



HAL
open science

Contributions to safety assurance of autonomous trains

Mohammed Chelouati

► **To cite this version:**

Mohammed Chelouati. Contributions to safety assurance of autonomous trains. Automatic Control Engineering. Université Gustave Eiffel, 2024. English. NNT : 2024UEFL2014 . tel-04692568

HAL Id: tel-04692568

<https://theses.hal.science/tel-04692568v1>

Submitted on 10 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



École doctorale MADIS 631

Thèse de doctorat d'INFORMATIQUE ET AUTOMATIQUE

Laboratoire ESTAS

Thèse présentée et soutenue le 05/06/2024, par **Mohammed CHELOUATI**

Contributions to safety assurance of autonomous trains

Contributions à l'assurance de sécurité des trains autonomes

Membres du Jury

Paola PELLEGRINI, DR,	Université Gustave Eiffel	Présidente
Jérémie GUIOCHET, Prof	Université de Toulouse III	Rapporteur
Philippe WEBER, Prof	Université de Lorraine	Rapporteur
Sana JABRI, Dr,	Thales	Examinatrice
Insaf SASSI, Dr,	IRT Railenium	Examinatrice
Abderraouf BOUSSIF, Dr	Université Gustave Eiffel	Encadrant
Julie BEUGIN, Dr	Université Gustave Eiffel	Encadrante
El-Miloudi EL KOURSI, DR	Université Gustave Eiffel	Directeur de thèse

Résumé

Le déploiement des trains autonomes soulève de nombreuses questions et défis, notamment ceux liés au niveau de sécurité visé, qui doit être globalement au moins équivalent à celui du système existant, ainsi que les moyens à mettre en œuvre pour l'atteindre. Conventionnellement, la mise en sécurité d'un système ferroviaire global ou d'un sous-système défini comprend une phase d'analyse des risques et une phase de maîtrise des situations dangereuses. Ainsi, pour tout système technique ferroviaire, qu'il soit classique, automatique ou autonome, un niveau de sécurité acceptable doit être assuré. Dans le contexte des trains autonomes, les défis liés à leur sécurité incluent les aspects émergents de l'intelligence artificielle, le transfert de tâches et de responsabilités du conducteur vers des systèmes décisionnels automatiques, ainsi que les aspects liés à l'autonomisation, tels que la transition entre les modes et la gestion des modes dégradés. La méthodologie de démonstration de sécurité des trains autonomes, doit ainsi prendre en compte les risques engendrés par l'ensemble de ces aspects. Autrement dit, elle doit définir l'ensemble des activités de sécurité (liées à l'introduction de l'autonomie et des Systèmes d'Intelligence Artificielle), complémentaires à la démonstration de sécurité conventionnelle.

Dans ce cadre, l'objectif de cette thèse est de contribuer à l'élaboration d'une démarche d'assurance de sécurité pour les trains autonomes. Concrètement, cette thèse propose trois contributions principales. Premièrement, nous proposons une méthodologie globale de haut niveau pour la structuration et la présentation de l'argumentation de sécurité pour les trains autonomes. La méthodologie est basée sur une approche orientée objectifs de sécurité (goal-based safety) en utilisant le formalisme graphique GSN (Goal Structuring Notation). Ensuite, nous proposons une modélisation de la conscience de situation (situational awareness) d'un système de conduite autonome d'un train, intégrant le processus de l'analyse dynamique des risques ferroviaires. Ce modèle permettra au système de conduite autonome de percevoir, de comprendre, d'anticiper et de s'adapter à des situations inconnues dans son environnement tout en prenant des décisions sûres. Le modèle est illustré à travers un cas d'étude concernant la détection et l'évitement d'obstacles sur la voie ferroviaire. Dernièrement, nous élaborons une approche de prise de décision basée sur l'évaluation dynamique des risques. L'approche utilise le Processus Décisionnel de Markov Partiellement Observable (POMDP) et vise à assurer une surveillance continue de l'environnement pour garantir la sécurité opérationnelle, en particulier la prévention des collisions. L'approche repose sur le maintien d'un niveau de risque acceptable grâce à une estimation et une actualisation continues de l'état opérationnel du train et des données de perception de l'environnement.

Mots clés: Assurance de sécurité des trains autonomes, Analyse dynamique des risques, Prise de décision basée sur le risque, Argumentaire de sécurité, Conscience de la situation des trains autonomes.

Abstract

The deployment of autonomous trains raises many questions and challenges, particularly concerning the required safety level, which must be globally at least equivalent to that of the existing systems, along with how to achieve it. Conventionally, ensuring the safety of a global railway system or a defined subsystem includes analyzing risks and effectively handling dangerous situations. Therefore, for any technical railway system, whether it is conventional, automatic, or autonomous, an acceptable level of safety must be ensured. In the context of autonomous trains, safety challenges include aspects related to the use of artificial intelligence models, the transfer of tasks and responsibilities from the driver to automatic decision-making systems, and issues related to autonomy, such as mode transitions and management of degraded modes. Thus, the safety demonstration methodology for autonomous trains must take into account the risks generated by all these aspects. In other words, it must define all the safety activities (related to the introduction of autonomy and artificial intelligence systems), complementary to conventional safety demonstration.

In this context, this dissertation proposes three main contributions towards the development of a safety assurance methodology for autonomous trains. Firstly, we establish a high-level framework for structuring and presenting safety arguments for autonomous trains. This framework is based on a goal-based approach represented by the graphical modeling Goal Structuring Notation (GSN). Then, we propose a model for the situational awareness of the automated driving system of an autonomous train, that integrating the process of dynamic risk assessment. This model enables the automated driving system to perceive, understand, anticipate and adapt its behavior to unknown situations while making safe decisions. This model is illustrated through a case study related to the obstacle detection and avoidance. Finally, we develop a decision-making approach based on dynamic risk assessment. The approach is based on Partially Observable Markov Decision Processes (POMDP) and aims to ensure continuous environmental monitoring to guarantee operational safety, particularly collision prevention. The approach is based on maintaining an acceptable level of risk through continuous estimation and updating of the train's operational state and environmental perception data.

Keywords: Safety of autonomous trains, Dynamic risk assessment, Risk-based decision-making, Safety argumentation, Situational awareness of autonomous trains.

Acknowledgements

The successful completion of this thesis would not have been possible without the support, guidance, and encouragement of many individuals to whom I am deeply grateful.

I would like to express my deepest gratitude to the chair of my thesis committee, Dr Paola PELLEGRINI, for her invaluable support and leadership throughout this process. Her guidance and encouragement have been essential to the completion of this work.

I extend my heartfelt thanks to the reviewers and thesis referees: Dr Philippe WEBER, Dr Jérémie GUIOCHET, Dr Sana JABRI and Dr Insaf SASSI. Their insightful feedback and constructive comments have greatly improved the quality of this thesis.

I am profoundly grateful to my thesis director, Dr El-Miloudi EL KOURSI, for his unwavering support, guidance, and encouragement throughout this journey. His expertise and advice have been instrumental in shaping this research.

I would also like to express my sincere thanks to my supervisors. To Dr Julie BEUGIN, thank you for your availability, valuable advices, and the human touch you brought to this journey. Your support has been immensely appreciated. To Dr Abderraouf BOUSSIF, I am especially indebted for your unwavering availability from day one to the last, your rigor and guidance, and the high standards you set. Your advice, both as a supervisor and a friend, has been crucial to the success of this work.

My heartfelt appreciation goes to my mother and my father, whose belief in me has been a constant source of motivation. Their unconditional support has been the foundation of my achievements.

Finally, I am deeply thankful to my wife for her ongoing support and encouragement. Her patience, understanding, and love have been my strength throughout this journey.

I would also like to extend my gratitude to my colleagues at IRT Railenium and University Gustave Eiffel for their camaraderie, assistance, and valuable discussions. Their support has made this journey much more enriching and enjoyable.

Thank you all for making this work possible!
Mohammed CHELOUATI
Lille - June 2024

Contents

1	Introduction	12
1.1	General context	12
1.2	Industrial context	16
1.3	Scientific context	16
1.4	Problem formalization	18
1.5	Contributions	18
1.6	Manuscript organization	19
2	Safety assurance of autonomous systems	21
2.1	Introduction	21
2.2	The autonomous driving system	22
2.3	Risk assessment	24
2.3.1	Risk assessment for conventional vehicles	26
2.3.2	Dynamic risk assessment for autonomous transportation systems	26
2.4	Risk models for autonomous vehicles	28
2.5	Decision-making for autonomous vehicles	32
2.5.1	Deterministic approaches	32
2.5.2	Non-deterministic approaches	33
2.6	Risk assessment in railways	40
2.6.1	Risk assessment for conventional trains	40
2.6.2	Risk assessment methods	40
2.6.3	Dynamic risk assessment for the autonomous train	42
2.7	Safety assurance of the autonomous train	43
2.7.1	Safety cases	44
2.7.2	Graphical safety argumentation	45
2.7.3	Goal Structuring Notation (GSN) application examples	47
2.8	Conclusion	48
3	Graphical argumentation using GSN for autonomous trains	50
3.1	Introduction	51
3.2	GSN for graphical safety argumentation	51
3.2.1	Key concepts	51
3.2.2	Development processes	52
3.2.3	Issues resolved with GSN	54
3.3	GSN-based safety cases in transportation systems	55
3.3.1	GSN in the automotive domain	55
3.3.2	GSN in the aviation domain	56
3.3.3	GSN in the railway domain	56
3.3.4	Toward using GSN for autonomous systems	58
3.4	Safety assurance approach of autonomous train	58
3.4.1	Overall system level	59

