PRISSMA Project
Plateforme de Recherche et d'Investissement pour la Sûreté
et la Sécurité de la Mobilité Autonome
04/2021 - 04/2024

[L8.2] HIGH-LEVEL SAFETY RULES AND IDENTIFICATION OF THEIR DOMAINS

REGLES DE SECURITE DE HAUT NIVEAU AVEC PERIMETRES IDENTIFIES

Main authors : P.-J. Meyer (Université Gustave Eiffel) and E. M. El Koursi (Université Gustave Eiffel)

Keywords: safety rules, safety assurance, autonomous driving

Abstract. This deliverable is related to Task 8.2 of the PRISSMA project. The main objectives of this task and deliverable are to evaluate the suitability of high-level safety requirements for autonomous vehicles in the existing literature and, when needed, to improve and refine their definition. The first chapter of this deliverable provides a presentation and literature review on the safety assurance of autonomous transport systems, while the second chapter lists and explains the identified safety requirements.

Résumé. Ce livrable est lié à la Tâche 8.2 du projet PRISSMA. Les principaux objectifs de cette tâche et ce livrable sont l'évaluation de l'adéquation des règles de sécurité de haut niveau pour les véhicules autonomes dans la littérature existante, et l'amélioration et le raffinage de leurs définitions si nécessaire. Le premier chapitre de ce livrable donne une présentation et un état de l'art sur l'assurance de la sécurité des systèmes de transport autonomes, et le second chapitre liste et explique les règles de sécurité identifiées.

Authors	Pierre-Jean Meyer (UGE), Mohammed Chelouati (UGE), Cédric Gava (SPHEREA), Florent Sovignet (STRMTG), El Miloudi El Koursi (UGE)
Document ID	PRISSMA/L8.2/V4
Date	28/03/2022
Type of document	Deliverable
Status	Final
Confidentiality	Confidential
WP allocation	WP8-T8.2
Distribution	PRISSMA partners
History Version 0 Version 1 Version 2 Version 3 Version 4	 05/07/2021 Creation 29/07/2021 Section 2.2 from Pierre-Jean Meyer 01/09/2021 Chapter 1 from Mohammed Chelouati, Section 2.3 from Cédric Gava, Section 2.6 and Appendix B from Florent Sovignet 07/10/2021 Switch to LaTeX template, Introduction and Sections 2.1 and 2.5 by El Miloudi El Koursi, Section 2.4 by Cédric Gava, Appendix A by Florent Sovignet 28/03/2022 Abstract, Conclusion and polishing of the content for the final version from Pierre-Jean Meyer.

Contents

INTRODUCTION

This document is related to task T8.2 of WP8. This task addresses the high level safety rules and identification of their domains. It has been prepared to provide an initial baseline of the safety requirements that have to be refined to ensure the safety assurance of autonomous vehicle within the context of PRISSMA (Plateforme de Recherche et Investissement pour la Sûreté et la Sécurité de la Mobilité Autonome) project.

The present document constitutes the final version of deliverable L8.2 “High level safety rules and identification of their domains”. It is structured in two main chapters, two appendices and references:

- Chapter 1 outlines the safety assurance of autonomous transport systems dealing with the safety case development. It presents the safety case elements based on safety requirements, objectives, evidences and arguments. It gives a literature review of the most widely used method GSN (goal structuring notation), as a graphical argumentation, to present a safety case in road, rail and aviation sectors.

- Chapter 2 identifies the high-level safety requirements for autonomous vehicles. These safety requirements should be traced to architecture elements that are responsible for the implementation of the measures preventing safety critical failures and malfunctions.

- The appendix A describes the GAME principle that aims to demonstrate that the whole system is globally-as-safe as a reference system.

- The appendix B describes the validation principles to validate the autonomous transport systems in a wide variety of use-cases.

1 SAFETY ASSURANCE OF AUTONOMOUS TRANSPORT SYSTEMS

1.1 Introduction

The key challenges facing the safety assurance of highly automated driving systems in open context are the apportionment and validation of adequate system safety requirements and the demonstration that these requirements will be guaranteed, within an environment that cannot be completely specified and continuously changes during operation of the system. The autonomous systems pose new issues, necessitating new design, development, and safety validation approaches. This brings the question of how autonomous systems vary from non-autonomous systems, and if the same method to safety assurance can be used to both. The conditions for acceptable functional safety for passenger vehicles are set by the international functional safety standard for road vehicles ISO 26262 and "Safety of the Intended Functionality" (SOTIF) focused on driver assistance, not highly automated driving systems.

With higher levels of driving automation, the system has to be able to control and monitor the vehicle as well as its environment. Due to the open road traffic space, the road traffic flow cannot be planned a priori, which is in contrast to aviation and railways. This underlines the novelty of the challenge of assuring the safety of higher levels of driving automation. The ISO 26262 standard requires the development of a safety case with a valid, evidence-based justification for a set of claims about the safety of a system for a given function over its operational context. ISO 26262 is the automotive specific adaption of IEC 61508 with requirements for the development of safety-relevant electrical and/or electronic (E/E) systems within road vehicles.

1.2 Safety cases

A safety case can be defined as a reasoned and compelling argument, supported by a body of evidence, that a system, service or organization will operate as intended for a defined application in a defined environment. According to [22], the safety case is based on tree principal elements: (i) safety requirement and objectives, (ii) safety evidence and (iii) safety argument (Figure 1). In safety cases, an argument is defined as a connected series of claims intended to establish an overall claim. In other terms, it communicates the relationship between evidence and objectives. Similarly, evidence should be used to support the claims. As a result, an argument is built through a hierarchy of claims. Both argument and evidence are essential components of the safety case and must be used in tandem. It's pointless to make an argument without evidence to back it up. It can be unclear whether (or how) safety objectives have been met if evidence is presented without argument.

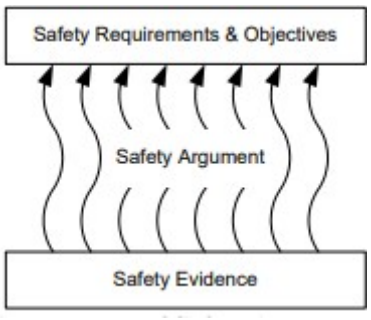

Figure 1: Role of safety argument

1.3 Safety argumentation

Demonstrating compliance with the safety standards involves collecting evidence that shows that the relevant safety criteria in the considered standards are met [10]. Although, safety standards describe the procedures for compliance that must be followed, which is often a challenging task to the system suppliers since these standards are presented in very large textual documents that are subject to interpretation [29]. In practice, a Safety Case will demonstrate that a given system is acceptably safe in each context or a given environment thought safety argumentation.

Argumentation is an approach that communicates the reasons why a system is considered acceptable [29]. The structure of the argument induces a specific way of structuring the evidence. The structure induced because of the argument can be expressed graphically or textually.

As presented in Figure 2, various graphical models have been used to illustrate an argumentation.

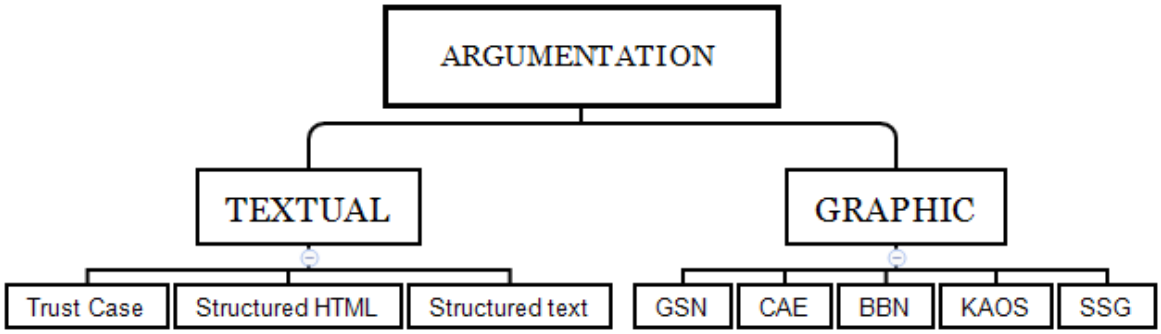

Figure 2: Safety argumentation

- **Goal Structuring Notation (GSN)**: A method that can be used to explicitly document the elements of an argument's structure and the argument's relationship to the evidence. In the GSN, the assertions of the argument are documented as the objectives and evidence are documented in the solutions [22, 1, 2].
- **Claims, Argument, and Evidence (CAE)**: Similarly to the GSN, the CAE is composed of a top-level claim asserted by an argument, a description of the arguments is presented to support the claim and a reference to the evidence is then presented to support the claim or the argument [4, 5, 20].
- **Bayesian Belief Network (BBN)**: Which induces a structure to be highlighted in a directed acyclic graph representing the conditional dependencies between them [6, 11, 15, 39].
- **Knowledge Acquisition in Automated Specification (KAOS)**: Which is a goal modelling language that is used for the Safety Case specification. This approach breaks down higher level goals using operator AND / OR in an argumentative way until proof of goal achievement is provided (i.e., goal directed requirements) [33, 7, 13].
- **Safety Specification Graph (SSG)**: Which are line graphs that represent a safety specification like nodes and proofs and relationships between them like edges [34].

From the literature reviewing, it seems that the GSN is the most widely used method, as a graphical argumentation, to present a safety case in safety critical systems. This is mainly since is respond to the challenges of safety cases based on its advantages as a structured approach to the development and presentation of safety evidence and its ability to present the individual elements of any safety case.

