintended omnidirectional movement due to manipulated configuration leads to a collision with objects or persons	S2 / Major

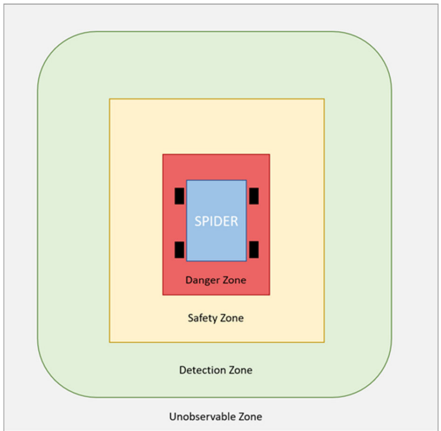


Fig. 5. Safety zones of the collision avoidance function, where in the safety zone the speed is reduced and in the danger zone an emergency brake is initiated

with the included patterns is analyzed. For the configuration of the collision avoidance function there are no additional failure modes and hence no safety impact. The failure modes *CAF not running* and *CAF running with false configuration* have already been part of the safety analysis (HARA). The failure mode of the firewall *configuration is wrongly blocked* leads to *CAF not running* and *wrong configuration is sent by the firewall* leads to *CAF running with false configuration*. Failure modes of the encryption also result in already existing failure modes. *Configuration cannot be read due to erroneous* encryption leads to *CAF not running* and *configuration file is corrupt due to incorrect encryption* leads to *CAF not running*. If CAF is running, there are no safety issues. If a problem is detected, the SPIDER is not operational, and no harm can be caused. The transmission of the configuration is not time critical; hence, additional software systems that probably slow down the communication do impact the performance but not safety.

However, if there is a problem with the CAF configuration, SPIDER is not available and operational and financial losses are the consequence. Therefore, the workflow has to be extended to include SOTIF [4].

Step 2 (S2): Design co-analysis, Step 3 (S3): Harmonization and Step 4 (S4) Trade-off are not applicable for the analyzed design patterns as there is no impact on safety.

Step 5: Argumentation (*Under development*). The Assurance Case provides a structured argument supported by a body of evidence, which in turn allows a valid assumption to be made, that a product is safe and secure for a particular use in a defined environment, in a convincing and complete manner. For such an argument structure the Goal Structuring Notation (GSN) [17] is used, which is a well-established graphical argumentation notation which is widely adopted within safety-critical domains for the presentation of safety arguments within safety cases (aviation, military, industry, healthcare and now also in the automotive). The OMG (Object Management Group) released a meta-model in the document "Structured Assurance Case Metamodel (SACM)" [18, 19], where SACM is developed based on concepts of GSN. First results of the argumentation part of the SPIDER use case are available, but the final results will be presented in a follow-up publication, because the whole concept of Assurance Case need to be understood and that goes beyond the scope of this publication at hand.

5. Conclusion and outlook

Safety and security co-engineering are of growing importance and has got increased attention of the automotive industry the last years, since it is known that the new cybersecurity standard ISO/SAE 21434 is obligatory. A main goal of safety and security co-engineering is to reduce effort by applying integrated approaches in which methods are combined. Further benefits of an integrated approach are that interdependencies of safety and security can be found easier, and the safety and security measures are coordinated. However, currently an overall integrated approach is still missing. Safety and security standards only provide notes where in the product life cycle coordination points exist. There are already some approaches targeting integration, whereas some of them are discussed in this paper.

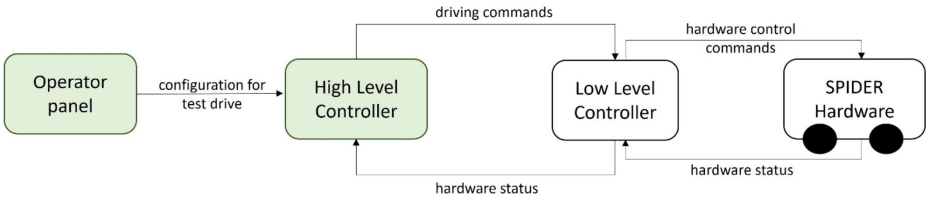


Fig. 6. Abstracted hardware architecture of SPIDER

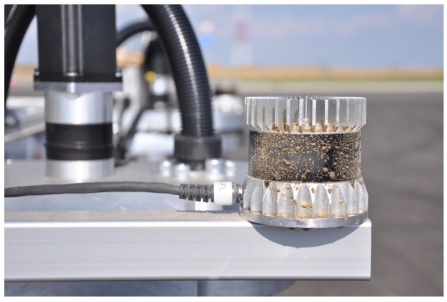


Fig. 7. LiDAR sensor covered with dirt to simulate reduced sensing performance

However, these methods are only some extracts and should be extended and combined to provide an overall integrated approach. We already applied safety and security co-engineering approaches at our mobile Hardware-in-the-Loop (HiL) platform SPIDER. SPIDER is safety and security relevant and is used to perform reproducible scenario-based tests on proving grounds. Due to its implemented safety mechanisms the safety of the user, other test engineers and the test equipment is ensured during the test execution. The SPIDER relies on several wireless interfaces, which can be easily attacked and lead to unauthorized access to the system or data manipulation, which can render the safety measures useless. Hence, SPIDER is an optimal platform to verify co-engineering approaches for HAVs functions. First experiences showed that safety and security co-engineering is helpful to reduce effort and costs and is absolutely needed to have a holistic view. However, existing methods and processes have to be adapted to be more efficient and applicable in the automotive context, e.g., argumentation of safety and security.

To have an even more holistic view it is also necessary to include methods specified in the SOTIF standard [4] which also considers the operational safety. In this standard, a major part are verification and validation strategies which will be analyzed and applied for SPIDER as a next step. For example, to test the robustness of the collision avoidance function, the sensor is covered with dirt (cp. Fig. 7), which reduces the perception capabilities. In case of reduced sensor performance, the collision avoidance function must react by triggering the emergency brake. The reduction of perception capabilities can be a safety (unintentional) or security (intentional) issue and can hence be part of both safety and security verification.

Acknowledgements

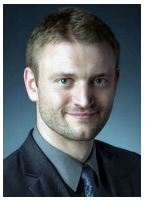
Virtual Vehicle Research GmbH has received funding within COMET Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action, the Austrian Federal Ministry for Digital and Economic Affairs, the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorized for the program management.

References

1. Schrammel, B., Schwarzl, C. (2018): Highly Automated Driving—The new challenges for Functional Safety and Cyber Security. White paper. https://www.v2c2.at/wp-content/uploads/2018/11/tuv-austria-white-paper-iv-highly-automated-driving_web.pdf.
2. Marko, N., Möhlmann, E., Nickovic, D., Niehaus, J., Priller, P., Rooker, M. (2020): Challenges of engineering safe and secure highly automated vehicles. White paper. arXiv: 2103.03544.
3. ISO (2018): ISO 26262 Road vehicles – functional safety.
4. ISO (2019): ISO/PAS 21448 Road vehicles—safety of the intended functionality.
5. ISO (2021): ISO/SAE FDIS 21434 Road vehicles—cybersecurity engineering.
6. Miller, C., Valasek, C. (2013): Adventures in automotive networks and control units. In DEF CON 21 hacking conference.
7. Leveson, N. (2004): A new accident model for engineering safer systems. In Safety science (Vol. 42, pp. 237–270). https://doi.org/10.1016/S0925-7535(03)00047-X.
8. Leveson, N., Thomas, J. P. (2018): STPA handbook. https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.
9. Preschern, C., Kajtazovic, N., Kreiner, C. (2015): Building a safety architecture pattern system. In Proceedings of the 18th European conference on pattern languages of program, EuroPLoP '13, New York (pp. 1–55). https://doi.org/10.1145/2739011.2739028.
10. Macher, G., Schmittner, C., Armengaud, E., Ma, Z., Kreiner, Ch., Martin, H., Brenner, E., Krammer, M. (2017): Integration of security in the development life cycle of dependable automotive CPS. In Solutions for cyber-physical systems (pp. 383–423). https://doi.org/10.4018/978-1-5225-2845-6.ch015.
11. Marko, N., Vasenev, A., Striecks, C. (2020): Collecting and classifying security and privacy design patterns for connected vehicles: SECREDAS approach. In Computer safety, reliability, and security. SAFECOMP 2020 workshops (pp. 36–53). https://doi.org/10.1007/978-3-030-55583-2_3.
12. Young, W., Leveson, N. G. (2014): An integrated approach to safety and security based on systems theory. Commun. ACM, 57(2), 31–35. https://doi.org/10.1145/2556938.
13. Triginer, J. C., Martin, H., Winkler, B., Marko, N. (2020): Integration of safety and cybersecurity analysis through combination of systems and reliability theory methods. In Embedded real-time systems.
14. Gamma, E., Helm, R., Johnson, R., Vlissides, J. (1995): Design patterns: elements of reusable object-oriented software. Boston: Addison-Wesley Longman Publishing Co.
15. Martin, H., Ma, Z., Schmittner, Ch., Winkler, B., Krammer, M., Schneider, D., Amorim, T., Macher, G., Kreiner, Ch. (2020): Combined automotive safety and security pattern engineering approach. In Reliability engineering and system safety (Vol. 198). https://doi.org/10.1016/j.ress.2019.106773.
16. Jiménez, V. J. E., Schwarzl, C., Martin, H. (2019): Evaluation of an indoor localization system for a mobile robot. In IEEE international conference on connected vehicles and expo (ICCVE) (pp. 1–5). https://doi.org/10.1109/ICCVE45908.2019.8965234.

17. Safety-Critical Systems Club (SCSC)—The Assurance Case Working Group (ACWG). GSN Community Standard. Version 2. January 2018 (SCSC-141B). Online. https://scsc.uk/gsn.
18. OMG (2021): Structured assurance case metamodel. https://www.omg.org/spec/SACM/About-SACM/.
19. Wei, R., Kelly, T. P., Dai, X., Zhao, S., Hawkins, R. (2019): Model based system assurance using the structured assurance case metamodel. J. Syst. Softw., 154, 211–233.

Authors

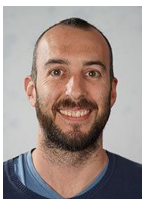


Christian Schwarzl
began his career after finishing the Information and Computer Engineering studies as researcher at Virtual Vehicle Research Center in Graz. He started his computer sciences dissertation in parallel, which he finished with honours in 2012. He became Lead Researcher in 2013 and managed the activities in the field of verification and validation. Since 2014 he is head of the Dependable Systems Group, which primarily focusses on the verification and validation of autonomous vehicles, functional safety and security. In February 2019 he has additionally taken the Position of Director Testing and Validation at ALP.Lab GmbH, where he is responsible for the development of test methods and process for automated driving functions. Since 2016 he is active member of the ISO/TC22/SC 32/WG 8 committee for functional safety and contributes to standards like ISO 26262 and ISO PAS 21448 – Safety of the intended functionality.

Nadja Marko
studied computer sciences at Alpen-Adria-Universität Klagenfurt where she received her master's degree. After her study, she worked as researcher and project coordinator in several national and European research projects at Virtual Vehicle. At the beginning her main research topics were model-based systems engineering with focus on requirements engineering methods and safety. In the last years, she extended her research fields with simulation-based verification, safety, and security of automated vehicles.

Helmut Martin
received his master's degree in electrical engineering from Graz University of Technology in 2004. After his studies he has been working in the automotive industry in the software development and as functional safety engineer/manager for automotive system development for six years. By 2011, he started as researcher and project coordinator at the Virtual Vehicle Research GmbH and worked in the EU Projects on the research topics of functional safety engineering for the automotive domain according to ISO 26262 and model-based embedded systems engineering. Today, he investigates further on safety and security concerns on different Automated Driving projects. In 2012, Mr. Martin founded the Functional Safety Community (FuSaCom), which address Functional Safety topics for different domains. In regular workshops, on different safety-related topics, the experience exchange between industry and research is fostered. Since December 2016 he has been an active member of ISO/TC 22/SC 32/WG 8 regarding functional safety and safety of the intended functionality (SOTIF) and ISO/TC 22/SC 32/WG 11 Cybersecurity for road vehicles. In 2019, Mr. Martin received his PhD from Graz University of Technology regarding functional safety and model-based safety engineering.

Víctor Expósito Jiménez
studied telecommunication engineering at the University of Seville in Spain and Graz University of Technology in Austria. Since 2018, he works at Virtual Vehicle Research GmbH in Graz as senior researcher position in which he has worked on numerous Austrian and European projects. In this time, the scope of his work has covered the development of automated driving functions, including its deployment, test, and validation. Previously, he worked at Joanneum Research Forschungsgesellschaft mbH for six years in different Austrian projects which included topics such as Internet of Things, Radio Frequency Identification (RFID), and sensor integration.



Joaquim Castella Triginer
is a researcher of the Dependable Systems group at Virtual Vehicle. He studied Computer Systems and Industrial Engineering at the Polytechnic University of Catalonia and Automotive Engineering at FH Joanneum Graz. Generally fascinated with the automotive industry and everything related to its technological advances in the electric and electronic (E/E) field of the vehicle, his main research interests include functional safety and cybersecurity, and systems architecture.

Bernhard Winkler
is senior researcher at the Virtual Vehicle Research GmbH. He received his M.Sc. degree in electrical engineering from Graz University of Technology, Austria. He worked in the automotive industry, especially in vehicle safety and testing. In the end of 2011 he joined the safety team at the Virtual Vehicle. His main research topics are model-based systems and functional safety engineering and safety, security and SOTIF analysis for ADAS with focus on concept and system level.

Robert Bramberger
is researcher at Virtual Vehicle Research GmbH. His research area is development, improvement and application of methods and processes to ensure functional safety of automated driving functions.

A.2 State of the Art Study of the Safety Argumentation Frameworks for Automated Driving System

State of the Art Study of the Safety Argumentation Frameworks for Automated Driving System

Ilona Cieslik[1], Víctor J. Expósito Jiménez[1], Helmut Martin[1(✉)], Heiko Scharke[2], and Hannes Schneider[2]

[1] Virtual Vehicle Research GmbH, 8010 Graz, Austria
{ilona.cieslik,victor.expositojimenez,helmut.martin}@v2c2.at
[2] AVL List GmbH, 8020 Graz, Austria
{heiko.scharke,hannes.schneider}@avl.com

Abstract. The automotive industry is experiencing a transition from assisted to highly automated driving. New concepts for validation of Automated Driving System (ADS) include amongst other a shift from a "technology based" approach to a "scenario based" assessment. The safety validation and vehicle type approval process of ADS are seen as the biggest challenges for the automotive stakeholders today. Considering a variety of existing white papers, standardization activities and regulatory approaches, manufacturers still struggle with selecting best practices that stay aligned with their Safety Management System and Safety Culture. A step forward would be to implement a harmonized and global safety assurance scheme that is compliant with relevant regulations, standards, and reflects locally accepted behavioural laws. This will ensure a common understanding of the safety and build needed trust around ADS.

Today many communities (regulatory bodies, local authorities, industrial stakeholders, and academia) work on proof-of-concept framework for the Safety Argumentation as an answer to this problem. Unfortunately, there is still no consensus on one definitive methodology and a set of safety metrics to measure ADS safety. An objective of this summary paper is to fill existing gaps in the literature reviews, concerning available methods and approaches for engineering frameworks, processes of scenario-based evaluation and a vendor- and technology-neutral Safety Argumentation approaches and tools. A particular focus is placed on safety metrics and emerging quantitative approaches.

Keywords: Safety metrics · Safety Argumentation · Automated Driving System (ADS) · Statistical approaches · Safety of the Intended Function (SOTIF)

1 The Need for a Safety Argumentation Framework

The automotive industry is experiencing a transition from assisted to highly automated driving. The regulatory Framework Document on Automated/Autonomous Vehicles under the United Nations Economic Commission of Europe (UNECE) WP.29 underlines that the certification process of the Automated Driving System (ADS) should prove

M. Trapp et al. (Eds.): SAFECOMP 2022 Workshops, LNCS 13415, pp. 178–191, 2022.
https://doi.org/10.1007/978-3-031-14862-0_14

that "a vehicle shall not cause any non-tolerable risk" [1]. Therefore, one of the biggest challenges to be solved before the release of ADS (i.e., Level 3 and higher according to ISO/SAE 22736 [2]) is their safety validation. One of the research questions is: "How can we argue the absence of unreasonable risk of ADS in its defined Operational Design domain (ODD)", without knowing an exact interpretation of "reasonable foreseeable", "preventable" or "tolerable" accidents?

The concepts of "WHAT" to audit and assess when introducing ADS on the public roads reached mature consensus [3, 4], comparing to the proof-of-concept frameworks for a Safety Argumentation (i.e., provisions on "HOW" to rational that the ADS is safe enough). Even though many industrial initiatives and research groups are active in this field, the latter still opens enormous room for stakeholders' discussions. Currently there is no agreement on one definitive methodology with qualitative and quantitative metrics to measure safety, related risks, and social acceptance.

The automotive community agrees that the introduction of ADS requires new assessment and test methods. New approaches for certification and standardization are arising to cover the growing needs of such complex systems. New concepts for validation of ADS include amongst other a shift from a "technology based" approach to a "scenario based" assessment that would allow various types of interventions. Scenario-based testing is already a known and established test approach within the industry for development, certification, and rating purposes [5] but needs to be adopted to the ADS validation purposes to foster public trust in this disruptive technology.

Finally, a step forward would be to implement a safety assurance framework that is embedded in the Safety Management System and aligned with the Safety Culture of the manufacturer. It will help not only to organize safety Verification & Validation (V&V) in an efficient, objective, transparent, and scalable manner but ensure dense coverage of traffic situations. Additionally, the manufacturer should demonstrate the compliance with relevant regulations, laws, standards, and good practices [6].

The project MASA (Methodology for KPI based ADAS/AD Safety Assessment and Argumentation) is a 2-year long project that started in 2021. Its main objective is to better understand and enhance existing methodologies and numerical approaches for V&V Safety Argumentation frameworks of ADS. The rest of the paper is structured as follows. The safety and engineering frameworks for ADS considering the regulatory regime, stakeholders' know-how and market demand are summarised in the Sect. 2. In the Sect. 3 attention is given to the elaboration of the Safety Case approach which is considered by many stakeholders at the center of the Safety Argumentation. The Sect. 4 covers the overview of novel performance models and state-of-the-art safety metrics that support the evidence creation process. This covers metrics proposed by the Contracting Parties to the UNECE and research and industrial projects. Section 5 summarises recommendations on existing Safety Argumentation frameworks and next steps for MASA project. The latest processes prescribed by the ISO 21448 - the Safety of The Intended Functionality (SOTIF) [7] on the risk/hazards and triggering conditions analysis are at the future research focus.

The collection of materials and access to the latest communities' updates were threefold, through i) list of publications, reports, and industry white papers selected according to the MASA key words ii) (online) attendance at sectorial conferences and webinars

organized by recognized and eligible organizations in area of ADAS/ADS and iii) participation and memberships of both project partners in the Working Groups (WGs) and regulatory/standardization committees. This methodology for identifying the state-of-the-art (SOTA) resulted in more than one hundred short listed publications that provide a valuable and robust overview over current safety framework challenges for ADS. Moreover, participation in around 60 webinars, workshops, community online meetings, including mid-terms reviews and project showcases with dissemination networking opportunities took place (e.g., event platforms like ScaleUp 360°, IEEE Mobility Practice, PAVE Virtual Panel, ITU-T Focus Group events, V&V Methoden, Set4to5, CertiCAV, VeriCAV Final Showcases, ERTRAC/EUCAR/ERTICO workshops, among others). Additionally, the experts from both partnering organizations, namely AVL and Virtual Vehicle participated in the discussions of relevant WGs, exchanged views and channeled feedback to shape the regulatory and standardization landscape. In some of the WGs and projects, the partners have very active roles (e.g., UNECE GRVA VMAD/FRAV, the JRC EC, ISO FuSa/SOTIF/co-engineering national mirror groups, ASAM, IAMTS Alliance, PEGASUS project family, etc.). In other groups the partners acted as observers to find some convergent points (IEEE, SAE, etc.).

This paper provides the expert-driven selection of relevant groups, standards, and projects. The complete report with an extensive SOTA list was reported to the board of the COMET K2 Competence Centers for Excellent Technologies in Austria. Nominal safety and hazard/risk analysis and emerging topics of SOTIF were in the main scope of the MASA project. Functional safety, the presence of artificial intelligence (AI)/Human Machine Interaction (HMI)/V2X communication components, security-informed safety, cybersecurity, or software updates, indeed pose additional challenges for the safety argumentation, nevertheless they were not in the project scope.

To sum up, it could be argued that the safety assurance approaches are lagging the technology development. V&V of AD/ADAS brings novel challenges that try to be tackled by new standards and regulatory draft documents capturing best practices. The paper goal is to fill existing gaps in the literature reviews with a particular focus on safety metrics.

2 A New Safety Assurance Scheme and Global Initiatives

In the literature there are defined different qualitative and quantitative ways to categorize ADS safety: i) safety as a process/framework, ii) safety as a measurement, and iii) safety as a threshold [8]. In the first stage, MASA focused on understanding assurance frameworks and good practices that consist of interacting methods and tools.

The Safety Argumentation supports the process of gaining knowledge and trust on the operating conditions of ADS. Its main objective is to deliver in a structured way the right evidence to prove safety of the vehicle. Decomposition of the safety requirements prescribed by the regulators and standardization bodies (see Sect. 2.1 and 2.2) and implementation of methods for deriving test requirements (see Sect. 2.3) help to arrive to credible evidence (Sect. 4). A well-structured framework could accelerate the exchange of information and requirements between stakeholders and optimize the tests amount to a manageable minimum [9].

The current automotive and transport communities that dialogue on enablers for industrialization of ADS could be split into i) regulatory Informal Working Groups (IWG), ii) standardization activities, and iii) research and industry initiatives.

2.1 Regulatory Informal Working Groups for ADS Safety Topics

In 2019 a new Working Party on Automated/Autonomous and Connected Vehicles (GRVA) of the UNECE was created. The WP was tasked with defining provisions for a "safe vehicle" and its approval. Its two IWGs play a significant role in the ADS Safety Argumentation, namely Functional Requirements for Automated and Autonomous Vehicles (FRAV) and Validation Method for Automated Driving (VMAD).

The starting point of the FRAV discussion was a "guardrails approach", where the regulatory bodies decided to not prescribe driving manoeuvres or values. The stakeholders recognise that more than one behaviour could be safe and do not want to hamper the technology development at so early stage of regulatory process. As an outcome of the FRAV sessions, a robust list of "General Safety Requirements" was concluded and reported to the GRVA committee [10]. The novelty of this list is that it does not focus on each technology and features separately but addresses the diversity of ADS applications. Moreover, it reflects a global and consolidated view provided by the Member States and the industry representatives covering the full range of anticipated safety needs. Nowadays, the FRAV members supersede where appropriate the list with technical specifications (limits, criteria, formulas, etc.).

On the other hand, the VMAD tackled the certification and audit/assessment process of ADS. The outcome of the discussions is the master document "New Assessment and Test Methods" (NATM) which describes existing test platforms and their interactions. Moreover, the group works on the documentation requirements needed during the audit and assessment phase. The goal is to demonstrate a combination of validation, engineering rigor, post-deployment feedback, and safety culture [5].

2.2 Global Standardization Activities

There are many standards that support the Safety Argumentation in different product development phases following known V-cycle: from the specification, design, integration, to V&V, product release and monitoring. The purpose of those documents is threefold: i) providing the terms, definition, and summarising existing state of the art of scenario-based validation, ii) defining processes, methods, and tools, and iii) sharing check lists or taxonomy/ontology trees. It is relevant to understand their scope and possible contributions to a global safety argumentation framework.

The following standards addressing existing ADS-equipped vehicle safety approaches were considered in MASA roadmap (Table 1). Their content is intended to be applied to ADS Level 3 and higher according to [2]. Currently, many discussions are taking place between standardization organizations like ISO, IEEE, SAE, BSI to align the terminology and look for a common vision for ADS deployment (e.g., the alignment activities for ODD specification under the umbrella of new ISO 3450x series). The dialogue between many member states, who express different perspectives and safety goals, should be seen as a long-lasting process, during which conflicting cases require

further elaboration (e.g., mentioned in Sect. 4 safety targets and measurable criteria which strongly depend on local rules and societal acceptance). It is evident that a lot of effort was done towards standardization, nevertheless the ADS ecosystem stakeholders still need to invest further resources to speak a common language.

Table 1. Global standardization activities relevant for the Safety Argumentation framework.

Name of the standard, year of publication	Relevance for the ADS Safety Argumentation	Goal: Terms & Definitions (TD)/Process & methods (PM), Check list (CL)
ISO/PAS 21448:2019 Safety of the intended functionality (SOTIF), rev. in 2022	Limitations and shortcomings of the technology and the misuse of the function, iterative process of improving the acceptance criteria by triggering conditions and risk/hazards evaluation	TD, PM
ISO/SAE PAS 22736:2021 Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles	Definitions, taxonomies, and best practices for six levels of driving automation, inc. Terms for the dynamic driving task (DDT), DDT fallback, minimal risk conditions, etc. (aligned with SAE J 3016-2021)	TD
ISO 26262:2018 Functional safety (12 parts)	E/E faults and failures as a source of safety problems and calls for a safety case	TD, PM
ISO 3450x series inc. ISO 34501-34505 (u. development)	Terminology and definition of test scenarios for ADS, the overall scenario-based safety evaluation process, a hierarchical taxonomy for ODD, tags for scenarios categorization	TD, PM, CL
ISO/AWI TS 5083 Safety for ADS - Design, verification, and validation (u. development)	Safety by design, V&V methods based on safety goals and principles, safety case, def. of positive risk balance and avoidance of unreasonable risk (proceed from an ISO/TR 4804)	PM
SAE J3131:2022 Automated Driving Reference Architecture	ADS reference software architecture that contains functional modules (does not dictate configuration, neither design requirements)	TD

(*continued*)

Table 1. (*continued*)

Name of the standard, year of publication	Relevance for the ADS Safety Argumentation	Goal: Terms & Definitions (TD)/Process & methods (PM), Check list (CL)
IEEE P2846:2022 Formal Model for Safety Considerations in Automated Vehicle Decision Making	Set of assumptions and foreseeable scenarios for the development of safety-related models, an extension of the safety envelope violation concept (RSS)	PM
ASAM OpenX standards (concept/implementation)	Foundation of definitions, language, formats, and interfaces for V&V in simulation platforms	TD, PM, CL
ANSI/UL 4600: 2022 Standard for Safety for the Evaluation of Autonomous Products	Safety case approach goal-based and technology-agnostic, a catalogue of good practices in the whole ADS cycle, inc. safety metrics strategy	CL

2.3 Research and Industry Communities, Initiatives, and Projects

The R&D projects, industrial consortia and local initiatives try to discuss varied approaches related to ADS performance. The task of the MASA was to analyse their unique concepts and what new they bring to the Safety Argumentation. These initiatives not only contribute with the publications and white papers, besides provide valuable input to the regulatory/standardization landscape (see Sect. 2.1 and 2.2):
V&V Methods (Germany), a part of PEGASUS project family[1] supported by The Federal Ministry for Economic Affairs and Climate Action (BMWK). The project develops a method, called criticality analysis, which analyzes the open context of urban traffic. It introduces definitions and concepts that support the Safety Argumentation (criticality analysis, criticality phenomenon, criticality metric) [11].
The CertiCAV Safety Assurance Framework for CAVs (UK) executed by the Connected Places Catapult on behalf of the Department for Transport. The projects developed a framework that has number of novel ideas like a Highly Automated SuperSystem (HASS), the concept of Deployment Risk Classifications, and a foundation for developing requirements with performance indicators [12].
The SAKURA Safety Assurance KUdos for Reliable Autonomous Vehicles (Japan) funded by the Ministry of Economy, Trade, and Industry (METI). The project proposed a very robust framework that goes further than the nominal driving conditions. It elaborates

[1] PEGASUS, Set Level 4 to 5, V&V Methods are projects developed by the VDA Leitinitiative autonomous and connected driving, https://www.vvm-projekt.de/en/.

on safety-relevant disturbances that an ADS may face in a real traffic (called traffic, perception, and vehicle motion disturbance scenarios) [13].
AURORA/UBER Argumentation Concept (USA), a commercial initiative. Till now the stakeholders were reluctant to openly share safety cases and reveal the format. Aurora shared their version of the self-driving Safety Case Framework [14].