3.4.2	AI-based component level	60
3.4.3	AI techniques level	60
3.5	Use case	63
3.5.1	Anti-collision function	63
3.5.2	Discussion	65
3.6	Conclusion	68
4	SA & DRA framework for autonomous trains	69
4.1	Introduction	69
4.2	Context and concepts	70
4.2.1	Autonomous Driving System (ADS)	70
4.2.2	Situation Awareness (SA)	73
4.2.3	Complementarity between SA and DRA concepts	75
4.3	A DRA and SA framework for autonomous trains	76
4.3.1	Perception module	76
4.3.2	Understanding & prediction module	78
4.3.3	Decision-making module	79
4.4	Illustrative case: anti-collision function	79
4.4.1	Perception module	81
4.4.2	Understanding & prediction module	82
4.4.3	Decision-making module	84
4.5	Conclusion	84
5	POMDP-based decision-making process of autonomous trains	85
5.1	Introduction	85
5.2	Toward the use of POMDPs in ADS	86
5.2.1	Handling uncertainties in decision-making processes	86
5.2.2	Benefits of POMDP in decision-making processes	87
5.3	Decision-making related to the train’s anti-collision function	88
5.3.1	DRA of the anti-collision function	89
5.3.2	Structuring risk profiles with the DRA framework	91
5.4	Methodology	92
5.4.1	POMDP definition	92
5.4.2	POMDP modeling of the train anti-collision system	93
5.5	Simulation and results	100
5.5.1	Perceived state	100
5.5.2	Obstacle generation function	100
5.5.3	Belief updater	101
5.5.4	Solver choice	101
5.5.5	Variables initialization	102
5.5.6	Risk formulation	102
5.5.7	Results	103
5.6	Conclusion	108
6	Conclusion and perspectives	109
6.1	Conclusions	109
6.2	Perspectives	110

List of Figures

1.1	Railway grades of automation and basic functions	14
1.2	Levels of driving automation by SAE International’s new standard J3016 (SAE, 2016)	15
2.1	A simplified illustration of the Autonomous Driving System (ADS) in automotive	23
2.2	On-board Autonomous Driving System (ADS) components and interactions	25
2.3	Illustration of the Markov Decision Process (MDP)	34
2.4	Illustration of Partially Observable Markov Decision Process (POMDP) . .	35
2.5	Process of risk assessment related to phases 3 and 4 of the life-cycle (EN-50126, 2017)	41
3.1	Six-step process for top-down developing goal structure	52
3.2	Bottom-up process for developing goal structure	54
3.3	The three hierarchical system levels	59
3.4	A system-level argument pattern for the autonomous train safety assurance	61
3.5	Main steps for building GSN safety argument patterns for autonomous trains	62
3.6	Decomposition of high-level goal: Ensure safety of the on-board ADS	64
3.7	Safety argumentation structure for addressing static obstacles	66
3.8	Safety argumentation structure for addressing dynamic obstacles	67
4.1	A high-level architecture of the on-board ADS of the autonomous train with a main focus on the decision-making process	71
4.2	Three-level model of Situation Awareness Endsley (1995)	74
4.3	The autonomous train situational awareness framework	77
4.4	Decision-making flowchart of the anti-collision function for autonomous trains	80
4.5	Framework for obstacle detection and avoidance (anti-collision) of autonomous trains	83
5.1	A simplified architecture of the ADS with a main focus on the DRA layer, strengthening the decision-making task	87
5.2	Generic illustration of the anti-collision function	89
5.3	Illustrative representation of the anti-collision function	90
5.4	The autonomous train dynamic risk assessment framework	91
5.5	A generic illustration of the POMDP model.	93
5.6	A generic spatial discretization of Cartesian plan into adaptive grid map for autonomous train navigation	94
5.7	The evolution of the actual state, perceived state, and the chosen action (setup 1)	103
5.8	The evolution of the actual state, perceived state, and the chosen action (setup 2)	104
5.9	The evolution of rewards over time (setup 1)	105

5.10	The evolution of rewards over time (setup 2)	105
5.11	The risk estimation over time (setup 1)	106
5.12	The risk estimation over time (setup 2)	107
5.13	The evolution of the observed distance to obstacle over time (setup 1)	107
5.14	The evolution of the observed distance to obstacle over time (setup 2)	108

List of Tables

2.1	An overview of risk models used for autonomous transportation systems . .	30
2.2	Summary of non-deterministic methods for Autonomous decision-making .	39
2.3	Graphical safety argumentation methods	46
3.1	The main GSN elements. (GSN-WG, 2021)	53
3.2	The use of the GSN method in automotive, aviation, and railway domains .	57
4.1	Comparison of Various ADS Architectures	72
5.1	POMDP solvers choice	100
5.2	Variables initialization	102

List of Acronyms

ADS Autonomous Driving System. 7, 14, 19, 22–25, 58, 59, 62–64, 70, 71, 73–76, 79, 85, 88–90, 99

AI Artificial Intelligence. 6, 17, 22, 23, 27, 33, 50, 59, 60, 71, 75, 85

AVs Autonomous Vehicles. 21–24, 26–29, 32–35, 37, 38, 75

CAE Claims Argument Evidence. 46

CSM-RA Common Safety Method-Risk Assessment. 59

DRA Dynamic Risk Assessment. 20, 26, 27, 42, 69, 70, 75, 76, 78, 79, 84, 90, 91, 93

GoA Grade of Automation. 14, 15, 22, 58

GSN Goal Structuring Notation. 7, 9, 19, 20, 45–48, 51–56, 58–60, 62, 63, 65, 68, 109

KAOS Knowledge Acquisition in Automated Specification. 45, 46

MDP Markov Decision Process. 33, 34, 36, 38

ML Machine Learning. 17, 23, 26, 27, 33, 40, 60, 74, 75

ODD Operational Domain Design. 22, 32, 59, 60, 79

POMDP Partially Observable Markov Decision Process. 20, 34–36, 38, 78, 86–88, 92–96, 99–101, 103, 108, 109

RL Reinforcement Learning. 36, 37, 74

SA Situational Awareness. 69, 70, 73–76, 78, 84

SACM Structured Assurance Case Metamodel. 45, 46

SNCF Société Nationale des Chemins de Fer Français. 16

SOTIF Safety Of The Intended Functionality. 60

SSG Safety Specification Graph. 45, 46

TASV Train Autonome Service Voyageurs. 16

List of author's publications

Journal articles

1. Mohammed Chelouati, Abderraouf Boussif, Julie Beugin, El-Miloudi El Koursi. **Graphical safety assurance case using Goal Structuring Notation (GSN) — challenges, opportunities and a framework for autonomous trains.** *Reliability Engineering & System Safety*, vol. 230, 108933, 2023. <https://doi.org/10.1016/j.ress.2022.108933>
2. Mohammed Chelouati, Abderraouf Boussif, Julie Beugin, El-Miloudi El Koursi. **A Risk-Based Decision-Making Process for Autonomous Trains Using POMDP: Case of the Anti-Collision Function.** *IEEE Access*, vol. 12, pp. 1-18, 2024. <https://doi.org/10.1109/ACCESS.2023.3347500>

Conference proceedings

1. Mohammed Chelouati, Abderraouf Boussif, Julie Beugin, El-Miloudi El Koursi. **A framework for risk-awareness and dynamic risk assessment for autonomous trains.** In *32nd European Safety And Reliability (ESREL) Conference*, 2022.
2. Mohammed Chelouati, Abderraouf Boussif, Julie Beugin, et al. **Argumentaire de sécurité graphique pour l'assurance de sécurité des trains autonomes.** In *Congrès Lambda Mu 23 "Innovations et maîtrise des risques pour un avenir durable" - 23e Congrès de Maîtrise des Risques et de Sûreté de Fonctionnement*, 2022.
3. Abhimanyu Tonk, Mohammed Chelouati, Abderraouf Boussif, Julie Beugin, Miloudi El Koursi. **A safety assurance methodology for autonomous trains.** *Transportation Research Procedia*, vol. 72, pp. 3016-3023, 2023. <https://doi.org/10.1016/j.trpro.2023.11.849>
4. Mohammed Chelouati, Abderraouf Boussif, Julie Beugin, El-Miloudi El Koursi. **Une approche orientée risques pour la prise de décision dans les trains autonomes : Cas de la fonction anti-collision.** In *Congrès Lambda Mu 24 "Innovations et maîtrise des risques pour un avenir durable" - Congrès de Maîtrise des Risques et de Sûreté de Fonctionnement*, 2024. (abstract accepted)

Chapter 1

Introduction

Contents

1.1	General context	12
1.2	Industrial context	16
1.3	Scientific context	16
1.4	Problem formalization	18
1.5	Contributions	18
1.6	Manuscript organization	19

1.1 General context

Throughout history, the railway system has significantly contributed to the transformation of modern transportation. With the establishment of the Stockton and Darlington Railway in 1825 ([Smiles, 1904](#)), often regarded as the world's first passenger railway, railroads played an important role in revolutionizing the efficient movement of passengers and goods over wide geographical areas. Their reputation for resilience, reliability, and environmentally responsible operations has made them an integral part of the global transportation network. However, along with the outstanding efficiency of railway systems, they also present a unique set of challenges, the first of which is safety. Over the years, railways have had accidents, some attributed to human errors, while others are due to environmental factors or aging infrastructure ([Liu et al., 2021a](#)). Ensuring the safety of passengers, goods and the environment is essential in this domain.

In recent years, there has been a growing interest in the potential for technological-driven solutions aimed to improve various aspects of railway operations. For instance, automated/autonomous railway systems are mainly designed to operate with minimal human intervention, with a primary focus on reducing operational costs, increasing availability, optimizing railway traffic, and minimizing energy consumption. While enhancing safety can be an inherent advantage of autonomous systems due to the potential risk mitigations associated with human errors and other safety-related challenges, the primary goals remain cost reduction and improved operational efficiency.

The desire to reduce costs, energy consumption, and obviously accidents are the driving forces behind this pursuit of automation and autonomy. In fact, some researches have shown that accidents are often related to human factors ([National Transportation Safety Board, 2002](#)), including misinterpretation of risks ([Khurana and Das, 2009](#)), slow or incorrect responses, and lapses in attentions ([Khurshid and Faisal, 2012](#)). Merging from that, the railway industry is acutely aware of the need to address these challenges.