1.4 Goal Structuring Notation (GSN)

GSN is a graphical argument notation which can be used to document explicitly the elements and structure of an argument and the argument's relationship to evidence. In GSN, the claims of the argument are documented as goals and items of evidence are cited in solutions [38]. The principal symbols of the notation are shown in Figure 3. When the elements of GSN are linked together in a network they are described as a *goal structure*. The principal purpose of any goal structure is to show how goals (claims about the system) are successively broken down into sub-goals until a point is reached where claims can be supported by direct reference to available evidence (solutions). An example of GSN is given in Figure 4, which is derived from the Hazard Avoidance Pattern [23]. Actually, GSN enables to structure and represent the objectives to be achieved: goal (G1); the supporting goal evidence: solutions (Sn1); strategy for braking down a goal to sub-goals (S1), context (C1), etc. In addition to that, a notation of assumption is also used (A1) as we presented previously.

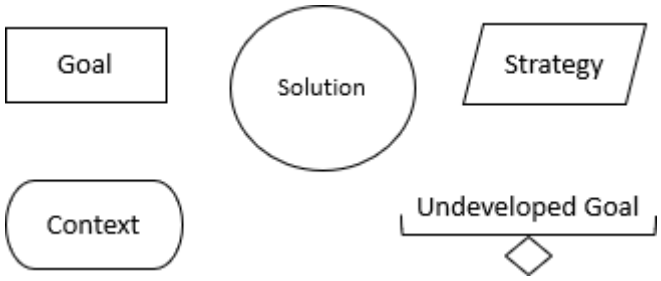

Figure 3: Role of safety argument

1.5 Literature review

Tables 1-3 below show the use of GSN approach in road, rail, and aviation sector.

GSN method was used in automobile domain for different reasons to structure the content of safety case elements, such as product and structure generic safety case modules to handle the complexity of the system representation and safety argument, or for reusable structure of safety assurance. The GSN was also used to create safety cases complying with ISO 26262 standard and representing patterns and procedures of safety assurance. Firstly, it provides an objective approach to capture concepts and relations used in the target safety standard, company, or project by keeping traceability of the relationships between the resulting conceptual model input and data. Secondly, it enables the explicit connection between conceptual models and safety cases to ensure that certification data is built properly and can be reused efficiently. Finally, using GSN, it enables well-structuring and well-controlling the content of safety case elements which reduces mistakes and misunderstanding between the different roles involved in producing, assessing, and using the safety case.

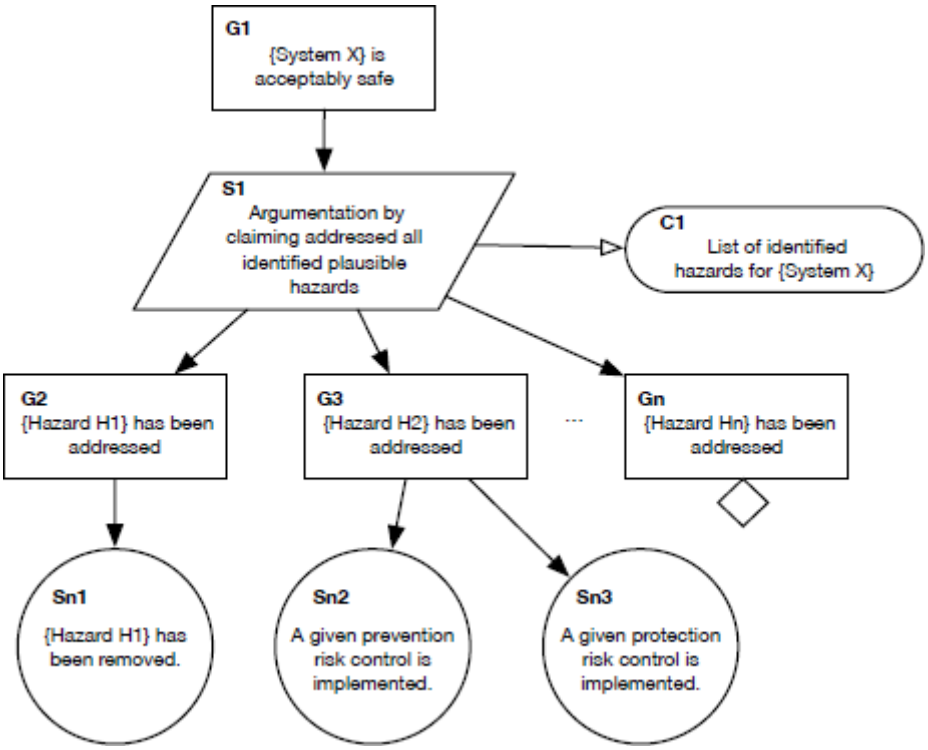

Figure 4: An example of GSN decomposition of high-level goal

Ref	Theoretical study	Process safety case	Product safety case	Compliance with standards	System level	Subsystem level	Applications
[40]			x	ISO 26262		x	
[32]	x	x		ISO 26262	x		
[16]	x	x		ISO 26262	x		
[26]			x	ISO 26262		x	Power Window system
[35]			x	ISO 26262	x		Vehicle Motion Control
[25]	x	x		ISO 26262	x		
[8]	x	x		ISO 26262		x	
[27]		x		ISO 26262		x	Hight Voltage System
[12]			x	ISO 26262		x	Fuel Level Estimation
[30]	x	x	x	ISO 26262	x		
[19]		x	x	ISO 26262		x	Air Suspension System

Table 1: Use of GSN in road sector, in compliance with ISO 26262 series.

Ref	Theoretical study	Process safety case	Product safety case	Compliance with standards	System level	Subsystem level	Applications
[24]		x	x		x		Gas Turbine Aero Engine Control System
[42]		x			x		Managing Mid-Air Collision Risk
[18]		x		DO-178C	x		Unmanned Aircraft System case study
[9]		x		Subcommittee F38.01 ASTM International	x	x	Small Unmanned Aircraft System
[14]		x			x	x	Unmanned Aircraft System mission
[28]	x	x			x		
[3]	x	x			x		Unmanned Aircraft System case study

Table 2: Use of GSN in aviation sector, in compliance with DO-178C standards.

Ref	Theoretical study	Process safety case	Product safety case	Compliance with standards	System level	Subsystem level	Applications
[37]		x		CENELEC EN 50126	x		Traceability Information Model
[36]	x			CENELEC EN 50129 ISO 26262	x		
[17]	x			CENELEC EN 50128 CENELEC EN 50129	x		MDSafeCer
[41]			x	CENELEC EN 50129	x	x	Wheel Slide Protection System
[21]		x	x		x		Train Door Controller
[31]		x	x	EMC Directive (2014/30/EU)	x	x	EM equipment/large machines

Table 3: Use of GSN in rail sector, in compliance with CENELES standards and ISO 26262.

2 HIGH-LEVEL SAFETY REQUIREMENTS

Before detailling the high-level safety requirements we highlight the context for autonomous mobility that is considered both in this section and in the PRISSMA project in general.

We are interested in autonomous road-based transportation systems with autonomy level SAE 4: the autonomous vehicle has high driving automation and can handle unexpected events or system failures, but the vehicle can only operate in self-driving mode in a limited area. The system of interest is a system of systems containing:

- the autonomous vehicle, equipped with artificial intelligence modules necessary for the autonomous driving (e.g. for perception, prediction or decision making);
- the infrastructure, which is augmented with additional perception or communication capacities. The capabilities of this infrastructure may vary depending on whether the vehicle evolves in a controlled environment (WP3) or in real conditions (WP4);
- a remote supervisor, which may be a human operator or a non-human supervision system, whose role is to intervene when necessary.

These safety requirements should be refined and traced to architecture elements that are responsible for the implementation of the measures preventing safety critical failures and malfunctions.

The structure and safety requirements in this section use the PFA report [45] as a baseline inspiration, from which some safety rules where reformulated, analyzed and refined when needed to better fit the context of PRISSMA, and new safety requirements were added. Note however that the safety rules from [45] on Operational Design Domain and Autonomous Driving mode are omitted in this deliverable to avoid overlaps with the dedicated PRISSMA deliverables L8.9 and L8.13 on these topics.

Each safety requirement below follows a similar structure, where we first provide a unique code for the requirement, followed by the statement of the requirement. Then, when needed or applicable, we provide some additional comments and analysis on this safety rule. Finally, for some of them, we also include (in blue) a list of references and some quotes taken from these references that we used to obtain the statement and analysis of our safety requirements.

2.1 Safety assurance

PRISSMA-SA-001 **A safety case shall be developed, in accordance with the safety plan, in order to provide the argument for the achievement of functional safety. In the case of a distributed development, the safety case of the item can be a combination of the safety cases of the customer and of the suppliers, which references evidence from the work products generated by the respective parties. Then the overall argument of the item is supported by arguments from all parties.**

The ISO 26262 [60] standard requires the development of a safety case with a valid, evidence-based justification for a set of claims about the safety of a system for a given function over its operational context. ISO 26262 is the automotive specific adaption of IEC 61508 with requirements for the development of safety-relevant electrical and/or electronic (E/E) systems within road vehicles. In practice, a Safety Case will demonstrate that a given system is acceptably safe in each context or a given environment thought safety argumentation.

#EndReq

PRISSMA-SA-002 A safety management system (SMS) shall be established, documented, and implemented to ensure the safety process of autonomous vehicle.