3 Safety Case as a Core Approach to Argue the Safety of ADS

The safety case could be seen as a starting point, as an easy and understandable tool to support the Safety Argumentation. A safety case approach is not a new tool, and many communities refer to it. There are available manuals and standards (e.g., IEC 61508, U.K. Defence Standard 00-55, HSE Railway Safety Case Regulations) which center on a safety case. The good practices from aeronautics, railway, marine or healthcare could have a value in the automotive sector [15]. Literature from R&D projects (see Sect. 2.3), University of York community [16] and Edge Case Research studies [17] bring additional input towards the safety case and its format.

3.1 Safety Case Content and Format

In the literature exist many complementary definitions for the safety case. ISO 26262:2018 for example defines a safety case in the context of the functional safety as "an argument that functional safety is achieved for items, or elements, and satisfied by evidence compiled from work products of activities during development." [18].

The elements of the safety case are i) goal, ii) argumentation, and iii) evidence. When it comes to the argument, it may be deterministic, probabilistic, or qualitative; the evidence may be design, process, or historic experience (with a proof of completeness for coverage of test cases). In the Safety Argumentation it is essential to decompose each goal (claim) further into tangible and measurable performance indicators (results and records). (Safety) metrics are a measurement used to evaluate and track safety performance [19] (see Sect. 4.1). Finally, as an outcome of V&V activities on different test platforms, relevant amount of supporting data should be collected to produce quantitative evidence of safety case credibility.

A format for structuring argumentation should be notation agnostic (i.e., textual, tabular, graphical). Some manuals like widely used the Goal Structuring Notation (GSN) [16], MISRA Safety Case Guideline [20], Structured Assurance Case Metamodel (SACM™) [21] or Claims, Arguments and Evidence (CAE) notation [22] support the arguments creation and documentation process. Using an agreed format enables a consistent and traceable decomposition from claims down to V&V methods and keeping the safety case evolving over the life cycle (a living document).

3.2 Safety Goals

At the beginning of the safety argumentation the manufacturer needs to define the Safety Goals. The current data bases of crashes and statistics often provide basic data on a high and aggregated level, without diving into the cause of the crash. For ADS there should

be different types of the safety goals that are included in the argumentation process [23] and that support accident investigation and decomposition tasks:
Functional Safety Goals - according to ISO 26262 compliance - behavior in case of system or component failure (hazards) that may arise by the functionality in the E/E system within its ODD are identified and assigned to the required ASIL level [18],
SOTIF Goals - according to ISO 21448 compliance - triggering conditions (including foreseeable misuses) and obtained test results purpose is to reduce the residual risk to an acceptable level and to improve ADS nominal performance [7],
Ethics Goals and Societal Expectations - according to ethical standards (e.g., ISO 39003 u. development), human errors and intentions, that could endanger other actors lives or damage properties, discriminate road users by age, gender, clothes, handicap,
Laws and Regulations Goals - according to UNECE regulations, national and traffic law under Road Traffic Code and locally accepted behavioural laws respectively,
Cybersecurity/Software updates Goals - according to ISO/SAE 21434, SAE J3061, UN Regulation No. 155/UN Regulation No. 156 respectively.
The MASA project focuses mainly on prescriptive requirements coming from laws, regulations, and standards (see Sect. 4.1), and investigation of novel approaches to support SOTIF processes argumentation (see Sect. 4.2).

3.3 Acceptance Criteria

Acceptance criteria for ADS are based on a statement that "safe enough" is not just a number, it is an argument. In literature there are many strategies and accompanying argumentation patterns, e.g., PAS 1881 [24] includes references to ALARP (Reducing risk as low as reasonably practicable), safety case architecture as in UL 4600 [25] or ISO 21448 [7] (under revision) additionally refer to GAMAB ("generally at least as good as") and MEM (Minimum Endogenous Mortality), among others. Some of acceptance criteria widely used in other sectors are summarized in [26]. Those are mentioned ALARP, GAMAB, MEM criteria and other like MISRI (Minimum Industry Safety Return on Investment) or RAPEX (Rapid Exchange of Information).

Currently, "a positive risk balance" criteria gain more attention. In this context, ADS should generate a "statistical positive risk balance" such that ADS demonstrate superior performance when compared statistically against human driving performance. One of the objectives of a new ISO 5083 standard (under development) [27] is to structure a holistic safety approach with the safety case pattern for ADS. The recommendations are based on the German Ethics Commission report [28].

4 Safety Evidence: An Attempt to "Quantify" the ADS Safety

It is recognized that the safety case is one of elements to a complete safety assurance framework for ADS. Particularly challenging is to collect enough valid evidence to support the safety goals and arguments (Sect. 3.2) and to claim the safety case completeness and sufficient V&V coverage. The safety metrics with data collection strategy should be introduced to result in an acceptable and strong safety case.

4.1 Requirements-Based Testing with Numerical Approaches

The Sect. 2 explains the need to quantify the product safety, its social acceptance and liability. Indeed, the introduction of ADS requires new type of i) performance models and ii) safety metrics. Consolidation of safety related metrics was the main objective of MASA, for ADS none-safety related metrics (network efficiency, energy emission, drive quality, costs and public health or comfort), the reader could consult [29].

Performance Models are based on "roadmanship" concept, which means the ability to drive on the road safely without creating hazards and responding correctly to hazards created by others. They reflect the ADS's situational awareness, time to response, speed adjustment, the vehicle's physics, driving culture and laws and diverse driving scenarios [8]. In the literature, ADS performance models are named as mathematical models for trajectory planning, safety envelope or escape path. Some prominent models that are part of the ADS regulations and standards are:

Competent and Careful Human Driver's Performance Model (C&C) proposed by the Japanese delegation to the UNECE. The model is included in the UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS), Annex 3. Its main assumption is that traffic accidents are split into rationally foreseeable and preventable [30].

The Responsibility-Sensitive Safety Model (RSS) by Intel (US). It is a white-box mathematical model. It formalizes the "duty of care", which means that a road actor should exercise "reasonable care" while performing acts that could harm others [31]. The safety envelope concepts are reflected in IEEE 2846 standard [32].

Fuzzy Safety Model (FSM) of the Joint Research Centre of the European Commission. This model builds on RSS findings. Its characteristics are based on fuzzy logic that would not require the vehicle to decelerate very sharply or very often. The model is considered next to the C&C model in the draft of ALKS extension (Annex 3) under the leadership of SIG UNR157 Task Force of UNECE [33].

The reader should be aware of other models existing in the literature, that were analysed under MASA but currently are rather in the concept phase: Safety Force Field (SFF), Instantaneous Safety Metric (ISM), Criticality Metric using Model Predictive Trajectory Optimization [19], among others.

Safety Metrics and Their Thresholds - till today the universal metrics for road safety were historical crash data like frequency and severity. The existing test procedures and protocols for assisted functions (ADAS) like Euro NCAP consumer tests or regulatory documents (e.g., UN No. 79, UN No. 131) introduce simple but comprehensive metrics to select situations out of traffic events to reduce the amount of test effort.

The need for harmonization and standardization of terms and techniques for ADS safety measures has been recognized over the last decade [34]. The metrics that refer to testing in nominal conditions are used under exchangeable terminologies: Behavioral Safety Measures [35], Safety Performance Assessment Metrics [36], Proximal Surrogate Indicators, Temporal/Spatial-based Conflict Indicators [37], Criticality/Risk Metrics [11]. They evaluate the criticality of the traffic situation.

Mentioned above references define important properties of each metric: definition, taxonomy, data source from off or onboard sources, observable variables, formulation: mathematical model, assumptions/thresholds when applicable, origin, limitations and

advantages, reason for inclusion, research examples, type of scenario, manoeuvre collision type suitability. The section "16. Metrics and Safety Performance Indicators (SPIs)" of [25] gives additional guidelines on metrics as a part of the Safety Culture. The manufacturer should present a metric strategy with collection, evaluation, and improvement processes. Further, the safety metrics could be categorized into [8]:

Prior/Predictive (Leading Metrics) - including general performance characteristics, associated with vehicle kinematics (longitudinal and lateral distance), perception and assessment of Object and Event detection specification (OEDR) [38], safe motion control metrics etc. Leading metrics are particularly important for ADS because their events happen more frequently than lagging measures.

An Outcome (Lagging Metrics) - covering post deployment, longer term metrics like for driver disengagement [39], ODD metrics [40], violation of road rules and crash severity and frequency. Assessing a correlation of leading metrics to safety outcomes should be used to drive improvement of the metrics and thresholds [19].

The metrics could only be implemented successfully considering the threshold and the pass-fail criteria definition. Their definition is not an easy task, neither standardized (e.g., ADAS systems use fixed rules of thumb like the two-second rule for establishing safety envelope [41]). Additionally, [25] states that the thresholds could be a desired value, limit, or incident frequency. The authors recommend varied approaches for consolidating claims of different stakeholders when selecting values/targets: i) technology aspects (the state-of-the-art technology limitations), ii) human driver aspects (response to traffic events: perception, recognition, decision), iii) social aspects (socially acceptable behaviours) iv) legal aspects (historical decisions of the court jurisdiction). Moreover, to allow ADS deployment, the thresholds should support federal, state, and local laws and could be a function of several parameters such as vehicle capabilities, road user type, and speed of the ego vehicle [19].

4.2 Quantitative Hazard and Risk Analysis as a Part of the SOTIF Processes

The safety metrics and performance models could support argumentation for known and nominal conditions. When it comes to edge and rare traffic scenarios, the probabilistic methods with the use of virtual testing environment could play a crucial role. A challenge for ADS today is a sound and systematic methodology for the identification and quantification of scenarios that are likely to exhibit hazardous behaviour.

SOTIF standard focuses on the limitations of the technology and the misuse of the function. One of its work products is to discover "potential triggering events" with the purpose of improving the defined acceptance criteria and minimizing the known/unknown hazard scenarios with each iteration [7]. The MASA established exemplary Triggering Conditions (TC) database and categorized them into 5 groups: i) Environmental TC (weather state, illumination, quality of the road surface), ii) Infrastructural TC (geometry of the road, road furniture, objects on the road/surroundings), iii) Communication and interferences, iv) Other road actors (adverse traffic behaviour, non-standard actors, surrounding vehicles), v) Ego vehicle behaviour (perception, operation/maneuverers). What is novel in MASA, selected triggering events will be parametrized, aligned with ASAM OpenXOntology standard, and demonstrated in industrial settings with AVL SCENIUS™ V&V Toolchain.

Parametrization process of the SOTIF Triggering Conditions is a robust task as it requires a new type of statistics (unit/scale, boundary values, ground truth measurement process, source of potential statistics, etc.). Real world occurrence likelihoods of today's data (traffic and crash data, labelled data sets from the naturalistic driving, weather forecast records, or infrastructure maintenance reports) require a totally new way of looking at them. The triggering conditions dependencies and dependencies between triggering events and the scenarios constitute another challenge [42].

Currently, the authors investigate and compare preliminary statistical approaches for SOTIF safety argumentation, and the outcome will be reported in the second year of the project. Some of the prominent approaches taken into consideration include [35, 42, 43]. The main objective of the MASA project in the next phase is to enhance the existing workflow of the SOTIF standard with novel numerical approaches.

5 Conclusions and Outlook

The MASA project examined concepts for safety metrics, formal performance models, taxonomies, and process approaches for measuring the safety of ADS-equipped vehicles. The analysis of the available safety frameworks revealed the current research needs and existing gaps between regulatory and industrial pace. Today's regulatory documents on ADS V&V type approval leave a lot open to interpretation.

Determining metrics may help to demonstrate safety as a part of holistic approach for assessing/evaluating aspects of ADS safety. Unfortunately, no one has presented a fully suitable set of metrics for arguing safety of ADS across the range of its functions and features, use cases, and ODDs. The manufacturer could follow many available safety practices to decide which fit the best in their safety case, but still depend on the Type Approval Authority opinion of whether it complies with requirements.

On the other hand, the global consensus was reached on the need to develop ADS strategy for safety metrics. It is seen as a joint work of stakeholders from industry, academia, authorities, and consultation with the civil society. The members of the working groups and research projects recognize both quantitative (product-oriented test results and records) and qualitative approaches (the Safety Culture and Management Systems oriented on processes) as valid to understand the level of safety and duty of care. It still needs to be explored how to leverage a mix of those approaches.

Further, it is suggested that for level 3 ADS-equipped vehicles, the current practices coming from the "technology based" regulatory and standardization documents should be extended with novel standards, such as ISO 26262, ISO 21448, ISO/SAE 21434. In this paper, we only give an overview of the state of the art around statistical approaches for hazard and risk analysis to identify rare and unknown cases.

The goal of the next period of the MASA project is to combine and extend established techniques for hazard analysis and risk assessment. The challenge is to supersede traffic scenario databases and test case frameworks with low probability but high consequence events (in literature so called triggering conditions [7], disturbance scenarios [13], or criticality phenomena [11]) that are not captured in the existing database of global functional scenarios [5]. Unfortunately, there is still lack of publicly acceptable and systematic identification method for triggering events.

To conclude, the development of reliable safety measures will be a significant achievement that expands the current V&V methods. Extending nominal and well-known scenarios with SOTIF approach creates an additional trust in ADS technology.

Acknowledgement. The publication was written at Virtual Vehicle Research GmbH in Graz, Austria. The authors would like to acknowledge the financial support within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorised for the programme management. They would furthermore like to express their thanks to their supporting industrial project partner, namely AVL List GmbH.

References

1. WP.29 UNECE: ECE/TRANS/WP.29/2019/34/Rev.1. Revised Framework Document on Automated/Autonomous Vehicles (2019)
2. ISO: ISO/SAE PAS 22736:2021 - Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (2021)
3. UNECE IWG VMAD, Subgroup 3 - Audit and in-service monitoring: VMAD-SG3-22-02 Audit Pillar (2022)
4. The EC: Draft regulations of the EC laying down rules for the application of Regulation (EU) 2019/2144 as regards uniform procedures and technical specifications for the type-approval of the automated driving system (ADS) of fully automated motor vehicles (2022)
5. UNECE IWG VMAD: ECE/TRANS/WP.29/GRVA/2022/2, the New Assessment/Test Method for Automated Driving (NATM) - Master Document (2022)
6. California PATH Program, University of Berkeley: Safety Assurance to earn public trust - Formalizing the Safety Case for ADS. In: V&V Methods Mid-term meeting (2022). https://www.vvm-projekt.de/midterm-docs. Accessed 02 June 2022
7. ISO: ISO/PAS 21448: 2022 - Road vehicles - Safety of the intended functionality (2022)
8. Blumenthal, M.S., Fraade-Blanar, L., Best, R., Irwin, J.L.: Safe Enough: Approaches to Assessing Acceptable Safety for AVs. RAND Corporation, Santa Monica, CA (2020)
9. Galbas, R.: How to systematically release AD systems? In: V&V Methods Mid-term Meeting (2022). https://www.vvm-projekt.de/midterm-docs. Accessed 02 June 2022
10. UNECE IWG FRAV: GRVA-12-23, Guidelines and Recommendations concerning Safety Requirements for ADS (2022)
11. Neurohr, C., Westhofen, L., Butz, M., Bollmann, M.H., Eberle, U., Galbas, R.: Criticality analysis for the verification and validation of automated vehicles. IEEE Access **9**, 18016–18041 (2021). https://doi.org/10.1109/ACCESS.2021.3053159
12. CATAPULT, CertiCAV Paper: A framework approach for assuring the behaviour of highly automated vehicles (2021)
13. Japan Automobile Manufacturers Association: Automated Driving Safety Evaluation Framework Ver 2.0 (2021)
14. Aurora: Aurora unveils first-ever Safety Case Framework that addresses the safety of both autonomous trucks and passenger vehicles. https://aurora.tech/blog/aurora-unveils-first-ever-safety-case-framework. Accessed 28 Apr 2022
15. Bishop, P., Bloomfield, R.: A methodology for safety case development. In: Redmill, F., Anderson, T. (eds) Industrial Perspectives of Safety-Critical Systems. Springer, London (1998). https://doi.org/10.1007/978-1-4471-1534-2_14

16. Kelly, T., Weaver, R.: The goal structuring notation–a safety argument notation. In: Proceedings of the Dependable Systems and Networks 2004 Workshop on Assurance Cases (2004)
17. Koopman, P., Osyk, B.: Safety argument considerations for public road testing of AVs. In: WCX SAE World Congress Experience (2019). https://doi.org/10.4271/2019-01-0123
18. ISO: ISO 26262:2018 - Road vehicles - Functional safety (2018)
19. Automated Vehicle Safety Consortium™: AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of ADS (2021)
20. The MISRA Guidelines for automotive safety arguments. https://www.misra.org.uk/misra-safety-argument/. Accessed 28 Apr 2022
21. Structured Assurance Case Metamodel (SACM™). https://www.omg.org/spec/SACM/2.2/About-SACM/. Accessed 03 June 2022
22. Claims, Arguments and Evidence (CAE) notation. https://www.adelard.com/asce/choosing-asce/cae.html. Accessed 03 June 2022
23. Schittenhelm, H.: How to ensure a safe operation of an automated driving system by a methodological approach? In: V&V Methods Mid-term Meeting (2022). https://www.vvm-projekt.de/midterm-docs. Accessed 02 June 2022
24. BSI: PAS 1881:2022 Assuring the operational safety of automated vehicles - Specification (2022)
25. ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products (2021)
26. Rae, A.: Acceptable Residual Risk - Principles, Philosophies and Practicalities. pp. 26–31 (2007). https://doi.org/10.1049/cp:20070436
27. ISO: ISO/AWI TS 5083 Safety for automated driving systems - Design, verification and validation (under development)
28. BMVI: Ethics Commission - Automated and Connected Driving, Report extract (2017)
29. VTT: Key performance indicators for assessing the impacts of automation in road transportation Results of the Trilateral key performance indicator survey (2018)
30. ECE/TRANS/WP.29/2020/81: A new UN Regulation on uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping System (2020)
31. Shalev-Shwartz, S., Shammah, S., Shashua, A.: On a formal model of safe and scalable self-driving cars. arXiv preprint arXiv:1708.06374 (2017)
32. IEEE: IEEE 2846-2022, Standard for Assumptions in Safety-Related Models for Automated Driving Systems (2022)
33. SIG UNR157 TF: UNR157-09-03r1 Performance models of ALKS, Annex 3 (2021)
34. NIST Workshop: Consensus Safety Measurement Methodologies for ADS-Equipped Vehicles. www.nist.gov/news-events/events/2019/06/consensus-safety-measurement-methodologies-ads-equipped-vehicles. Accessed 28 Apr 2022
35. Kramer, B., Neurohr, C., Büker, M., Böde, E., Fränzle, M., Damm, W.: Identification and quantification of hazardous scenarios for automated driving. In: Zeller, M., Höfig, K. (eds.) IMBSA 2020. LNCS, vol. 12297, pp. 163–178. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58920-2_11
36. Wishart, J., et al.: Driving safety performance assessment metrics for ads-equipped vehicles. SAE Technical Paper **2**, 2020-01-1206 (2020)
37. Mahmud, S.S., Ferreira, L., Hoque, M.S., Tavassoli, A.: Application of proximal surrogate indicators for safety evaluation. IATSS Res. **41**(4), 153–163 (2017)
38. Hoss, M., Scholtes, M., Eckstein, L.: A review of testing object-based environment perception for safe automated driving. Automot. Innov. **5**, 223–250 (2022). https://doi.org/10.1007/s42154-021-00172-y
39. Dixit, V.V., Chand, S., Nair, D.J.: Autonomous vehicles: disengagements, accidents and reaction times. PLoS ONE **11**(12), e0168054 (2016)

40. ASAM OpenODD: Concept Paper. www.asam.net/index.php?eID=dumpFile&t = f&f = 4544&token=1260ce1c4f0afdbe18261f7137c689b1d9c27576. Accessed 28 Apr 2022
41. Koopman, P., Osyk, B., Weast, J.: Autonomous vehicles meet the physical world: RSS, variability, uncertainty, and proving safety. In: Romanovsky, A., Troubitsyna, E., Bitsch, F. (eds.) SAFECOMP 2019. LNCS, vol. 11698, pp. 245–253. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26601-1_17
42. De Gelder, E., Elrofai, H., Saberi, A.K., Paardekooper, J.P., Den Camp, O.O., De Schutter, B.: Risk quantification for automated driving systems in real-world driving scenarios. IEEE Access **9**, 168953–168970 (2021)
43. Karunakaran, D., Worrall, S., Nebot, E.: Efficient statistical validation with edge cases to evaluate highly automated vehicles. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–8. IEEE (2020)

A.3 Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions

Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions

Víctor J. Expósito Jiménez[1(✉)], Helmut Martin[1], Christian Schwarzl[1], Georg Macher[2], and Eugen Brenner[2]

[1] Virtual Vehicle Research GmbH, Inffeldgasse 21a, 8010 Graz, Austria
{victor.expositojimenez,helmut.martin,christian.schwarzl}@v2c2.at
[2] Graz University of Technology, Rechbauerstraße 12, 8010 Graz, Austria
{georg.macher,brenner}@tugraz.at

Abstract. Safety in the automotive domain is a well-known topic, which has been in constant development in the past years. The complexity of new systems that add more advanced components in each function has opened new trends that have to be covered from the safety perspective. In this case, not only specifications and requirements have to be covered but also scenarios, which cover all relevant information of the vehicle environment. Many of them are not yet still sufficient defined or considered. In this context, Safety of the Intended Functionality (SOTIF) appears to ensure the system when it might fail because of technological shortcomings or misuses by users.

An identification of the plausibly insufficiencies of ADAS/ADS functions has to be done to discover the potential triggering conditions that can lead to these unknown scenarios, which might effect a hazardous behaviour. The main goal of this publication is the definition of an use case to identify these triggering conditions that have been applied to the collision avoidance function implemented in our self-developed mobile Hardware-in-Loop (HiL) platform.

Keywords: Triggering conditions · SOTIF · ADAS · Automated Driving Systems

1 Introduction

The validation of the Advanced Driver-Assistance Systems (ADAS)/Automated Driving Systems (ADS) has been a topic for numerous research works due to the complexity of functions that have been exponentially growing over the years, since there are more components and software included in each function and their relationships between them (e.g. AI algorithms,...) are getting more challenging . Reported public cases [1,2] with human deaths have occurred in the last years, which have given an unreliable picture of the current status of the automated driven cars to the public view. Many standards have been developed to provide

M. Trapp et al. (Eds.): SAFECOMP 2022 Workshops, LNCS 13415, pp. 11–22, 2022.
https://doi.org/10.1007/978-3-031-14862-0_1

a common framework and to overcome these accidents as much a possible in the future in which all aspects related to system dependability have to be covered. The dependability of a function can be split into three main parts according to the origin of the hazard. Functional Safety hazards, effected by malfunctions based on hardware and systematics faults, are covered in the ISO26262 [3]. Cybersecurity is covered in the ISO/SAE 21434 [4], which is focused on external threads. The ISO21448, Safety of the Intended Functionality (SOTIF) [5], standard is being developed to cover the gap that has been added due to the new requirements and safety standards that ADAS/AD functions developments have to face. With more functions dependent on complex sensors and algorithms, more unknown scenarios could occur in which the function is not designed for. The main goal of the SOTIF standard is to minimize the unknown hazardous scenarios due to technical shortcomings or misuses, specially the ones that can lead to hazardous situations. Many researches [6,7] have been trying to reuse and link part of the processes described in the ISO26262 into the SOTIF standard, such as Hazard Analysis and Risk Assessment (HARA) or System-Theoretic Process Analysis (STPA). Another important point for SOTIF is the identification of the limitation of each technology used for automated driving [8], which gives more knowledge in order to identify and understand the cause of these unknown scenarios. To make the steps more explicit for the safety argumentation, the UL4600 [9] provides a checklist of necessary elements to ensure that a function is safe in all aspects.

The complexity of covering all possible variations and scenarios that can occur, make validation of ADAS/AD functions a hard task and there are different approaches to face this goal. For example, the validation approach based only on real driven miles data would be unpractical and the cost of the process would be unacceptable [10]. Another way to support this task would be through simulations. During the last years, new environment simulators specifically designed for this domain, such as CARLA [11] or SVL Simulator [12], have been developed. The main issue with this kind of validation is the creation of the realistic high-fidelity sensor models, which are highly complex to develop and also need huge computational resources and high-performance machines [13] as well as the generation of scenarios [14] to cover all possible scenario variations. Authors in [15] also include machine learning approaches to get the conclusion that the lack of meaningful trained data makes this approach not ready to fully cover the whole SOTIF argumentation yet. On the other hand, to avoid the issues previously commented, the authors in [16] present an approach purely statistical, which could provide a quantitative argument of functions validation. Since some concepts could be reused in our research, in this case, the direction of our research is focused on the decomposition of the function to identify the triggering conditions according to the function design.

One of the main step of a SOTIF analysis is the definition of the Operation Design Domain (ODD), according to the standards such as ISO21448 or UL4600. An ODD can be defined as the intended behavior within defined environmental condition that the function has been designed to work on. The definition of the

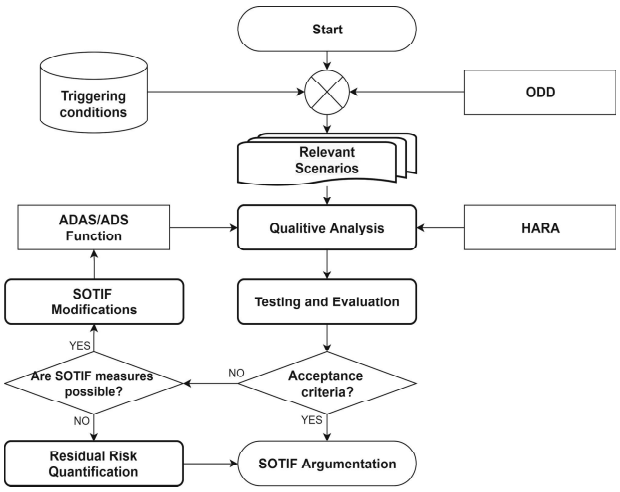


Fig. 1. Flowchart of the proposed triggering conditions methodology

ODD must include all necessary parameters, which make the definition meaningful such as constraints to road conditions. There are different approaches for the definition of a ODD, from the described in the PEGASUS project [17] in which the ODD definition is split into six different layers according to the functionality of the parameters to the description given in the SAKURA project [18] or the provided by ASAM in ASAM OpenODD [19]. The authors in [20] present an overview of available approaches for the most common ODD definition. These approaches provide us with the necessary taxonomy to give the basis to create a triggering conditions list. This publication gives an overview of triggering conditions in automated driving [21], but in our approach, it is based on scenario level and not the system as it is described.