In the European context, two programs stand out in advancing autonomous railway systems. “Horizon 2020” (2014-2020)¹, the EU’s primary funding program for research and innovation, emphasizes the development of smart and sustainable transport technologies. Shift2Rail² is an initiative under Horizon 2020 that specifically focuses on railway improvements through automation (Ristić-Durrant et al., 2018) and digitalization (Steele and Roberts, 2022). Europe’s rail initiative under the “Horizon Europe” program (2021-2030) now succeeds the Shift2Rail initiative, aiming at advancing the efficiency, sustainability, and safety of European railways. These initiatives lay the foundation for various European projects that aim to integrate autonomous technologies into the railway sector. As part of these projects, we can list :

- **X2RAIL-2/4**³, dedicated to advancing railway signalling and automation by researching and developing key technologies critical for the next generation of rail management systems. The project’s aspirations include enhancing communication systems to support advanced automation, increasing track capacity with the integration of Automatic Train Operation (ATO) and Moving Block systems, and updating signalling infrastructure towards a decentralized model. All these components are planned to be integrated within the European Train Control System (ETCS)/ERTMS (European Rail Traffic Management). Furthermore, X2Rail seeks to improve energy efficiency and punctuality using ATO systems, foster innovation in laboratory testing environments, and reinforce the security of railway signalling and communication networks against cyber threats (Stickel et al., 2022);
- **SMART-1/2**⁴, primarily focused on increasing the efficiency, capacity, and safety of rail freight services on European railways through automation. This project has set two key objectives: the first is the development of a prototype autonomous obstacle detection system for improved safety in rail transport. The second objective is to create a real-time management system for marshalling yards, aimed at optimizing the assembly and disassembly of freight trains to enhance the overall effectiveness of rail freight operations (Ristić-Durrant et al., 2020);
- **TAURO**⁵, focused on identifying and analyzing foundational technologies for the future European automated and autonomous rail transport, to be further developed, certified and deployed through the activities planned for Europe’s Rail. The project encompasses several key areas: environment perception for automation, remote driving and command systems, automatic status monitoring and diagnostics for autonomous trains, and the development of technologies that support the migration to Automatic Train Operation (ATO) over the European Train Control System (ETCS) (Chouchani and Zinkunegi, 2022);
- **R2DATO**⁶, aims to take advantage of digitalization and automation to develop the next generation of Automatic Train Control (ATC). The project is dedicated to delivering scalable solutions for Digital and Automatic Train Operation (DATO) capabilities, extending up to fully autonomous operations. The objective is to enhance the capacity and efficiency of existing rail networks, paving the way for smarter, more reliable rail services that cater to increasing transport demands.

¹https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en

²<https://rail-research.europa.eu/about-shift2rail/>

³https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-1

⁴https://projects.shift2rail.org/s2r_ip5_n.aspx?p=SMART

⁵https://projects.shift2rail.org/s2r_ipx_n.aspx?p=tauro

⁶<https://projects.rail-research.europa.eu/eurail-fp2/>

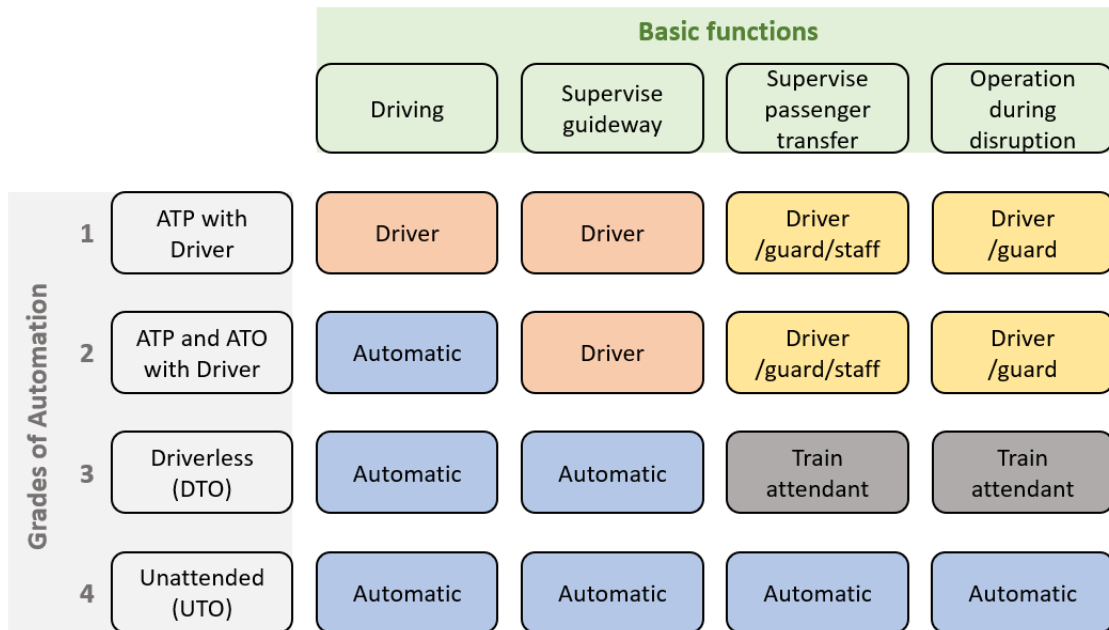


Figure 1.1: Railway grades of automation and basic functions

At the national level (in France), Tech4Rail⁷ represents a pivotal initiative in the country's strategic engagement with the railway sector. The primary objective of Tech4Rail is to develop cutting-edge technological solutions to increase the capacity and reliability of the French railway network (Trentesaux et al., 2018). This includes the integration of advanced digital systems for improved traffic management and predictive maintenance of infrastructure and rolling stock. Furthermore, Tech4Rail is committed to promoting sustainable mobility by reducing the carbon footprint of rail transport, notably through the electrification of small lines (very low-density lines) and the use of renewable energy sources.

Additionally, the focus has been oriented towards the Innovative Light Train program⁸, which has been the foundation for the TELLi⁹ and Draisy¹⁰ projects. These initiatives collectively underscore the national ambition to unlocking the potential of advanced rail infrastructure and logistics through the development of intelligent systems that can anticipate maintenance needs, automated traffic flows, and dynamically adjust to the real-time demands of rail traffic management. TELLi, focusing on telecommunication innovations, seeks to augment rail network connectivity, while Draisy aims to revolutionize rail yard operations with a digitally-integrated approach to cargo handling.

Ensuring the safety of these autonomous systems requires a particular attention, specifically ensuring the transfer of operational responsibility from human operators to an Autonomous/Automatic Driving System (ADS) with various Grades of Automation (GoA) (Niestadt et al., 2019a). Standard IEC 62267 (IEC-62267, 2009) has defined four grades of automation for guided transportation systems (see Figure 1.1). Similarly, the Society of Automotive Engineers¹¹ (SAE) defines six levels of driving automation ranging from 0 (fully manual) to 5 (fully autonomous) in the context of motor vehicles and their operation

⁷<https://www.sncf.fr>

⁸<https://www.sncf.com/fr/innovation-developpement/innovation-recherche/mobilite-pour-tous-dans-les-territoires>

⁹<https://www.sncf.com/fr/innovation-developpement/innovation-recherche/train-leger-innovant>

¹⁰<https://www.sncf.com/fr/innovation-developpement/innovation-recherche/draisy>

¹¹<https://www.sae.org>

on roadways, as presented in Figure 1.2.

The GoA provides a standardized framework for classifying the level of automation in a railway system. This classification allows a clear understanding of the roles and responsibilities of automation systems and human operators in various railway operations (Lemonnier et al., 2023). As railway systems advance towards higher levels of automation, establishing a clear GoA becomes essential for assessing risk, determining safety measures, and ensuring a reliable transition from manned to unmanned operations (Brandenburger and Naumann, 2019; Brandenburger et al., 2021). By clearly defining the extent of automation, stakeholders can establish a safety assurance process that is consistent throughout the entire life-cycle of the railway system, from the initial design phase to the full operational phase (Tonk et al., 2022). Consequently, the clear categorization of automation levels informs and directs the systematic verification and validation activities required to maintain operational safety in autonomous railway systems. In this way, the GoA is not merely a technical specification but a fundamental component of the safety architecture within the rail industry, guiding the rigorous evaluation and management of automated rail operations.

The development of railway systems is driven by the needs of a growing economy and the pursuit of improved transport efficiency. At an industrial level, this development is leading to concrete measures to improve railway operations. Companies are now focused on bringing advanced technologies into real-world environments and highlighting the practical benefits of these innovations in everyday rail transportation. The next section will delve into these industrial developments, highlighting specific advancements and their impacts on the operations and management of railway systems.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human driver monitors the driving environment						
0	No Automation	the full-time performance by the <i>human driver</i> of all aspects of the <i>dynamic driving task</i> , even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the <i>driving mode</i> -specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the <i>driving mode</i> -specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	System	Human driver	Human driver	Some driving modes
Automated driving system ("system") monitors the driving environment						
3	Conditional Automation	the <i>driving mode</i> -specific performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> with the expectation that the <i>human driver</i> will respond appropriately to a <i>request to intervene</i>	System	System	Human driver	Some driving modes
4	High Automation	the <i>driving mode</i> -specific performance by an automated driving system of all aspects of the <i>dynamic driving task</i> , even if a <i>human driver</i> does not respond appropriately to a <i>request to Intervene</i>	System	System	System	Some driving modes
5	Full Automation	the full-time performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> under all roadway and environmental conditions that can be managed by a <i>human driver</i>	System	System	System	All driving modes

Figure 1.2: Levels of driving automation by SAE International's new standard J3016 (SAE, 2016)

1.2 Industrial context

This PhD thesis is a part of the research project TASV (Autonomous Train Passenger Service, 2018-2023), more specifically the safety demonstration Working Package, led by SNCF. TASV project is part of a larger constellation of research initiatives at the railway research and technological institute (IRT Railenium). As the core of French railway research and development, Railenium embodies a strategic partnership among academia, industry, and public stakeholders, fostering collaborative projects to improve and digitalize the railway systems both nationally and internationally. It is within this direction that the TASV project has been established, representing a cutting-edge endeavor within Railenium's ambitious Autonomous Train program¹². The TASV project, alongside the complementary TFA (Autonomous Freight Train) and TCRail (for remote train control) projects aim to define and validate new paradigms for automated train operations by addressing a spectrum of objectives from enhancing onboard system intelligence to ensuring interoperability across diverse rail networks. In the context of TASV, several research studies have been initiated to explore the advantages of digitalization and automation (Gadmer et al., 2022). For example, there is ongoing research works to contribute to the safety demonstration process (Boussif et al., 2023), signal and track obstacle detection (Jourdan et al., 2022), operational risk assessment (Boussif et al., 2023), combining sophisticated algorithms with real-time data analytics to predict and prevent potential incidents (Hathat et al., 2022). Additionally, the qualification, verification and validation, and certification of AI systems are also aspects of interest in the TASV project (Boudardara et al., 2023, 2022). In parallel, the TFA project undertakes rigorous simulation-based assessments to evaluate the robustness of autonomous train operations under varied scenarios (Collart-Dutilleul et al., 2019; Plissonneau Duquene, 2023), while TCRail investigates the challenge behind the remote train control connectivity of freight trains (Masson et al., 2019).