The SMS operates under the direction of the overall organisation safety policy, to ensure its consistency and objectives, to review its performance and to adapt it constantly to the changing environment (changing technology, organisational structures, and societal criteria). In order to emphasise the dynamic nature of a good SMS is organised into a planning and risk control system and a learning system to ensure that the effect of any such change is appropriately safely managed. As an example, in the railway sector the Safety Management System (SMS) must be approved by an authorising entity (a National Safety Agency (NSA) The French NSA, EPSF (Etablissement Public de Sécurité Ferroviaire) guideline [43] or the European Union Agency for Railways [44]). [58] AVSC published safety management system guidance for autonomous vehicle development adapted from similar frameworks used in the aviation, rail and nuclear industries, highlighting a systematic approach to testing and evaluating Automated Driving Systems (ADS) SAE level 4 and 5. The Safety assurance activities are a pilar of the Safety Management System (SMS) which is an important issue in all safety critical sectors that have formalised their SMS through their respective international organisations. To ensure the safety assurance of highly automated driving systems in open context requires to ensure that the adequate system safety requirements are implemented, traced, and validated by set up the high-level rules presented in the following sections in relation to the main rules.

#EndReq

2.2 Technical rules

PRISSMA-T-001 All high-level safety requirements for autonomous vehicles should not conflict with existing traffic laws and regulations for non-autonomous vehicles

All existing guidelines for the safe design of autonomous vehicles from administrative, legal or industrial entities [48, 49, 50, 53] agree that the first prerequisite for the deployment of autonomous vehicles is that they follow existing guidance, regulations and standards previously designed for the safety of non-autonomous road users. These guidance are established by standard-developing organizations such as the International Standard Organization (ISO) in particular with ISO/26262 on Road vehicles - Functional safety [60], the US Federal Motor Vehicle Safety Standards (FMVSS), or the SAE International.

References:

- Intel [48] Section 2.1.1
- Rand [53] “Measure Category 1”
- Transport Canada [50] “Outcome 4”
- US-DOT [49] “1-System Safety”

#EndReq

PRISSMA-T-002 The autonomous vehicle should communicate to other road users its autonomous-driving mode

The knowledge of the activation of the autonomous driving mode may influence how other road users behave around this autonomous vehicle (whether these road users consider au-

tonomous vehicles as more or less trustworthy than human drivers). It is thus important that an autonomous vehicle communicates its current driving mode to other users as clearly as possible to avoid possible misinterpretation [48, 50]. This should be the case for default autonomous driving mode, but also for other emergency driving modes that may require other road users to take additional safety precautions, such as during Minimum Risk Maneuvers as in Section 2.4. Knowledge of the autonomous driving mode may also be useful for other autonomous vehicles.

References:

- from Intel [48], one of the 12 principles of AD: Behavior in traffic. "The behavior of the automated function needs to not only be easy-to-understand for surrounding (vulnerable) road users, but also predictable and manageable."
- Transport Canada [50] Outcome 7 "Vehicle controls are accessible to users (that is, intuitive and easy to understand). The vehicle can communicate critical messages to occupants and other road users when needed, taking into account relevant accessibility factors, needs of different occupants, and the intended use of the vehicle. This outcome statement aims to prevent safety hazards that could arise from the accidental misuse or misinterpretation of the ADS features."

#EndReq #PRISSMA-MRM

PRISSMA-T-003 Vehicles with an automation level of 3 or lower should ensure that the driver remains vigilant and alert during autonomous-driving mode

As required in any autonomous vehicle with automation level of 3 or lower, the driver needs to remain alert and capable of taking over the driving when an autonomous-driving mode comes to abruptly stop. This need is highlighted by the fact that a large number of high-level safety requirements in this section explicitly involve the driver.

The system thus needs to be able to determine the state of alertness of the driver in order to determine whether it is safe to deactivate the autonomous-driving mode for the driver to take over [48].

The system should also be equipped with a "dead-man" component to ensure that the vehicle can safely reach a full stop when driver takeover is required while the driver alertness is not sufficient.

References:

- Intel [48] 12 principle "User responsibility" and Section 2.2.2.15: "To promote safety, the user's state (i.e. state of alertness) must be suitable for a responsible takeover procedure. The system should be able to recognize the user's state and keep them informed about their responsibilities concerning the required user's task. It should also be able to inform the respective operator about safety-relevant driving situations in unmanned driving services."
- US-DOT [49] section "6. Human Machine Interface": "For example, in a Level 3 vehicle, the driver always must be receptive to a request by the system to take back driving responsibilities. However, a driver's ability to do so is limited by their capacity to stay alert to the driving task and thus capable of quickly taking over control, while at the same

time not performing the actual driving task until prompted by the vehicle. Entities are encouraged to consider whether it is reasonable and appropriate to incorporate driver engagement monitoring in cases where drivers could be involved in the driving task so as to assess driver awareness and readiness to perform the full driving task."

#EndReq #PRISSMA-TR

PRISSMA-T-004 The vehicle design shall allow the driver to take over vehicle control at any time, according to the takeover procedure

The automation level of the considered vehicles in this project are limited both in terms of their partial autonomy and of the operational conditions in which these vehicles can be partially autonomous. It is thus natural to include the possibility for the driver to interrupt the autonomous-driving mode and take over at any time they deem necessary, even when the vehicle had not detected a specific need for this handover.

This requirement corresponds to the technical rule T-02 in [45, 46]. In addition, reference [48] also mention that this take-over procedure "shall require an explicit interaction from the vehicle operator, indicating a high confidence of intent". This sub-requirement would ensure that this take-over is never requested by mistake.

More detailed requirements on the takeover procedure can be found in Section 2.3 regarding transitions to and from autonomous-driving mode.

References:

- PFA [45] and VMAD [46]: rule T-02
- Intel [48] 12 principles "VEHICLE OPERATER-INITIATED HANDOVER": "Engaging and disengaging the automated driving system shall require an explicit interaction from the vehicle operator, indicating a high confidence of intent."

#EndReq #PRISSMA-TR #PFA-T-02

PRISSMA-T-005 The current driving mode should always be clearly identified and communicated to the driver

Since, even in autonomous-driving mode, the driver still needs to remain vigilant of the driving conditions in case of an unforeseen handover request from the vehicle, they need to be clearly and unambiguously informed of the current autonomy status of the vehicle or of a change in driving mode [48, 50]. Based on technical rules T-03, T-04 and T-05 in [45, 46], these information should include:

- whether the vehicle is in autonomous-driving mode or not
- which autonomous-driving modes are currently active (to reduce the driver's confusion when the vehicle is equipped with several levels or functions of autonomy)
- the vehicle behavior in autonomous-driving mode and the limits of this behavior
- the driver's remaining responsibilities while in autonomous-driving mode

- the procedure to comply with (such as possible handover request from the vehicle) and possible consequences if the driver does not comply.

Reference [49] also gives a list of the following five status considered as the minimum amount of information that the vehicle should be able to communicate to the driver: AD mode is functioning properly; AD mode is currently engaged; AD mode is currently unavailable for use; AD mode is experiencing a malfunction; requesting control transition from the AD mode to the operator.

References:

- PFA [45] and VMAD [46]: rules T-03 "The driver shall be clearly informed that the vehicle is in AD mode or not."
- PFA [45] and VMAD [46]: rules T-04 "In case of cohabitation on a single vehicle of several driving modes with different delegation levels, the necessary measures must be taken to control driver mode confusion risks (e.g. driver erroneously thinking he can stop to monitor vehicle and environment)."
- PFA [45] and VMAD [46]: rules T-05 "The driver shall be clearly informed of: the vehicle behavior in AD mode and the limits of this behavior ; his own responsibilities, the procedures to comply with (e.g. takeover procedure) and possible consequences if he does not comply."
- Intel [48] 12 principle User Responsibility,
  - subsection Responsibilities: "The aspects of the driving task which remain under the user's responsibility must be clear to the user."
  - subsection Mode awareness: "The automated function must ensure that the currently active driving mode can be recognized explicitly and unmistakably at any time. In addition, a change in driving mode must be clearly apparent to the user as well."
- Transport Canada [50] Outcome 7 "Vehicle controls are accessible to users (that is, intuitive and easy to understand). The vehicle can communicate critical messages to occupants and other road users when needed, taking into account relevant accessibility factors, needs of different occupants, and the intended use of the vehicle. This outcome statement aims to prevent safety hazards that could arise from the accidental misuse or misinterpretation of the ADS features."
- US-DOT [49], 6. Human Machine Interface:
  - "Understanding the interaction between the vehicle and the driver, commonly referred to as "human machine interface" (HMI), has always played an important role in the automotive design process. New complexity is introduced to this interaction as ADSs take on driving functions, in part because in some cases the vehicle must be capable of accurately conveying information to the human driver regarding intentions and vehicle performance. This is particularly true for ADSs in which human drivers may be requested to perform any part of the driving task."

  - "An ADS should be capable of informing the human operator or occupant through various indicators that the ADS is: • Functioning properly; • Currently engaged in ADS mode; • Currently "unavailable" for use; • Experiencing a malfunction; and/or • Requesting control transition from the ADS to the operator."

#EndReq #PFA-T-03 #PFA-T-04 #PFA-T-05

PRISSMA-T-006 The driver shall be educated and trained regarding the behaviors, capabilities and limits of the different autonomous-driving modes, as well as their own responsibilities, the procedures to comply with (e.g. takeover procedure) and possible consequences if they do not comply

While the previous requirement interpreted technical rule T-05 from [45, 46] as a set of information that needs to be provided to the driver during the use of the vehicle, another interpretation of rule T-05 is on the required education and training of owners and users of such autonomous vehicles, before using the vehicle and its autonomous-driving modes. Indeed, similarly to the requirement of obtaining the appropriate driving license before using any non-autonomous vehicle, users of autonomous vehicles need to receive the appropriate training regarding the autonomous-driving capabilities and behaviors, its limits, and the remaining responsibilities left to the driver [49, 50].