The structure of the paper at hand is as follows: our approach is explained in the next section. Section 2 goes through all blocks of the proposed use case and its implementation in the collision avoidance function of our Hardware-in-Loop (HiL) platform. Finally, Sect. 3 shows the conclusions and the future outlook of our research.

2 Use Case and Methodology

The proposed methodology is shown in Fig. 1 in which the used blocks and their relationships are depicted. The main goal of this publication is to show the key ideas and process of our use case to give the elements to provide a SOTIF argumentation. This is described as a brief basis and further description will be given in following research work as a more mature methodology.

2.1 ADAS/ADS Function Description

In this first block, the function and its functionality as well as relevant technical aspects are given in order to be able to know the behaviour of the function. The use case example in this publication is the collision avoidance function implemented in our self-developed mobile Hardware-in-Loop (HiL) platform called SPIDER [22,23]. A collision avoidance function could work in two ways, as an Automatic Emergency Brake (AEB) or providing an alternative way/evasive maneuver to reach the target location. For simplicity, only the AEB is implemented at this moment on the robot and it will be used in the analysis in this paper. The function uses four 16-lines lidar sensors located at each corner (Front-Left, Front-Right, Back-Left, Back-Right) to provide redundancy where each point is seen, at least, by two sensors and leave the centre of the robot for different flexible setups in the future as Fig. 2 shows. The data from the lidars are fused to generate an occupancy grid, which shows the objects surrounding the robot and sends the activation signal to start the emergency brake in case one object enters within the delimited zone as the danger zone. According to the SAE J3016 Standard [24], it is a Level 4 function since the dynamic driving task fallback has no human interaction.

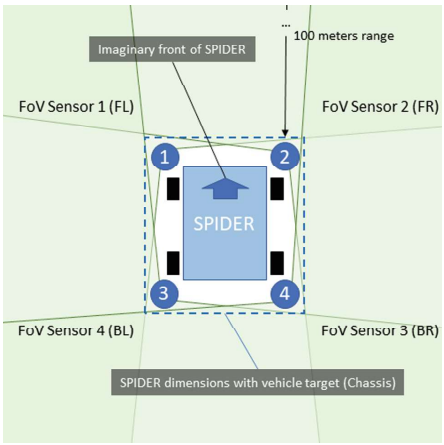


Fig. 2. Hardware setup of the collision avoidance function on the HiL Platform (Field of View - FoV, Front Left - FL, Front Right - FR, Back Right - BR, Back Left - BL)

To guarantee that there is no collision on the nominal scenario, the Responsibility-Sensitive Safety (RSS) [25] model was implemented as the minimum distance before the emergency brake is activated, which formula was adapted for our use case. RSS model was chosen since it provides a proven [26] safe distance according to our robot parameters. Ego vehicle is defined as the main actor of interest in the scenario, sometimes is also refereed as Vehicle Under

Test (VUT). On the other hand, the target vehicle is considered as an element of the scenario and its behaviour is described in the scenario definition. The assumption of zero speed of the vehicle located at front of the ego vehicle(target vehicle) was set to zero due to the ODD is only taking into account static objects. Therefore, the RSS equation of the used function is given as follows:

$$d_{min} = \left[v_r \rho + \frac{1}{2} a_{max,accel} \rho^2 + \frac{(v_r + \rho a_{max,accel})^2}{2 a_{min,brake}} \right]_+$$
$$[x]_+ := max\{x, 0\}. \tag{2.1}$$

- d_{min}: Minimum distance to ensure that there is no crash with the obstacle.
- v_r: Current velocity of the SPIDER (m/s)
- ρ : Response time in seconds.
- $a_{max,accel}$: Maximum acceleration of the robot (m/s^2).
- $a_{max,brake}$: Minimum braking acceleration of the robot (m/s^2).

2.2 Operational Design Domain

The defined ODD for our use case is a simple one-way road as shown in Fig. 3 in which there is a static object located at the end of the road within the intended driving path of the robot vehicle. In the definition of this ODD, clear weather conditions as well as dry asphalt are assumed, but these conditions would be modified as a part of the test scenarios in the following sections to identify the triggering conditions of the function. The ego-vehicle, our robot, drives with constant velocity towards the object.

- D_{object}: Distance between the static object and the robot.
- D_{rss}: Distance calculated by using the RSS equation from (2.1).
- D_{ρ}: The distance that the robot travels until it perceives the object, makes a decision, and sends an actuation signal.
- D_{act}: The distance between the robot activates the actuation process and it is finished.
- D_{brake} the distance necessary to brake the robot until velocity is zero. $D_{brake} = D_{act} + D_{\rho}$. In nominal situations, $D_{brake} = D_{rss}$.
- $D_{perception}$: the perception range of the sensors. $D_{perception}$ should be longer than D_{object} and D_{brake}.
- μ: the friction factor of the road.

According to the technical parameters of the function (v_r: 50 km/h, ρ: 1 s, $a_{max,accel}$: 2.0 m/s^2, and $a_{max,brake}$: 5.0 m/s^2), and the used RSS model, is assumed that the maximum RSS minimum distance (D_{rss}) will always be smaller than the range of the perception sensors ($D_{rss,max} < D_{perception}$).

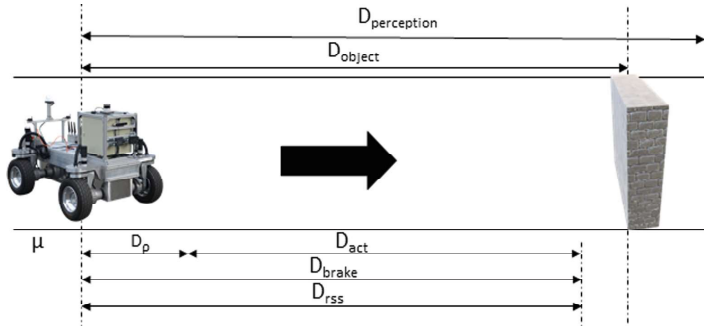


Fig. 3. Operational design domain (ODD) for the selected use case

2.3 Triggering Conditions List

According to the ISO21448 [5] the triggering conditions are the specific conditions of a scenario that serve as an initiator for a subsequent system reaction leading to hazardous behaviour. To identify the possible conditions that can bring up the function to unknown hazards, parameters of the defined ODD will be modified. An own database with the potential triggering conditions based on the taxonomy described in Annex C of the ISO21448 [5] and the BSI standard [27] was developed for this purpose. This database includes categories and different levels of subcategories, which adds granularity to each subcategory. For example, the main category is at the top (e.g. environmental conditions, road conditions,...) and the more granular conditions are at the bottom (e.g. heavy snow). The main category, environmental conditions in this example, follows other subcategories such as the *weather conditions* or *illumination* and, then, these subcategories are split again into other subcategories. In the case of *weather conditions*, this subcategory is split into rain, snow, fog and other climate conditions to continue with more fined granularity, which includes different intensities as light, medium and heavy.. The main purpose of using this approach is to provide granularity to cover the maximum variety of situations, which should help to discover more shortcomings of function during the analysis.

2.4 Relevant Scenarios

In our use case, a scenario is defined as the combination of the previously defined ODD combined with potential triggering conditions. Based on this definition, the relevant scenarios are the compilation of scenarios that includes the selected triggering conditions in which the output of the function will be analysed to know whether the selected triggering condition could lead to a hazardous behaviour. In this step, only the triggering conditions that makes senses for the defined ODD will be taken into account. For simplicity, in a first iteration only a single triggering condition is added to the ODD to better identify the cause and origin of the hazard.

2.5 Hazard Analysis and Risk Assessment

HARA methodology is always performed in Functional Safety processes in which associated hazards to the system are analysed and classified according to the level of danger and caused injuries. This classification used the Automotive Safety Integrity Level (ASIL) which uses the parameters, Exposure (E), which is the possibility of occurrence of the hazard within the operation situation; Severity (S), which is the estimation of the harm that the hazard might cause; and Controllability (C) that is ability to avoid the harm (see ISO262626 [3] part 3). But unlike the process implemented in fuctional safety standard, the ASIL level is not used in the SOTIF analysis process and only the identification of the hazards are used.

As a part of the HARA analysis and the given scenario, the following identified hazards can be extracted: *"a collision due to the inability to completely brake before reaching an obstacle"* and *"activation of the emergency brake due to a false object detection"*

2.6 Qualitative Analysis

In this step, the impact of the defined scenarios on the function has to be analysed to know if they could lead to a performance limitation which could further effect a hazard. To do this, first, the function is split into the three main categories: perception, decision, and actuation. The perception part is also divided into perception sense and perception algo, a shorter reference to perception algorithms. This differentiation is done to separate the part in which perception only provides data from the perceived reality through sensor input and perception algo in which information is extracted based on that perceived data to get any environmental information, e.g. objects, driving road. In this use case, the perception sense includes the sensor drivers, data fusion, and some data post-processes like filtering ground from the point clouds. The block perception algo includes the costmap generator, which generates an occupancy grid from the filtered point cloud to determine the objects that surround the SPIDER. Decision part includes the node, called collision detector, that goes through the generated occupancy grid and detects the closest object and its distance to the robot. The output of the node is the trigger of the emergency brake in case the object violates the calculated allowed minimum distance or the object distance if an object is detected but the emergency brake is not yet necessary. Finally, the actuation part is done by the commands to steering, braking and propulsion. In the given use case it includes the emergency brake from the SPIDER. The main reason for this categorization is to identify the components and their location in the system architecture in which the performance limitations could occur. For this reason, collaboration with domain experts and developers is crucial to better identify the origin of the hazards. A discussion has to be done for each selected scenario to identify how it could affect the function output. For example, for the scenario that includes medium or high snow level, these specific conditions may affect the perception sense part of the function, since it can include ghost points in the

perceived data, and the severity of the effect of the limitation is further rated by NO/Low/Medium/High (S0-S3). Therefore, this triggering condition could lead to the previously identified hazard *"activation of the emergency brake due to false object detection"*. Another specific scenario is the one in which the surface is slippery due to heavy rain. This kind of surface reduces the friction of the wheels (actuation part) and increases the distance necessary to completely stop the robot. This triggering condition has an impact on the surface friction of the ODD and could lead to the hazard *"a collision due to the inability to completely brake before reaching an obstacle"*. Similar discussions have to be done through the whole list of selected scenarios to identify the potential hazardous scenarios. An excerpt of this analysis can be shown in Fig. 4. In this use case, it was assumed the controllability (C) of the function is always C3 since it is a level 4 function and no backup operator is implemented.

			Risk Assessment				
Triggering Conditions	**Cause**	**Impact / Consequences on vehicle level**	**Severity (S0 - S3)**	**Exposure (E1-E4)**	**Intended reaction (System)** (Behaviour that can be realised)	**Deviating Behavior [Y/N]** (Does the intended Reaction not correspond to the Common Reaction? If yes, then a SOTIF measure is necessary.)	**SOTIF Measure** (Additional security measure to minimise the risk. The aim is to minimise exposure.)
F01_IC_PER_SEN_01							
[F01_L01] Limitation F01-01_Maximum_perception_range							
Heavy Snow (visibility (PR20) < 0.5 km AND precipitation rate (PR18) > 5,0 mm/hr)	The visibility of the lidar sensos might be drastically reduce to make that $D_{perception} < D_{object}$	a collision due to the inability to completely brake before reaching an obstacle	S2	E1	The robot should detect the object and brake before a collision with the object	Y	- Regarding the current HW setup, this is a limitation due to the selected sensor technology. Redundance with radars could increase the realibility of the function in this scenario
[F01_L02] Limitation F01-02_Reduction_detection_accuracy							
Heavy Fog (visibility (PR20) < 0,2 km)	The fog particles could be detected as objects, which generate false positives	activation of the emergency brake due to a false object detection	S1	E2	The function should be able to differenciate between object and enviromental noise to avoid the creating of false positives	Y	- Improving the algorithms for a more precise object detection. Include sensor redundance to make the fuction stronger to this effect
F02_IC_ACT_01							
[F02_L01] Limitation F02-01_Reduction_actuation_eficiency							
Heavy Rain (precipitation rate (PR18) > 7,6 mm/hr)	The rain wets the asphalt and could increase the necessary distance to completely stopped the robot ($D_{brake} > D_{ccc}$)	a collision due to the inability to completely brake before reaching an obstacle	S2	E1	The Drss guarantees the robot completely stops before reaching the object, avoiding the collision	Y	- Improving the algorithms to detect rainy scenarios and adapt the D_rss parameters to the current situation (e.g. decrease the $a_{min,brake}$)

Fig. 4. Triggering conditions analysis sheet excerpt

2.7 Testing and Evaluation

The first step in this process is the definition of the Key Performance Indicators (KPIs), which captures the behaviour in a nominal scenario and set the considered acceptable criteria to release the function. This nominal scenario, sometimes also called functional scenario, is the one used in optimal conditions without the inclusion of any of the triggering conditions. An example of a KPI could be Time To Collision (TTC) [28], which represents the time that the ego vehicle needs until it reaches the target vehicle or an obstacle according to the current ego and target velocities and its distance between the two vehicles. This metric can be defined by the formula (2.2). Similar to RSS formula (2.1), it is assumed that $v_{target} = 0$ since there is a static object instead of a target vehicle defined in the ODD.

$$TTC = d/(v_{ego} - v_{target}). \tag{2.2}$$

Afterwards, the results obtained in the different levels of the verification and validation procedures have to be compared with the nominal conditions and ensure they are within the defined acceptable limits.

2.8 SOTIF Modifications

The main goal of applying SOTIF measures is to improve the output of the function and minimise the risk level that applied triggering conditions might lead to. Following the goal of the principle of ALARP (As Low As Reasonably Practicable) [29], which assumes that the efforts to achieve zero risk could be not feasible and it has to be taken into account the necessary efforts needed to reduce the risk as much as reasonable. For example, there are some cases in which it is not possible to further risk reduction or the associated efforts to reduce it are not worth it. Returning to the proposed use case, in that case where a weather conditions is included as potential triggering conditions in the scenario as heavy snow, the visibility is drastically reduced. There is a chance that the snow could limit the perception range of the sensors and it can lead to the previously defined hazard: *"a collision due to the inability to completely brake before reaching an obstacle"* in which the relationship $D_{perception} < D_{brake}$ could apply to lead to the specified hazard. According to the situation, modifications have to be applied to improve the output on the specified scenario. For example, diversity of sensor technologies, which are stronger against the described issue or an external hardware to measure the specific triggering condition (fog level on the environment) could be applied to mitigate the occurrence of the hazard. In cases in which the effectiveness of the SOTIF measures are not enough to pass the defined acceptance criteria level, these scenarios have to be considered and evaluated within the *"Residual Risk Evaluation"* block.

2.9 Residual Risk Quantification

In case the maximum efforts to minimize the risks are reached and there is not other possible way or feasible way to further improve the function, the evaluation of the existing residual risk has to be calculated. The risk could be expressed in a function as follows:

$$R = f(S, O)$$

According to the given formula, the risk (R) can be defined by the severity (S) and probability of occurrence (O). In this function, the severity is the specified in the analysis for the associated hazard. On the other hand, finding the right occurrence probabilities of each triggering condition presents a more complex problem since there are a wide variety of situations from weather to road conditions. Moreover, finding or calculating these probabilities could change between countries, where the availability of some of them may not be publicly accessible, which makes the problem even harder. At the end of this block, a quantitative evaluation of each identified risk ($R_0 ... R_N$) should be provided. For example,

the number of kilometres or the number of hours until the function could lead to a hazard.

2.10 SOTIF Argumentation

Finally, the output of the previous blocks, which includes the carried out measures covered by functional modification and verification measures, such as simulations and testing, as well as the remained evaluated risks for each triggering condition, are included as inputs for the results of the SOTIF argumentation where both, reports (e.g. identified triggering conditions, relevant scenarios...) and metrics (e.g. number of hours until hazards behavior appears, KPIs,...), should be included.

3 Outlook

In this publication, a methodology to extract and evaluate the triggering conditions of an ADAS/ADS function is described. Since driven kilometres or simulations may not be enough in some situations, by using the introduced approach, a profound understanding of the system as well as the discovery of unknown hazardous scenarios could be provided. The addition of the triggering conditions analysis in theuse case helps the engineers to discover the ones that might lead to a hazard, and also provide more parameters in order to evaluate the risks of the function. For this purpose, the paper goes through all blocks of the proposed approach in which a description and a goal for each one is given as well as the relationship between them. The use case is illustrated by using the collision avoidance function implementation in our mobile robot HiL platform.

In future research works, the residual risk evaluation will be investigated in more detail to be able to provide a quantitative argument at the end of the process, which can provide specific metrics (e.g. validated number of kilometres) that will be used to characterize the safety of a function to achieve the demanded acceptance criteria. Moreover, in the following research, we want to extend this methodology in different approaches of the collision avoidance function to see how different sensor setups (e.g., RADAR+LIDAR, RADAR+LIDAR+CAMERA) can affect the function validation by combination of simulation and real-world testing.

Acknowledgments. Research leading to these results has been performed in the project FRACTAL A Cognitive Fractal and Secure EDGE based on an unique Open-Safe-Reliable-Low Power Hardware Platform Node, under grant agreement No. 877056. The project is co funded by grants from Germany, Austria, Finland, France, Norway, Latvia, Belgium, Italy, Switzerland, and Czech Republic and -Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU). The publication was written at Virtual Vehicle Research GmbH in Graz and partially funded within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG).

References

1. National Transportation Safety Board (NTSB). Collision between a sport utility vehicle operating with partial driving automation and a crash attenuator, Mountain View, California, 23 March 2018 (2020). www.ntsb.gov/investigations/AccidentReports/Reports/HAR2001.pdf
2. Bonnefon, J.-F.: 18 the Uber accident, pp. 93–98 (2021)
3. ISO26262: Road vehicles - Functional safety, International Organization for Standardization, Geneva, CH, Standard (2018)
4. ISO/SAE 21434:2021 Road vehicles - Cybersecurity engineering, International Organization for Standardization, Standard (2021)
5. ISO 21448: Road vehicles - Safety of the intended functionality, International Organization for Standardization, Geneva, CH, Standard (2021)
6. Khatun, M., Glaß, M., Jung, R.: Scenario-based extended hara incorporating functional safety and sotif for autonomous driving. In ESREL-30th European Safety and Reliability Conference, November 2020
7. Kinalzyk, D.: SOTIF process and methods in combination with functional safety. In: Yilmaz, M., Clarke, P., Messnarz, R., Reiner, M. (eds.) EuroSPI 2021. CCIS, vol. 1442, pp. 612–623. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85521-5_41
8. Martin, H., Winkler, B., Grubmüller, S., Watzenig, D.: Identification of performance limitations of sensing technologies for automated driving. In: 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE), pp. 1–6 (2019)
9. Ul4600 standard for safety: Evaluation of autonomous products, International Organization for Standardization, Standard (2021)
10. Kalra, N., Paddock, S.M.: Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability? Transp. Res. Part A Policy Pract. **94**, 182–193 (2016). www.sciencedirect.com/science/article/pii/S0965856416302129
11. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: an open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning, pp. 1–16 (2017)
12. Rong, G., et al.: Lgsvl simulator: a high fidelity simulator for autonomous driving, arXiv preprint arXiv:2005.03778 (2020)
13. Schlager, B., et al: State-of-the-art sensor models for virtual testing of advanced driver assistance systems/autonomous driving functions. SAE Int. J. Connected Autom. Veh. **3**(3), 233–261 (2020). https://doi.org/10.4271/12-03-03-0018
14. Pilz, C., Steinbauer, G., Schratter, M., Watzenig, D.: Development of a scenario simulation platform to support autonomous driving verification. In: 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE), pp. 1–7 (2019)
15. Abdulazim, A., Elbahaey, M., Mohamed, A.: Putting safety of intended functionality sotif into practice. In: SAE WCX Digital Summit. SAE International, April 2021. https://doi.org/10.4271/2021-01-0196
16. Vaicenavicius, J., Wiklund, T., Grigaite, A., Kalkauskas, A., Vysniauskas, I., Keen, S.D.: Self-driving car safety quantification via omponent-level analysis. SAE Int. J. Connected Autom. Veh. **4**(1), 35–45 (2021). https://doi.org/10.4271/12-04-01-0004

17. Scholtes, M., et al.: 6-layer model for a structured description and categorization of urban traffic and environment. IEEE Access **9**, 59131–59147 (2021)
18. Jama and Sakura. Automated driving safety evaluation framework ver. 1.0 guidelines for safety evaluation of automated driving technology (2021). www.jama-english.jp/publications/Automated_Driving_Safety_Evaluation_Framework_Ver1.0.pdf
19. ASAM e. V. ASAM OpenODD (2021). www.asam.net/project-detail/asam-openodd/
20. Ito, M.: Odd description methods for automated driving vehicle and verifiability for safety. JUCS J. Univ. Comput. Sci. **27**(8), 796–810 (2021). https://doi.org/10.3897/jucs.72333
21. Mekki-Mokhtar, A., Blanquart, J.-P., Guiochet, J., Powell, D., Roy, M.: Safety trigger conditions for critical autonomous systems. In: 2012 IEEE 18th Pacific Rim International Symposium on Dependable Computing, pp. 61–69 (2012)
22. Expósito Jiménez, V.J., Schwarzl, C., Martin, H.: Evaluation of an indoor localization system for a mobile robot. In: 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE), pp. 1–5 (2019)
23. Virtual Vehicle Research GmbH. SPIDER: mobile platform for the development and testing of autnomous driving functions (2021). www.v2c2.at/spider/
24. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, SAE International, Standard (2021)
25. Shalev-Shwartz, S., Shammah, S., Shashua, A.: On a formal model of safe and scalable self-driving cars (2018)
26. Gassmann, B., et al.: Towards standardization of AV safety: C++ library for responsibility sensitive safety. In: 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 2265–2271 (2019)
27. BSI PAS 1883:2020-operational design domain (ODD) taxonomy for an automated driving system (ADS). Specification, The British Standards Institution, Standard (2020)
28. Balas, V.E., Balas, M.M.: Driver assisting by inverse time to collision. In: 2006 World Automation Congress, pp. 1–6 (2006)
29. Malekzadeh, M., Bate, I.: Making an ALARP decision of sufficient testing. In: 2014 IEEE 15th International Symposium on High-Assurance Systems Engineering, pp. 57–64 (2014)

A.4 Safety of the Intended Functionality Concept Integration into a Validation Tool Suite

Safety of the Intended Functionality Concept Integration into a Validation Tool Suite

Víctor J. Expósito Jiménez*, *Bernhard Winkler, Joaquim M. Castella Triginer
Virtual Vehicle Research GmbH, Graz, Austria; email: {victor.expositojimenez,bernhard.winkler, joaquim.castellatriginer}@v2c2.at
Heiko Scharke, Hannes Schneider
AVL List GmbH, Graz, Austria; email: {heiko.scharke, hannes.schneider}@avl.com
Eugen Brenner, Georg Macher
Graz University of Technology, Graz, Austria; email: {brenner, georg.macher}@tugraz.at

Abstract

Nowadays, the increasing complexity of Advanced Driver Assistance Systems (ADAS) and Automated Driving (AD) means that the industry must move towards a scenario-based approach to validation rather than relying on established technology-based methods. This new focus also requires the validation process to take into account Safety of the Intended Functionality (SOTIF), as many scenarios may trigger hazardous vehicle behaviour. Thus, this work demonstrates how the integration of the SOTIF process within an existing validation tool suite can be achieved. The necessary adaptations are explained with accompanying examples to aid comprehension of the approach.

Keywords: SOTIF, Safety of the Intended Functionality, Scenario Validation, ADAS/ADS

1 Introduction

Scenario validation plays a significant role in the entire vehicle validation process as an increasing number of features and safety systems rely on sensors. Unlike functional safety [1] or cybersecurity [2], which covers failures and malfunctions, and external attacks respectively, the Safety of the Intended Functionality (SOTIF) standard [3] focuses on the technical shortcomings and human misuses that may result in hazardous behaviour at vehicle level. Its focus is to increase the identification of hazardous scenarios to be validated as well as to minimise the area in which unknown hazardous scenarios could appear. Figure 1 shows the cause-and-effect model in which is depicted how a potential triggering condition could result in a hazardous behaviour at the end of the process. According to the ISO21448, a triggering condition is a *"specific condition of a scenario that serves as an initiator for a subsequent system reaction contributing to either a hazardous behaviour or an inability to prevent or detect and mitigate a reasonably foreseeable indirect misuse. The concept of 'triggering' includes the possibility that there can be multiple conditions that can gradually happen, leading to hazardous behaviour or the inability to prevent or detect and mitigate a reasonably foreseeable misuse. The term "potential triggering condition" can be used when the ability to initiate a corresponding reaction is not yet established"*. Another concise definition is given in [4], where a triggering condition is defined as *"an external condition (relative to ego-vehicle) in a scenario that triggers one or multiple functional insufficiencies and further results in hazardous behaviour. They are system-dependent was well"*. The SOTIF standard also defines a performance insufficiency as a *"limitation of the technical capability contributing to a hazardous behaviour or inability to prevent or detect and mitigate reasonably foreseeable indirect misuse when activated by one or more triggering conditions*. Examples of performance insufficiencies could be the limitation of the actuation or the perception range of the sensor used detect objects. Consequently, a functional insufficiency is defined as an *insufficiency of specification or performance insufficiency*. Finally, the definition of hazard is adapted from the given in the ISO26262, *"potential source of harm caused by malfunctioning behaviour of the item"*. SOTIF standard replaces the word "malfunctioning" by "hazardous" and the phrase "of the item" by "at the vehicle level" in comparison with the given by the ISO26262 to adapt the definition to the scope of the standard. For clarification, the insufficiencies of specification are out of the scope of the related project to this work, therefore, a functional insufficiency is considered the same as a functional insufficiency in this work as is shown in the box of the project scope in Figure 1. The inclusion of a triggering condition could start a reaction in the system that could activate a functional insufficiency and could finally result in a hazardous behaviour. A hazardous behaviour is defined based on the result of the Key Performance Indicators (KPIs) or Safety Performance Indicators (SPIs). These metrics are used to discern if the result of the tests is within a defined tolerable value or, on the other hand, is outside the tolerable windows and is set as a hazardous behaviour. A KPI is a metric that is used for measure a specific parameter of the system. In a similar way, a SPI defines a metric

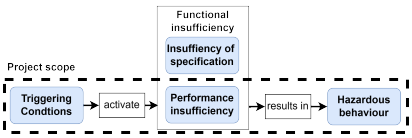


Figure 1: Cause-and-effect model from ISO21448

but focused on the safety domain such as the Minimum Safe Distance Violation (MSDV) or the Time-to-Collision (TTC) [5]

The Operational Design Domain (ODD) is a crucial concept in the scenario validation process. While it is defined in the SOTIF standard, the definition provided in the UL4600 standard [6] outlines more precisely what an ODD constitutes from our perspective. As per this standard, an ODD refers to *"the set of environments and situations the item is intended to operate within. This includes not only direct environmental conditions and geographic restrictions, but also a characterization of the set of objects, events, and other conditions that will occur within that environment"*. The scenario development for the ODDs utilises the widely accepted methodology of The 6-Layer Scenario Model [2] [7]. This model splits the definition of each scenario into six layers, each concentrating on the context of the scenario. The layers and their definitions are:

- *Layer 1 – Road network and traffic guidance objects*: e.g. road markings, and traffic signs and traffic lights.
- *Layer 2 – Roadside structures*: e.g. buildings, vegetation, streets lamps, and advertising boards.
- *Layer 3 – Temporary modifications of L1 and L2*: e.g. roadwork signs, temporary markings, and covered markings.
- *Layer 4 – Dynamic Objects*: e.g. vehicles (moving and non-moving), pedestrians (moving and non-moving), trailers, and animals.
- *Layer 5 – Environmental Conditions*: e.g. illumination, precipitation, and road weather.
- *Layer 6 – Digital information*: e.g. state of traffic lights, switchable traffics signs, and V2X messages.