The pursuit of an autonomous railway system encompasses not only the design and implementation of advanced control technologies but also an encompassing view of the safety regulations and standardization imperatives to the rail industry. Within this context, the dissertation addresses critical safety challenges, building on the foundational work established by TASV to advance the safe automation of train control. The goal is to ensure that the progressive steps towards automation do not merely end with technical capability but extend to establish robust and safe operations within the rail sector. By focusing on decision-making in safety-critical situations and safety argumentation during the design and the operation phases, this dissertation aims to support the potential of automation and autonomy in improving rail safety. Indeed, ensuring adequate decisions in hazardous operational contexts and enhancing the traditional approaches to safety assurance, i.e., activities and processes in place to assess, ensure, and justify safety, are pivotal challenges for autonomous trains to guarantee their seamless and safe integration into the current railway overall system. The contributions herein align with these safety challenges, which were also highlighted in the TASV project.

1.3 Scientific context

Ensuring the overall safety is a fundamental objective of the railway industry. In fact, before and during the deployment of any railway system, a comprehensive safety assessment should be carried out through a structured safety demonstration process. This assessment usually leads to the use of technical and functional safety measures like Automated Train Protection (ATP) and operational measures such as rules and procedures. Together, these measures are essential for assuring the safe operations of the railway system. Safety

¹²<https://railenium.eu/train-autonome/>

assurance in autonomous railway systems deals with challenges in both design and operations. In Europe, railway safety has been traditionally justified by following safety argumentation, in line with European standards CENELEC EN 50126 (EN-50126, 2017), EN 50128 (EN-50128, 2011), and EN 50129 (EN-50126, 2017). However, this approach faces challenges in fully addressing the safety demonstration needs of systems that use AI and complex Machine Learning (ML) algorithms (De la Vara and Panesar-Walawege, 2013). These algorithms focus on recognizing patterns and making decisions based on data (Bishop, 2006), while including a broader range of computational methods that give systems the ability to reason and solve problems (Russell et al., 2016).

The railway systems using AI and ML technologies operate in dynamic railway environments, characterized by intrinsic uncertainties in train states and surroundings (Shafaei et al., 2018). Ensuring their safety demands a paradigm shift in safety assurance methodologies. Moreover, proving compliance with such systems, which inherently adapt and learn, poses safety challenges. Consequently, new approaches are needed to handle these complexities, ensuring safety assurance throughout both the design and operational phases. Safety assurance within the context of autonomous railways can be divided into two primary phases: design-time and run-time. Each of these phases presents its own set of challenges and considerations.

Safety assurance during design-time

The safety assurance of railway systems during the overall lifecycle in general, and particularly, the design-time is guided by the standard EN 50126. In addition, the railway sector has relied upon textual argumentation as a written evidence for safety assurance. Rooted in European standard EN 50129, this approach entails presenting safety cases that methodically demonstrate compliance with safety requirements and standards. However, with the advent of autonomous train systems, incorporating comprehensive decision-making processes and dynamic risk assessments, this method grapples with limitations in effectively tackling these complex aspects. Indeed, autonomous systems introduce a level of complexity and dynamics not previously encountered in rail transportation. With the integration of sophisticated decision-making algorithms and the necessity for continuous risk evaluation, the textual methods of safety argumentation struggle to encapsulate the unpredictable scenarios presented by these autonomous systems. The limitations become clear as one considers the nature of autonomous trains, which require a proactive and predictive safety approach, contrasting with the preventive and reactive ones of current standards.

Safety assurance during run-time

Ensuring the safety of autonomous trains during their current operations is a critical concern. This encompasses the need for dependable run-time safety measures and the application of real-time safety assurance techniques. Importantly, the inherent possibility of safety incidents during autonomous train operations requires a highly sophisticated and reliable set of safety measures. The proactive aspect of such safety measures involves not only anticipating and preventing possible faults and failures, but also ensuring swift and effective action in response to unexpected hazardous events and situations. Furthermore, the continuous monitoring and dynamic response capabilities of these safety measures are essential. They must be designed to adapt quickly to changing conditions and operational demands, ensuring that the highest level of safety is maintained at all times. This means that the safety systems not only react to current conditions but also adapt their strategies proactively, staying aware of potential risks and ensuring uninterrupted, safe operation of autonomous trains.

Finally, it is worth stressing that the safety assurance at both the design and run-time phases for autonomous railway systems is imperative to deal with the challenges identified. On the one hand, design-time safety assurance, traditionally anchored to standards such as EN 50126/28/29, has to adapt to the complexity of autonomous technologies. This adaptation involves not only reinterpreting existing standards, but also developing new approaches to address the unpredictability of autonomous operations. On the other hand, for run-time safety, the focus shifts to dynamic risk assessment and the need for continuous monitoring and responsive safety management systems. Here, the real-time operational context demands that safety measures are proactive, flexible, and able to mitigate risks as they occur. Both phases are crucial in maintaining the integrity of autonomous rail systems.

1.4 Problem formalization

To develop contributions to help the safety demonstration process of the autonomous train during the design time, as well as its safe operation during run-time, the focus is oriented primarily on three key facets:

1. **Safety argumentation:** In the safety assurance process of a system, justifying how the system complies with safety requirements, consists in creating a safety case. This is a structured document that presents safety claims, supported by arguments and evidence, to justify the safety of this system. In the traditional approach, a textual safety argumentation is employed. For autonomous trains, the safety requirements' compliance also needs to address dynamic risk assessments and risk-based decision-making processes. Therefore, the key is to develop a clear and effective safety argumentation through innovative methodologies and approaches;
2. **Situational awareness and dynamic risk assessment:** Autonomous trains have to operate in dynamic and unpredictable railway environments, requiring real-time risk assessment capabilities. For a train driver, such capabilities refer to as situational awareness. This involves continuously monitoring the surroundings, identifying potential hazards, and executing timely actions to prevent accidents. For the autonomous train to continuously assess its environment for potential hazards, the key is to establish a dynamic risk assessment framework able to structure the architecture and the data organization in the decision-making process of the autonomous driving system. The aim is to ensure that the onboard system will be designed to effectively identify and evaluate risk in real-time.
3. **Risk-based decision-making process:** The key approach is then to develop the risk-based decision-making process. This involves defining an approach that can effectively interpret the risk data provided by the dynamic risk assessment while helping the autonomous train in making safe decisions. Addressing this challenge is essential for enabling the train to adapt its actions in response to varying risk levels, thereby maintaining safety during its operation. Furthermore, the risk-based decision-making process should consider the risk level based on the current operational conditions. This ongoing assessment is essential for identifying and avoiding potential hazards in the train's surrounding environment.

1.5 Contributions

Having identified the safety challenges associated with autonomous trains, the next step is to define the research questions to be addressed, and the intended objectives. The problem

formalization discussed previously underscores the safety assurance challenges posed by autonomous trains from their initial design to their operational phases.

- **Research question 1:** How can safety argumentation in autonomous train systems be structured effectively to accommodate their complexities while complying with existing safety standards?
- **Research question 2:** What approaches, and real-time risk assessment models, are needed to enhance the efficiency and effectiveness of safety assurance and decision-making for autonomous trains operating in dynamic environments?
- **Research question 3:** What strategies and functions can be developed to ensure timely preventive actions are taken to maintain safety during autonomous train operations?

Based on the previous research questions, we fixed the following corresponding objectives :

- **Objective 1:** Develop a comprehensive safety argumentation approach to address the complexities involved in constructing the safety case of autonomous trains. This approach focuses on safety considerations at the design phase, ensuring that a comprehensive safety argumentation is integrated from the beginning of the autonomous train's development;
- **Objective 2:** Propose a situational awareness process integrating a dynamic risk assessment framework for autonomous trains. This framework allows the autonomous driving system to continuously and effectively monitor the environment, accurately identify potential hazards, and takes decisions while assessing the associated risk in real-time;
- **Objective 3:** Develop a risk-based decision-making model for autonomous trains. This process should be capable of understanding and interpreting the dynamic risk assessment information to continuously provide the level of risk under varying operational conditions. The objective is to enable the autonomous train to make informed and safe decisions that effectively prevent hazards and ensure safe operation through the surrounding environment.

In conclusion, these objectives contribute to an integrated and comprehensive safety demonstration process for autonomous trains. By addressing both design-time safety argumentation and run-time dynamic risk assessment, autonomous trains are equipped to operate safely and efficiently within the dynamic and unpredictable railway environment.

1.6 Manuscript organization

This dissertation is organized as follows:

- **Chapter 2 - Safety assurance of autonomous systems.** This chapter provides an extensive literature review essential to our research. It encompasses various critical aspects, including an examination of the Autonomous Driving System (ADS), risk assessment methodologies for both conventional and autonomous vehicles, risk models and decision-making approaches, and a detailed exploration of risk assessment within the railway context. Additionally, the chapter investigates safety assurance methods, such as safety case elaboration with graphical safety argumentation methods, and the significance of Goal Structuring Notation (GSN). This comprehensive review forms a solid foundation for our subsequent research contributions.