References:

- PFA [45] and VMAD [46]: rule T-05
- Transport Canada [50] "Outcome 8": "Expected outcome: Concrete actions have been taken to ensure awareness of the capabilities and limitations of the ADS features of the vehicle, as well as the vehicle's safe fallback conditions. Drivers/users are aware of what is expected of them in relation to the dynamic driving task under different conditions and of the vehicle and ADS features maintenance requirements. Drivers/users will be informed of any changes in these expectations that arise following a system update."
- US-DOT [49] "11. Consumer Education and Training": "Education and training is imperative for increased safety during the deployment of ADSs. Therefore, entities are encouraged to develop, document, and maintain employee, dealer, distributor, and consumer education and training programs to address the anticipated differences in the use and operation of ADSs from those of the conventional vehicles that the public owns and operates today. Such programs should consider providing target users the necessary level of understanding to utilize these technologies properly, efficiently, and in the safest manner possible.

  Consumer education programs are encouraged to cover topics such as ADSs' functional intent, operational parameters, system capabilities and limitations, engagement/disengagement methods, HMI, emergency fallback scenarios, operational design domain parameters (i.e., limitations), and mechanisms that could alter ADS behavior while in service. They should also include explicit information on what the ADS is capable and not capable of in an effort to minimize potential risks from user system abuse or misunderstanding."

#EndReq #PFA-T-05

PRISSMA-T-007 The autonomous vehicle should be able to safely handle malfunctions and failures of some of its components

Since the autonomy of the vehicle relies on a large number of hardware and software components that are each subject to possible failures, the vehicle needs to be equipped with strategies to preserve the safety of the autonomous-driving mode when these failures occur, or to safely disengage the autonomous-driving mode if these failures are too critical [48, 49, 50]. In [49], they consider two types of strategies to mitigate possible failures of the system.

Fail-Operational mechanisms are strategies that allows the autonomous-driving mode to continue to function despite the failure of one or more components. These strategies include: hardware and software redundancies, as particularly emphasized for the sensing and perception components in [48, 54] and in technical rule T-01 of [45, 46]; adaptive compensation; or driving in a degraded mode (e.g. reduced speed, reduced autonomy capabilities) to mitigate the safety risks.

Fail-Safe mechanisms are strategies to safely disengage from the autonomous-driving mode when the failure affects the safety of this mode. The possible strategies involve either requesting the driver to take over [53] (see the high-level safety requirements for transitions to/from autonomous-driving mode in Section 2.3 of this document), or to engage in Minimal Risk Maneuvers (see the high-level safety requirements for MRM in Section 2.4 of this document) or Minimal Risk Conditions [49, 50] to safely stop the vehicle unless the driver takes over control. Technical rule T-06 of [45, 46] also adds that if such a safety-critical failure occurred and forced the deactivation of autonomous-driving mode, then it should not be allowed to re-enable this mode until the vehicle has gone through proper maintenance operations to identify and fix the failure.

General references

- Intel [48] Safe operation/Dealing with degradations: “If safety-related functions or system components become hazardous (e.g. unavailable), the automated driving system shall:
  - Be capable of compensating and transferring the system to a safe condition/state (with acceptable risk).
  - Ensure a sufficient time frame for the safe transition of control to the vehicle operator.”
- Transport Canada [50], 6. SAFETY SYSTEMS: “Expected outcome: The vehicle is equipped with safety systems with appropriate redundancies that continuously monitor system performance, perform fault detection and hazard analysis, signal any malfunctions, and ultimately take corrective actions or revert to a minimal risk condition when needed.”
- US-DOT [49] (second file), Chapter 6, section Failure mitigation strategies. “Based on the general failure modes identified, potential failure mode responses and strategies were identified. This effort focused on FS strategies for cases where the ADS cannot continue to operate due to a significant failure, and FO strategies for cases where the ADS could continue to operate even in the face of a failure.

  - Fail-Safe Mechanisms. The primary goal of an FS strategy is to rapidly achieve an MRC where the vehicle and occupants are safe. Three candidate FS mechanisms were considered for further evaluation.
    * Transition to fallback-ready user control
    * Safely stop in lane of travel
    * Safely move out of travel lane and stop
  - Fail-Operational Mechanisms. FO strategies allow the ADS to continue to function, even in the event of one or more failures. It is important to note that this operation may only be supported for a limited duration, or potentially with a reduced set of capabilities. Three primary FO mechanisms were considered for further analysis.
    * Hardware/software redundancy
    * Adaptive compensation
    * Degraded operations
      · Reduced top speed
      · Reduced level of automation
      · Reduced ODD
      · Reduced maneuver capabilities
      · Reduced OEDR capabilities"

More specific on perception and sensor redundancy

- PFA [45] and VMAD [46]: rules T-01 "A single perception malfunction without failure should not induce a hazardous event. Consequently, the set of sensors used for the perception of a safety relevant environmental feature shall not be based on a single physical principle."
- Intel [48] Section 2.2.2.1 ENVIRONMENT PERCEPTION SENSORS: "In the unlikely event of severe sensor degradation or E/E faults, the sensor arrangement needs to be laid out such that it enables the safe capturing of relevant elements in degraded mode until the safe state is reached."
- Mercedes [54]: "By optimizing the design and/or providing redundant components or systems we can best ensure the automated vehicle's essential functions remain operational, even when a malfunction occurs."

More specific on AD failure and disengaging AD mode

- PFA [45] and VMAD [46]: rules T-06 "In case of failure impacting safety in AD mode, an appropriate degradation concept shall be to inhibit AD mode until next vehicle switch off and vehicle proper operation has been verified either by self-diagnostic or by maintenance."
- Rand corporation [53] on Measure category 2: "A disengagement is the deactivation of the autonomous mode when a failure of the autonomous technology is detected or when the safe operation of the vehicle requires that the autonomous vehicle test driver disengage the autonomous mode and take immediate manual control of the vehicle. (California Code of Regulations, undated.)"

- Transport Canada [50] Outcome 9. USER PROTECTIONS DURING COLLISIONS OR SYSTEM FAILURES: "The vehicle will be brought to a safe state following a collision or system failure, and will convey safety critical information to passengers, first responders, and emergency services."
- US-DOT [49] 4. Fallback (Minimal Risk Condition): "Entities are encouraged to have a documented process for transitioning to a minimal risk condition when a problem is encountered or the ADS cannot operate safely. ADSs operating on the road should be capable of detecting that the ADS has malfunctioned, is operating in a degraded state, or is operating outside of the ODD. Furthermore, ADSs should be able to notify the human driver of such events in a way that enables the driver to regain proper control of the vehicle or allows the ADS to return to a minimal risk condition independently."

#EndReq #PRISSMA-TR #PRISSMA-MRM #PFA-T-01 #PFA-T-06

2.3 Transitions to/from autonomous driving mode

Analysis of High-Level requirements from [45]. The use case of this document is focused on autonomous vehicle with a Driver.

- SAE Level 3 or 4
- passenger car
- driver on-board, in driver seat
- driver able to drive (driving capability as well as driving license)
- driving on motorways or "equivalent" roads (no crossroads, central safety guard, rare pedestrian).

This use case is updated to fit a system of autonomous driving vehicle, which main actor can be remote. This system of systems includes:

- Autonomous vehicle
- Infrastructure
- Command/Control center
- Enabling systems (test systems, integrated logistic support, . . .)

Actors : The Driver is replaced by an Operator which can be Local Operator (inside the vehicle), or Remote Operator in the Command/Control. Theses operators are required for ensuring the completion of the functions requiring human oversight [63].

PRISSMA-TR-001 A deliberate operator action is required to activate AD mode. OR The activation of AD mode shall be Human In The Loop (HITL) function with the operator [63].

The AD vehicle shall start MRM in case of disconnection with the Command/Control center. The monitoring of connection with control/command center, which is a security function, should comply with the state of the art rules to insure highest safety level:

- Redundancy, preferably with asymmetric technologies
- Non AI function, which would prevails over other AD functions

The remote operator shall be able to trigger MRM if deactivation of AD fails
#EndReq #PFA-TR-01

PRISSMA-TR-002 The operator actions to takeover shall be identical in both following cases: the operator takes over from his/her own (without prior system request); the system request the operator to takeover.

This requirement, which makes sense for a local operator inside the vehicle, mays not be applicable for any use case of autonomous vehicle (for accountability reasons for example: a crash analysis involving an autonomous vehicle shows that the remote operator was driving the vehicle when the crash occurs. If the operator has takeover on system request, his legal charges may be less than if he decided by himself to takeover).
#EndReq #PFA-TR-02

PRISSMA-TR-003 When the operator takes over vehicle control on her/his own (without prior system request), the vehicle shall not disturb the remote operator takeover by an inappropriate action (e.g. by switching headlamps off, at night).

The scenario management process shall identify all the acceptable/inappropriate actions of takeover.
#EndReq #PFA-TR-03

PRISSMA-TR-004 When the operator takes over after a system request, the system shall give back the control to the operator with a vehicle configuration maximizing operator controllability (e.g. wipers ON in case of rain, headlamps ON by night).