In this context, attempts have been made to define a taxonomy that can describe most scenarios in the most detailed manner. For example, the one from the British Standards Institution [8] or the taxonomy from the Society of Automotive Engineers (SAE) [9] are widely recognized. Additionally, Annex B of the SOTIF standard also addresses this matter.

Once the main concepts have been outlined, the following section describes the integration process into the validation tool. Finally, section 3 provides a summary of all the work presented in this publication and establishes the direction for future research.

2 SOTIF Concept Integration Process

The integration has been implemented within the AVL SCENIUS™ [10] tool suite that is developed with the scenario-based validation approach in mind. The complete validation process is covered, from scenario design to scenario management, test case generation, test allocation, and result reporting. The suite is based on three main tools modules. First, the scenario designer allows the user to handle all aspects of the scenario including the parametrisation. It fully supports ASAM OpenScenario [11] and OpenDrive [12]. All scenarios are immediately verified for standard conformity as well as by the enhanced data and logic checks. Then, the user could manage all stored scenarios in the scenario data manager. All the elements relevant for the sufficient description of scenarios such as road content, traffic content, and other environmental data are managed and stored in a central database. Finally, the test case generator provides the user the possibility of defining test orders in a simulation or transfer to another different execution environment. The implemented smart testing algorithms enable the automatic reduction of the vast amount of test cases and parameter variations. In addition to the main benefits provided by the tool suite such as time-cost saving, efficiency, fast integration and traceability; the inclusion of the SOTIF concept extends and improves the identification and validation of both unknown and known hazardous scenarios of a ADAS/AD function to obtain a more precise safety argumentation.

An ontology is used to describe the scenarios that will be used for testing. ASAM OpenXOntology [13] is used as a reference, but modifications are included to better fit the requirements of the tool chain. Internally, the ontology and its relationships are defined by using four kind of entities:

- *Node:* A node is the entity in which the hierarchy of the ontology is built. It can be a child of another node, or a parent for Enums or Params. Examples of nodes are the ambient or weather conditions of the scenario, which are the parents of scenario parameters such as rain or illumination parameter.
- *Enum:* Defines a list of values that are related to each other. For example, an enum is the snowfall condition, which is defined by three different levels of severity: heavy snow, light snow, and moderate snow.
- *Value:* Defines an entity that is an abstraction of a phenomenon, but it is not yet modelled. In this case, the phenomenon already exits in the system but has not yet been parametrised. As example, this type of entity are the one previously mentioned: heavy snow, light snow, and moderate snow.
- *Param:* Defines an entity that can be quantified. Each one is associated with an unit to be measured. For example, scenario illuminance, which is associated with lux units.

The integration of the SOTIF concept requires the addition of a new node in the system, which is the parent of all

Figure 2: Updated cause-and-effect model including the scenarios constraints

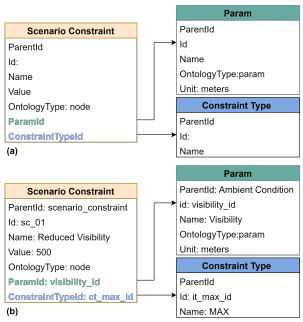


Figure 3: Definition of a scenario constraint

triggering conditions defined in the validation tool. The next integration step is to linking the triggering conditions with the existing defined ontology to be able to parametrise each triggering condition and include them into the test scenario. Therefore, an intermediate block has to be added to the cause-and-effect model, which is shown in Figure 2. In this new approach, a scenario constraint block is added, which connects the triggering conditions with the performance insufficiencies. By using this approach, we are able to define each triggering conditions as a combination of one or many scenario constraints. Following the scenario constraint parametrisation shown in Figure 4(a), a scenario constraint is a node entity that is linked to a *Param* and a new entity called *Constraint Type*, which sets the type of constraint such as a maximum or a minimum value. The *Param* entity is associated to the existing ontology of the system. For example, a heavy snow condition is defined as a visibility limitation up to 500 meters according to BSI PAS 1883 standard [8]. This scenario constraints is defined in Figure 4(b), where relationship and value is given in the *Node* entity that, in turn, is associated to the *Param* from the ontology (visibility) and the type of constraint (MAX). The complete hierarchy and relationship tree of the system for this specific triggering condition is given in Figure 4.

The SOTIF concept is designed with scalability in mind due to the test scenario shall increase the complexity of the triggering conditions and their parametrization. The Figure 5 shows the triggering condition *Heavy Snow during Night-time* as an example of this increasing complexity. These types of triggering conditions are treated as the combination of two independent triggering conditions: *Heavy Snow* + *Night-time*. In contrast to the previous example, the weather condition *Heavy Snow* is more finely parametrised. In this context, not only is the impact in the visibility is considered, but also the effect on the scenario illumination and the asphalt friction. Following the standards again, the illuminance in a heavy snow scenario can be parametrised from 1 lux to 2000 lux. Additionally, a reduction factor of 0.8 is applied on the asphalt friction in this potential triggering condition. However, this particular triggering condition occurs during night-time, therefore, the illuminance condition due to night-time is also applied (illuminance less than 1 lux according to the standards). Therefore, there are two illuminance constraints in this triggering condition parametrization. In this situation, the most limiting conditions is applied. It means that in this definition, the illuminance parametrised in night-time overrides the value of the illuminance parametrised in the heavy snow condition.

Finally, when one or more potential triggering conditions are selected in the scenario for testing, the scenario constraints (e.g., limited visibility, reduced friction...) associated to each potential triggering condition are also included in the generated test cases. The resulting metrics of the matrix test cases show the impact of the selected potential triggering conditions on the function, which are compared with the nominal performance of the function (i.e., no potential triggering conditions included) to determine them not longer as potential but triggering conditions for the function, and to identify the thresholds at which they are relevant to impact and effect on the function output.

3 Conclusions and future work

In this publication, the integration of a SOTIF concept has been explained, where some adaptations and parametrisation of the scenario constraints have to be done in order to integrate triggering conditions into an existing scenario ontology. As a first step, an extensive list of potential triggering conditions has been investigated based on current state-of-the-art and available standards. They are then parametrised by using an existing system ontology, which is used to model the scenarios and defining the needed entities and relationship to be able to link the triggering conditions, scenario constraints, and the existing ontology.

As a future task, we will define the triggering conditions that cannot be parametrised using existing standards. Furthermore, we are researching a methodology that will allow us to capture the majority of potential triggering conditions from the perception side based on the performance insufficiencies due to the infinite number of triggering conditions in the real world, which are not possible to cover manually.

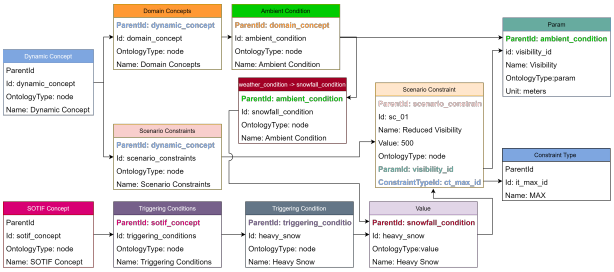


Figure 4: Heavy Snow triggering condition in SCENIUS tool suite

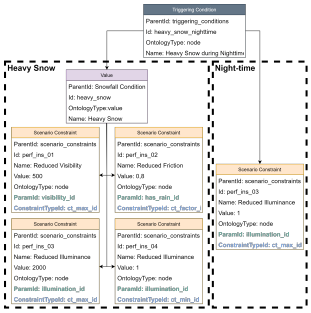


Figure 5: Extended Triggering Condition: "Heavy Snow during Night-time"

4 Acknowledgement

The publication was written at Virtual Vehicle Research GmbH in Graz, Austria. The authors would like to acknowledge the financial support within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Labour and Economy (BMAW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorised for the programme management. They would furthermore like to express their thanks to their supporting industrial project partner, namely AVL List GmbH.

References

[1] "ISO26262: Road vehicles — Functional safety," standard, International Organization for Standardization, Geneva, CH, 2018.

[2] "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," standard, SAE International, 2021.

[3] "ISO 21448:2022 Road vehicles — Safety of the intended functionality," standard, International Organization for Standardization, Geneva, CH, 2022.

[4] Z. Zhu, R. Philipp, C. Hungar, and F. Howar, "Systematization and identification of triggering conditions: A preliminary step for efficient testing of autonomous vehicles," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, pp. 798–805, 2022.

[5] J. Vaicenavicius, T. Wiklund, A. Grigaite, A. Kalkauskas, I. Vysniauskas, and S. D. Keen, "Self-driving car safety quantification via component-level analysis," *SAE International Journal of Connected and Automated Vehicles*, vol. 4, pp. 35–45, mar 2021.

[6] "Ul4600 standard for safety: Evaluation of autonomous products," standard, International Organization for Standardization, 2021.

[7] M. Scholtes, L. Westhofen, L. R. Turner, K. Lotto, M. Schuldes, H. Weber, N. Wagener, C. Neurohr, M. H. Bollmann, F. Körtke, J. Hiller, M. Hoss, J. Bock, and L. Eckstein, "6-layer model for a structured description and categorization of urban traffic and environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021.

[8] "BSI PAS 1883:2020-Operational design domain (ODD) taxonomy for an automated driving system (ADS). Specification," standard, The British Standards Institution, 2020.

[9] "AVSC Best Practice for Describing an Operational Design Domain: Conceptual Framework and Lexicon," standard, SAE International, 2020.

[10] AVL List GmbH, "AVL SCENIUS." `https://www.avl.com/en/testing-solutions/automated-and-connected-mobility-testing/avl-scenius`, 2023. Accessed: 2023-09-17.

[11] "ASAM OpenScenario," standard, ASAM e. V., 2022.

[12] "ASAM OpenDrive," standard, ASAM e. V., 2022.

[13] "ASAM OpenXOntology," standard, ASAM e. V., 2022.

A.5 SOTIF Validation for ADS by using Perception Performance Insufficiencies Injection

vehicles

MDPI

Article

Safety of the Intended Functionality Validation for Automated Driving Systems by Using Perception Performance Insufficiencies Injection

Víctor J. Expósito Jiménez [1,*], Georg Macher [2,*], Daniel Watzenig [1,3] and Eugen Brenner [2]

[1] Virtual Vehicle Research GmbH, 8010 Graz, Austria; daniel.watzenig@tugraz.at
[2] Institute of Technical Informatics, Graz University of Technology, 8010 Graz, Austria; brenner@tugraz.at
[3] Institute of Computer Graphics and Vision, Faculty of Computer Science and Biomedical Engineering, Graz University of Techhnology, 8010 Graz, Austria
* Correspondence: victor.expositojimenez@v2c2.at or victor.expositojimenez@student.tugraz.at (V.J.E.J.); georg.macher@tugraz.at (G.M.)

Abstract: System perception of the environment becomes more important as the level of automation increases, especially at the higher levels of automation (L3+) of Automated Driving Systems. As a consequence, scenario-based validation becomes more important in the overall validation process of a vehicle. Testing all scenarios with potential triggering conditions that may lead to hazardous vehicle behaviour is not a realistic approach, as the number of such scenarios tends to be unmanageable. Therefore, another approach has to be provided to deal with this problem. In this paper, we present our approach, which uses the injection of perception performance insufficiencies instead of directly testing the potential triggering conditions. Finally, a use case is described that illustrates the implementation of the proposed approach.

Keywords: SOTIF; scenario-based validation; performance insufficiencies; triggering conditions; ADS

Citation: Expósito Jiménez, V.J.; Macher, G.; Watzenig, D.; Brenner, E. Safety of the Intended Functionality Validation for Automated Driving Systems by Using Perception Performance Insufficiencies Injection. *Vehicles* **2024**, *6*, 1164–1184. https://doi.org/10.3390/vehicles6030055

Academic Editors: Sijing Guo, Bin Wang, Quan Zhou and Bin Shuai

Received: 19 April 2024
Revised: 18 June 2024
Accepted: 20 June 2024
Published: 4 July 2024

1. Introduction

Automated Driving Systems present a new challenge in the field of safety argumentation due to the complexity of validation, because this requires covering not only malfunctions but also scenario conditions and complex algorithms, greatly increasing the effort involved in obtaining quantitative evidence that ensures the safety of systems. The validation of ADAS/AD functions is shifting from component-based validation to a more scenario-based validation. Unlike component-based validation, which ensures that all components are working properly (i.e., no faults or malfunctioning), scenario-based validation adds the focus in cases when everything works as intended but different situations and components of the scenario could create a situation that may lead to hazardous behaviour.

1.1. Safety Validation

New regulations [1] require the provision of evidence from the validation process to obtain authorisation for driving on public roads and, more importantly, to avoid accidents [2–5] that occurred in the past. To obtain a better picture of safety and validation processes in the domain of autonomous vehicles, the authors [6,7] give an overview of the current situation, describing the main requirements and concepts. Standards such as UL4600 [8] also provide a list of all the required evidence to ensure the validation of a system. Other standards are being defined to explain this new safety domain in which scenarios are much more relevant. The Safety Of The Intended Functionality (SOTIF) is defined in the standard ISO21448:2022 [9] and covers the validation of hazards that are not initiated by a malfunction in the system but by misuse and technical shortcomings. The

standard introduces the concept of potential triggering conditions, which are the conditions from a scenario that could cause the system to exhibit hazardous behaviour. As part of a SOTIF validation, potential triggering conditions must be covered, but the huge number of possible scenario situations makes testing all possible potential triggering conditions an unmanageable task. In order to fully understand the work described here, some concepts and terminology need to be clarified. A triggering condition is defined in [9] as a specific condition of a scenario that starts a reaction in the system contributing to hazardous behaviour. Potential could be included as a prefix when it is not yet validated, but experts see evidence that it could turn out to be a triggering condition in the end. A functional insufficiency is defined as an insufficiency of specification or performance insufficiency. The scope of the work is only focused on performance insufficiencies; therefore, the insufficiencies of specification, which are the initiators of unintentional misuse, are not considered in this publication. A performance insufficiency is defined as a limitation of technical capability contributing to hazardous behaviour when activated by one or more triggering conditions. An output insufficiency is an insufficiency on a functional level and, like the other insufficiency, can be activated by one or more functional insufficiencies or triggering conditions. Hazardous behaviour is defined as the behaviour of a system that is not within the specified acceptance criteria. The acceptance criteria could be defined on the basis of different Key Performance Indicators (KPIs) or, in the case of safety, Safety Performance Indicators (SPIs). An overview of the currently available metrics to define these criteria is given in [10]. Additionally, the authors in [11] present a survey of the current standards related to safety in the automated driving domain, including the definition of common perception failures and relevant metrics to evaluate perception systems. Other authors [12] also give an overview of the current standards but focus on SOTIF. This research work shows the relationship between SOTIF and other standards as well as its implementation in the verification and validation process. Another safety standards overview is provided in [13] with a focus on object-based environment perception. The relationship between these concepts is shown in Figure 1, which illustrates how a triggering condition may start in a potential triggering condition (heavy fog) and lead to hazardous behaviour (unintended braking). Another example could occur if a vehicle leaves a tunnel. If an ADS only relies on a camera as a perception sensor, this camera has some moments with extremely high-contrast images that may impact the behaviour of the ADS. In an attempt to cover the topic of triggering conditions, the authors in [14] present a systematisation and identification of triggering conditions, providing a categorisation to better handle them, which also gives us a better understanding of the concept in this context. This topic was also covered in our previous work [15], where the process of testing the triggering conditions was explained, but the realisation of the impossibility of covering all triggering conditions made us change the direction of our research to validate an ADS by using perception performance insufficiencies instead.

Accounting for the scenario side, one of the main goals of the SOTIF is to minimise the scenarios that could be hazardous. Unlike ASIL methodology from the Functional Safety Standard (ISO26262) [16], a SOTIF validation does not provide a classification according to a specific metric. In SOTIF, the validation of the ADS should improve in each iteration due to minimising the already known hazardous scenarios or discovering new hazardous scenarios, which change with every iteration. Figure 2 shows how the scenarios are divided into four main areas: from safe and known scenarios to the worst-case scenario, unsafe and unknown scenarios. As depicted in this figure, each validation iteration must reduce the number of unsafe unknown scenarios, firstly validating the already known unsafe and safe scenarios and secondly discovering new scenarios that could be unsafe and unknown. Unfortunately, the area of hazardous unknown scenarios cannot be completely accounted for, as such scenarios can always occur. The rest have to be treated as a residual risk of ADS.

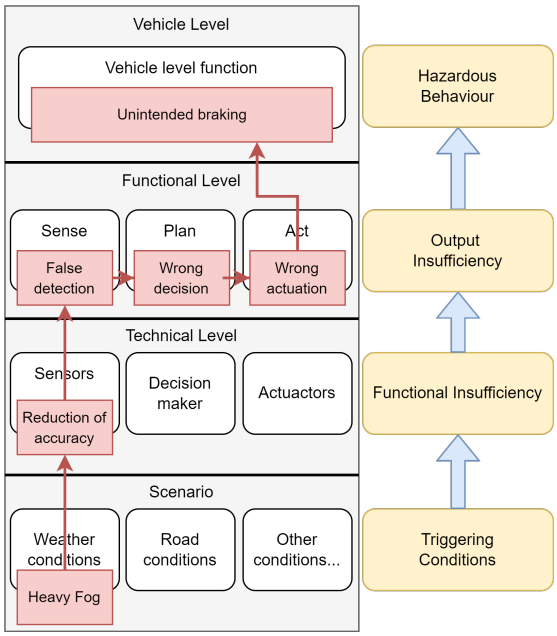


Figure 1. Cause and effect model between potential functional insufficiencies and triggering conditions.

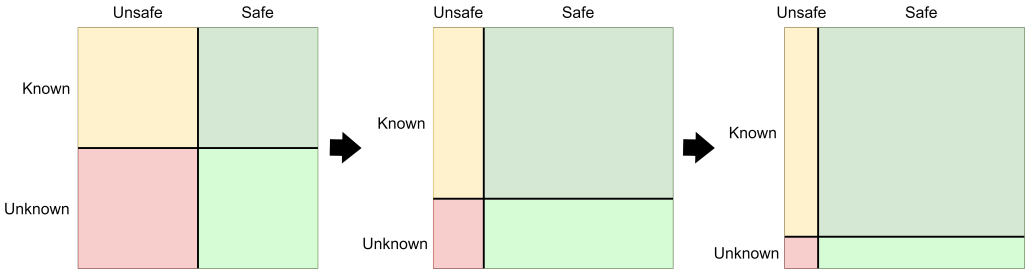


Figure 2. Evolution of the scenario categories through the SOTIF validation iterations.

This approach is not only focused on the performance insufficiencies injection but also on finding a way to discover and identify edge scenarios. A further classification of scenarios [17] is also given according to their occurrence or hazardous level. For example, corner scenarios are scenarios that occur in rare conditions with normal operational parameters (e.g., low sun angle, ice-covered road, etc.). On the other hand, an edge case is also a scenario that occurs in rare situations but with the presence of extreme values. Thus, not all corner cases are edge cases and vice versa. A nominal scenario means a traffic scenario containing situations that reflect regular and non-critical driving manoeuvres according to [18], which also defines a critical scenario as one that needs an emergency manoeuvre to avoid harm or react to a system failure.

A key concept arises in scenario-based validation, where the description of the scenario as well as the domain in which the system works properly has to be defined. In this context, the Operational Design Domain (ODD) defines all scenario situations in which an ADS is designed to work safely. ISO34502 [19] provides the principles to define an

ODD, where a fine-grained description of the scenario plays a crucial role in this task. Many taxonomies [20,21] together with the previously cited [9,19] have tried to reduce the gap presented by this issue, describing various aspects from different levels of weather conditions (wind, snow, etc.) to road topologies. The authors in [22,23] describe how to maximise ODD coverage in the scenario validation process. The Pegasus project [24] proposed the six-layer scenario model, which categorises each scenario into layers according to the kind of actors and their functionality in each case. Furthermore, ASAM OpenODD [25] provides the necessary syntax to include a defined ODD in the software simulations and be able to carry out the validation. The syntax also allows different definitions to be reused, shared and combined, providing greater flexibility and collaboration between partners in the development of ADS as required.

1.2. Sensor Models

Although there are many Hardware-In-Loop (HIL) platforms [26,27] for collecting data to validate Automated Driving Systems, two main questions arise when using this approach. The first one is starting to think how much real-world data would be enough to validate the ADS [28]; the second is the impossibility of collecting data for all types of possible scenarios, such as different types and levels of weather conditions [29,30]. Therefore, virtual validation is the more feasible way to validate an ADS. Typically, an Automated Driving System is built on three main blocks: sense, plan, and act. Figure 3 shows each block and the relationship between them. The sense block carries out the perception of the environment using sensors such as cameras, lidar or radar to perceive what is happening in the current environment. In our approach, the sense block is also split into two different sub-blocks: the perception and algo blocks. The sense-perception block models the observed reality (e.g., the environment) based on sensor inputs. For example, the point cloud generated by a lidar sensor is based on the perceived environment. On the other hand, the sense-algorithm block is responsible for extracting information from the input of the perception block (e.g., an object list from the generated point cloud). The plan block is in charge of the decision making according to the input from the perception sensors and the defined functionality, triggering the necessary actions according to the situation. Finally, the act block executes the actions decided upon, for example, braking or turning. This model is also referred to by different names such as Sense–Decide–Act or Perception–Decision–Actuation, but the meaning of each block remains the same.

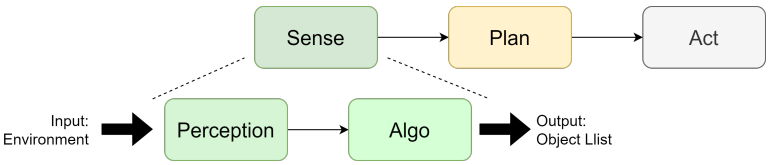


Figure 3. Sense—Plan—Act Model.

There are many research works focused on the study of the behaviours of diverse sensor technologies in different harsh environments. The authors in [31] present the current state of sensor models for virtual validation, including an explanation of the different types of sensor model fidelities. In this context, many research works have focused on the behaviour of different sensor technologies and on developing error models that model sensor performance in harsh environments. For example, the authors in [32–34] provide an in-depth analysis of the performance of camera and lidar technologies in adverse weather conditions such as foggy environments. A fog error model for a point cloud generated by lidar sensors was developed in [35]. The same authors have continued the work by modelling lidar performance insufficiencies in snowfall conditions [36]. Another example can be found in the following work [37], where the author develops a library to edit the point cloud generated by a lidar, including effects such as cropping or added reflection.

This research is also developed in the context of security, as these additions can be used as a spoofing attack by external actors. Machine learning has also been used to develop sensor error models, as shown by [38]. The authors explain their process for incorporating rain into the image frames produced by a camera. The aforementioned research works and tools could be used as a component of our approach to include perception performance insufficiencies into the system to be validated. One of the issues of using high-fidelity models is the increased computing demand of the simulation; the proposed approach in [39] mitigates this problem, albeit by using low-fidelity models, that show similar behaviour with lower computing costs. Andrea Piazzoni et al. in [40–42] conduct extensive research regarding perception error models. Similar to this paper, the authors remark on the importance of considering perception errors in the virtual validation process and its absence in current techniques. They proposed a perception error model and the guidelines to be included in the simulation pipeline. Moreover, the model is implemented using different sensor configurations, showing how each configuration is related to system safety. Perception errors are also utilised for virtual validation by using adversarial attacks [43] in the approach proposed in [44], concluding that these attacks, although seemingly harmless, have an impact on the final behaviour of the system. Another approach of RGB-camera perception error models is proposed in [45] to estimate rare failure probabilities used to learn high-likelihood failure trajectory distributions.

1.3. Risk Evaluation for Autonomous Vehicles

With regard to quantitative risk assessment, there have been recent approaches in this area. The first related work is proposed in [46], which includes a well-described list of deficiencies in the standard and possible corrections. It also gives a brief idea of how they could use a statistical approach for SOTIF validation. The authors in [47] propose an approach in which they can give quantitative values to each category of the HARA analysis (exposure, controllability and severity) from ISO26262 and calculate the risk of the ADS. Based on these values and statistical approaches, they can define the probability of risk of some extracted scenarios with and without triggering conditions. Unlike our approach, which is focused on full virtual validation, this approach focuses on using real-world data for validation. Another approach to calculating the risk is given in [48], where the authors use a fault tree analysis and HARA to provide a quantitative metric of an ADS. The approach given in [49] uses one-side binomial and Poisson distribution, but the authors recognise that the given analysis is greatly simplified to a model of a specific ADS. Moreover, no false positive or perception triggering conditions are included. In addition to the benefits of facilitating modular design, this approach makes it possible to demonstrate that sufficient safety conditions are met at the component level, using data sets of reduced size and therefore cost compared with those required for validation by vehicle-level road tests. As a disadvantage, the very specific scenario complicates the inclusion of the methodology beyond the described scenario. This paper [50] proposes another statistical validation method that uses reinforcement learning to identify the scenarios that lead to a system outcome outside the acceptance criteria. This approach reduces the number of necessary scenario simulations needed to validate the collision avoidance system from the publication. A perception validation methodology using failure rate probabilities is given in [51]. Similar to our approach, it uses the Responsibility-Sensitive Safety (RSS) [52] area as a main area to focus on in validation, but the approach only takes into account the perception component and not the impact in the complete system. Another research work in this field is [53]. The authors in this publication develop a system to monitor, quantify and mitigate SOTIF risks. The methodology is validated through an HIL platform that uses AI algorithms in the perception system.

1.4. Structure of the Article

The structure of the publication is as follows: Section 2 explains how perception performance insufficiencies are implemented in our approach, providing a classification of

these performance insufficiencies according to the impact on the system and the perception technology. The next section describes the risk evaluation, where performance insufficiency injection is used to provide a quantitative metric that can be used to evaluate the system. Section 4 illustrates two use cases in which the proposed approach is put into practice. In the first use case, the risk evaluation for a system with a visibility insufficiency is described, showing how this insufficiency impacts the output of the function and the associated risk. Additionally, the second use case shows the impact and the associated risk of an accuracy insufficiency in the same system. Finally, Section 5 summarises the methodology proposed in this publication and outlines the next steps of this research.

2. Perception Performance Insufficiencies Injection

As previously stated, validating all potential triggering conditions of the system is an unmanageable task. Therefore, this work focuses on the performance insufficiencies rather than the reasons behind them. The approach is centred on the impact on the ADS: for example, validation of an ADS when a scenario includes a tunnel. The validation strategy could involve generating test cases that involve all possible types of tunnels, also accounting for sizes and materials. Based on this validation strategy, including all definitions of tunnels and scenario conditions creates an almost infinite number of test cases, making the validation infeasible. Therefore, our approach goes directly to the impact that triggering conditions could generate in the system: extremely high-contrast images for a short period of time in a camera-based ADS. This strategy validates not only the specific triggering condition but also the unknown scenarios that could affect it in a similar way. This contributes to reaching one of the main goals of SOTIF, which is to minimise the unknown scenarios that are both hazardous and non-hazardous.

The perception component of an ADS includes different sensors and technologies that can vary the perceived reality. In our approach, performance insufficiencies are given at a high level of abstraction. Therefore, when a performance insufficiency is injected, it is related to the raw data that the sensor provides as input to the system. For example, if a lidar performance insufficiency is implemented, it uses point cloud messages for the injection, or image frames if a camera is used. The architecture of the approach is shown in Figure 4, which shows that the injection is added to the raw data of the sensor before it is included in the ADS system. Although in our approach only perception performance insufficiencies are considered, the effect of the injection could appear in any block of the ADS of Figure 3: sense, decision and actuation.