- **Chapter 3 - Graphical safety argumentation for autonomous trains.** This chapter provides a comprehensive review of safety cases, highlighting their role in the assurance of safety. Additionally, it investigates the practical implementation of graphical safety argumentation, specifically within the framework of the Goal Structuring Notation (GSN). This examination includes various transportation domains, such as automotive, aviation, and railways, showing the versatility of these approaches. Moreover, we establish an application of the GSN method for structuring the argumentation of safety functions, such as the anti-collision function and the safe train stopping function, offering practical insights into the real-world application of these graphical approaches.
- **Chapter 4 - Dynamic risk assessment and situational awareness of the autonomous train.** This chapter reviews Dynamic Risk Assessment (DRA) specifically for autonomous trains and introduces a new framework for situational awareness and dynamic risk assessment. The chapter starts with explanations of DRA and situational awareness and their importance for autonomous train operations. It then presents a new safety framework allowing the ADS to continuously monitoring the environment, accurately identifying hazards, and making decisions while assessing risks in real-time. A key feature of this chapter is a use case illustrating the application of the DRA framework to the obstacle detection function in autonomous trains, showing how it helps in making safe decisions by anticipating and adapting to changing conditions. This chapter provides a practical look at how DRA can ensure safety in autonomous train operations and sets the stage for further research in this area.
- **Chapter 5 - Risk-based decision-making process for autonomous trains using POMDPs.** This chapter develops a risk-based decision-making process employing Partially Observable Markov Decision Processes (POMDPs) for continuous environmental monitoring in autonomous train systems. The primary objective of this chapter is to ensure the safe operation of autonomous trains by consistently assessing and adapting to risk levels associated with the surrounding environment. This approach involves the estimation and update of risk, considering all the uncertainties inherent in the perceived environment. The chapter provides a comprehensive review of how POMDPs can be effectively used to make informed decisions that prioritize safety. Particular attention is given to the application of this process in the context of preventing collisions in autonomous trains.
- **Chapter 6 - Conclusion.** This chapter provides conclusion remarks regarding the dissertation and draws future research directions.

Chapter 2

Safety assurance of autonomous systems

Contents

2.1	Introduction	21
2.2	The autonomous driving system	22
2.3	Risk assessment	24
2.3.1	Risk assessment for conventional vehicles	26
2.3.2	Dynamic risk assessment for autonomous transportation systems	26
2.4	Risk models for autonomous vehicles	28
2.5	Decision-making for autonomous vehicles	32
2.5.1	Deterministic approaches	32
2.5.2	Non-deterministic approaches	33
2.6	Risk assessment in railways	40
2.6.1	Risk assessment for conventional trains	40
2.6.2	Risk assessment methods	40
2.6.3	Dynamic risk assessment for the autonomous train	42
2.7	Safety assurance of the autonomous train	43
2.7.1	Safety cases	44
2.7.2	Graphical safety argumentation	45
2.7.3	Goal Structuring Notation (GSN) application examples	47
2.8	Conclusion	48

2.1 Introduction

The development of Autonomous Vehicles (AVs) is an ambition that arose a few years ago in the road sector and is now growing in the same way in the rail sector due to the expected benefits (Fagnant and Kockelman, 2015; Bagloee et al., 2016). Indeed, autonomous transportation creates completely new possibilities for optimizing the railway network capacity, expanding mobility, creating new economic opportunities for jobs and investment, and increasing environmental benefits (Wang et al., 2016b; Martínez-Díaz and Soriguera, 2018; Yin et al., 2017; Read et al., 2019; Singh et al., 2021). Efficient management of energy consumption is another significant additional benefit. Such advantages contribute to reduce transport costs but also show the important role that autonomous trains can play in achieving EU policy objectives of sustainable development (EU-Commission, 2017; Niestadt et al., 2019b).

A central challenge is confronted in the pursuit of autonomous rail systems, which is ensuring their safety. The process of ensuring the safety of autonomous trains includes tasks such as defining autonomy levels (GoA) (Ramos et al., 2019a), specifying the Operational Design Domain (ODD) (Torens et al., 2023; Tonk and Boussif, 2022), identifying hazards related to autonomy (Ventikos et al., 2020), implementing fail-safe mechanisms (Pek and Althoff, 2019), qualifying AI systems, rigorous testing, verification, and validation, (Corso and Kochenderfer, 2020), using dynamic risk analysis (Reich and Trapp, 2020), and real-time monitoring (Machin et al., 2016). Equally significant is the systematic documentation, substantiation, and structured organization of safety-related results, established in the autonomous train safety case. This rigorous process is vital for demonstrating compliance with the associated railway safety standards, building a body of evidence that supports established safety requirements and criteria.

In this chapter, we provide a review of the autonomous transportation systems, starting with an examination of existing autonomous driving systems used for AVs. We also present the essential aspects of risk assessment, risk models, and decision-making processes, focusing on the challenges and potential solutions in the context of AVs. Subsequently, we transition the discussion to the equally important aspects of autonomous trains. In conclusion, we provide a general overview of AVs before shifting focus towards railways, with a particular attention on autonomous trains.

2.2 The autonomous driving system

The effort to enhance transportation efficiency and safety has positioned ADSs as a key area of innovation. These systems are essential for new developments in transportation, significantly affecting both the operation and safety of AVs (Frigerio et al., 2021).

The autonomous driving system or automated driving system is defined, according to ISO 34501 standard (ISO-34501, 2022), as a “*hardware and software that are collectively capable of performing the entire Dynamic Driving Task (DDT) on a sustained basis, regardless of whether it is limited to a specific Operational Design Domain (ODD)*”. Similarly, it is defined according to the Society of Automotive Engineers (SAE) as a “*type of motor vehicle equipment that is capable of performing some or all aspects of the Dynamic Driving Task (DDT) without human intervention*” (SAE, 2016). Notice that, in this context, the DDT is referred to as: “*all the real-time functions required to operate a vehicle in on-road traffic, excluding election of final and intermediate destinations*” (SAE, 2016). Indeed, the two definitions signify the evolution of transportation technology, where ADS plays a central role in augmenting vehicle autonomy and diminishing dependency on human drivers.

In fact, to understand the function of ADS, a look into its foundational principles and architectural components is necessary. These core aspects, which include sensor fusion, perception, control and decision-making, and mapping and localization, collectively shape the capabilities of ADSs. Figure 2.1 represents a simplified illustration of the ADS for AVs, divided into three primary functional layers: *Perception*, *Planning*, and *Control*. In the perception layer, the vehicle uses an array of sensors (e.g., GPS, cameras, LiDAR, etc.) to gather data about its surroundings. This information is crucial for detecting and interpreting the environment, which includes other vehicles, pedestrians, road signs, and lane markings. Moreover, the planning layer is subdivided into three main components: behavior planning, motion planning, and route planning. Behavior planning determines how the vehicle should behave in response to specific situations, such as when to give way or pass. In addition, motion planning calculates the vehicle’s trajectory based on the desired behavior, ensuring safe and efficient operation around obstacles. Route planning involves mapping out a path to the destination, guided by the localization of the vehicle

and its task objectives. Finally, the control layer is where the vehicle’s actuators (i.g., acceleration, braking and steering mechanisms) are managed. This layer translates the planned trajectories into actionable commands, executing the driving maneuvers necessary to follow the planned route while maintaining safety and comfort.

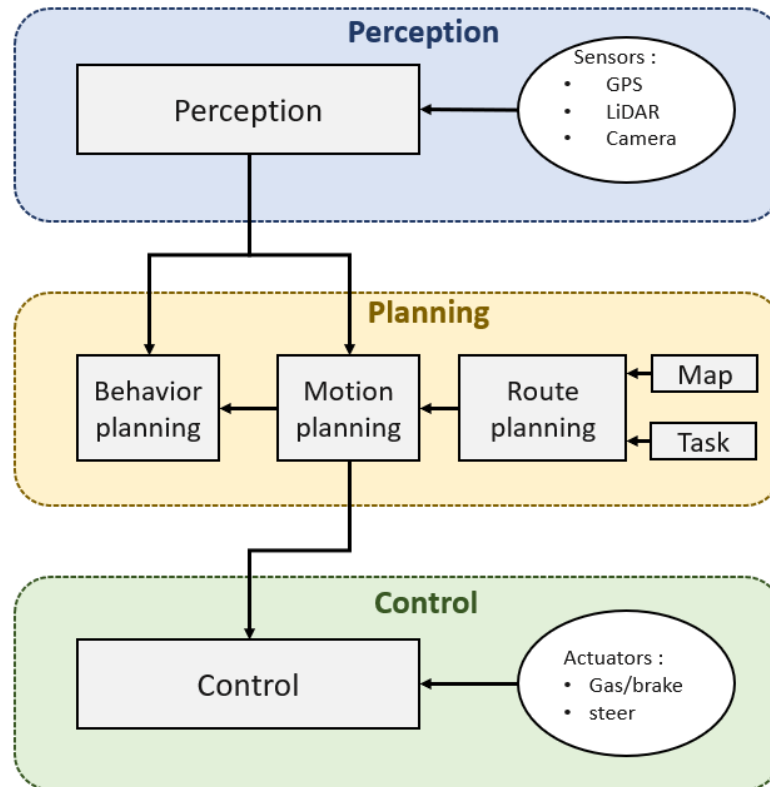


Figure 2.1: A simplified illustration of the Autonomous Driving System (ADS) in automotive

One of the basic components of the ADS is **perception**. In fact, perception is the ability of autonomous systems to interpret sensory data and extract meaningful insights (Bogdoll et al., 2023). In this context, advanced ML and computer vision algorithms play a pivotal role. This role is to provide to AVs the ability to recognize objects, pedestrians, traffic signs, and other elements of the environment (Chen et al., 2015). Through AI techniques, AVs become able to identify patterns, predict the behavior of surrounding entities, and adapt its behavior to dynamic scenarios (Hou et al., 2023).

Another main component in ADS is **sensor fusion**, which is a process that combines data/information from multiple sensors, each with a specific purpose (Kocić et al., 2018). Sensors, such as LiDAR, radar, cameras, and GPS, serve as the sensory organs of AVs, capturing a wide array of information from their surroundings. Furthermore, sensor fusion techniques enable these systems to synthesize this data into a cohesive situational awareness (Reich and Trapp, 2020). By combining the strengths of various sensors, autonomous systems gain a comprehensive understanding of their environment, a vital capability for safe navigation.