The scenario management process shall identify all the actions that improves the operator controllability.
#EndReq #PFA-TR-04

PRISSMA-TR-005 If the operator does not takeover vehicle control after a system request, the system shall start execution of a MRM. If the remote operator still does not takeover during the MRM, the vehicle will be stopped (refer to MRM requirements).

#EndReq #PFA-TR-05 #PRISSMA-MRM

PRISSMA-TR-006 The AD mode deactivation (end of vehicle longitudinal and lateral control) shall only be performed when system has verified that the operator has taken over vehicle control. This verification shall at least include a criterion on vehicle lateral control (except if the vehicle is already stopped).

#EndReq #PFA-TR-06

PRISSMA-TR-007 In AD mode, if situation would be difficult to control by the operator (taking into account vehicle technical status and urgency level) the vehicle: shall manage the situation without requesting the operator to takeover; shall inform the operator.

Comment: The process shall maximize the number of Scenario where this situation occurs : the ADS system manages hazardous situation without requesting driver to takeover.
#EndReq #PFA-TR-07

PRISSMA-TR-008 "Non Driving activities" allowed in AD mode shall be consistent with the available delay for operator takeover after a system request. The operator has to be informed that he must be at any time in a situation, which enables him to answer to the requests of the system within the requested time period.

#EndReq #PFA-TR-08

PRISSMA-TR-009 "Non Driving activities" allowed in AD mode and available through vehicle systems shall be: only available in AD mode; interrupted, with a specific HMI, when the vehicle requests the operator to takeover or when the driver takes the control on her/his own.

#EndReq #PFA-TR-09

2.4 Minimum risk maneuvers (MRM)

PRISSMA-MRM-011 The list of MRM of a particular autonomous driving system shall be defined based on its ODD and accepted by certification authorities.

The original requirement states: "For an automated function, which consists to operate at moderate speed, in dense traffic, on highway with a driver on-board, a possible MRM is to slow down the vehicle and stop it in its lane". This requirement has poor value because it does not state what shall be done but what could be done in a general way, for a specific kind of autonomous system in a particular environment. Instead, an high level requirement is proposed, which targets the safety assessment process as a whole.
#EndReq #PFA-MRM-01

PRISSMA-MRM-012 If an autonomous driving system has more than one MRM, each MRM shall be associated to one particular subset of the autonomous driving system ODD including ranges outside its ODD. The union of all these subsets shall cover the whole ODD and their outside ranges, and their intersection shall be empty to demonstrate the ability of the system's ODD to select one and only one particular MRM based on its operational conditions.

The focus here is to enforce that the discrimination between two possible MRM suffers no ambiguity at the decision stage. During operation, the best MRM can be different : if the autonomous driving system is close to the emergency lane, or if it doesn't, or if it cannot assess the distance with emergency lane leads to different situations. Moreover, if the AD systems state goes beyond its ODD, the MRM maneuvers triggered should also be detailed.
#EndReq #PFA-MRM-01

PRISSMA-MRM-013 Each MRM of a given autonomous driving system shall be qualified by an independent safety authority accountable of the validity of the MRM based on the ODD it's associated with.

#EndReq #PFA-MRM-01

PRISSMA-MRM-002 During the whole MRM, the operator can takeover in usual ways (refer to TR requirements).

#EndReq #PFA-MRM-02 #PRISSMA-TR

PRISSMA-MRM-003 In MRM phasis, when the vehicle is stopped, it shall signals itself to other drivers by flashing hazard (or stop) lights.

#EndReq #PFA-MRM-03

PRISSMA-MRM-004 At the end of the MRM:

- **for a short period of time (typically 15 s), vehicle is maintained standing still without operator action (e.g. despite a slope) and driver can takeover in the usual way (refer to TR requirements)**
- **after this period or if driver decide to immobilize the vehicle:**
  - **vehicle is definitively immobilized (Parking brake AND Gearbox N or P) OR (Gearbox on P)**
  - **AD mode is deactivated**

#EndReq #PFA-MRM-04 #PRISSMA-TR

2.5 Monitoring, reporting and learning

PRISSMA-MRL-001 The OEM should specify procedures for monitoring, reporting and learning system within the scope of autonomous vehicle operations and in compliance with standards and regulations.

Audit and monitoring enable OEMs (Original Equipment Manufacturer) to get a feedback on vehicle performances prior to an accident or incident (as opposed to investigation). They are also crucial to the continual review and upgrade of the autonomous vehicle, and the fostering of a learning system. Generally, a learning system is supported by a database where the records of events investigations and corrective actions are kept. This type of database is generally developed by extending the system hazard logs, and is maintained by OEMs. Hence the autonomous vehicle performances should be monitored, and any near-miss, incident or accidents that vehicle is unable to prevent should be rigorously investigated. PFA [45] and VMAD [46] recommend that an OEM shall have processes to monitor incident/accidents with their vehicles overtime and react appropriately. Reference [58] AVSC published safety management system guidance for autonomous vehicle development adapted from similar frameworks used in the aviation, rail and nuclear industries, highlighting a systematic approach to testing and evaluating Automated Driving Systems (ADS) SAE level 4 and 5. This framework based on four pilars (promoting

continuous learning, taking ownership, scaling safety management and transferring best practices) aims to identify, track, and trace potential safety risks at a holistic level.

References

- PFA [45] and VMAD [46]: rule SC-01 "The OEMs shall set up a common process to create and maintain a common catalogue of scenario, including misuses, to be used for safety argumentation during design and verification/validation phases. The catalogue will be enriched continuously. This set up shall be made in compliance with laws (e.g. competitive laws)."
- PFA [45] and VMAD [46]: rule AS-01 " The OEM shall set up internally, a process to collect, analyse and treat incidents/accidents faced by the customers, and if necessary, update the vehicles."
- PFA [45] and VMAD [46]: rule AS-02 " The OEMs shall share the lessons learnt from field experience, including safety- related events occurring in real life vehicle use, in order to enrich a common scenario catalogue (in compliance with laws - e.g. competitive laws)."

#EndReq #PFA-SC-01 #PFA-AS-01 #PFA-AS-02

2.6 Safety targets

PRISSMA-ST-001 Automated Vehicle deployment shall improve the road safety.

PRISSMA-ST-002 The automated vehicle is free from unreasonable risk.

Two references were used as input for this section:

- [45]: This paper deals with the following use case: SAE Level 3 or 4 passenger car, driver on-board in driver seat, driving on motorways or "equivalent" roads (no crossroads, central safety guard, rare pedestrian). This use case is different from the use cases focused on by PRISSMA. Nevertheless, author's opinion is that, although the principles have been defined for a specific use case, they can be applied (in particular the qualitative safety principles and field experience principles) in other use cases, possibly after adaptation.
- [46]: This paper is not limited to use case SAE L3-4 passenger cars on highways. It describes what could be the French approach for safety validation of AD vehicles (co-written by authorities and ecosystem). It emphasizes the fact that it is not a formal nor definitive position from French authorities nor French industry on regulatory options.

Both references address the following safety targets:

- A: "Automated Vehicle deployment shall improve the road safety"
- B: "The automated vehicle is free from unreasonable risk"

[45] goes a step further and defines à SOTIF "validation acceptance criteria" intended to be consistent with safety target A and derived from risk level on highways. The GAME (at least as good as) method was chosen to define risk acceptability (the compared system being a vehicle with a human driver, for which the risk level can be determined through road safety statistics). The final results are as follows: SOTIF overall "validation acceptance criteria":

- P(accident with fatalities) $\leq 10^{-8}$ /h (with a safety factor of 10)
- P(accident with light or severe injuries) $\leq 10^{-7}$ /h (round up to keep a one order of magnitude gap with regards to accident with fatalities).

The author's opinion is that these "validation acceptance criteria" ensure that the introduction of a highly automated vehicle on highways will not increase the level of risk for the public.

These quantitative safety targets are the safety levels that must be reached by the system as a whole for a given hazard. They can be expressed as tolerable hazard rates (THR) objectives that are apportioned during the safety analysis process to systems functions whose failures and malfunctioning lead to a given identified hazardous situation.

Note that in France, it is plausible that the authorities will set some quantitative safety targets. Discussions on this topic already took place within the GAME working group by defining probabilistic acceptance criteria based on accident probabilities, for a general use case. The accidents probabilities have been estimated on the basis of French Data Base of accidents data, and average mileage of 3 categories of road vehicles: light vehicles, busses and coaches. The results still have to be validated, but the orders of magnitude are the same as the risk acceptance criteria presented here. There is also ongoing legislative work at the EU level, which could also result in the setting of safety targets for ADS.

#EndReq

CONCLUSION

This document is the final deliverable related to Task 8.2 in WP8 of the PRISSMA project on the identification of high-level safety requirements for autonomous vehicles. We first provided a detailed description of the concept of safety assurance for autonomous transport systems, covering safety cases, safety arguments, and a literature review of the most widely used method GSN (Goal Structuring Notation). The second chapter of this deliverable focuses on defining and explaining the safety requirements identified in Task 8.2. This report also contains two appendices on important topics related to the safety of autonomous vehicles: the GAME principle; and safety validation principles.