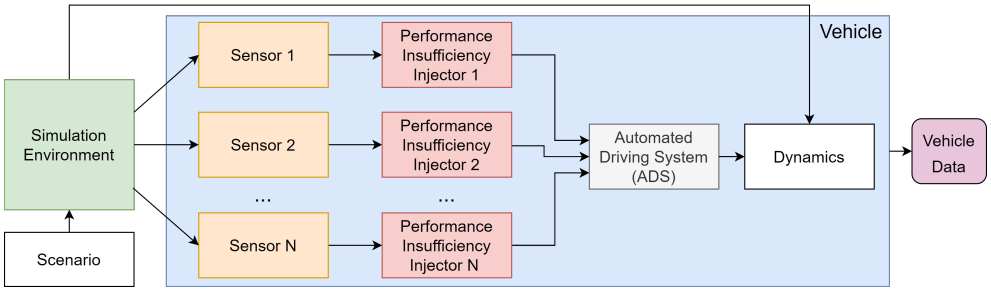


Figure 4. Architecture of the injection approach.

To implement this approach, a classification of performance insufficiencies is defined first, with a main category defining the general impact of the insufficiency. Then, each of these categories serves as a parent of the performance insufficiencies defined by a specific technology as well as the insufficiencies modelled for specific triggering conditions. Thus, the performance insufficiencies in our approach are classified as follows:

- Generic Performance Insufficiency (GPI): This refers to a general performance insufficiency that is not related to any specific sensor technology but rather the impact on sensor perception. It is used as a general category for performance insufficiencies. Table 1 shows an excerpt of some identified performance insufficiencies, also describing their impact on the sensor.
- Technology Performance Insufficiency (TPI): In these insufficiencies, the defined generic performance insufficiencies are modelled for a specific technology. For example, the reduction of field of view performance insufficiency from the GPI table could be defined for the lidar technology as cropping in the point cloud message provided by the lidar sensor function. Thus, if the visibility of the sensor is limited to a specified distance, the injector will remove the points farther than this distance. Table 2 shows an excerpt of the performance insufficiencies for the lidar technology and how they are modelled in the system.
- Triggering Condition Performance Insufficiency (TCPI): This is a performance insufficiency that was modelled for a specific triggering condition and technology, such as the lidar snowfall modelling from [36] or camera rain models from [38]. This category also includes the defined taxonomies from the standards (SAE [54], BSI [21], SOTIF [9], etc.) that could be set as triggering conditions in the validation process. For example, visibility in a heavy snow scenario is limited to 500 m according to the SAE [54]. Note that these performance insufficiencies are not system-independent; therefore, they have to be included in all available sensors simultaneously. In this context, if a triggering condition is validated for ADS, then this includes a radar, camera, and lidar sensor; all performance insufficiency injections for all sensors must be included at the same time and at the same fidelity level to avoid inaccurate results.

A test case function (f_{TC}) is defined without any performance insufficiency injection, resulting in values that fall within the acceptance criteria as $f_{TC}() \in \epsilon_{acceptance-criteria}$. The same function, including the performance insufficiency (PI) injector, is defined in Equation (1) for all insufficiency levels (S).

$$f_{TC}(PI_i) \quad \forall \quad 1 \leq i \leq S \tag{1}$$

If many performance insufficiencies (N) are validated at the same time, the previous equation could be expressed as Equation (2), where each added performance insufficiency (PI_j) is included independently of from each other and with different levels of intensity (S) (e.g., in cases when limited visibility and illuminance is validated at the same time in the system).

$$f_{TC}(PI_{ji}) \quad \forall \quad 1 \leq i \leq S \quad \text{and} \quad 1 \leq j \leq N \tag{2}$$

As discussed in [55], a triggering condition could be defined by one or many performance insufficiencies. For example, a heavy fog triggering condition could not be only parameterised by a visibility performance insufficiency as defined in [21], but also by illuminance and accuracy insufficiencies. Therefore, these three insufficiencies should be injected into the system at the same time. Consequently, a Triggering Condition Performance Insufficiency (TCPI) is defined as one or many performance insufficiencies that are injected at the same time and insufficiency level. This can then be formalised as

$$f_{TC}(TCPI) = f_{TC}(PI_{1i}, PI_{2i}, ..., PI_{ji}) \quad \forall \quad 1 \leq i \leq S \quad \text{and} \quad 1 \leq j \leq N \tag{3}$$

Consequently, the equation to inject one or many TCPIs (M) into the test case is defined as follows:

$$f_{TC}(TCPI_k) \quad \forall \quad 1 \leq k \leq M \tag{4}$$

Table 1. Generic performance insufficiencies excerpt list.

GPI ID	Generic Performance Insufficiency (GPI)	Impact
PI-01	Reduction of Field of View (FoV)	The visual range of the sensor is reduced from the nominal sensor performance.
PI-02	Light disturbance	An external light source affects the sensor perception.
PI-03	Misalignment	The position of the sensor was changed from the calibrated sensor position.
PI-04	Reduction of resolution	Sensor resolution is reduced according to the nominal performance provided by the manufacturer.
PI-05	Reduction of accuracy	Sensor accuracy decreases according to the nominal performance.
PI-06	Reduction of luminous intensity	The luminous intensity of the sensor is reduced according to the technical specifications.
PI-07	Slower processing time	Sensor processing time is slower than the maximum processing time in nominal conditions.

Table 2. Lidar technology performance insufficiency excerpt list.

Technology Performance Insufficiency (TPI)	Parent Generic Performance Insufficiency (GPI)	Potential Triggering Conditions	Performance Insufficiency Injection
Reduction of Field of View (FoV)	PI-01	Snowfall, fog conditions, etc.	Crop the raw point cloud (vertical and horizontal cropping) generated by the lidar sensor.
Light Disturbance	PI-02	Mirrors, water on the street, etc.	Add random points into the point cloud message.
Misalignment	PI-03	Wrong calibration, earthen or gravel roads, potholes, etc.	Change the position of the sensor.
Reduction of accuracy	PI-05	Sensor cover, housing dirtiness, occlusion, etc.	Include noise into the point cloud message.
Slower Processing Time	PI-07	Driving in urban areas, etc.	Include random objects into the point cloud message.

3. Risk Quantification

It is essential to perform a quantitative evaluation to ensure objective validation of the function. Thus, the next stage of our methodology involves quantifying the risk for validating the ADS, which enables a comparison of the results with newer iterations of the ADS to introduce improvements. Figure 5 demonstrates the correlation between risk and cause and effect in accordance with the SOTIF standard. ISO26262 [16] defines exposure, controllability and severity as part of the Hazard Analysis and Risk Assessment (HARA) methodology. The HARA methodology assigns a safety level to each defined hazard called the Automotive Safety Integrity Level (ASIL). The calculated ASIL of a hazard is based on three main variables:

- Severity (S): the level of injury to the driver and passengers.
- Controllability (C): if the hazard could be controlled by the driver.
- Exposure (E): how often the hazard occurred during the driving time.

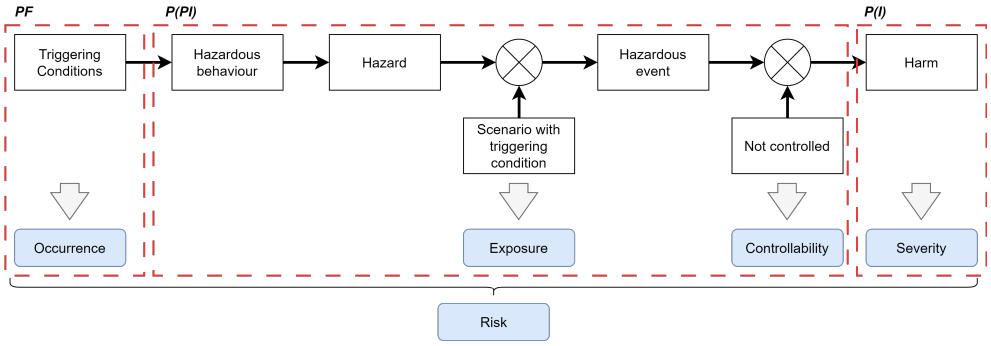


Figure 5. Risk quantification from our approach and compared with ISO21448.

Each variable is assigned a value ranging from 0 (mildest) to 4 (worst). The sum of these three variables (e.g., $S0 + C2 + E1$) is used to determine the ASIL level. ASIL A is assigned to the lowest risk, while ASIL D is assigned to the highest risk. In contrast, ISO21448 [9] defines risk as the product of controllability and severity. Thus, the primary objective is to reduce controllability and severity to prevent harm or achieve the safety goal, considering a previously defined residual risk. In ideal scenarios, controllability and severity should be zero ($C = 0$ or $S = 0$) to achieve optimal outcomes.

$$Risk = PF * P(PI) * P(I) \tag{5}$$

This approach calculates the risk based on Equation (5). This equation is based on two different probabilities. $P(PI)$ represents the likelihood that a performance insufficiency may have an impact on the nominal performance of the system. This means that the results once the performance insufficiency is injected are outside a defined acceptable window. Nominal performance is calculated based on the values from Montecarlo simulations of the same scenario without the injection of any performance insufficiency. The tolerable window is calculated based on the standard deviation of this simulation and a given factor. As Figure 5 shows, if $P(PI)$ is greater than zero, this implies that the ADS is susceptible to the injected insufficiency and it is also uncontrolled by ADS and, therefore, it is relevant for the system. On the other hand, $P(I)$ is considered the probability of injury. It is assumed that both probabilities are independent for the sake of simplicity; however, dependent probabilities will be considered in further research to achieve more accurate quantification results. The probability of $P(I)$ is determined using the methodology explained in [47,56]. The authors employ the model developed by Kusano and Gabler [57] to calculate the likelihood of injury. This model assesses the probability of injury as being greater when the injury level is equal to or greater than level 2 according to the Maximum Abbreviated Injury Scale (MAIS) [58]. MAIS level 2 is defined by moderate injuries with a low probability of death (1–2%). This value is considered in our approach as severity greater than zero ($S > 0$) and, therefore, takes the risk into consideration. Otherwise, the probability of severity is zero as well as the risk. In this case, although the performance insufficiency still affects the system, SOTIF modifications may be applied in the ADS to enhance system reliability. The model [59] utilised to calculate the probability of injury is defined as follows:

$$P(I) = \begin{cases} \frac{1}{1+e^{-(\beta_0+\beta_1\Delta v+\beta_2)}} & \text{MAIS} \geq 2 \\ 0 & \text{MAIS} < 2 \end{cases} \tag{6}$$

A Plausibility Factor (PF) was also added to the risk calculation to adjust the different levels of performance insufficiency injection. This factor decreases as the injection level becomes more extreme. For instance, if the system experiences reduced visibility,

it is more plausible that visibility is not drastically reduced than that sensor visibility is almost lost. The plausibility coefficient has the same value as the probability of occurrence when a triggering condition is injected. However, it is not the same when a performance insufficiency is injected. Currently, the value of this factor is determined through expert judgement; however, further research will be conducted to calculate this factor accurately in the future. Finally, the equation that defines the risk for a performance insufficiency is given in Equation (7) for all levels of injection (S).

$$Risk_{PI} = \sum_{i=0}^{S} PF_i * P(PI_i) * P(I_i) \quad \forall \quad 1 \leq i \leq S \tag{7}$$

Consequently, the risk evaluation of an ADS is obtained by summing all calculated risks for the considered performance insufficiencies (N), as shown in Equation 8.

$$Risk_{ADS} = \sum_{i=0}^{N} Risk_{PI_j} \quad \forall \quad 1 \leq j \leq N \tag{8}$$

The risk calculated by this methodology provides a quantitative metric that must be evaluated by the stakeholder to determine if it is within the acceptance criteria for its SOTIF evaluation. The As Low As Reasonably Practicable (ALARP) [60] principle or similar should then be applied to reduce the risk to the lowest possible level. This also aligns with one of the main goals of SOTIF, according to which each validation iteration improves the system's reliability and safety.

4. Use Case

This section presents two use cases to illustrate the proposed approach outlined in previous sections. In the first use case, the ADS is subjected to a limited visibility performance insufficiency by including a generic performance insufficiency with varying levels of intensity. In the second use case, the lack of accuracy is included in the system as a performance insufficiency, injecting different reflection levels. The environment simulator used was the open-source CARLA Simulator [61] in which all elements of the map were removed, leaving only the road and the vehicles (ego and target vehicles). The object detection of the ADS is based on the cluster detection from the Autoware [62] software stack. The ADS uses a lidar sensor located at the top of the vehicle as a unique perception sensor. The vehicle controller of the ADS was developed in ROS [63]. Figure 6 depicts the architecture for these use cases. It is assumed that the performance insufficiency injection does not introduce a significant delay into the system and that it does not affect the results. The scenario used is a deceleration scenario, where the ego vehicle (green) reduces speed or stops to avoid a collision with the target vehicle (red) located in front. The waypoint is straight without turning in any direction. The initial speed of the vehicles is zero with the maximum speed reached by the ego vehicle being 80 km/h. The deceleration scenario is shown in Figure 7. The top left picture shows the CARLA simulation; the top right shows the visualisation from the vehicles. A logical view is given below these pictures.

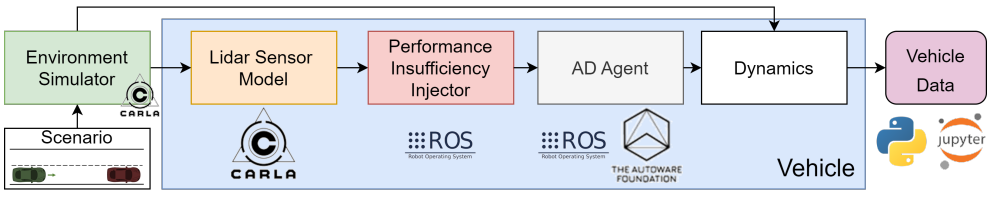


Figure 6. Use cases architecture.

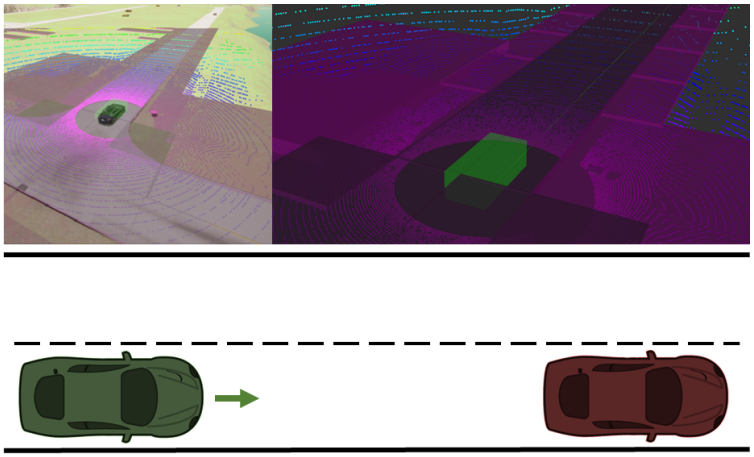


Figure 7. Deceleration Scenario. Up-left: CARLA visualisation. Up-right: vehicle visualisation. Down: logical view.

The nominal performance of the ADS is shown in Figure 8, where the distance travelled is plotted on the x-axis over time. Black lines show the performance of the ADS of each simulation with the mean value of the simulations also displayed with an overlapping cyan line. The execution time tolerable window is shown with vertical blue lines. Similarly, green lines are used to limit the tolerance window of the ADS distance travelled. The tolerable window is calculated based on the mean value from the nominal performance simulations in which the standard deviation is multiplied by a factor set to the upper and lower limits. In this use case, one hundred simulations ($B = 100$) were performed to obtain the probabilistic values. Whether execution time or distance travelled is outside the defined tolerance window, this is considered hazardous behaviour.

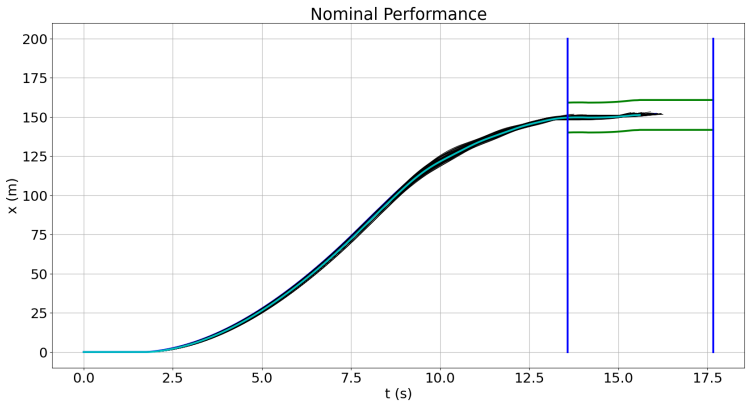


Figure 8. Nominal performance results.

4.1. Performance Insufficiencies Injection

The initial aim is to translate the generic performance insufficiency into a technological performance insufficiency. Since the perception sensor in this ADS is a lidar, the reduction of the field of view has been modelled as the cropping of the point cloud generated from

the lidar sensor. In order to increase the performance of the simulations and reduce the test cases, only the Responsibility-Sensitive Safety (RSS) [52] area (A_{RSS}) is considered in the tests. This decision was made because, based on the model, only this area is relevant for the safety of the vehicle. In our approach, the velocity of the target vehicle was set to zero in the original RSS equation to be more conservative.

$$D_{RSS} = \left[v_r \rho + \frac{1}{2} a_{max,accel} \rho^2 + \frac{(v_r + \rho a_{max,accel})^2}{2 a_{min,brake}} \right]_+ \qquad [x]_+ := max\{x, 0\} \tag{9}$$

The meaning and the given value for each parameter of the formula are outlined below:

- D_{RSS}: Minimum distance to ensure that there is no crash with the obstacle.
- v_r: Max ego vehicle velocity (m/s) in the test scenario. Value: 22.22 m/s (80 km/h).
- ρ: Response time in seconds: 0.5 s.
- $a_{max,accel}$: Maximum acceleration of the robot (m/s^2). Value: 5.5 m/s^2.
- $a_{min,brake}$: Minimum braking acceleration of the robot (m/s^2). Value: 4.5 m/s^2.

The minimum distance is only used for the longitude value; for the latitude value, the standard width for a highway is applied instead, 3.75 m [64]. The RSS distance, calculated using the given values, is 81.09 m (D_{RSS}). Vehicle length is assumed to be five metres. The area (A_{RSS}) considered for analysis is depicted in Figure 9 and can be described as follows:

$$\begin{aligned} V_{length}/2 < X < V_{length}/2 + D_{RSS} \\ -R_{width}/2 < Y < R_{width}/2 \end{aligned} \tag{10}$$

The reduced visibility levels for the limited visibility performance insufficiency injected into the system are 80, 60, 45, 30, 20 and 15 m. Figure 10 shows the results of these injections. In this use case, level 0 (80 m), level 1 (60 m) and level 2 (45 m) do not have any impact on the output of the ADS. From level 3 (30 m) on, the injections do have an impact on the system, resulting in collisions. At these levels, the difference between nominal performance (cyan line) and the output with the injection is clear, with collisions shown as vertical red lines. Collisions occur earlier in each injection, since detection of the target is delayed due to visibility insufficiency, leading to delayed braking and, finally, a collision. Table 3 shows the probabilities of hazardous behaviour and collision of each injection level. In this case, all hazardous behaviours lead to a collision, although this behaviour does not always occur. As shown in the charts, the impact of the performance insufficiency starts to be relevant at level 3. At this level, two thirds of all simulations are outside of the tolerance windows and lead to a collision. Then, the injections have a full impact on the outcome of the ADS. These results follow the cause and effect model shown in Figure 1, where the visibility reduction injected at the technical level leads to an output insufficiency at the functional level. There is a late detection from the sensor block that generates a lag on the actuation block, followed by a late braking at the vehicle level, which ends in hazardous behaviour.

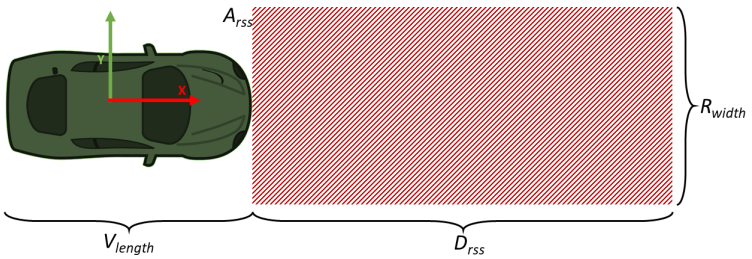


Figure 9. Considered RSS area in the test cases.

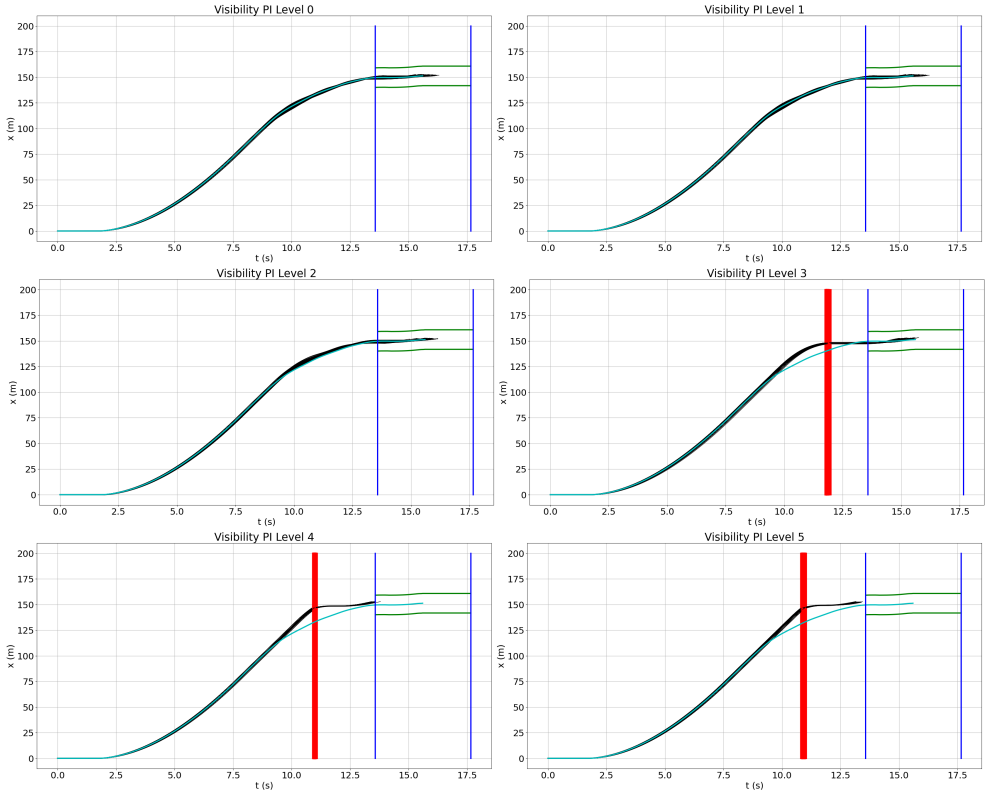


Figure 10. Results of the visibility performance insufficiency injections.

Table 3. Simulation results for each visibility performance insufficiency level.

PI_{vis} Level (Meters)	Hazardous Behaviour $P(HB)$	Collision $P(C)$	$P(PI)$
Level 0 (80 m)	0.00	0.00	0.00
Level 1 (60 m)	0.00	0.00	0.00
Level 2 (45 m)	0.00	0.00	0.00
Level 3 (30 m)	0.66	0.66	0.66
Level 4 (20 m)	1.00	1.00	1.00
Level 5 (15 m)	1.00	1.00	1.00

4.2. Field of View Reduction

Quantitative Risk Evaluation

As previously described in Section 3, the risk depends on the Plausibility Factor (PF), the probability of performance insufficiency ($P(PI)$), and the probability of injury ($P(I)$). The Plausibility Factor was calculated based on the values from an exponential distribution with a given lambda value ($\lambda = 1$) and a random variable (X) set by expertise judgement for each level. The plausibility values for this use case for all levels are shown in Table 4.

Table 4. Plausibility factor for the visibility performance insufficiency.

PI_{vis} Level	Visibility Limitation	X	Given PF
Level 0	80 m	P(X≥0)	$PF_{vis_{80}}$ = 1.00000
Level 1	60 m	P(X≥1)	$PF_{vis_{60}}$ = 0.36788
Level 2	45 m	P(X≥2)	$PF_{vis_{45}}$ = 0.13534
Level 3	30 m	P(X≥3)	$PF_{vis_{30}} = 4.979 \times 10^{-2}$
Level 4	20 m	P(X≥4)	$PF_{vis_{20}} = 1.832 \times 10^{-2}$
Level 5	15 m	P(X≥5)	$PF_{vis_{15}} = 6.74 \times 10^{-3}$

Based on the risk calculation given in Section 3, the following equation shows the evaluated risk for the lowest level of the reduced visibility performance level. The Plausibility Factor for this injection is the maximum, $PF_{vis_{80}} = 1.0000$, because it could occur with a high probability in many scenarios: if this happens, its impact must not be minimised. The probability of performance insufficiency is zero, which is expected since the reduced visibility is close to the nominal field of view of the sensor. The probability of injury is also zero, since there are no collisions. Consequently, the risk for this performance insufficiency level is zero, as shown in Equation (11).

$$Risk_{vis_{80}} = PF_{vis_{80}} * P(PI_{vis_{180}}) * P(I_{vis_{80}}) = 1.00 * 0.00 * 0.00 = 0.00 \quad (11)$$

Unlike the risk evaluation previously calculated, the risk for level 3 of injection ($Risk_{vis_{30}}$) is not zero. In this case, the given Plausibility Factor is more restrictive, as this kind of performance insufficiency level does not occur as regularly. The probability of this injection level is not zero, and it has an impact on two thirds of the simulations. The probability of injury is not zero, since there are collisions that could cause moderate injuries to drivers.

$$Risk_{vis_{30}} = 0.04979 * 0.66 * 1.22966 \times 10^{-2} \quad (12)$$

As expected, the values of the probability of injury increase when stricter levels are injected because there is less time for the vehicle to brake, and thus the crash velocity is higher for each level. On the other hand, the overall risk for each level is not always higher than the previous level because of the given Plausibility Factor. Finally, the quantitative risk evaluation for this reduced visibility performance insufficiency based on the results from Table 5 is as follows:

$$\begin{aligned} Risk_{PI_{vis}} = Risk_{vis_{80}} + Risk_{vis_{60}} + Risk_{vis_{45}} + \\ Risk_{vis_{30}} + Risk_{vis_{20}} + Risk_{vis_{15}} = 1.36557 \times 10^{-3} \end{aligned} \quad (13)$$

Table 5. Risk evaluation for each visibility performance insufficiency level.

PI_{vis} Level (Meters)	PF	$P(PI)$	$P(I)$	Risk
Level 0 (80 m)	1.00000	0.00	0.00	0.00
Level 1 (60 m)	0.36788	0.00	0.00	0.00
Level 2 (45 m)	0.13534	0.00	0.00	0.00
Level 3 (30 m)	0.04979	0.66	1.22966×10^{-2}	4.04083×10^{-4}
Level 4 (20 m)	0.01832	1.00	3.83674×10^{-2}	7.02891×10^{-4}
Level 5 (15 m)	0.00674	1.00	3.83675×10^{-2}	2.58597×10^{-4}

This quantitative risk evaluation provides a reference point for the minimisation of risk in subsequent iterations of the SOTIF validation. It is noted that these results could be used to validate the function for specific triggering conditions defined in the standards. For instance, the SAE standard [20] classifies fog into six levels based on system visibility.