Moreover, **mapping and localization** are also essential for the accurate positioning of AVs within their environment (Zheng et al., 2023). Through high-definition maps and advanced localization techniques, vehicles can be informed about their location with a certain level of accuracy. Simultaneously, they can update these maps in real-time, adapting to changes in the environment and ensuring optimal route planning (Chalvatzaras

et al., 2022).

Finally, **control and decision-making** algorithms are the brain behind the operation of autonomous systems. These algorithms receive processed sensory data and create a plan for action (Li et al., 2018). Whether it's steering, accelerating, braking, or executing complex maneuvers, control and decision-making algorithms ensure that AVs respond effectively and safely to the information gathered from their sensors. By continuously evaluating data and predicting potential outcomes, these algorithms enable safe and efficient navigation (Schwartz et al., 2018).

After discussing the essential components of ADS for AVs, which are crucial for their operation and safety, we now focus on the proposed architecture of the on-board ADS for autonomous trains. The proposed architecture is designed to respond to the challenges faced by autonomous trains, including environment detection, accurate positioning, and effective decision-making in various railway scenarios. This architecture aims to integrate ADS technologies into the railway domain, outlining the adjustments needed to meet the high safety and reliability requirements of rail transport.

The fundamental architecture of the ADS used in AVs globally remains unchanged when adapted to autonomous trains. Although the operating environment shifts from roadways to railways, the key components of the ADS, include perception, mapping and localization, decision-making, and control. Figure 2.2 illustrates the comprehensive architecture (in left) and the main functions (in right) of the on-board ADS for autonomous trains. Central to the system are the autonomous driving tasks, which are primarily composed of perception, sensor fusion, control and decision-making, mapping and localization, and other critical functions such as obstacle avoidance. These tasks are interconnected with various aspects of the train's operation and management systems. For instance, the ADS is responsible for motion planning, fault diagnosis, and ensuring vehicle cybersecurity, all of which are crucial for the safe and efficient functioning of autonomous trains. Additionally, the ADS interfaces with external elements like trackside ADS, infrastructure, the remote control center, and the train attendant, demonstrating the system's integration with the larger operational environment. Moreover, maintenance management is an integral part of this architecture, ensuring the reliability of the ADS. The figure also acknowledges the influence of surrounding environment on the system, highlighting the adaptive nature of the ADS to real-world conditions. Other elements, which may include Automated Train Operation (ATO) and European Rail Traffic Management System (ERTMS) standards, provide an additional layer of operational parameters that the AADS must comply with, ensuring both regional and international interoperability.

From the foundational elements of autonomous vehicle technology presented in the previous section, the importance of understanding its foundation becomes clear. To understand how these vehicles might safely integrate the railway overall environment, a review of risk assessment is required. The focus now shifts from the general functioning of the autonomous driving system to a detailed examination of how potential risks are assessed by these vehicles.

2.3 Risk assessment

The promise of AVs in reducing traffic, lowering accident rates, and improving driving experience relies on their ability to make efficient and safe decisions without human intervention (Costantini et al., 2020). Central to this capability is the process of risk assessment. A clear understanding of how risk is perceived and evaluated by AVs is crucial to ensuring their safe deployment. This section outlines risk assessment for AVs, focusing on its significance, methodologies, and dynamic aspects.

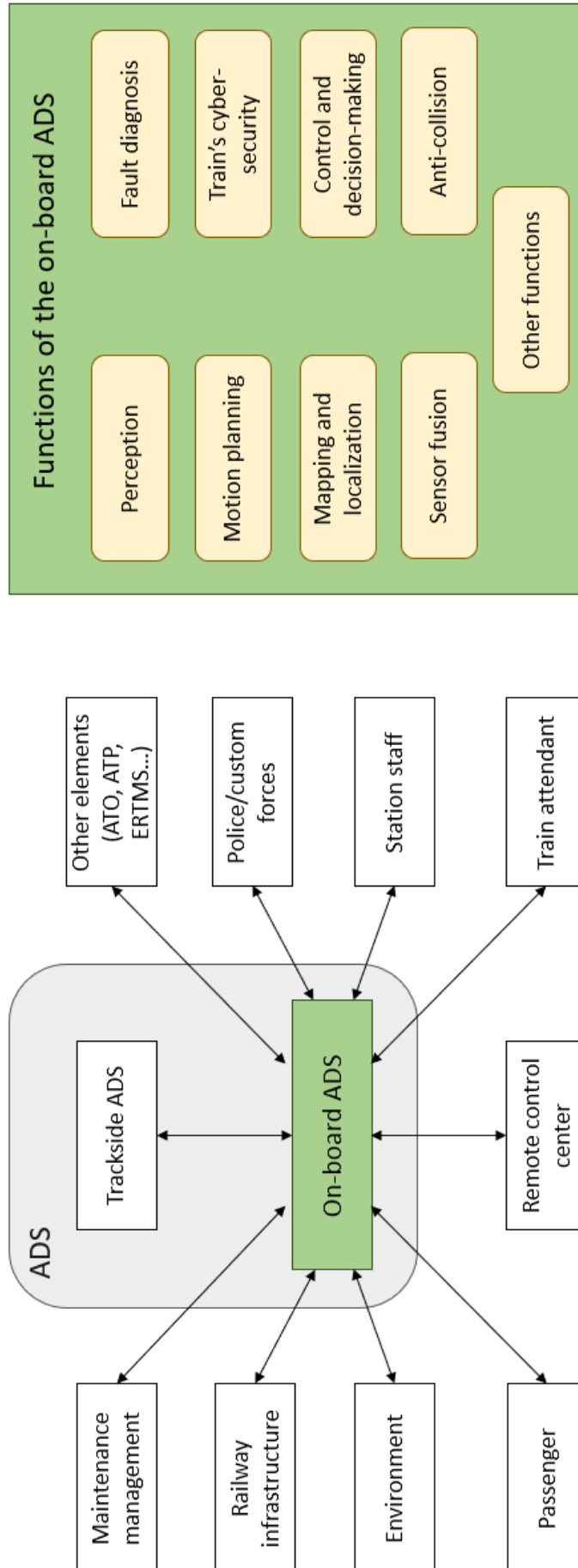


Figure 2.2: On-board Autonomous Driving System (ADS) components and interactions

2.3.1 Risk assessment for conventional vehicles

In the context of automotive engineering, *risk* is defined as “*a combination of the probability of occurrence of harm and the severity of that harm*”, according to the ISO 26262 standard (ISO-26262, 2018). Based on this definition, the risk often pertains to the probability of an adverse event occurring due to the vehicle’s operation, such as a system malfunction, accident, or any event that compromises safety.

According to the ISO 31000 standard, which provides guidelines on risk management, risk assessment is defined as the “*overall process of risk identification, risk analysis, and risk evaluation*” (ISO-31000, 2018). In the automotive domain, this includes a structured approach to identify potential hazards, determine the likelihood of those hazards occurring, assess the severity of their consequences, and decide on appropriate mitigation measures.

For conventional vehicles (with drivers), the risk assessment process is mainly focused on three aspects: the vehicle’s mechanical and electronic integrity (Agrawal et al., 2021), the environment in which it operates (Islam et al., 2016), and the driver’s capabilities (Macher et al., 2016). Standards and regulations typically emphasize vehicle maintenance, roadworthiness tests, and driver training as key components to mitigate risk (European-Commission, 2014). Additionally, passive and active safety systems, such as airbags and anti-lock braking systems, respectively, play a role in risk reduction.

In fact, while risk assessment principles guide conventional vehicles’ (with human drivers) safety assurance, they become more complicated when applied to AVs. In reality, AVs introduce new dimensions of complexity, especially when it comes to interpreting and predicting human behavior, managing, in real-time, large amounts of data from all sensors, and ensuring the reliability of complex algorithms and software. This complexity is likely to exceed the risks for conventional vehicles. Transitioning from risk assessment methods for conventional vehicles, the discussion advances to DRA for AVs.

2.3.2 Dynamic risk assessment for autonomous transportation systems

The approach of assessing risk while taking into account time-varying and/or environmental changes is known as Dynamic Risk Assessment (DRA), which is generally based on probabilistic methods. The DRA is a proactive risk assessment method that integrates real-time data, adapts to current situational changes, and updates risk estimations accordingly (Zio, 2018). Unlike traditional risk assessment methods, which often rely on historical data and predefined scenarios, DRA is structured to accommodate the evolving and unpredictable nature of operational conditions (Feth et al., 2018). In fact, DRA provides a continuous, iterative process of risk evaluation that is especially pertinent in high-speed, complex, or rapidly changing contexts where the static models of yesteryears may no longer be efficient.

While the concept of DRA is recognized for its adaptability and real-time response, methodologies can differ based on the specific application, technological tools available, and the nature of risks involved. Nevertheless, the core principle remains consistent: a continuous, real-time evaluation of risks with an adaptive response strategy (Aven, 2016). Hereafter, we present a survey of DRA used in transportation systems, highlighting the diversity and innovation in risk assessment approaches that are essential for managing complexities of modern mobility systems.

Indeed, AVs requires robust decision-making capabilities and rigorous risk assessment procedures to ensure safety and reliability. In their study, Liu et al. (2021b) explored advanced technologies for making decisions in AVs by focusing on ML and data-driven algorithms. By emphasizing learning-based methods, they not only explored current applications but also suggested the future directions of these methodologies. This shift towards ML and data-driven techniques is becoming increasingly central in the automotive

domain. The need for efficient data use became more evident through the work of [Yu et al. \(2021\)](#). Their research combined data-driven risk assessment with graphical augmentation, presenting an innovative approach to gauge and predict autonomous vehicle decisions in diverse scenarios. Such integrative methods shed light on the potential intersections of technology and real-world environments. Shifting focus towards software components, [Reich and Trapp \(2020\)](#) proposed a dynamic risk assessment module that focuses on situation awareness. As vehicles become more autonomous, understanding and adapting to immediate surroundings becomes paramount, making such innovations a necessity. Extending the discussion to autonomous mobile robots, [Müller et al. \(2022\)](#) underscored the importance of situational risk assessments. These robotic platforms, though distinct from conventional vehicles, share the overarching need for safety protocols, hinting at the expansive nature of automotive safety. [Feth et al. \(2018\)](#) then conducted a comprehensive analysis of the potential of deep learning, developing a dynamic risk assessment framework for higher-level AVs. Their focus on ML once again emphasizes the automotive industry's move towards cutting-edge computational techniques. Lastly, [Chia et al. \(2022\)](#) conducted a comprehensive survey on the methodologies that drive risk assessment in autonomous driving. Their insights offer a summary of the safety practices that support the future of transportation.