REFERENCES

[1] Althammer, E., Schoitsch, E., Sonneck, G., Eriksson, H., Vinter, J., 2008. Modular certification support — the DECOS concept of generic safety cases, in: 2008 6th IEEE International Conference on Industrial Informatics. Presented at the 2008 6th IEEE International Conference on Industrial Informatics, pp. 258–263. https://doi.org/10.1109/INDIN.2008.4618105

[2] Ankrum, T.S., Kromholz, A.H., 2005. Structured assurance cases: three common standards, in: Ninth IEEE International Symposium on High-Assurance Systems Engineering (HASE'05). Presented at the Ninth IEEE International Symposium on High-Assurance Systems Engineering (HASE'05), pp. 99–108. https://doi.org/10.1109/HASE.2005.20

[3] Asaadi, E., Denney, E., Menzies, J., Pai, G., Petroff, D., 2020. Dynamic Assurance Cases: A Pathway to Trusted Autonomy. Computer 53, 35–46. https://doi.org/10.1109/MC.2020.3022030

[4] Bishop, P., Bloomfield, R., 2000. A Methodology for Safety Case Development. Saf. Reliab. 20, 34–42. https://doi.org/10.1080/09617353.2000.11690698

[5] Bloomfield, R., Bishop, P., 2010. Safety and Assurance Cases: Past, Present and Possible Future – an Adelard Perspective, in: Dale, C., Anderson, T. (Eds.), Making Systems Safer. Springer, London, pp. 51–67. https://doi.org/10.1007/978-1-84996-086-1_4

[6] Bouissou, M., Martin, F., Ourghanlian, A., 1999. Assessment of a safety-critical system including software: a Bayesian belief network for evidence sources, in: Annual Reliability and Maintainability. Symposium. 1999 Proceedings (Cat. No.99CH36283). Presented at the Annual Reliability and Maintainability. Symposium. 1999 Proceedings (Cat. No.99CH36283), pp. 142–150. https://doi.org/10.1109/RAMS.1999.744110

[7] Brunel, J., Cazin, J., 2012. Formal verification of a safety argumentation and application to a complex UAV system, in: International Conference on Computer Safety, Reliability, and Security. Springer, pp. 307–318.

[8] Burton, S., Gauerhof, L., Heinzemann, C., 2017. Making the Case for Safety of Machine Learning in Highly Automated Driving, in: Tonetta, S., Schoitsch, E., Bitsch, F. (Eds.), Computer Safety, Reliability, and Security, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 5–16. https://doi.org/10.1007/978-3-319-66284-8_1

[9] Clothier, R., Denney, E., Pai, G.J., 2017. Making a Risk Informed Safety Case for Small Unmanned Aircraft System Operations, in: 17th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2017-3275

[10] Council, N.R., Sciences, D. on E. and P., Board, C.S. and T., Systems, C. on C.D.S., 2007. Software for Dependable Systems: Sufficient Evidence? National Academies Press.

[11] Dahll, G., 2000. Combining disparate sources of information in the safety assessment of software-based systems. Nucl. Eng. Des. 195, 307–319. https://doi.org/10.1016/S0029-5493(99)00213-7

[12] Dardar, R., 2014. Building a Safety Case in Compliance with ISO 26262 for Fuel LevelEstimation and Display System.

[13] Dardenne, A., van Lamsweerde, A., Fickas, S., 1993. Goal-directed requirements acquisition. Sci. Comput. Program. 20, 3–50. https://doi.org/10.1016/0167-6423(93)90021-G

[14] Denney, E., Pai, G., Whiteside, I., 2019. The Role of Safety Architectures in Aviation Safety Cases. Reliab. Eng. Syst. Saf. https://doi.org/10.1016/j.ress.2019.106502

[15] Fenton, N., Littlewood, B., Neil, M., Strigini, L., Sutcliffe, A., Wright, D., 1998. Assessing dependability of safety critical systems using diverse evidence. IEE Proc. - Softw. 145, 35–39. https://doi.org/10.1049/ip-sen:19984895

[16] Gallina, B., 2014. A Model-Driven Safety Certification Method for Process Compliance, in: 2014 IEEE International Symposium on Software Reliability Engineering Workshops. Presented at the 2014 IEEE International Symposium on Software Reliability Engineering Workshops, pp. 204–209. https://doi.org/10.1109/ISSREW.2014.30

[17] Gallina, B., Gómez-Martínez, E., Benac-Earle, C., 2017. Promoting MBA in the rail sector by deriving process-related evidence via MDSafeCer. Comput. Stand. Interfaces 54, 119–128. https://doi.org/10.1016/j.csi.2016.11.007

[18] Guarro, S., Yau, M.K., Ozguner, U., Aldemir, T., Kurt, A., Hejase, M., Knudson, M., 2017. Risk Informed Safety Case Framework for Unmanned Aircraft System Flight Software Certification, in: AIAA Information Systems-AIAA Infotech @ Aerospace, AIAA SciTech Forum. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2017-0910

[19] Habli, I., Ibarra, I., Land, J., Rivett, R., Kelly, T., 2010. Model-Based Assurance for Justifying Automotive Functional Safety. https://doi.org/10.4271/2010-01-0209

[20] Hamilton, V., 2011. Accounting for Evidence: Managing Evidence for Goal Based Software Safety Standards, in: Dale, C., Anderson, T. (Eds.), Advances in Systems Safety. Springer, London, pp. 41–51. https://doi.org/10.1007/978-0-85729-133-2_3

[21] Hirata, C., Nadjm-Tehrani, S. (2019, September). Combining GSN and STPA for Safety Arguments. In International Conference on Computer Safety, Reliability, and Security (pp. 5-15). Springer, Cham.

[22] Kelly, T., Weaver, R., 2004. The goal structuring notation–a safety argument notation. Proc Dependable Syst Netw. Workshop Assur. Cases.

[23] Kelly, T.P., McDermid, J.A., 1997. Safety case construction and reuse using patterns, in: Safe Comp 97. Springer, pp. 55–69.

[24] Kurd, Z., Kelly, T., McDermid, J., Calinescu, R., Kwiatkowska, M., 2009. Establishing a Framework for Dynamic Risk Management in 'Intelligent' Aero-Engine Control, in: Buth, B., Rabe, G., Seyfarth, T. (Eds.), Computer Safety, Reliability, and Security, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 326–341. https://doi.org/10.1007/978-3-642-04468-7_26

[25] Luo, Y., Saberi, A.K., den Brand, M. van, 2019. Safety-Driven Development and ISO 26262, in: Dajsuren, Y., van den Brand, M. (Eds.), Automotive Systems and Software Engineering: State of the Art and Future Trends. Springer International Publishing, Cham, pp. 225–254. https://doi.org/10.1007/978-3-030-12157-0_10

[26] Luo, Y., van den Brand, M., Engelen, L., Klabbers, M., 2015. A Modeling Approach to Support Safety Assurance in the Automotive Domain, in: Selvaraj, H., Zydek, D., Chmaj, G. (Eds.), Progress in Systems Engineering, Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, pp. 339–345. https://doi.org/10.1007/978-3-319-08422-0_50

[27] Martin, H., Krammer, M., Bramberger, R., Armengaud, E., 2016. Process-and product-based lines of argument for automotive safety cases, in: ACM/IEEE 7th International Conference on Cyber-Physical Systems, ICCPS.

[28] McDermid, J.A., Jia, Y., Habli, I., 2019. Towards a Framework for Safety Assurance of Autonomous Systems [WWW Document]. Artif. Intell. Saf. 2019. URL http://eprints.whiterose.ac.uk/150187/ (accessed 4.30.21).

[29] Nair, S., de la Vara, J.L., Sabetzadeh, M., Briand, L., 2014. An extended systematic literature review on provision of evidence for safety certification. Inf. Softw. Technol. 56, 689–717. https://doi.org/10.1016/j.infsof.2014.03.001

[30] Palin, R., Ward, D., Habli, I., Rivett, R., 2011. ISO 26262 safety cases: Compliance and assurance, in: 6th IET International Conference on System Safety 2011. Presented at the 6th IET International Conference on System Safety 2011, pp. 1–6. https://doi.org/10.1049/cp.2011.0251

[31] Pissoort, D., Bultinck, T., Boydens, J., Catrysse, J., 2019. Use of the Goal Structuring Notation (GSN) as Generic Notation for an "EMC Assurance Case," in: 2019 International Symposium on Electromagnetic Compatibility - EMC EUROPE. Presented at the 2019 International Symposium on Electromagnetic Compatibility - EMC EUROPE, IEEE, Barcelona, Spain, pp. 465–469. https://doi.org/10.1109/EMCEurope.2019.8872009

[32] Rudolph, A., Voget, S., Mottok, J., 2018. A consistent safety case argumentation for artificial intelligence in safety related automotive systems, in: ERTS 2018, 9th European Congress on Embedded Real Time Software and Systems (ERTS 2018). Toulouse, France.

[33] Sabetzadeh, M., Falessi, D., Briand, L., Alesio, S.D., McGeorge, D., Åhjem, V., Borg, J., 2011. Combining Goal Models, Expert Elicitation, and Probabilistic Simulation for

Qualification of New Technology, in: 2011 IEEE 13th International Symposium on High-Assurance Systems Engineering. Presented at the 2011 IEEE 13th International Symposium on High-Assurance Systems Engineering, pp. 63–72. https://doi.org/10.1109/HASE.2011.22

[34] Saeed, A., de Lemos, R., Anderson, T., 1995. On the safety analysis of requirements specifications for safety-critical software. ISA Trans. 34, 283–295. https://doi.org/10.1016/0019-0578(95)00019-V

[35] Schmid, T., Schraufstetter, S., Wagner, S., Hellhake, D., 2019. A Safety Argumentation for Fail-Operational Automotive Systems in Compliance with ISO 26262, in: 2019 4th International Conference on System Reliability and Safety (ICSRS). Presented at the 2019 4th International Conference on System Reliability and Safety (ICSRS), pp. 484–493. https://doi.org/10.1109/ICSRS48664.2019.8987656

[36] Stålhane, T., Myklebust, T., 2016. The Agile Safety Case, in: Skavhaug, A., Guiochet, J., Schoitsch, E., Bitsch, F. (Eds.), Computer Safety, Reliability, and Security, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 5–16. https://doi.org/10.1007/978-3-319-45480-1_1

[37] Taguchi, K., Daisuke, S., Nishihara, H., Takai, T., 2014. Linking Traceability with GSN, in: 2014 IEEE International Symposium on Software Reliability Engineering Workshops. Presented at the 2014 IEEE International Symposium on Software Reliability Engineering Workshops, pp. 192–197. https://doi.org/10.1109/ISSREW.2014.79

[38] The Assurance Case Working Group (ACWG), 2018. GSN Community Standard (Version 2) 82.