- Level 5: 0 m $\leq$ visibility < 61 m
- Level 4: 61 m $\leq$ visibility < 244 m
- Level 3: 244 m $\leq$ visibility < 805 m
- Level 2: 805 m $\leq$ visibility < 1609 m
- Level 1: visibility $\geq$ 1609 m

Therefore, the system is validated for the SAE fog scale up to level 4 (visibility > 60 m), ensuring zero risk at those levels in the system:

$$Risk_{PI_{visSAELevel4}} = Risk_{vis_{60}} = 0.00 \quad (14)$$

4.3. Accuracy Reduction

A reduction in the accuracy of the perception component is included in the system based on the classification from Table 1 in this use case. The primary objective of this use case is to demonstrate how reflections at different density levels can affect the system's object detection and resulting behaviour.

4.3.1. Performance Insufficiencies Injection

Table 2 shows how this performance insufficiency is modelled for lidar-specific technology in which random points are injected into the point cloud from the message. Similar to the first use case, only the area (A_{RSS}) from the RSS is considered because it is the relevant safety area for the test case. Different levels of point density are injected in the system, where density is calculated based on the number of points in the A_{RSS} based on sensor resolution and the number of injected points for this area ($injection\_density = number\_injected\_points/number\_A_{RSS}\_points$).

Figure 11 shows two levels of injection from the vehicle perspective. The left picture shows level 2 of injection, where we observe the false negatives produced by the random reflections in the point cloud message. The right picture shows a higher level of injection, where the amount of false detections makes the ego vehicle stop completely. The impact of each injection level is shown in Figure 12. Levels 0 and 1 do not have any relevant impact on the outcome of the function, but from level 3, the impact of the outcome is remarkable. Level 2 makes the ADS still follow the path but with successive stops due to the false negatives, while at levels 3 and 4, the ego vehicle remains stopped. Noteworthy is level 5, where the huge amount of reflections causes the target vehicle not to be considered as an object, eventually leading to a collision. Unlike the previous use case, in this case, hazardous behaviour does not always generate a collision, as Table 6 depicts.

Table 6. Simulation results for each performance insufficiency level.

PI_{acc} Level (Injection Density in %)	Hazardous Behaviour $P(HB)$	Collision $P(C)$	$P(PI)$
Level 0 (0.15%)	0.00	0.00	0.0
Level 1 (0.30%)	0.00	0.00	0.00
Level 2 (0.75%)	1.00	0.00	1.00
Level 3 (1.49%)	1.00	0.00	1.00
Level 4 (2.99%)	1.00	0.00	1.00
Level 5 (5.97%)	1.00	1.00	1.00

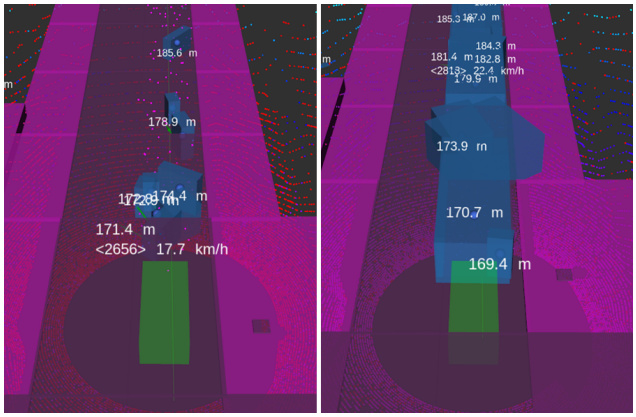


Figure 11. Reduced accuracy performance insufficiency simulation.

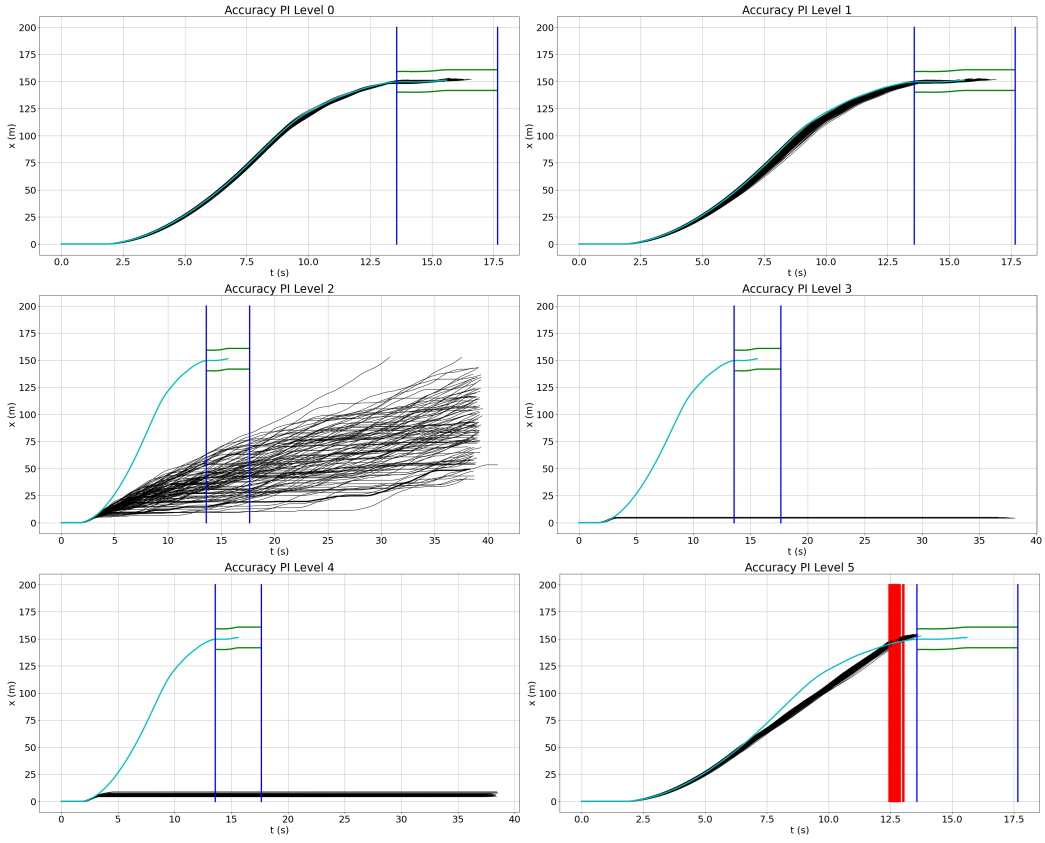


Figure 12. Results of the accuracy performance insufficiency use case.

Table 6 displays the results of the injections. It is noteworthy that although this injection has an impact on the system at many levels, the included reflections do not affect the system at levels 0 to 2. However, levels 3 and 4 include false positives due to the injected reflections, which activate the braking system and prevent the vehicle from following the defined path. On the other hand, all hazardous behaviours lead to a collision at the most restrictive level of injection.

4.3.2. Quantitative Risk Evaluation

After calculating the probability of performance insufficiency, the risk can be evaluated. As with the previous use case, the Plausibility Factor was determined using an exponential distribution and values based on expert judgement. Table 7 displays the risk values for each level with only the final level posing a risk in the ADS due to the occurrence of collisions. It is important to note that there are many injections where the performance insufficiency has an impact on the behaviour of the system, but there are no collisions. Despite this, SOTIF measures should be implemented to minimise the probability of hazardous behaviours. For example, including a wide range of perception sensor technologies could be beneficial as each sensor technology has its advantages and disadvantages in specific environmental situations, which could help mitigate the impact of certain scenarios.

Table 7. Risk evaluation for each accuracy performance insufficiency level.

PI_{acc} Level (Injection Density in %)	PF	$P(PI)$	$P(I)$	Risk
Level 0 (0.15%)	1.00000	0.00	0.00	0.00
Level 1 (0.30%)	0.36788	0.00	0.00	0.00
Level 2 (0.75%)	0.13534	1.00	0.00	0.00
Level 3 (1.49%)	0.04979	1.00	0.00	0.00
Level 4 (2.99%)	0.01832	1.00	0.00	0.00
Level 5 (5.97%)	0.00674	1.00	1.26752×10^{-2}	8.54309×10^{-5}

The equation below shows the calculated risk for the reduction of accuracy, where only the last injection has an impact on the final risk quantification for this performance insufficiency; even though most of the performance insufficiency injections have an impact on the system output, only the last injection has an impact on the risk quantification. The aim is to indicate that both risk and SOTIF modifications are relevant in the validation process.

$$Risk_{ADS} = \sum_{i=0}^{5} Risk_{PI_j} = 8.54309 \times 10^{-5} \quad \forall \quad 0 \leq j \leq 5 \tag{15}$$

5. Conclusions and Future Work

This document describes a methodology for validating perception performance insufficiencies in automotive driving systems. Due to the impossibility of validating all possible triggering conditions of a scenario, the evaluation focuses on the impact of performance insufficiencies in the perception component of the system and its effect on the output of the entire system in order to determine whether it may lead to hazardous behaviour and, finally, possibly cause harm. In the document, a classification of the performance insufficiencies is given, showing the impact of the defined insufficiency in the system. Then, a model for each perception insufficiency is used to inject the insufficiency into the system and determine whether the specified performance insufficiency does have an impact on the ADS output. Based on the results from the injections, a Plausibility Factor and a probability of injury are calculated; when the injection leads to a collision, a quantitative risk could be calculated. Since the risk is based on the severity of injuries, if there are no collisions

or the injuries caused by a collision are light, the risk is set to zero. In these situations, SOTIF measurement should be considered to minimise the probability of performance insufficiency. The calculated risk provides us with a quantitative metric that serves as a reference for improvement in further validation iterations. Finally, this text describes two use cases that show how the proposed methodology can be applied. The first use case validates a limited visibility performance insufficiency in which most of the hazardous behaviours lead to risk due to collisions. The second use case validates a reduction in accuracy, where most levels of injection do not result in a collision, indicating no risk, but they still have an impact on the system. Therefore, SOTIF measures should still be applied to improve the ADS against this type of performance insufficiency.

This research has raised several questions that require further investigation. Although the risk evaluation provides a quantitative metric, it is not related to any specific measure, such as the number of hours driven or kilometres travelled. Future research should address this issue to better link the obtained metric with real-world measurements. Another important question that needs to be addressed is how to determine when a performance insufficiency has been fully validated, including accurate models for each performance insufficiency. Additionally, a more effective approach to identifying edge cases in performance insufficiency injection testing should be implemented in the future.

Author Contributions: Conceptualization, V.J.E.J.; Methodology, V.J.E.J.; Software, V.J.E.J.; Validation, V.J.E.J.; Formal analysis, V.J.E.J.; Investigation, V.J.E.J.; Resources, V.J.E.J.; Data curation, V.J.E.J.; Writing—original draft, V.J.E.J.; Writing—review and editing, D.W., G.M. and E.B.; Visualization, V.J.E.J.; Supervision, G.M. and E.B. All authors have read and agreed to the published version of the manuscript.

Funding: Open Access Funding by the Graz University of Technology.

Data Availability Statement: Data are contained within the article. The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Acknowledgments: The publication was written at Virtual Vehicle Research GmbH in Graz, Austria. The authors would like to acknowledge the financial support within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Labour and Economy (BMAW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorised for the programme management. Supported by TU Graz Open Access Publishing Fund.

Conflicts of Interest: Author Víctor J. Expósito Jiménez and Daniel Watzenig were employed by the company Virtual Vehicle Research GmbH. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADAS	Advanced Driver-Assistance System
ADS	Automated Driving System
ALARP	As Low As Reasonably Practicable
ASIL	Automotive Safety Integrity Level
GPI	Generic Performance Insufficiency
HARA	Hazard Analysis and Risk Assessment
HIL	Hardware-In-Loop
KPI	Key Performance Indicator
MDPI	Multidisciplinary Digital Publishing Institute
ODD	Operational Design Domain

PF	Plausibility Factor
PI	Performance Insufficiency
RSS	Responsibility-Sensitive Safety
SOTIF	Safety Of The Intended Functionality
SPI	Safety Performance Indicator
TCPI	Triggering Condition Performance Insufficiency
TPI	Technology Performance Insufficiency

References

1. *(EU) 2022/1426*; Commission Implementing Regulation (EU) 2022/1426—Commision Implementing Act AD v4.1. European Commission: Brussel, Belgium, 2022.
2. National Transportation Safety Board (NTSB). Collision between a Sport Utility Vehicle Operating with Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018. 2020. Available online: https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR2001.pdf (accessed on 12 May 2023).
3. Bonnefon, J.F. 18 The Uber Accident. In *The Car That Knew Too Much: Can a Machine Be Moral?*; The MIT Press: Cambridge, MA, USA, 2021; pp. 93–98.
4. Shah, S.A. Safe-AV: A Fault Tolerant Safety Architecture for Autonomous Vehicles. Ph.D. Thesis, McMaster University, Hamilton, ON, USA, 2019.
5. AI Incident Database. Incident 293: Cruise's Self-Driving Car Involved in a Multiple-Injury Collision at an San Francisco Intersection. 2022. Available online: https://incidentdatabase.ai/cite/293/ (accessed on 3 March 2024).
6. Ballingall, S.; Sarvi, M.; Sweatman, P. Standards relevant to automated driving system safety: A systematic assessment. *Transp. Eng.* **2023**, *13*, 100202. [CrossRef]
7. Koopman, P. *How Safe Is Safe Enough?: Measuring and Predicting Autonomous Vehicle Safety*; Amazon Digital Services LLC: Seattle, WA, USA, 2022.
8. *UL4600*; Standard for Safety: Evaluation of Autonomous Products. Standards and Engagement Inc.: Evanston, IL, USA, 2021.
9. *ISO 21448:2022*; Road vehicles—Safety of the Intended Functionality. International Organization for Standardization: Geneva, Switzerland, 2022.
10. Westhofen, L.; Neurohr, C.; Koopmann, T.; Butz, M.; Schütt, B.; Utesch, F.; Neurohr, B.; Gutenkunst, C.; Böde, E. Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art. *Arch. Comput. Methods Eng.* **2022**, *30*, 1–35. [CrossRef]
11. Sun, C.; Zhang, R.; Lu, Y.; Cui, Y.; Deng, Z.; Cao, D.; Khajepour, A. Toward Ensuring Safety for Autonomous Driving Perception: Standardization Progress, Research Advances, and Perspectives. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 3286–3304. [CrossRef]
12. Wang, H.; Shao, W.; Sun, C.; Yang, K.; Cao, D.; Li, J. A Survey on an Emerging Safety Challenge for Autonomous Vehicles: Safety of the Intended Functionality. *Engineering* **2024**, *33*, 17–34. [CrossRef]
13. Hoss, M.; Scholtes, M.; Eckstein, L. A Review of Testing Object-Based Environment Perception for Safe Automated Driving. *Automot. Innov.* **2022**, *5*, 223–250. [CrossRef]
14. Zhu, Z.; Philipp, R.; Hungar, C.; Howar, F. Systematization and Identification of Triggering Conditions: A Preliminary Step for Efficient Testing of Autonomous Vehicles. In Proceedings of the 2022 IEEE Intelligent Vehicles Symposium (IV), Aachen, Germany, 4–9 June 2022; pp. 798–805. [CrossRef]
15. Expósito Jiménez, V.J.; Martin, H.; Schwarzl, C.; Macher, G.; Brenner, E. Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions. In *International Conference on Computer Safety, Reliability, and Security*; SAFECOMP 2022 Workshops; Trapp, M., Schoitsch, E., Guiochet, J., Bitsch, F., Eds.; Springer: Cham, Switzerland, 2022; pp. 11–22.
16. *ISO 26262:2018*; Road Vehicles—Functional Safety. International Organization for Standardization: Geneva, Switzerland, 2018.
17. Koopman, P.; Kane, A.; Black, J. Credible Autonomy Safety Argumentation. In Proceedings of the 27th Safety-Critical Systems Symposium 2019, Bristol, UK, 5–7 February 2019.
18. United Nations Economic Commission for Europe (UNECE). New Assessment/Test Method for Automated Driving (NATM)—Master Document (Final Draft). 2021. Available online: https://unece.org/sites/default/files/2021-01/GRVA-09-07e.pdf (accessed on 9 February 2024).
19. *ISO/DIS 34505*; Road Vehicles—Test Scenarios for Automated Driving Systems—Scenario Based Safety Evaluation Framework. International Organization for Standardization: Geneva, Switzerland, 2022.
20. *AVSC00002202004*; AVSC Best Practice for Describing an Operational Design Domain: Conceptual Framework and Lexicon. Society of Automotive Engineers (SAE) International: Pittsburgh, PA, USA, 2020.
21. *BSI PAS 1883:2020*; Operational Design Domain (ODD) Taxonomy for an Automated Driving System (ADS). Specification. The British Standards Institution: London, UK, 2020.
22. Weissensteiner, P.; Stettinger, G.; Khastgir, S.; Watzenig, D. Operational Design Domain-Driven Coverage for the Safety Argumentation of Automated Vehicles. *IEEE Access* **2023**, *11*, 12263–12284. [CrossRef]
23. Weissensteiner, P.; Stettinger, G.; Rumetshofer, J.; Watzenig, D. Virtual Validation of an Automated Lane-Keeping System with an Extended Operational Design Domain. *Electronics* **2022**, *11*, 72. [CrossRef]

24. Scholtes, M.; Westhofen, L.; Turner, L.R.; Lotto, K.; Schuldes, M.; Weber, H.; Wagener, N.; Neurohr, C.; Bollmann, M.H.; Körtke, F.; et al. 6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment. *IEEE Access* **2021**, *9*, 59131–59147. [CrossRef]
25. ASAM e.V. ASAM OpenODD. 2021. Available online: https://www.asam.net/project-detail/asam-openodd/ (accessed on 5 December 2023).
26. Virtual Vehicle Research GmbH. SPIDER: Mobile Platform for the Development and Testing of Autnomous Driving Functions. 2021. Available online: https://www.v2c2.at/spider/ (accessed on 5 December 2023).
27. AVL List GmbH. AVL DRIVINGCUBE. 2023. Available online: https://www.avl.com/en/testing-solutions/automated-and-connected-mobility-testing/avl-drivingcube(accessed on 7 February 2024).
28. de Gelder, E.; Paardekooper, J.P.; Op den Camp, O.; De Schutter, B. Safety assessment of automated vehicles: how to determine whether we have collected enough field data? *Traffic Inj. Prev.* **2019**, *20*, S162–S170. [CrossRef]
29. Linnhoff, C.; Hofrichter, K.; Elster, L.; Rosenberger, P.; Winner, H. Measuring the Influence of Environmental Conditions on Automotive Lidar Sensors. *Sensors* **2022**, *22*, 5266. [CrossRef]
30. Fang, J.; Zhou, D.; Zhao, J.; Tang, C.; Xu, C.Z.; Zhang, L. LiDAR-CS Dataset: LiDAR Point Cloud Dataset with Cross-Sensors for 3D Object Detection. *arXiv* **2023**, arXiv:cs.CV/2301.12515. http://arxiv.org/abs/2301.12515.
31. Schlager, B.; Muckenhuber, S.; Schmidt, S.; Holzer, H.; Rott, R.; Maier, F.M.; Saad, K.; Kirchengast, M.; Stettinger, G.; Watzenig, D.; et al. State-of-the-Art Sensor Models for Virtual Testing of Advanced Driver Assistance Systems/Autonomous Driving Functions. *SAE Int. J. Connect. Autom. Veh.* **2020**, *3*, 233–261. [CrossRef]
32. Bijelic, M.; Gruber, T.; Mannan, F.; Kraus, F.; Ritter, W.; Dietmayer, K.; Heide, F. Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11679–11689. [CrossRef]
33. Dreissig, M.; Scheuble, D.; Piewak, F.; Boedecker, J. Survey on LiDAR Perception in Adverse Weather Conditions. *arXiv* **2023**, arXiv:cs.RO/2304.06312. http://arxiv.org/abs/2304.06312.
34. Minh Mai, N.A.; Duthon, P.; Salmane, P.H.; Khoudour, L.; Crouzil, A.; Velastin, S.A. Camera and LiDAR analysis for 3D object detection in foggy weather conditions. In Proceedings of the 2022 12th International Conference on Pattern Recognition Systems (ICPRS), Etienne, France, 7–10 June 2022; pp. 1–7. [CrossRef]
35. Hahner, M.; Sakaridis, C.; Dai, D.; Van Gool, L. Fog Simulation on Real LiDAR Point Clouds for 3D Object Detection in Adverse Weather. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021.
36. Hahner, M.; Sakaridis, C.; Bijelic, M.; Heide, F.; Yu, F.; Dai, D.; Van Gool, L. LiDAR Snowfall Simulation for Robust 3D Object Detection. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, US, 18–24 June 2022; pp. 16343–16353. [CrossRef]
37. Skender, I. Robustness Test for ADAS Function. Master's Thesis, Graz University of Technology, Graz, Austria, 2022.
38. Pizzati, F.; Cerri, P.; de Charette, R. Model-Based Occlusion Disentanglement for Image-to-Image Translation. In *Proceedings of the Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Cham, Switzerland, 2020; pp. 447–463.
39. Sadeghi, J.; Rogers, B.; Gunn, J.; Saunders, T.; Samangooei, S.; Dokania, P.K.; Redford, J. A Step Towards Efficient Evaluation of Complex Perception Tasks in Simulation. *arXiv* **2021**, arXiv:cs.LG/2110.02739. http://arxiv.org/abs/2110.02739.
40. Piazzoni, A. *Modeling Perception Errors in Autonomous Vehicles and Their Impact on Behavior*; Nanyang Technological University: Nanyang, China, 2023. [CrossRef]
41. Piazzoni, A.; Cherian, J.; Slavik, M.; Dauwels, J. Modeling perception errors towards robust decision making in autonomous vehicles. In Proceedings of the IJCAI'20: Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 11–17 July 2021.
42. Piazzoni, A.; Cherian, J.; Dauwels, J.; Chau, L.P. PEM: Perception Error Model for Virtual Testing of Autonomous Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 670–681. [CrossRef]
43. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing properties of neural networks. *arXiv* **2014**, arXiv:cs.CV/1312.6199. http://arxiv.org/abs/1312.6199.
44. Sadeghi, J.; Lord, N.A.; Redford, J.; Mueller, R. Attacking Motion Planners Using Adversarial Perception Errors. *arXiv* **2023**, arXiv:cs.RO/2311.12722. http://arxiv.org/abs/2311.12722.
45. Innes, C.; Ramamoorthy, S. Testing Rare Downstream Safety Violations via Upstream Adaptive Sampling of Perception Error Models. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May –2 June 2023 2023; pp. 12744–12750. [CrossRef]
46. Putze, L.; Westhofen, L.; Koopmann, T.; Böde, E.; Neurohr, C. On Quantification for SOTIF Validation of Automated Driving Systems. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; pp. 1–8. [CrossRef]
47. de Gelder, E.; Elrofai, H.; Saberi, A.K.; Paardekooper, J.P.; Op den Camp, O.; de Schutter, B. Risk Quantification for Automated Driving Systems in Real-World Driving Scenarios. *IEEE Access* **2021**, *9*, 168953–168970. [CrossRef]

48. Kramer, B.; Neurohr, C.; Büker, M.; Böde, E.; Fränzle, M.; Damm, W. Identification and Quantification of Hazardous Scenarios for Automated Driving. In *International Symposium on Model-Based Safety and Assessment*; Zeller, M., Höfig, K., Eds.; Springer: Cham, Switzerland, 2020; pp. 163–178.
49. Vaicenavicius, J.; Wiklund, T.; Grigaite, A.; Kalkauskas, A.; Vysniauskas, I.; Keen, S.D. Self-Driving Car Safety Quantification via Component-Level Analysis. *SAE Int. J. Connect. Autom. Veh.* **2021**, *4*, 35–45. [CrossRef]
50. Karunakaran, D.; Worrall, S.; Nebot, E.M. Efficient Statistical Validation with Edge Cases to Evaluate Highly Automated Vehicles. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; pp. 1–8.
51. Chu, J.; Zhao, T.; Jiao, J.; Yuan, Y.; Jing, Y. SOTIF-Oriented Perception Evaluation Method for Forward Obstacle Detection of Autonomous Vehicles. *IEEE Syst. J.* **2023**, *17*, 2319–2330. [CrossRef]
52. Shalev-Shwartz, S.; Shammah, S.; Shashua, A. On a Formal Model of Safe and Scalable Self-driving Cars. *arXiv* **2017**, arXiv:1708.06374.
53. Peng, L.; Li, B.; Yu, W.; Yang, K.; Shao, W.; Wang, H. SOTIF Entropy: Online SOTIF Risk Quantification and Mitigation for Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 1530–1546. [CrossRef]
54. *ISO/SAE PAS 22736:2021*; Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. Society of Automotive Engineers (SAE) International: Pittsburgh, PA, USA, 2021.
55. Expósito Jiménez, V.J.; Winkler, B.; Castella Triginer, J.M.; Scharke, H.; Schneider, H.; Brenner, E.; Macher, G. Safety of the Intended Functionality Concept Integration into a Validation Tool Suite. *Ada User J.* **2023**, *44*, 244–447. [CrossRef]
56. Zhao, D.; Huang, X.; Peng, H.; Lam, H.; LeBlanc, D.J. Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 733–744. [CrossRef]
57. Kusano, K.D.; Gabler, H.C. Potential Occupant Injury Reduction in Pre-Crash System Equipped Vehicles in the Striking Vehicle of Rear-end Crashes. In *Annals of Advances in Automotive Medicine*; Annual Scientific Conference; Association for the Advancement of Automotive Medicine: Chicago, L, USA, 2010; Volume 54, pp. 203–214.
58. Gennarelli, T.A.; Wodzin, E. AIS 2005: A contemporary injury scale. *Injury* **2006**, *37*, 1083–1091. [CrossRef]
59. Kusano, K.D.; Gabler, H.C. Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1546–1555. [CrossRef]
60. *EN50126-2*; Railway Applications—The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) Part 2: Systems Approach to Safety. CENELEC—European Committee for Electrotechnical Standardization: Brussel, Belgium, 2017.
61. Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V. CARLA: An Open Urban Driving Simulator. In Proceedings of the 1st Annual Conference on Robot Learning, Mountain View, CA, USA, 13–15 November 2017; pp. 1–16.
62. The Autoware Foundation. Autoware. 2021. Available online: https://www.autoware.org/autoware (accessed on 3 May 2023).
63. Open Source Robotics Foundation, Inc. ROS—Robot Operating System. 2023. Available online: https://www.ros.org/ (accessed on 3 May 2023).
64. European Road Safety Observation—European Commission. Motorways 2018. 2018. Available online: https://road-safety.transport.ec.europa.eu/system/files/2021-07/ersosynthesis2018-motorways.pdf (accessed on 7 September 2023).