Railway systems, being essential for mass transportation, have incorporated a set of processes and procedures aimed at ensuring the safety of their passengers. In recent years, the railway domain has seen a surge in implementing advanced methodologies for real-time risk assessment, especially in the light of high-speed train projects. For instance, a study by [Xue et al. \(2020\)](#) proposed a risk coupling model based on system dynamics, emphasizing high-speed rail systems and their complex factors. This model helps explain how different risk factors in high-speed train projects are interconnected. Urban railway networks, including metro systems, are now focusing on assessing overcrowding risks. One significant contribution in this area is the work of [Alawad et al. \(2020b\)](#), who highlighted the potential of the Adaptive Neuro-Fuzzy Inference System (ANFIS) to estimate the risk levels of overcrowding in railway stations. This study outlines how using AI can improve safety at busy and crowded train stations. Furthermore, [Alawad et al. \(2020a\)](#) made significant progress by incorporating deep learning into evaluating railway risks. Their method shows how complex ML models can play a crucial role in analyzing detailed data, helping railways to make informed and safe decisions. Derailment risks are also critical in the railway domain. [Appoh and Yunusa-Kaltungo \(2022\)](#) introduced a dynamic hybrid model focusing on comprehensive risk assessment. Their case study focused on train derailment caused by coupler failures, showing how using different methods together can give a better understanding. Moreover, rail transport of hazardous materials is another segment requiring detailed risk analysis. For example, [Zarei et al. \(2022\)](#) presented a dynamic domino effect risk analysis model in this regard. This research underscores the critical impact of risks in rail transport, especially when hazardous materials are involved.

The research and techniques reviewed across various fields highlight a unified progression towards predictive, proactive, and data-centric strategies. As we advance into a future where human and algorithmic decision-making increasingly converge, DRA become essential in establishing safe and efficient outcomes for every industry.

After discussing the methods and considerations of risk assessment and dynamic risk assessment for AVs, the next step is to summarize these processes within specific approaches. These are represented by risk models designed for AVs. These models offer a systematic method to identify potential hazards, establish safety measures, and define decision-making processes. In the following section, we examine these risk models, understanding their fundamental principles and importance to autonomous driving.

2.4 Risk models for autonomous vehicles

Risk models are key tools in risk assessment and management, enabling a systematic and analytical framework for quantifying risks. They integrate various variables and scenarios to predict potential outcomes, thus facilitating the development of strategic responses (Katrakazas et al., 2019). For autonomous vehicles, the application of risk models is critical. Indeed, the operation of AVs involves complex decision-making processes, interaction with unpredictable environments, and compliance with safety standards, which necessitates the use of advanced risk models. These models assess a wide range of factors affecting AVs, from software algorithms to sensor reliability, ensuring all potential risks are thoroughly identified and mitigated.

Table 2.1 provides a comprehensive overview of a variety of risk assessment models applied across diverse domains, predominantly in automotive, aviation, and railway systems. Chronologically arranged, the table illustrates the evolution of risk models emphasizing dynamic risk management and real-time risk assessment. Feth et al. (2018) is notable for introducing a method designed for the automotive sector, using Convolutional Neural Networks (CNN) and relying on data-driven inputs like the current driving situation and various obstacle parameters, with applications centered on collision risk metrics. In contrast, Ma et al. (2019) explored risk models in the railway domain, employing a Bayesian network to assess high-speed catenary risks under varying conditions, aiming to predict catenary flashovers during adverse weather. Subsequent models, such as those proposed by Cheng et al. (2021), focused on Unattended Train Operation, incorporating elements such as Probabilistic Hybrid Automata (PHA) and Model Predictive Control (MPC). Maritime and autonomous surface vehicles have also been explored, with Hagen et al. (2022) employing Model Predictive Control (MPC) for collision avoidance and Hartsell et al. (2021) applying risk management principles to autonomous systems, utilizing Bow Tie Diagrams (BTD) and incorporating sensor observations and fault diagnosis. Similarly, Kufoalor et al. (2020) employed MPC for collision avoidance in Unmanned Surface Vehicles, integrating sensor observations and runtime monitoring. Chia et al. (2021) developed a Recursive Risk Assessment Framework for the automotive sector, combining safety levels with predictive risk numbers and employing both Physics-based and Data-based methods. Reich's works in Reich and Trapp (2020) and Reich et al. (2021) focused on automotive applications, employing Bayesian networks for dynamic risk assessment and collision avoidance, utilizing perception information to influence vehicle behavior maneuvers. Katrakazas et al. (2019) integrated Interaction-aware motion models and Dynamic Bayesian Networks (DBNs) for real-time risk assessment in AVs, utilizing sensor measurements and vehicle kinematics. Eggert (2018) contributed by developing a dynamic risk map for the automotive domain, focusing on predictive driving and utilizing situation classification and trajectory predictions. Lastly, Li et al. (2022) explored decision-making in AVs through a probabilistic model and Deep Q-network (DQN), applying this in scenarios with static and dynamic obstacles.

In conclusion, risk models are crucial for enhancing risk management strategies, providing a methodical way to quantify and analyze risks. The variety of models listed in Table 2.1, from CNNs (Convolutional Neural Networks) to BNs (Bayesian Networks), demonstrates the comprehensive nature of risk modeling. In the context of AVs, the complexity of autonomous decision-making, interactions with dynamic and changing environments, and the need for safety compliance require the use of sophisticated and adapted risk models. The development and variety of these models, as indicated in the table, shows ongoing progress in the domain, highlighting the continuous efforts to improve safety and mitigate risks across various sectors.

Examining the diverse characteristics of risk models reveals the nature of assessing and

managing uncertainties in autonomous transport systems. These models provide crucial insights into real-time risk assessment, forming a foundation for developing and enhancing safety requirements. However, while risk models are key in identifying and quantifying potential hazards, it is important to examine the decision-making processes addressing these risks more closely. Therefore, we move from the theoretical frameworks of risk models to practical decision-making strategies for AVs. In the following sections describe how AVs, guided by risk models, prioritize safety when making decisions under uncertain and changing conditions. This study aims to demonstrate the way risk assessment is integrated in the process of decision-making to ensure safety of AVs. Notice that, the ‘P-based’ and ‘D-based’ terms used in the table denote probabilistic-based and deterministic-based models, highlighting the distinction between approaches that use statistical probabilities and those that apply physical models and fixed variables to assess risks.

Table 2.1: An overview of risk models used for autonomous transportation systems

Reference	Concepts	Domain/System	Risk Model	P-based	D-based	Inputs	Outputs	Applications
Feth et al. (2018)	Resilience, Dynamic Risk Management, Dynamic risk assessment	Automotive	CNN	x		Current driving situation, Distance to obstacle, Obstacle speed, Obstacle trajectory, Obstacle position	Current risk confidence	Collision risk metrics
Ma et al. (2019)	High-speed Catenary, Dynamic risk assessment, Bayesian network	Railway	BN	x		Current catenary conditions, leakage current and conductivity of the pollution layer	Probability of "risk"/dynamic risk level	Catenary flashover in bad weather
Cheng et al. (2021)	Unattended Train Operation, Probabilistic hybrid automata, Online quantitative safety monitoring	Railway	Safety Constraint Computation Algorithm	x		Stochastic events, Maximum reachable probability of forbidden state	Safety status of UTO	UTO system
Hagen et al. (2022)	Collision Avoidance	Maritime/ Autonomous surface vehicle	Model Predictive Control (MPC)	x		Obstacle tracking system, current course and speed	Desired course, Desired speed	Polar 845 Sport vessel Telemetron
Hartsell et al. (2021)	System risk management, Dynamic risk case, Bow-Tie diagram	Autonomous vehicles	Bow Tie Diagrams (BTD)	x		Sensor observations, diagnosis, run-time monitoring	Risk estimation	Unmanned underwater vehicle
Kufoalor et al. (2020)	Collision Avoidance	Unmanned Surface Vehicle	Model Predictive Control (MPC)	x		Sensor observations, diagnosis, run-time monitoring	Cost function	Polar 845 Sport vessel Telemetron

Reference	Concepts	Domain/System	Risk Model	P-based	D-based	Inputs	Outputs	Applications
Chia et al. (2021)	Recursive Risk Assessment Framework, Predictive risk number, Collision Avoidance, Safety level	Automotive	Safety levels	x	x	Positioning GNSS, Sensor data, AV database, Vehicle operation	Safety Goals	Illustrative example
Reich and Trapp (2020)	Run-time safety, Situation awareness, Dynamic risk assessment	Automotive	BN		x	State of the traffic	Vehicle behavior maneuvers	Framework for situation-aware dynamic risk assessment of autonomous vehicles
Reich et al. (2021)	Run-time safety, Situation awareness, Dynamic risk assessment, Collision Avoidance	Automotive	BN		x	Perception information	Time To Collision, Collision probability	CARLA AV Simulator
Katrakazas et al. (2019)	Real-time risk assessment, Collision estimation, Vehicle-based risk estimation	Autonomous vehicles	Interaction-aware motion models, Dynamic Bayesian Networks (DBN)	x	x	Sensor measurements, Kinematics of the vehicles	Vehicle Level Risk, Network Level real-time collision risk	Car scenarios
Eggert (2018)	Risk modelling, Risk maps, Predictive driving	Automotive	Dynamic risk map			Situation classification, trajectory predictions	Risk map	Driving behavior evaluation and risk-avoiding trajectory planning
Li et al. (2022)	Autonomous vehicles, decision-making, ADAS, Reinforcement learning	Automotive	Probabilistic model, position uncertainty, distance-based safety metrics, Deep Q-network (DQN)	x	x	DQN-based algorithm	Best policy	CARLA (2 scenarios: static and dynamic obstacles)

2.5 Decision-making for autonomous vehicles

The development of Decision-Making Systems (DMS) in Autonomous Vehicles (AVs) represents significant progress and knowledge, built on past developments and driven by constant innovation. Starting with early efforts in the 1990s, such as Autonomous Land Vehicle in a Neural Network (ALVINN) system by Pomerleau (1988) and the rule-based frameworks by Dickmanns et al. (1994), the field has seen rapid growth. Notably, the series of the Defense Advanced Research Projects Agency (DARPA), especially Grand Challenges (Buehler et al., 2009), has been crucial in promoting significant contributions and defining the current research and applications.