[39] Vreeswijk, G.A.W., 2005. Argumentation in Bayesian Belief Networks, in: Rahwan, I., Moraïtis, P., Reed, C. (Eds.), Argumentation in Multi-Agent Systems, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 111–129. https://doi.org/10.1007/978-3-540-32261-0_8

[40] Wagner, S., Schätz, B., Puchner, S., Kock, P., 2010. A Case Study on Safety Cases in the Automotive Domain: Modules, Patterns, and Models, in: 2010 IEEE 21st International Symposium on Software Reliability Engineering. Presented at the 2010 IEEE 21st International Symposium on Software Reliability Engineering, pp. 269–278. https://doi.org/10.1109/ISSRE.2010.31

[41] Wang, R., Guiochet, J., Motet, G., Schön, W., 2017. Modelling Confidence in Railway Safety Case. Saf. Sci. 110. https://doi.org/10.1016/j.ssci.2017.11.012

[42] Williams, B.P., Clothier, R., Fulton, N., Johnson, S., Lin, X., Cox, K., 2014. Building the Safety Case for UAS Operations in Support of Natural Disaster Response, in: 14th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2014-2286

[43] Guide relatif aux tâches de sécurité autres que la conduite des trains, 2015, n.d. . EPSF. URL https://securite-ferroviaire.fr/actualites/publication-du-guide-relatif-aux-taches-de-securite-autres-que-la-cor

[44] ERA, European Railway Safety Culture Model 7. URL https://www.era.europa.eu/activities/safety-culture_en

[45] PFA Position paper - AD Safety WG - 2019-V1.0.pdf

[46] VMAD-05-12 AD safety validation - french views - Vdef.pdf, 2020

[47] EC : guidance for application of article 20 of vehicle safety Directive, 2019

[48] Intel, Aptive et al : safety first for autonomous driving, 2019

[49] US-DOT : autonomous driving vision for safety and framework for ADS testable cases and scenarios, 2018

[50] Transport Canada : safety assessment for ADS in Canada, 2018

[51] Japan MLIT : autonomous driving safety guidelines, 2018

[52] OICA : future certification of ADS, 2018

[53] Rand Corporation, AV safety measurement, 2018

[54] Mercedes and Bosch, reinventing safety : a joint approach to automated driving systems, 2018

[55] Mobileye : Responsibility-Sensitive Safety (RSS) : a mathematical model for autonomous vehicle safety and implementing in NHTSA pre-crash scenarios, 2018

[56] NL-RDW : software driving license for autonomous cars, 2017

[57] NHTSA-Critical Reasons for Crashes Investigated in National Motor Vehicle Crash 1 Causation Survey DOT-HS-812-115 Feb 2015

[58] AVSC- WARRENDALE, Pa. (July 22, 2021) –AVSC Information Report for Adapting a Safety Management System (SMS) for Automated Driving System (ADS) SAE Level 4 and 5 Testing and Evaluation, https://www.sae.org/news/press-room/2021/07/automated-vehicle-safety-consortium-publishes-safety-management-syste

[59] ASFA – Analyse des accidents corporels sur autoroutes concédées – Année 2015 2016

[60] ISO - Road vehicles - Functional safety ISO/26262, 2018

[61] ISO - Road vehicles - Safety Of the Intended Functionality ISO/PAS21448 – Jan 2019

[62] Systèmes de transport public guidés urbains de personnes - Principe “GAME” document STRMTG – V2 2011

[63] High-Level Expert Group on Artificial Intelligence, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, European Commission, 2020, https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assess

A GAME, ALARP AND MEM PRINCIPLES

Different principles and methods exist to determine the acceptable level of risk for a system in case of explicit risk estimation. Legal frameworks can set a specific principle to follow for designated systems or sectors. In such cases, to be compliant with relevant legal framework, the demonstration that the level of risk achieved is acceptable needs to follow the set principle and method.

Regulatory frameworks in the EU for the railway sector use three methods: ALARP (UK), GAME (FR) and MEM (DE). These methods are briefly described below.

Authorities may use these principles and methods for the legal frameworks that are currently being set in the field of automated driving:

- In France, the legal requirements for automated road transport systems (ARTS) provides that such systems have to follow the GAME principle: “Any automated road transport system or any part of an existing transport system is designed, put into service and, if necessary, modified in such a way that the overall level of safety with regard to users, operating personnel and third parties is at least equivalent to the existing level of safety or to that resulting from the implementation of systems or subsystems providing comparable services or functions taking into account the state of the art, the experience feedback concerning them, and reasonably foreseeable traffic conditions on the route or traffic zone considered.” (décret n°2021-873)
- At the EU level, a legal framework for ADS is being prepared. There are ongoing discussions about the use of GAME and/or MEM principles in this framework.

A.1 GAME (Globalement Au Moins Equivalent) principle

The French legal framework on guided public transports (décret n°2017-440) requires the use of the GAME approach.

The principle is that any new guided public transport system/subsystem or any modification of an existing system/subsystem should be globally as safe or safer than existing systems providing similar services.

The words “at least” indicate that the safety performance of the new or modified system must not be worse than the safety performance of similar systems used as reference. The word ”globally” means that the level of safety must be considered at the level of the whole studied system/subsystem. It implies that a structural ”insufficiency” of the system can, subject to justifications, be compensated by a gain at the level of other structural devices or else be made acceptable by means of operational measure (maintenance or particular operating criteria, for example).

Guidance on the application of the GAME principle in the guides public transports can be found here: `http://www.strmtg.developpement-durable.gouv.fr/documents-generau html`

Inspired by the legal framework on guides systems, the French legal framework on ARTS (décret n°2021-873) also makes the GAME principle mandatory. Guidance on how to apply this

principle to innovative systems such as ARTS is being prepared in working groups coordinated by the STRMTG.

Schematically, three applicable safety demonstration approaches can be used in order to apply the GAME principle to ARTS:

- Type 1 approach: compliance with technical and safety regulations or compliance with a technical reference;
- Type 2 approach: comparison with an existing ARTS system (”gap approach”);
- Type 3 approach: detailed risk analysis with respect to each hazardous event using standardized/recognized methods.

The safety demonstration should be based on one of these approaches or a combination of the three, respecting the following priorities:

- First, if an applicable regulatory framework exists, it must be applied. The reference is in this case imposed by the applicable regulatory provisions.
- Secondly, when there is no applicable regulatory framework or applicable technical framework, the comparison approach with an existing benchmark system (“gap approach”) can be considered. The system taken as reference must be perfectly understood as well as its limits and conditions of use so as to allow the identification of any deviations with the new system, both technically and in terms of use and maintenance conditions. For each deviation identified, it must be demonstrated that the measures implemented guarantee that the level of safety of the new system does not fall below to the level of safety of the system taken as reference. The evolution of the rules of the art compared to those of the reference system, must be taken into account.
- Finally, if the criteria of the type 2 approach cannot be met and there is therefore no reference system (in the case of innovations and new designs for example), a detailed risk analysis must be carried out.

For types 2 and 3 approaches, several analysis methodologies are possible (quantitative, qualitative, etc.). The nature of the analysis depends directly on the specificities and scope of the project (at the level of a system, a subsystem, a function, a component, etc.).

In any case, the ARTS French legal framework provides that the safety demonstration methods are subject to an assessment by a designated third party.

A.2 ALARP (As Low As Reasonably Practicable) principle

ALARP means “As Low As Reasonably Practicable”. The principle is that the residual risk shall be reduced as far as reasonably practicable. In UK legal framework, it is equivalent to SFAIRP (”so far as is reasonably practicable”).

For a risk to be ALARP, it must be possible to demonstrate that the cost involved in reducing the risk further would be grossly disproportionate to the benefit gained.

In essence, making sure a risk has been reduced ALARP is about weighing the risk against the sacrifice needed to further reduce it. The decision is weighted in favor of health and safety because the presumption is that the duty-holder should implement the risk reduction measure. To avoid having to make this sacrifice, the duty-holder must be able to show that it would be grossly disproportionate to the benefits of risk reduction that would be achieved. Thus, the process is not one of balancing the costs and benefits of measures but, rather, of adopting measures except where they are ruled out because they involve grossly disproportionate sacrifices. (source `https://www.hse.gov.uk/managing/theory/alarpglance.htm`).

In most situations, the application of established good practice, including formal codes of practice, can often be considered to be a suitable demonstration that risk is reduced ALARP. In cases where there is no relevant good practice, a more detailed comparison has to be undertaken, for example via quantified risk assessment and Cost Benefit Analysis (CBA). Guidance on CBA can be found here: `https://www.hse.gov.uk/managing/theory/alarpcba.htm`

It is important to note that the ALARP principle does not take into account absolute risk levels or considerations of the tolerability of risk. It is purely based on a comparison of the costs of a measure with the risk reduction it achieves.