A.6 Systematic SOTIF-aligned approach to explore residual risk in autonomous driving systems

Systematic SOTIF-aligned approach to explore residual risk in autonomous driving systems

H. Scharke[1], H. Goossens[2], S. Kalisvaart[2], V.J. Expósito Jiménez[3]

1: AVL List GmbH, Hans-List-Platz 1, 8020 Graz
2: TNO, HNK building, Radarweg 60, 1043 NT Amsterdam, The Netherland
3: Virtual Vehicle Research GmbH, Inffeldgasse 21A, 8010 Graz, Austria

Abstract: The validation and homologation of autonomous vehicles (AVs) requires the proof that AVs can handle all the critical situations that they may encounter in their Operational Design Domain (ODD). The ISO 21448 (Safety Of The Intended Functionality - SOTIF) standards provides guidance and a framework to ensure the safety of autonomous vehicles in complex and dynamic traffic. The key elements in SOTIF are: 1. the specification of the ODD and possible hazards, 2. the identification and evaluation of triggering conditions to cause possible hazardous behaviours and 3. comprehensive safety evaluation and risk assessment and reporting. This paper and presentation explains the workflow and building blocks of a comprehensive safety assessment toolchain for automated vehicles, which allows a systematic, ODD based and SOTIF-aligned, test planning including sophisticated methods to explore the residual risk of the autonomous driving system. The toolchain exploration is based on AVL SCENIUS™ software. The presented risk quantification approach combining real work driving scenarios and statistics together with scenario-based validation in virtual domains.

Keywords: SOTIF toolchain, systematic safety assessment, risk quantification, real world statistics

1. Introduction

Autonomous vehicle safety development frameworks are designed to ensure that self-driving cars are safe and reliable. One of the most important frameworks is ISO 21448 SOTIF standard [1], which was developed to address the new safety challenges that autonomous vehicle software developers are facing, such as the interactions of the complex vehicle systems with the complex behaviour of the surrounding environment. The SOTIF standard ensures the safety of autonomous vehicles in complex and dynamic traffic and environmental situations even when there are no faults or failures in the system. SOTIF addresses the challenges that are not covered by e.g. Functional Safety - ISO 26262 [2]. Rather SOTIF focuses on:

- functional insufficiencies and limitations, such as sensors and perception algorithms cannot detect or classify an object,
- reasonable foreseen misuse by persons, such as when a driver overrides or ignores a warning from the system or when driving in excluded areas, and
- general argumention framework and guidance on measures to ensure the SOTIF.

The SOTIF workflow can be summarized into six main blocks (Figure 1):

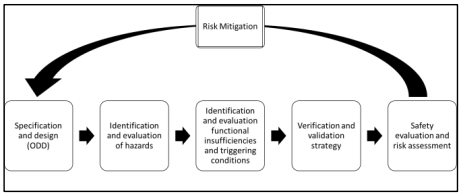


Figure 1: Overview of six main blocks of SOTIF workflow

Specification and design
It describes the objectives and requirements for specifying the functionality and designing the system architecture. The functionality is specified at the vehicle level, considering the use cases, the ODD, the level of driving automation, the driver role, and the driving policy.

Identification and evaluation of hazards
It describes the objectives and requirements for identifying the hazards resulting from functional insufficiencies, evaluating the risk of hazardous behaviours, and specifying the acceptance criteria for the residual risk.

Identification and evaluation of potential functional insufficiencies and triggering conditions
It describes the objectives and requirements for identifying and evaluating the potential functional insufficiencies and triggering conditions that can lead to hazardous behaviours, such as a sudden change in weather or lighting that affects the sensor performance or visibility.

Verification and validation strategy
It describes the objectives and requirements for specifying the integration and testing strategy for verifying and validating the SOTIF. The verification and validation activities include testing, simulation,

analysis, inspection, review, and audit. It aims to provide evidence that the SOTIF-related risks have been reduced to an acceptable level, and that the system behaves as intended in the ODD.

Safety evaluation and risk assessment
It describes the objectives and requirements for validating the system's response to known/unknown scenarios and evaluating the residual risk due to known/unknown hazardous scenarios.

Risk mitigation
This is the final step in SOTIF and describes the objectives and requirements for defining and implementing measures to improve the SOTIF and reduce the risk of the hazardous behaviour. Risk mitigation can involve updating specification and design changes, system modification, functional enhancements, restrictions in ODD, handing over authority, or addressing misuse.

The main targets of SOTIF approach are (Figure 2):
- to increase the identification of hazardous scenarios to be validated
- to minimize the area in which unsafe behaviour could appear
- to minimize the area in which unknown hazardous scenarios could appear and
- to improve the defined acceptance criteria level and minimize the residual risk with each new iteration.

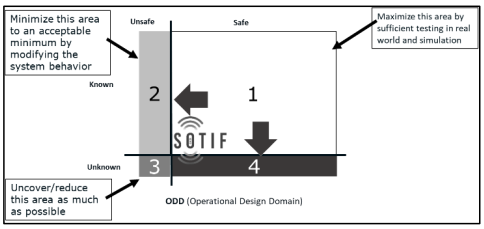


Figure 2: SOTIF main targets

2. Safety Validation Toolchain

In this chapter, we will explore the building blocks and methods of a comprehensive safety validation toolchain based on AVL SCENIUS™ software. A robust safety validation toolchain should adhere to the SOTIF approach, which is crucial for the safety development, release, and homologation process. The toolchain must encompass and trace all components, from requirement and safety data management to systematic test catalogue generation and optimization. It should ensure seamless test transfer into various execution environments, followed by safety result collection, safety and risk argumentation, and documentation.
Figure 3 shows a general overview of the elements of a comprehensive SOTIF conform validation toolchain.

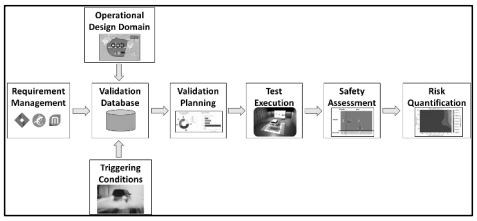


Figure 3: Elements of the safety validation toolchain

Requirement Management System
Requirement management, such as per the Automotive SPICE standard, is crucial for automotive development. Requirements must be identified, documented, and tracked throughout the development process to meet industry quality standards, resulting in safer, more reliable, and environmentally friendly vehicles. The toolchain interfaces to requirement management systems are vital because they ensure seamless integration and traceability of requirements throughout the development and validation process.

Operational Design Domain
A major requirement in SOTIF validation is the definition of an ODD in which it can be ensured that the autonomous vehicle works as intended and safely. The toolchain currently incorporates ASAM OpenXOntology® concept to the scenario "Layer 1 - Road network and traffic guidance objects ", "Layer 4 – Dynamic Objects" and "Layer 5 – Environmental Conditions". There are the three corresponding concepts defined (Figure 4a): the static, the dynamic, and the actor concept (Figure 4b), facilitating a structured categorization of parameters, which is crucial for simulating and analysing the myriad of scenarios an autonomous vehicle might encounter. A transfer to ASAM OpenODD approach will be required as soon as the standard is adopted.

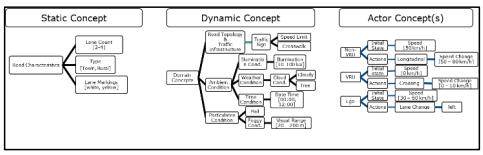


Figure 4a: Description of the ODD

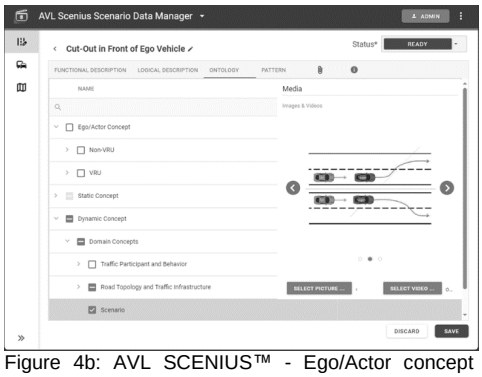


Figure 4b: AVL SCENIUS™ - Ego/Actor concept implementation

Triggering Conditions

Triggering conditions cause scenario constraints to activate functional insufficiencies in automated vehicles which needs to be validated to avoid hazardous behaviour. For example, if the ODD includes that the system should work in snowy and/or rainy conditions, then different levels of snow and rain (light, medium, heavy) must be included in the tests. The environment triggering condition is snowfall and consequently bad lightning. The below Figure 5 below, shows the cause effect model for snowfall and poor lightning causing scenario constraints, such as reduced visibility, low illuminance and lower friction on the road. The constraints need to be activated during testing of the scenario in order to check if the system behaves hazardously or not according to the ODD definition, e.g. in simulation the reduced sensor visibility range is tested between 0 and 500m and the road with different degree of friction.

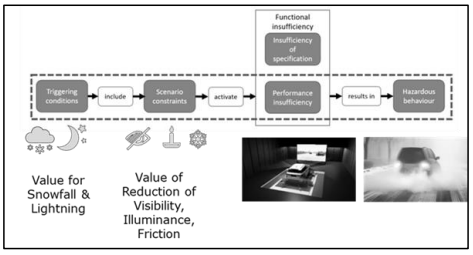


Figure 5: Cause/effect illustration for snow fall

In the AVL SCENIUS™ approach, the SOTIF triggering condition integration is done by adding two extensions to the ODD ontology concept (Figure 6):

1. A fourth ontology concept, called “SOTIF Concept” is added, which includes a row for the “Triggering Conditions” and
2. an extension to the "Dynamic Concept" is added to describe "Scenario Constraints".

More detailed information on the approach is described in [8].

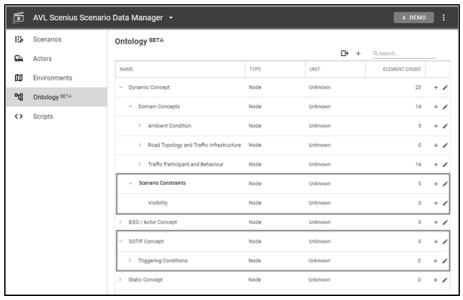


Figure 6: AVL SCENIUS™ - Scenario constraints and triggering conditions implementation

Validation Database

Testing and validation of autonomous vehicles requires the creation and/or extraction of complete scenario space in and partially outside the defined ODD. Those scenarios can be derived from various sources such as real-world data, simulations, regulatory mandates, accident scenarios or expert insights. Real-world data extraction plays a pivotal role in the validation of autonomous vehicles. For that, AVL SCENIUS™ can interface with TNO StreetWise. Automated driving type approval regulations, such as UNE 157 ALKS [3] and EU 2022/1426 [4], require approaches to create realistic validation scenario for the given ODD and residual risk assessment based on statistical data of occurrence of scenarios and parameters.

AVL SCENIUS™ stores and traces all scenario artefacts in a central database: test scenario (ASAM OpenScenario®), test environments (ASAM OpenDrive® and OpenScenGraph®), different actor model and ODD description. Figure 7 shows the test scenario database.

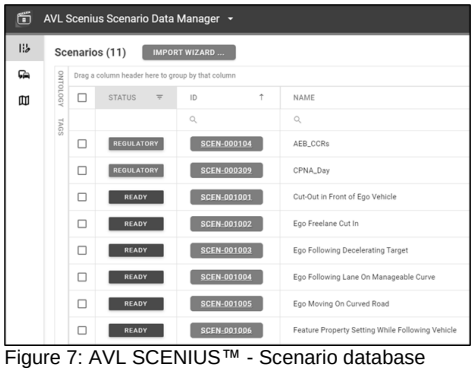


Figure 7: AVL SCENIUS™ - Scenario database

Validation Planning
A central part of an SOTIF conform validation toolchain is the systematic test planning and the implementation of different optimization strategies for the different test environments. The AVL SCENIUS™ structured planning approach connects and traces all test artefacts: the requirements, the ODD elements, the test scenarios and road segments together with a hierarchical project management.
Figure 8a shows an example project structure divided into: (P) Vehicle variant project, (F) Autonomous vehicle feature, (UC) Requirement use cases, (RQ) Specific requirement and (TO) Individual test order. The entire structure can be derived and linked to typical ALM tools.

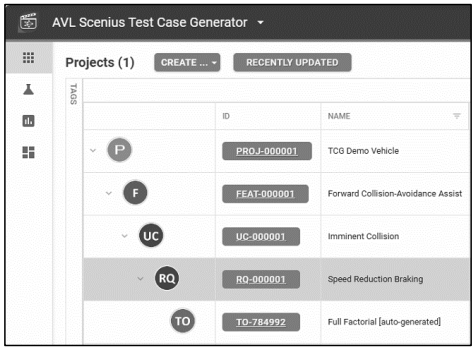


Figure 8a: AVL SCENIUS™ - Structured validation planning

As shown in Figure 8b, various data types are assigned to each project level to allows systematic evaluation of the autonomous vehicles.

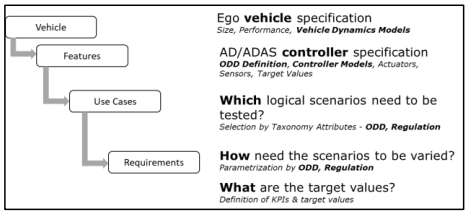


Figure 8b: Structured validation planning

In (P), the vehicle variant properties are specified, such as dimensions, sensor setup information or general vehicle models required for virtual validation. On the feature level (F), different ODD's can be created and managed together with the parameters and Key Performance Indicators (KPIs), such as ODD1 = 100km/h on German highways and sunny weather with TTC >= 4s. In the development and release process, it allows assignments, optimization and efficient testing of sub-ODD parameters and later improve the system performance for larger or complete ODD.
Use Cases (UC) are linked to scenario classes (logical scenarios) together with parameters and triggering conditions.
A requirements (RQ) maps to a specific situation (scenario) that has to be verified - and defines a specific variation of parameters derived from a use case (UC). The parameter variation ranges, validation criteria and target KPI values are derived and adapted from the feature level or specified separately for each single requirement.
The final outcome for this structured validation planning process are individual test orders which are required to be validated in the defined test environments.

Test Execution
A detailed explanation of AVL SCENIUS™ test execution building block and workflow is beyond of the scope of this paper. Autonomous vehicles validation requires extensive testing in various environments to ensure safety and acceptable residual risk. Systematic validation, especially in virtual environments, is crucial as it is cost-effective and time-efficient, reducing the need for expensive physical tests. However, virtual validation must correlate with real-world conditions to be effective. Different models are needed for various tests; for example, perception testing requires a high-fidelity environmental model, while motion control needs an accurate vehicle model.
AVL SCENIUS™ ASAM standard compatibility and open RestAPI ensures consistency and seamless integration into various testing environments and stages. The results of the test execution in different testing environments are collected and aggregated in a central database to allow sophisticated evaluation, dashboarding and reporting.

Safety Assessment
The basis of a solid safety assessment is structured validation planning linked to the requirements and definition of safety goals on feature level (F) or Requirement level (RQ). The AVL SCENIUS™ toolchain integrates four major parts to for advanced safety assessment:
1. Integrated KPI scripting engine based on Python modules
2. Flexible KPI assignment on requirement level (Figure 9a)

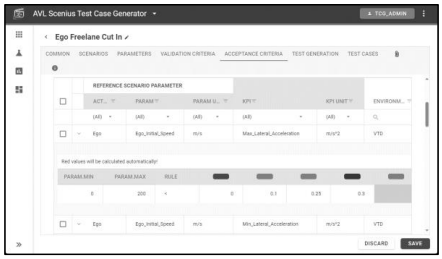


Figure 9a: AVL SCENIUS™ - Safety KPI definition

3. Safety result and test progress monitoring on each project levels (Figure 9b)

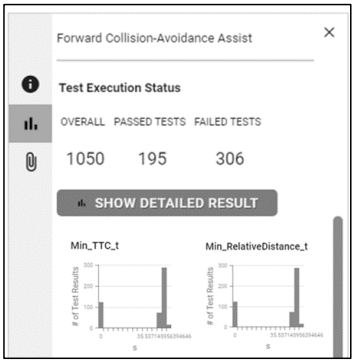


Figure 9b: AVL SCENIUS™ - safety and test progress monitoring

4. Detailed test analyzer of recorded signal of each executed test cases

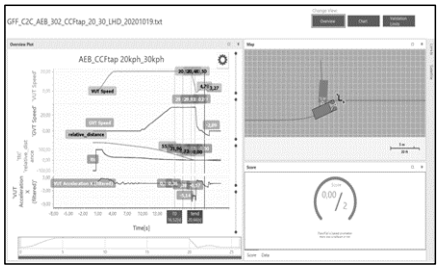


Figure 9c: AVLSCENIUS8TM)

3. Risk Quantification

The scope of this section is the safety evaluation required for the argumentation that an autonomous vehicle is free of reasonably foreseeable and preventable safety risks (as formulated in the ALKS regulation [5]). The challenge is to quantify a safety risk and indicate what is considered to be “reasonably foreseeable” and “reasonably preventable”, considering all the realistically possible combinations of situations and conditions that an ADS may encounter on the road. Because human experts cannot oversee the interactions of such a system with the environment and the safety implications of these interactions, a purely expert-based safety evaluation is no longer feasible. Therefore safety evaluation of advanced L2 and L3/L4 systems must be data driven as well.

Based on the validation testing and safety assessment, which is described in the previous chapter, a SOTIF risk quantification is carried out to identify critical risk aspect during development and homologation of autonomous driving systems. The risks quantification relates to limitations or intended functionality of the system in its ODD. The implemented approach is based on ISO26262 [2] and the approach presented in the risk quantification paper [6] and is defined by formula (1).

$$\sum_{i=1}^{n} (Exposure \times Severity \times Controllability) \quad (1)$$

Exposure signifies the probability of scenario category determined in real-world capturing.
Severity signifies the human or material harm that the autonomous driving function may lead to, which for example can be expressed in collision probability.
Controllability is mostly focused on human driving and excluded in autonomous driving functions.

Figure 10 shows the end-to-end toolchain integration that seamlessly connects real-world scenario capturing with virtual safety validation providing a data-driven safety risk assessment. The automated workflow includes the extraction of scenarios from data of the TOD, the parameter identification and the collection of required scenario statistics – to systematically cover all TOD/ODD elements. The link of the real-world scenario capturing to the virtual scenario testing allows the traceability of virtual testing back to the original real-world *exposure* to triggering conditions and enables the creation of safety evidence in the form of aggregate level, statistical safety risk estimates.

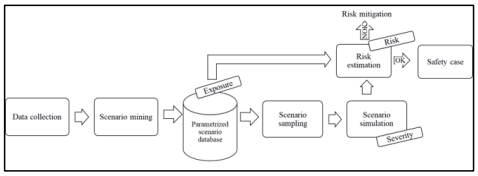


Figure 10: Toolchain for scenario-based risk quantification

Data collection
The data collection module contains object level data for both the sensor setup and the perception stack (including object recognition and tracking and sensor fusion). The original sensor data sources could be instrumented prototype vehicles, commercially operated vehicles, roadside cameras or drone images. The most important elements of the output data are the timeseries of positions of the ego vehicle and the other road users. Lane marking information is essential for locating the different road users with respect to the local road layout. Moreover, measurement data on lighting and weather conditions are recommended for a more detailed safety analysis.

Scenario mining
This module basically detects the pattern of each scenario category in the input data and measures the parameters of each observed concrete scenario. A scenario category is a collection of scenarios sharing similar characteristics. A scenario category refers to a qualitative description of a scenario [7]. The scenario mining module therefore takes the characterizing pattern of each scenario category as input and marks the timestamps in the data where the object data matches the scenario patterns. For each scenario category, scenarios are subsequently parameterized using a set of characterizing parameters. A concrete observed scenario will be stored in the scenario database with the set of parameters associated with the corresponding scenario category. The scenario mining results in thousands of scenarios in a scenario category as observed on Dutch and German highways are depicted in Figure 11 (TNO StreetWise). For simplicity of the illustration is limited to three parameters only.

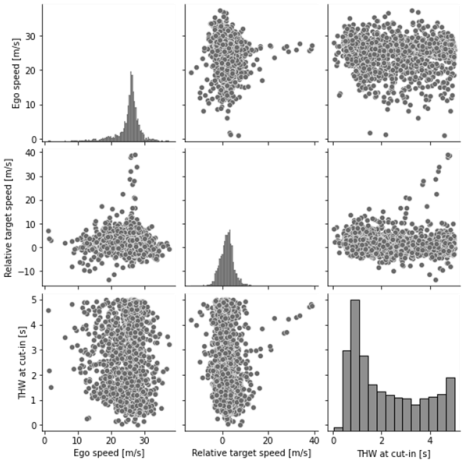


Figure 11: Scatter plots of three parameters for the scenario category

Parametrized scenario database
Each of the observed scenario as detected in the input data is stored in the database. As explained the observed scenarios are described as a combination of the scenario category and the parameters and tags that characterize each observed concrete scenario. Based on the thousands of concrete scenarios that are stored in this way in the database, the probability of occurrence of each scenario category as well as the Probability Density Function (PDF) of the scenario parameters can be estimated (Figure 11). These probabilities reflect the *exposure* values for the scenarios as observed in the actual TOD which will be used for the risk estimation step.

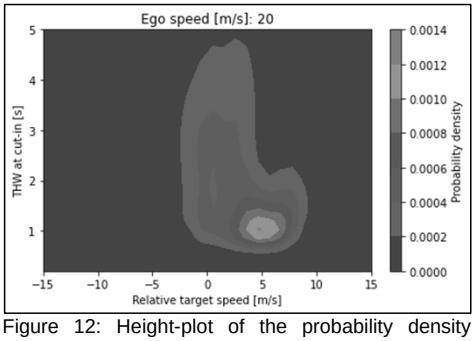


Figure 12: Height-plot of the probability density function of THW and relative target speed for ego speed slice at 20 m/s for cut-in scenarios

Scenario simulation
With scenario *exposure* values stored in the scenario database, the other main part of the safety assessment is the determination of the *severity* level. This step is performed by (validated) simulation in the AVL SCENIUS™ toolchain for a vast number of scenarios. The selection of the test scenarios to be simulated is performed by the scenario sampling, which is explained in the next section. The simulation delivers the *severity* level of each simulated test scenario. The *severity* level is expressed (at least) in terms of a binary representation of the outcome of the test scenario (collision or no collision), as well as safety criticality metrics selected by the use, such as TTC at near miss or relative velocity at impact. One might choose for performing detailed crash analyses, in which the *severity* is estimated from human models or from injury risk curves. Actually, even comfort parameters such as acceleration and jerk can be collected from the scenario simulations. This provides the possibility to use the same software pipeline for the statistical assessment of safety risk, perceived safety and even comfort.
The example Figure 13 shows the collision probability for cut-in scenarios for an advanced automated

driving function. In the simulation, varying values of the relative target speed and the distance between the ego and target vehicle at start of the cut-in (as Time Head Way) have been applied while the ego speed is shown at 20 m/s (other values are also sampled). Obviously, the collision probability is highest at large negative speed differences and short cut-in distance.

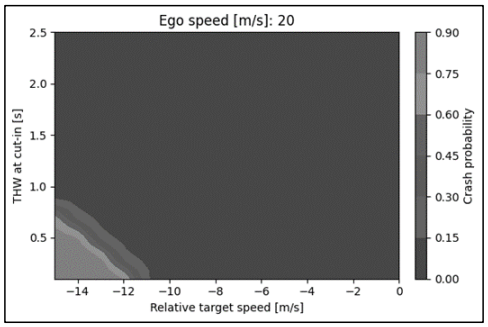


Figure 13: Collision probability for varying cut-in scenario parameter values

Scenario sampling
For a reliable safety assessment result, it is required that the TOD and ODD are simulated with sufficient coverage and in sufficient detail. A brute force (full factorial) approach would provide coverage over the parameter ranges but does not guarantee sufficient coverage in terms of granularity and quickly leads to an impractical amount of scenario simulations. Therefore, it is preferred to apply an iterative sampling approach in which more samples are drawn from areas with higher interest. The AVL SCENIUS™ toolchain supports different combinatorial and DoE sampling methods.

Risk estimation
The collision probability as shown in Figure 13 is not enough to estimate the actual risk for one scenario category as not every combination of parameter values is equally likely. The risk estimation module therefore combines the *severity* of the simulation outcome with the *exposure* values from the database. Figure 14 show the estimated risk by multiplying the estimated collision probability and the estimated occurrence probability density.
This total residual risk is subsequently obtained by integrating the risk over the entire parameter space. In case of the example cut-in scenario category the total residual risk is estimated to be $1.6 \cdot 10^{-8}$ per hour of driving. Provided that the scenario database is complete enough, the test coverage is sufficient, the simulation results are sufficiently validated, one can claim that the risk of a collision due to incorrect braking of the system under test in case of a cut-in scenario is represented by this value.

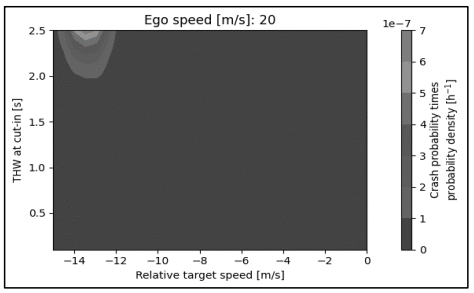


Figure 14: Estimated scenario risk

In a similar way, the risk of a fatal collision can be estimated, which value can be used to support a safety claim in the safety argumentation. When the residual risk is considered too high, the risk must be mitigated.
The final goal of SOTIF conforms safety and risk assessment is a solid and credible statement to homologation authorities of the safety performance and residual risk of the specific autonomous vehicle. The risk quantification reports provides a structured assessment of potential hazards associated with these technologies and if the systems meets the necessary safety standards and performance criteria, e.g. as outlined in UNECE Regulation 157 or the EU Regulation 2022/1426.

4. Conclusion

Ensuring the safety of autonomous vehicles is a critical challenge that requires a comprehensive and systematic approach. The SOTIF standard provides a robust framework to address this challenge by focusing on the specification of the Operational Design Domain (ODD), identifying potential hazards, evaluating conditions that could trigger hazardous behaviours to systematically investigate safety risks.
The use of advanced tools like AVL SCENIUS™ plays a pivotal role in this process. The tool fully supports the SOTIF workflow and facilitates a systematic and thorough approach to safety testing by combining real-world driving scenarios with virtual simulations. This dual approach allows for a detailed risk quantification, ensuring that all possible scenarios within the ODD are covered. Moreover, the workflow outlined in the document ensures end-to-end consistency and traceability, from the initial ODD description to the final validation and homologation of the system.
By adhering to the SOTIF standard and utilizing advanced testing tools like AVL SCENIUS™, developers can ensure that autonomous vehicles are capable of handling the complexities of real-world

traffic. This not only enhances the safety and reliability of these systems but also fosters public trust and acceptance of autonomous driving technology. As the industry continues to evolve, such rigorous safety frameworks and testing methodologies will be indispensable in achieving the goal of fully autonomous and safe transportation.
Further research and development work will focus on the three pillars: a) fidelity of the simulation models which must be adequate to deliver reasonable and trustable results, b) on greater support and coverage of different safety test types, such as functional safety tests or cybersecurity tests and c) on integrating approaches for risk assessment of unknown scenarios.

5. Acknowledgement

The paper was partially written at Virtual Vehicle Research GmbH in Graz, Austria. The authors would like to acknowledge the financial support within the COMET K2 Competence Centres for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Labor and Economy (BMAW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorized for the program management.

7. References

[1] International Organization of Standardization, "ISO/PAS 21448 Road vehicles - Safety of the intended functionality," International Organization of Standardization, 2022.

[2] International Organization of Standardization, "ISO 26262 Road vehicles - Functional safety," International Organization of Standardization, 2018

[3] UNECE, "UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS)," 01 April 2022. [Online]. Available: https://unece.org/transport/documents/2021/03/standards/un-regulation-no-157-automated-lane-keeping-systems-alks.

[4] EUR-Lex, "Commission Implementing Regulation (EU) 2022/1426 of 5 August 2022 laying down rules for the application of Regulation (EU) 2019/2144 of the European Parliament and of the Council as regards uniform procedures and technical specifications for the type-app," [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R1426. [Accessed 20 1 2024].

[5] UNECE, "UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS)," 01 April 2022. [Online]. Available: https://unece.org/transport/documents/2021/03/standards/un-regulation-no-157-automated-lane-keeping-systems-alks.

[6] E. D. Gelder, H. Elrofai, A. K. Saberi, J.-P. Paardekooper, O. O. D. Camp and B. D. Schutter, "Risk Quantification for Automated Driving Systems in Real-World Driving Scenarios," IEEE Access, 2021.

[7] E. de Gelder, J.-P. Paardekooper, A. Khabbaz, H. Elrofai, O. Op den Camp, S. Kraines, J. Ploeg and B. de Schutter, "Towards an Ontology for Scenario Definition for the Assessment of Automated Vehicles: An Object-Oriented Framework," IEEE Transactions on Intelligent Vehicles, pp. 1-1, 2022.

[8] Víctor J. Expósito Jiménez, Bernhard Winkler, Joaquim M. Castella Triginer, Heiko Scharke, Hannes Schneider, Eugen Brenner, and Georg Macher. 2024. Safety of the Intended Functionality Concept Integration into a Validation Tool Suite. Ada Lett. 43, 2 (December 2023), 69–72. https://doi.org/10.1145/3672359.3672369

8. Glossary

SOTIF: Safety of the Intended Function (ISO 21448)
ODD: Operational Design Domain
TOD: Target Operation Domain
ALM: Application Livecycle Management
PDF: Probability Density Function
KPI: Key Performance Indicator

Bibliography

[1] International Organization for Standardization, "ISO21448:2022 Road vehicles — Safety of the intended functionality," standard, International Organization for Standardization, Geneva, CH, 2022. → [pxiii], [p5], [p13], [p23], [p28]

[2] J. Bock, R. Krajewski, L. Eckstein, J. Klimke, J. Sauerbier, and A. Zlocki, "Data basis for scenario-based validation of HAD on highways," in *27. Aachen Colloquium Automobile and Engine Technology : October 8th-10th, 2018, Eurogress Aachen, Germany, 1-2*, pp. 8–10, Institute for Automotive Engineering, RWTH Aachen, Oct 2018. → [pxiii], [p14]

[3] Society of Automotive Engineers (SAE) International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," standard, Society of Automotive Engineers (SAE) International, 2021. → [pxiii], [p15], [p28]

[4] C. Schwarzl, N. Marko, H. Martin, V. Expósito Jiménez, J. Castella Triginer, B. Winkler, and R. Bramberger, "Safety and Security Co-engineering for Highly Automated Vehicles," *e & i Elektrotechnik und Informationstechnik*, vol. 138, pp. 469–479, nov 2021. → [p2], [p17], [p59]

[5] I. Cieslik, V. J. Expósito Jiménez, H. Martin, H. Scharke, and H. Schneider, "State of the art study of the safety argumentation frameworks for automated driving system," in *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops* (M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, eds.), (Cham), pp. 178–191, Springer International Publishing, 2022. → [p3], [p17], [p59]

[6] V. J. Expósito Jiménez, H. Martin, C. Schwarzl, G. Macher, and E. Brenner, "Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions," in *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops* (M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, eds.), (Cham), pp. 11–22, Springer International Publishing, 2022. → [p3], [p21], [p59]

[7] V. J. Expósito Jiménez, B. Winkler, J. M. Castella Triginer, H. Scharke, H. Schneider, E. Brenner, and G. Macher, "Safety of the Intended Functionality Concept Integration into a Validation Tool Suite," *ACM SIGAda Ada Letters*, vol. 43, pp. 69––72, jun 2024. → [p3], [p26], [p60]

[8] V. J. Expósito Jiménez, G. Macher, D. Watzenig, and E. Brenner, "Safety of the Intended Functionality Validation for Automated Driving Systems by Using Perception Performance Insufficiencies Injection," *Vehicles*, vol. 6, no. 3, pp. 1164–1184, 2024. → [p3], [p26], [p29], [p60]

[9] Heiko Scharke and Henk Goossens and Sytze Kalisvaart and Expósito Jiménez, Víctor J., "Systematic SOTIF-aligned approach to explore residual risk in autonomous driving systems," in *Book Of Proceedings - International Congress: SIA VISION 2024 - 16 and 17 OCTOBER 2024*, (Paris, France), pp. 185–192, Société des Ingénieurs de l'Automobile, 2024. → [p3], [p26], [p60]

[10] International Organization for Standardization, "ISO26262:2018 Road vehicles — Functional safety," standard, International Organization for Standardization, Geneva, CH, 2018. → [p5], [p6]

[11] Society of Automotive Engineers (SAE) International, "ISO/SAE 21434:2021 Road vehicles — Cybersecurity engineering," standard, Society of Automotive Engineers (SAE) International, 2021. → [p5]

[12] Association for the Advancement of Automotive Medicine, "Abbreviated Injury Scale." `https://www.aaam.org/abbreviated-injury-scale-ais/about-ais/`, 2008. Accessed: 2024-09-03. → [p6]

[13] Work group EGAS, "Standardized E-Gas monitoring concept for engine management systems of gasoline and diesel engines." `https://downloads.regulations.gov/NHTSA-2014-0108-0021/attachment_1.pdf`, 2008. Accessed: 2024-09-03. → [p8]

[14] Z. Zhu, R. Philipp, C. Hungar, and F. Howar, "Systematization and identification of triggering conditions: A preliminary step for efficient testing of autonomous vehicles," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, pp. 798–805, 2022. → [p9], [p18], [p22]

[15] Robert Bosch GmbH, "Sense, Think, Act: Was ein automatisiertes Fahrzeug können muss." `https://www.bosch-mobility.com/de/mobility-themen/automatisiertes-fahren-sense-think-act/`, 2024. Accessed: 2024-10-02. → [p10]

[16] P. Koopman, A. Kane, and J. Black, "Credible autonomy safety argumentation," in *27th Safety-Critical Systems Symposium 2019*, pp. 34–50, 2019. → [p13]

[17] United Nations Economic Commission for Europe (UNECE), "New Assessment/Test Method for Automated Driving (NATM) - Master Document (Final Draft)." `https://unece.org/sites/default/files/2021-01/GRVA-09-07e.pdf`, 2021. Accessed: 2024-02-09. → [p13]

[18] M. Scholtes, L. Westhofen, L. R. Turner, K. Lotto, M. Schuldes, H. Weber, N. Wagener, C. Neurohr, M. H. Bollmann, F. Körtke, J. Hiller, M. Hoss, J. Bock, and L. Eckstein, "6-layer model for a structured description and categorization of urban traffic and environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021. → [p13]

[19] M. Scholtes, L. Westhofen, L. R. Turner, K. Lotto, M. Schuldes, H. Weber, N. Wagener, C. Neurohr, M. H. Bollmann, F. Körtke, J. Hiller, M. Hoss, J. Bock, and L. Eckstein, "6-layer model for a structured description and categorization of urban traffic and environment," *IEEE Access*, vol. 9, pp. 59131–59147, 2021. → [p13]

[20] UL Standards and Engagement Inc., "UL4600 Standard for Safety: Evaluation of Autonomous Products," standard, UL Standards and Engagement Inc., 2021. → [p14]

[21] M. Ito, "ODD description methods for automated driving vehicle and verifiability for safety," *JUCS - Journal of Universal Computer Science*, vol. 27, no. 8, pp. 796–810, 2021. → [p14]

[22] JAMA and SAKURA, "Automated Driving Safety Evaluation Framework Ver. 3.0 -Guidelines for Safety Evaluation of Automated Driving Technology." `https://www.jama.or.jp/english/reports/docs/Automated_Driving_Safety_Evaluation_Framework_Ver3.0.pdf`, 2022. Accessed: 2024-09-09. → [p14]

[23] ASAM e. V., "ASAM OpenODD." `https://www.asam.net/project-detail/asam-openodd/`, 2021. Accessed: 2024-08-23. → [p14]

[24] International Organization for Standardization, "ISO34503:2023 Road Vehicles — Test scenarios for automated driving systems — Specification for operational design domain," standard, International Organization for Standardization, Geneva, CH, 2023. → [p14]

[25] Society of Automotive Engineers (SAE) International, "AVSC00002202004: AVSC Best Practice for Describing an Operational Design Domain: Conceptual Framework and Lexicon," standard, Society of Automotive Engineers (SAE) International, 2020. → [p15], [p23], [p48]

[26] The British Standards Institution, "BSI PAS 1883:2020 - Operational Design Domain (ODD) Taxonomy for ADS Specification," standard, The British Standards Institution, 2020. → [p15], [p23], [p28], [p38]

[27] European Commission, "Commission Implementing Regulation (EU) 2022/1426 - Commision implementing act AD v4.1," standard, European Commission, Brussel, BE, 2022. → [p17]

[28] Publications Office of the European Union, "Interpretation of EU Regulation 2022/1426 on the Type Approval of Automated Driving Systems, Publications Office of the European Union." `https://publications.jrc.ec.europa.eu/repository/bitstream/JRC136417/JRC136417_01.pdf`, 2024. Accessed: 2024-09-09. → [p17]

[29] UNECE, "UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS)," standard, United Nations Economic Commission for Europe, 2021. → [p17]

[30] National Transportation Safety Board (NTSB), "Collision between a Sport Utility Vehicle Operating with Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018." `https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR2001.pdf`, 2020. Accessed: 2021-12-05. → [p17]

[31] J.-F. Bonnefon, *18 THE UBER ACCIDENT*, pp. 93–98. The MIT Press, 2021. → [p17]

[32] S. A. Shah, *Safe-AV: A Fault Tolerant Safety Architecture for Autonomous Vehicles*. PhD thesis, McMaster University, 2019. → [p17]

[33] AI Incident Database, "Incident 293: Cruise's Self-Driving Car Involved in a Multiple-Injury Collision at an San Francisco Intersection." `https://incidentdatabase.ai/cite/293/`, 2022. Accessed: 2024-03-03. → [p17]

[34] S. Ballingall, M. Sarvi, and P. Sweatman, "Standards relevant to automated driving system safety: A systematic assessment," *Transportation Engineering*, vol. 13, p. 100202, 2023. → [p17]

[35] Philip Koopman, *How Safe is Safe Enough?* Independently Published, 2022. → [p17]

[36] International Organization for Standardization, "ISO34502:2022 Road vehicles — Test scenarios for automated driving systems — Scenario based safety evaluation framework," standard, International Organization for Standardization, Geneva, CH, 2022. → [p17]

[37] M. Khatun, M. Glaß, and R. Jung, "Scenario-Based Extended HARA Incorporating Functional Safety and SOTIF for Autonomous Driving," in *ESREL-30th European Safety and Reliability Conference*, November 2020. → [p17]

[38] D. Kinalzyk, "SOTIF Process and Methods in Combination with Functional Safety," in *Systems, Software and Services Process Improvement* (M. Yilmaz, P. Clarke, R. Messnarz, and M. Reiner, eds.), (Cham), pp. 612–623, Springer International Publishing, 2021. → [p17]

[39] A. Mekki-Mokhtar, J.-P. Blanquart, J. Guiochet, D. Powell, and M. Roy, "Safety trigger conditions for critical autonomous systems," in *2012 IEEE 18th Pacific Rim International Symposium on Dependable Computing*, pp. 61–69, 2012. → [p17]

[40] H. Martin, B. Winkler, S. Grubmüller, and D. Watzenig, "Identification of performance limitations of sensing technologies for automated driving," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 1–6, 2019. → [p17]

[41] C. Sun, R. Zhang, Y. Lu, Y. Cui, Z. Deng, D. Cao, and A. Khajepour, "Toward ensuring safety for autonomous driving perception: Standardization progress, research advances, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–19, 2023. → [p17]

[42] M. Hoss, M. Scholtes, and L. Eckstein, "A review of testing object-based environment perception for safe automated driving," *Automotive Innovation*, vol. 5, no. 3, pp. 223–250, 2022. → [p17]

[43] H. Wang, W. Shao, C. Sun, K. Yang, D. Cao, and J. Li, "A survey on an emerging safety challenge for autonomous vehicles: Safety of the intended functionality," *Engineering*, 2024. → [p18]

[44] L. Westhofen, C. Neurohr, T. Koopmann, M. Butz, B. Schütt, F. Utesch, B. Neurohr, C. Gutenkunst, and E. Böde, "Criticality metrics for automated driving: A review and suitability analysis of the state of the art," *Archives of Computational Methods in Engineering*, vol. 30, p. 1–35, Aug 2022. → [p18]

[45] J. C. Hayward, "Near-miss determination through use of a scale of danger," *51st Annual Meeting of the Highway Research Board*, vol. 43, no. 2, pp. 24–34, 1972. → [p18]

[46] V. E. Balas and M. M. Balas, "Driver assisting by inverse time to collision," in *2006 World Automation Congress*, pp. 1–6, 2006. → [p18]

[47] K. Ozbay, H. Yang, B. Bartin, and S. Mudigonda, "Derivation and validation of new simulation-based surrogate safety measure," *Transportation research record*, vol. 2083, no. 1, pp. 105–113, 2008. → [p18]

[48] J. Hillenbrand, A. M. Spieker, and K. Kroschel, "A Multilevel Collision Mitigation Approach—Its Situation Assessment, Decision Making, and Performance Tradeoffs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 4, pp. 528–540, 2006. → [p18]

[49] R. Schubert, K. Schulze, and G. Wanielik, "Situation Assessment for Automatic Lane-Change Maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 607–616, 2010. → [p18]

[50] F. Cunto and F. F. Saccomanno, "Calibration and validation of simulated vehicle safety performance at signalized intersections," *Accident Analysis and Prevention*, vol. 40, no. 3, pp. 1171–1179, 2008. → [p18]

[51] A. Abdulazim, M. Elbahaey, and A. Mohamed, "Putting safety of intended functionality sotif into practice," in *SAE WCX Digital Summit*, SAE International, apr 2021. → [p18]

[52] P. Weissensteiner, G. Stettinger, S. Khastgir, and D. Watzenig, "Operational Design Domain-Driven Coverage for the Safety Argumentation of Automated Vehicles," *IEEE Access*, vol. 11, pp. 12263–12284, 2023. → [p18]

[53] P. Weissensteiner, G. Stettinger, J. Rumetshofer, and D. Watzenig, "Virtual Validation of an Automated Lane-Keeping System with an Extended Operational Design Domain," *Electronics*, vol. 11, no. 1, 2022. → [p18]

[54] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016. → [p18]

[55] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, pp. 1–16, 2017. → [p18], [p38]

[56] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta, *et al.*, "LGSVL Simulator: A High Fidelity Simulator for Autonomous Driving," *arXiv preprint arXiv:2005.03778*, 2020. → [p18]

[57] F. Rosique, P. J. Navarro, C. Fernández, and A. Padilla, "A systematic review of perception system and simulators for autonomous vehicles research," *Sensors*, vol. 19, no. 3, 2019. → [p18]

[58] P. Kaur, S. Taghavi, Z. Tian, and W. Shi, "A survey on simulators for testing self-driving cars," in *2021 Fourth International Conference on Connected and Autonomous Driving (MetroCAD)*, pp. 62–70, 2021. → [p18]

[59] B. Schlager, S. Muckenhuber, S. Schmidt, H. Holzer, R. Rott, F. M. Maier, K. Saad, M. Kirchengast, G. Stettinger, D. Watzenig, and J. Ruebsam, "State-of-the-art sensor models for virtual testing of advanced driver assistance systems/autonomous driving functions," *SAE International Journal of Connected and Automated Vehicles*, vol. 3, pp. 233–261, oct 2020. → [p18], [p19]

[60] C. Pilz, G. Steinbauer, M. Schratter, and D. Watzenig, "Development of a scenario simulation platform to support autonomous driving verification," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 1–7, 2019. → [p18]

[61] Virtual Vehicle Research GmbH, "SPIDER: Mobile Platform for the Development and Testing of Autnomous Driving Functions." `https://www.v2c2.at/spider/`, 2021. Accessed: 2023-12-05. → [p18]

[62] V. J. Expósito Jiménez, C. Schwarzl, and H. Martin, "Evaluation of an indoor localization system for a mobile robot," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 1–5, 2019. → [p18]

[63] E. de Gelder, J.-P. Paardekooper, O. Op den Camp, and B. De Schutter, "Safety assessment of automated vehicles: how to determine whether we have collected enough field data?," *Traffic Injury Prevention*, vol. 20, pp. S162–S170, 2019. Peer-Reviewed Journal for the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV). → [p18]

[64] C. Linnhoff, K. Hofrichter, L. Elster, P. Rosenberger, and H. Winner, "Measuring the influence of environmental conditions on automotive lidar sensors," *Sensors*, vol. 22, no. 14, 2022. → [p18]

[65] J. Fang, D. Zhou, J. Zhao, C. Tang, C.-Z. Xu, and L. Zhang, "LiDAR-CS Dataset: LiDAR Point Cloud Dataset with Cross-Sensors for 3D Object Detection," 2023. → [p18]

[66] Drechsler, Maikol Funk, Sharma, Varun, Reway, Fabio, Schütz, Christoph, and Huber, Werner, "Dynamic vehicle-in-the-loop: A novel method for testing automated driving functions," *SAE International Journal of Connected and Automated Vehicles*, vol. 5, pp. 367–380, jun 2022. → [p19]

[67] AVL List GmbH, "AVL DRIVINGCUBE." `https://www.avl.com/en/testing-solutions/automated-and-connected-mobility-testing/avl-drivingcube`, 2023. Accessed: 2024-02-07. → [p19]

[68] K. Montalban, C. Reymann, D. Atchuthan, P.-E. Dupouy, N. Riviere, and S. Lacroix, "A quantitative analysis of point clouds from automotive lidars exposed to artificial rain and fog," *Atmosphere*, vol. 12, no. 6, 2021. → [p19]

[69] W. Dai, S. Chen, Z. Huang, Y. Xu, and D. Kong, "Lidar intensity completion: Fully exploiting the message from lidar sensors," *Sensors*, vol. 22, no. 19, 2022. → [p19]

[70] W. Sun, Y. Hu, D. G. MacDonnell, C. Weimer, and R. R. Baize, "Technique to separate lidar signal and sunlight," *Opt. Express*, vol. 24, pp. 12949–12954, Jun 2016. → [p19]

[71] D. Scheuble, C. Linnhoff, M. Bijelic, L. Elster, P. Rosenberger, W. Ritter, and H. Winner, "Simulating road spray effects in automotive lidar sensor models," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, pp. 659–666, 2024. → [p19]

[72] J. R. V. Rivero, T. Gerbich, B. Buschardt, and J. Chen, "The effect of spray water on an automotive lidar sensor: A real-time simulation study," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 1, pp. 57–72, 2022. → [p19]

[73] T. Goelles, B. Schlager, and S. Muckenhuber, "Fault detection, isolation, identification and recovery (fdiir) methods for automotive perception sensors including a detailed literature survey for lidar," *Sensors*, vol. 20, no. 13, 2020. → [p19]

[74] M. Segata, R. L. Cigno, R. K. Bhadani, M. Bunting, and J. Sprinkle, "A lidar error model for cooperative driving simulations," in *2018 IEEE Vehicular Networking Conference (VNC)*, pp. 1–8, 2018. → [p19]

[75] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11679–11689, 2020. → [p19]

[76] M. Dreissig, D. Scheuble, F. Piewak, and J. Boedecker, "Survey on lidar perception in adverse weather conditions," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1–8, 2023. → [p19]

[77] N. A. Minh Mai, P. Duthon, P. H. Salmane, L. Khoudour, A. Crouzil, and S. A. Velastin, "Camera and LiDAR analysis for 3D object detection in foggy weather conditions," in *2022 12th International Conference on Pattern Recognition Systems (ICPRS)*, pp. 1–7, 2022. → [p19]

[78] M. Hahner, C. Sakaridis, D. Dai, and L. Van Gool, "Fog Simulation on Real LiDAR Point Clouds for 3D Object Detection in Adverse Weather," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15263–15272, 2021. → [p19]

[79] M. Hahner, C. Sakaridis, M. Bijelic, F. Heide, F. Yu, D. Dai, and L. Van Gool, "LiDAR Snowfall Simulation for Robust 3D Object Detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16343–16353, 2022. → [p19], [p28]

[80] Ilhan Skender, "Robustness Test for ADAS Function," Master's thesis, Graz University of Technology, 2022. → [p19]

[81] F. Pizzati, P. Cerri, and R. de Charette, "Model-based occlusion disentanglement for image-to-image translation," in *Computer Vision – ECCV 2020* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), (Cham), pp. 447–463, Springer International Publishing, 2020. → [p19], [p28]

[82] J. Sadeghi, B. Rogers, J. Gunn, T. Saunders, S. Samangooei, P. K. Dokania, and J. Redford, "A step towards efficient evaluation of complex perception tasks in simulation," 2021. → [p19]

[83] A. Piazzoni, "Modeling perception errors in autonomous vehicles and their impact on behavior," 2023. → [p19]

[84] A. Piazzoni, J. Cherian, M. Slavik, and J. Dauwels, "Modeling perception errors towards robust decision making in autonomous vehicles," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI'20, 2021. → [p19]

[85] A. Piazzoni, J. Cherian, J. Dauwels, and L.-P. Chau, "PEM: Perception Error Model for Virtual Testing of Autonomous Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 670–681, 2024. → [p19]

[86] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," 2014. → [p19]

[87] J. Sadeghi, N. A. Lord, J. Redford, and R. Mueller, "Attacking motion planners using adversarial perception errors," 2023. → [p19]

[88] C. Innes and S. Ramamoorthy, "Testing rare downstream safety violations via upstream adaptive sampling of perception error models," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12744–12750, 2023. → [p19]

[89] L. Putze, L. Westhofen, T. Koopmann, E. Böde, and C. Neurohr, "On Quantification for SOTIF Validation of Automated Driving Systems," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1–8, 2023. → [p19]

[90] E. de Gelder, H. Elrofai, A. K. Saberi, J.-P. Paardekooper, O. Op den Camp, and B. de Schutter, "Risk quantification for automated driving systems in real-world driving scenarios," *IEEE Access*, vol. 9, pp. 168953–168970, 2021. → [p19], [p32]

[91] E. de Gelder and O. O. den Camp, "How certain are we that our automated driving system is safe?," *Traffic Injury Prevention*, vol. 24, no. sup1, pp. S131–S140, 2023. PMID: 37267005. → [p19]

[92] B. Kramer, C. Neurohr, M. Büker, E. Böde, M. Fränzle, and W. Damm, "Identification and quantification of hazardous scenarios for automated driving," in *Model-Based Safety and Assessment* (M. Zeller and K. Höfig, eds.), (Cham), pp. 163–178, Springer International Publishing, 2020. → [p19]

[93] J. Vaicenavicius, T. Wiklund, A. Grigaite, A. Kalkauskas, I. Vysniauskas, and S. D. Keen, "Self-driving car safety quantification via component-level analysis," *SAE International Journal of Connected and Automated Vehicles*, vol. 4, pp. 35–45, mar 2021. → [p20]

[94] D. Karunakaran, S. Worrall, and E. M. Nebot, "Efficient Statistical Validation with Edge Cases to Evaluate Highly Automated Vehicles," *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8, 2020. → [p20]

[95] J. Chu, T. Zhao, J. Jiao, Y. Yuan, and Y. Jing, "SOTIF-Oriented Perception Evaluation Method for Forward Obstacle Detection of Autonomous Vehicles," *IEEE Systems Journal*, vol. 17, no. 2, pp. 2319–2330, 2023. → [p20]

[96] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," *ArXiv*, vol. abs/1708.06374, 2017. → [p20], [p38]

[97] L. Peng, B. Li, W. Yu, K. Yang, W. Shao, and H. Wang, "SOTIF Entropy: Online SOTIF Risk Quantification and Mitigation for Autonomous Driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1530–1546, 2024. → [p20]

[98] U.S. Department of Transportation, "Federal Highway Administration: Road Function Classifications." `https://safety.fhwa.dot.gov/speedmgt/data_facts/docs/rd_func_class_1_42.pdf`, 2020. Accessed: 2024-08-23. → [p21]

[99] ASAM e. V., "ASAM OpenLabel." `https://www.asam.net/project-detail/asam-openlabel-v100/`, 2021. Accessed: 2024-05-31. → [p21]

[100] V. J. Expósito Jiménez, B. Winkler, J. M. Castella Triginer, H. Scharke, H. Schneider, E. Brenner, and G. Macher, "Safety of the Intended Functionality Concept Integration into a Validation Tool Suite," in *27th Ada-Europe International Conference on Reliable Software Technologies (AEiC 2023)*, jun 2023. → [p26]

[101] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. LeBlanc, "Accelerated evaluation of automated vehicles in car-following maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 733–744, 2018. → [p32]

[102] K. D. Kusano and H. C. Gabler, "Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1546–1555, 2012. → [p32]

[103] K. D. Kusano and H. C. Gabler, "Potential occupant injury reduction in pre-crash system equipped vehicles in the striking vehicle of rear-end crashes," *Annals of advances in automotive medicine. Association for the Advancement of Automotive Medicine. Annual Scientific Conference*, vol. 54, pp. 203–214, 2010. → [p32]

[104] T. A. Gennarelli and E. Wodzin, "AIS 2005: A contemporary injury scale," *Injury*, vol. 37, no. 12, pp. 1083–1091, 2006. Special Issue: Trauma Outcomes. → [p32]

[105] CENELEC - European Committee for Electrotechnical Standardization, "EN50126-2: Railway Applications – The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) Part 2: Systems Approach to Safety," standard, CENELEC - European Committee for Electrotechnical Standardization, Brussel, BE, 2017. → [p33]

[106] M. Malekzadeh and I. Bate, "Making an ALARP Decision of Sufficient Testing," in *2014 IEEE 15th International Symposium on High-Assurance Systems Engineering*, pp. 57–64, 2014. → [p33]

[107] AVL List GmbH, "AVL SCENIUS." `https://www.avl.com/en/testing-solutions/automated-and-connected-mobility-testing/avl-scenius`, 2023. Accessed: 2024-08-23. → [p36]

[108] ASAM e. V., "ASAM OpenScenario." `https://releases.asam.net/OpenSCENARIO/2.0-concepts/ASAM_OpenSCENARIO_2-0_Concept_Paper.html`, 2022. Accessed: 2024-07-29. → [p37]

[109] ASAM e. V., "ASAM OpenDrive." `https://www.asam.net/standards/detail/opendrive/`, 2022. Accessed: 2024-07-29. → [p37]

[110] ASAM e. V., "ASAM OpenXOntology." `https://www.asam.net/project-detail/asam-openxontology/`, 2022. Accessed: 2024-07-29. → [p37]

[111] Open Source Robotics Foundation, Inc, "ROS - Robot Operating System." `https://www.ros.org/`, 2023. Accessed: 2023-05-03. → [p38]

[112] The Autoware Foundation, "Autoware." `https://www.autoware.org/autoware`, 2021. Accessed: 2023-05-03. → [p38]

[113] Gassmann, Bernd and Oboril, Fabian and Buerkle, Cornelius and Liu, Shuang and Yan, Shoumeng and Elli, Maria Soledad and Alvarez, Ignacio and Aerrabotu, Naveen and Jaber, Suhel and van Beek, Peter and Iyer, Darshan and Weast, Jack, "Towards Standardization of AV Safety: C++ Library for Responsibility Sensitive Safety," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 2265–2271, 2019. → [p38]

[114] European Road Safety Observation - European Commission, "Motorways 2018." `https://road-safety.transport.ec.europa.eu/system/files/2021-07/ersosynthesis2018-motorways.pdf`, 2018. Accessed: 2023-09-07. → [p41]