A.3 MEM (Minimum Endogenous Mortality) principle

The Minimum Endogenous Mortality (MEM) rule is a way to set a risk acceptance level based on the natural death rate of human beings. This approach is based on the work of A. Kuhlmann in Germany in 1981.

The death rate of humans between 5 and 15 years of age in industrial countries is the chosen reference (2.10-4 fatalities per person and year).

The MEM requirement states that the additional overall hazard death rate caused by technical systems should not exceed this limit. Each individual being exposed to several systems, the MEM rule states that each system should not contribute to more than 5

This leads to the principle that a technical system should not lead to a risk of fatality of more than 10-5 fatality per person and per year.

B SAFETY VALIDATION PRINCIPLES

[46] describes the overall architecture of what could be the French approach for safety validation, gives some validation principles and goes into details on some aspects (not covered in this document).

Safety validation should combine two main axes:

- a process-centered axis, to be mainly scrutinized by public authorities through audit of conception and validation methods ;
- a performance-centered axis, to be mainly scrutinized by public authorities through tests.

The efficient combination of these two axis strongly relies on the management of driving scenarios for the conception and validation of automated driving systems.

The performance-based approach should be predominant for public validation. In a performance-based approach, manoeuvers or responses, and their sequences, should play a central role in use cases description and validation approaches. Risk scenarios management should be the frame for validation architecture and an important angle through which public authorities can scrutinize industry validation processes.

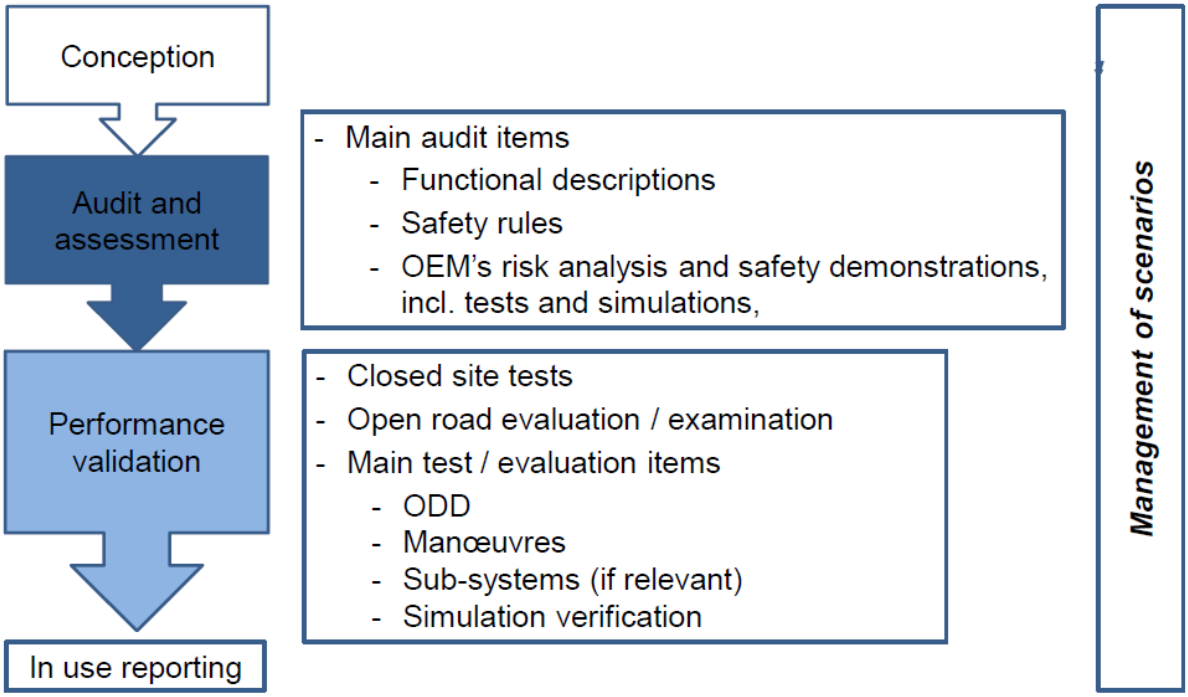

Figure 5: Overall approach architecture

Overarching rules (shared by [45] and [46])

- Overall intention: Automated Vehicle deployment shall improve the road safety (safety target, see also Section 2.6)
- Safety qualitative objective: The automated vehicle is free from unreasonable risk (safety target, see also Section 2.6)

- High-level safety rules: The vehicle shall comply with a set of high-level safety rules contributing to safety, whether or not their safety impact can be quantitatively assessed. A minimum set of high-level rules should be shared by all OEMs
- Scenario-based approach: design and verification/validation phases shall take into account relevant driving scenarios, including reasonably foreseeable misuses. A minimum set of these scenarios should be shared by all OEMs
- Field experience shall be taken into account to continuously improve vehicle safety. Lessons learned from the field should be shared as far as possible

Validation principles [46]

- Principle 1: Validation should handle a wide variety of use-cases (functions, ODDs, maneuvers)
- Principle 2: Validation should verify that reasonably foreseeable risks, combining system failures and driving hazards, are identified and addressed, and their impacts are minimized
- Principle 3: Public validation should combine physical or simulated tests and audits in order to assess results (or performance) and processes, based on a sufficient knowledge of the system's design
  - Principle 3.1: Validation by public authorities should focus on driving responses (manoeuvers) to systems failures and driving hazards and assess both:
    - critical maneuvers' safety, responding to edge scenarios
    - current maneuvers carefulness or roadmanship
  - Principle 3.2: Physical tests should combine:
    - a standardized approach, for a limited set of common functions or maneuvers
    - a use-case-specific approach, based on risk analysis, including randomly
  - Principle 3.3: Audits should be based on manageable and interpretable descriptions of:
    - ODDs
    - system architectures and description of safety-relevant functions
    - manoeuvers and their logical sequence
    - systems' and manoeuvers' overarching safety rules
    - scenario management, risk screening and scoring methods and relevant results, covering system failures and driving hazards scenarios
    - rationale for internal combination of different safety demonstration tools (e.g. tests, simulations with various Xs-in-the-loop)
    - safety demonstration methods, tools and facilities, with a focus on simulation tools' calibration.
    - behavior and perception studies

- Principle 4: Validation may look at specific functional sub-systems, whenever they are safety-critical and suitable for replicable validation tools. This is presumably the case for HD mapping, human-machine interfaces, V2X connectivity and objects/event detection and recognition.
- Principle 5: Transparency of managing risk scenarios for safety analysis is key to build a proper balance between internal validation processes and public validation scrutiny
- Principle 6: In-use data reporting should enrich common state of the art and public validation processes. Inter alia, public authorities should develop knowledge on critical driving situations, based on feedbacks by the industry.

Expectations towards audit [46]

Audits should mainly address the following concerns:

- verify that processes for conception / development / risk analysis / safety demonstrations, are documented and, provided so, comply with internal design rules ;
- assess the relevance of internal design rules as regard to state of the art ;
- assess the ability of internal risk analysis and safety demonstration processes, to cover the largest scope of reasonably foreseeable risks.

Expectations towards tests [46]

Tests and simulations should assess the ability of autonomous driving systems or sub-systems to perform safely in representative driving conditions. Based on scenarios analysis, validation approaches should combine different tools:

- simulation ;
- closed sites tests ;
- open road tests.

Which could be:

- use-case agnostic or use-case specific or endogenous ;
- pre-defined or randomized.

Expectations towards scenario management [46]

The main challenge of AD validation is to manage driving hazards (previously handled by the driver) in risk analysis. Relevant scenarios to design and validate ADS should be managed (because of their huge number) to comply SOTIF and identify the residual risk. Scenario management should be the frame for validation architecture and the main window through which public authorities can scrutinize industry validation processes.

Scenario management has a key role to play in reducing the gap between the potentially infinite combination of driving conditions x events and responses (both from the ego vehicle

and alter road users), and the capacity to address a limited number of scenario by the means of validation tools (either simulations or tests). Managing scenarios supports identification of both most likely and most critical scenarios. It helps reduce the probability of unidentified critical situations (“black-swans”). This process’ description and some of its outputs must be transparent to public output. Auditing this process should be part of public authority validation.

LIST OF ACRONYMS

AD Autonomous Driving.

ADS Autonomous Driving Systems.

ALARP As Low As Reasonably Practicable.

ARTS Automated Road Transport Systems.

BBN Bayesian Belief Network.

CAE Claims, Argument, and Evidence.

CBA Cost Benefit Analysis.

EPSF Etablissement Public de Sécurité Ferroviaire.

FMVSS US Federal Motor Vehicle Safety Standards.

GAME Globalement Au Moins Equivalent.

GSN Goal Structuring Notation.

HITL Human In The Loop.

HMI Human-Machine Interactions.

ISO International Standard Organization.

KAOS Knowledge Acquisition in Automated Specification.

MEM Minimum Endogenous Mortality.

MRM Minimal Risk Maneuvers.

NSA National Safety Agency.

ODD Operational Design Domain.

OEM Original Equipment Manufacturer.

SFAIRP So Far As Is Reasonably Practicable.

SMS Safety Management System.

SOTIF Safety Of The Intended Functionality.

SSG Safety Specification Graph.

THR Tolerable Hazard Rates.

US-DOT US Department of Transportation.