The effectiveness of a DMS relies on the integrated functioning of various interrelated components. These include precise localization and perception, understanding of operational parameters, and recognition of changing environment. The coordinated interaction of these elements is critical of making informed and reliable decisions, an essential requirement for the safe operation of AVs in complex environments.

Following the decision-formulation phase, attention moves to implementation, involving methods like motion planning and lower-level control mechanisms. These approaches are important for transforming formulated decisions into practical actions, thus ensuring that the vehicle functions effectively in its specified ODD.

Furthermore, Ulbrich et al. (2015) outline the following essential requirements for the operationalization of a DMS :

- **Rapidity:** swift decision-making is crucial for timely driving maneuvers;
- **Coherency:** the decision-making module should maintain consistency, avoiding unnecessary shifts in planning;
- **Providentness:** the system should foresee potential future scenarios and incorporate this foresight into decision-making;
- **Predictability:** decision-making should conform to human driver perception of safety and judgment.

In the subsequent subsections, we examine the historical developments, necessary components and execution strategies, along with a detailed review of the different natures of decision-making approaches (deterministic and non-deterministic) found in the literature. This analysis aims to provide understanding into the aspects of decision-making in AVs for future research and technological progress.

2.5.1 Deterministic approaches

During the initial phases of AVs development, classical approaches, particularly rule-based systems, were a key strategy. These systems relied on a set of predefined rules and logic to dictate the vehicles' reaction to certain scenarios and inputs (Ferber and Weiss, 1999). The central idea was that a well-developed and extensive rules could enable the vehicle to effectively and safely operate through a variety of operational conditions.

Exploring the history of rule-based systems reveals early efforts like ALVINN (Autonomous Land Vehicle In a Neural Network) (Pomerleau, 1988) and the models introduced by Dickmanns et al. (1994). These initiatives demonstrated the potential of using rule-based frameworks for AVs, with vehicles competently performing actions based on visual cues and events. The predictability and determinism of these early models mirrored the technological limitations and the emerging understanding of the field at that time.

Though rule-based systems initiated progress in autonomous vehicle technology, they had notable limitations. Their main objective was the inability to manage or adapt to

unknown situations or uncertainties. Limited by programmed (predefined) scenarios, these systems had difficulties with unpredictable events or changing environments (Zadeh et al., 1996). The adaptability and capacity of traditional methods were also questioned as driving conditions became more complex and varied.

In response to these limitations, the field shifted towards alternative strategies. In fact, the objective was to create decision-making frameworks that could adapt and respond to challenges. Advances in AVs technologies brought forward probabilistic models, strategies based on ML, and hybrid approaches. This aimed at addressing the complexities and uncertainties of operating in real-world scenarios. Furthermore, while traditional methods established the initial framework for AVs development, their limitations led a continuous search for more adaptable solutions. Insights from early models have significantly influenced research and development, leading to the complex decision-making systems in the current AVs.

2.5.2 Non-deterministic approaches

Shifting from deterministic methods that depend on rule-based systems, inadequate for unknown scenarios or uncertainties, led the research community to pursue alternative approaches. This shift resulted in the adoption of probabilistic methods, essential for dealing with the uncertainties related to changing environments and for the safe operation of AVs. The adoption of probabilistic or non-deterministic methods represents a significant change in AVs. These methods, based on probabilities, are designed to manage the uncertainties, support reasoning, and help the decision-making process. This allows AVs to adjust and react to new and changing situations (Kaelbling et al., 1998; Murphy, 2012). Some potential methods for the probabilistic approach can be given by :

1. **Markov Decision Processes (MDPs):** Markov Decision Processes (MDPs) are mathematical frameworks that capture the dynamic decision-making scenarios in environments with stochastic outcomes (Bellman, 1957; Puterman, 1990). Fundamentally, a MDP includes a collection of states, a series of actions, transition probabilities that outline the system dynamics, and a reward function. Within autonomous driving, states could indicate various scenarios on the road, actions could refer to potential maneuvers a car can perform, and the transition probabilities would account for the uncertainties in the environment's reaction to these actions White (1993).

Formally, a MDP is described by a tuple $\langle S, A, T, R \rangle$ where S is the set of states, A is the set of actions; T is transition probability of the system to a state s' from state s taking action a , R is the reward that the agent will receive depending on the state of the system and the action chosen. Figure 2.3 shows the main functioning of an MDP and its interaction with its environment. It states that executing an action $a \in A$, given the system defined by a state $s \in S$, is what will be called a policy $\pi : s \rightarrow a$. The goal of such a problem is to find an optimal policy (sequence of actions) π^* that maximizes the expected reward over the time horizon T (Howard, 1960).

The field of autonomous driving has seen a significant increase in research utilizing MDP to develop optimal and safe driving policies. Wu et al. (2022) explored the development of a hybrid driving decision-making system by combining Markov logic networks with AI based on neural networks, aiming to improve safety in AVs operations. Avilés et al. (2022) and collaborators provides a probabilistic logical description of an MDP to steer decisions in autonomous driving, especially targeting collision avoidance with other vehicles. In another insightful study, Ramanathan and Kartik (2021) discussed the potential of hidden Markov models and partially observable MDPs for intention-aware decision-making in self-driven vehicles.

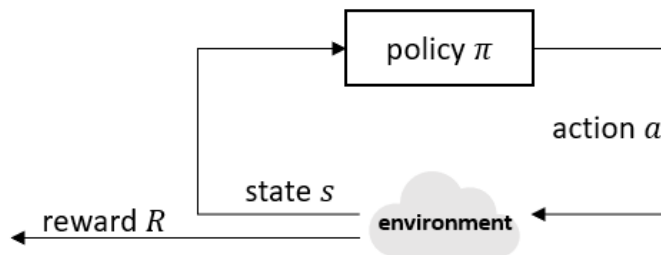


Figure 2.3: Illustration of the Markov Decision Process (MDP)

Adding a game-theoretic perspective, [Coskun and Langari \(2018\)](#) introduced a new decision-making method for autonomous driving by combining Markov games and MDPs. Addressing traffic scenarios, [Palanisamy et al. \(2020\)](#) examined the use of hierarchical structures within the MDP framework to facilitate autonomous driving decisions at intersections. In a related approach, but with a focus on highway scenarios, [Guan et al. \(2018\)](#) proposed a MDP-based method that sidesteps the reliance on human driving data or predefined rules.

Furhtermore, MDPs are adaptable, and their adaptability and flexibility was demonstrated by [Coskun et al. \(2019\)](#), who combined the ideas of Fuzzy logic and MDPs to develop a predictive Fuzzy Markov Decision Process (FMDP) model suited for autonomous driving scenarios. In a collaborative project with Toyota and Renault, [Laugier \(2019\)](#) highlighted how Bayesian and Machine Learning approaches enhance motion autonomy and safety in AVs. Driving decision-making, especially in situations like unmanned vehicle crossings, can be significantly augmented using the Markov model as outlined by [Feng et al. \(2014\)](#). Lastly, acknowledging the cooperative and competitive nature of traffic, [Cheng et al. \(2021\)](#) explored how AVs can learn policies and establish social norms in traffic through the lens of Markov games and deep reinforcement learning.

2. **Partially Observable Markov Decision Processes (POMDPs):** extend the MDPs to scenarios where the environment is dynamic and complete information about the current state is not always available ([Monahan, 1982](#)). Given the scenarios faced by AVs, where sensors might not always provide accurate information about the surroundings due to occlusions, malfunctions, or challenging environmental conditions, the use of POMDPs can be crucial. POMDPs offer a methodological approach to weigh the available information and help the autonomous driving systems take optimal decisions in real-world scenarios ([Silver and Veness, 2010](#)).

Examining its mathematical foundation, a POMDP is represented using a tuple $\langle S, A, O, T, Z, R \rangle$. Here, S encompasses a set of states and A represents a set of actions. Concurrently, the set O designates the possible observations. Transition probabilities, denoted by T , provide the probability of state transitions upon the execution of an action. Meanwhile, observation probabilities, described by Z , give insights into the likelihood of receiving a certain observation following an action, leading to a specific state. Lastly, the function R is an indicator of the anticipated reward for executing a particular action within a given state. Given the inherent unpredictability, agents operate with a *'belief state'*, a probability distribution spanning states. This belief state is updated with every action and subsequent observation, relying on the principles of Bayes' rule ([Schulman, 1984](#)).

The illustration presented in Figure 2.4 captures the essence of POMDPs when applied to scenarios typical of autonomous driving. The figure underscores states,

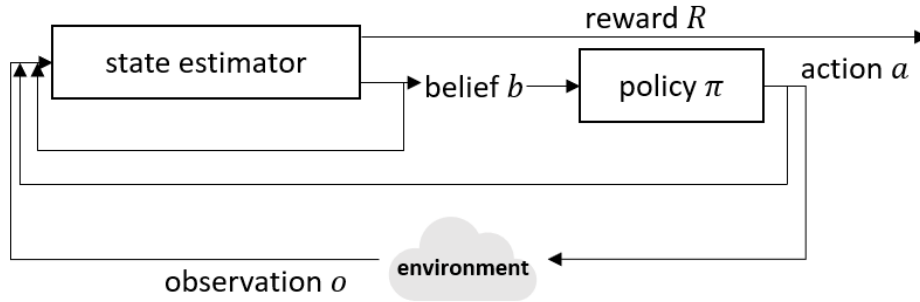


Figure 2.4: Illustration of Partially Observable Markov Decision Process (POMDP)

potential actions, and consequential probabilistic observations, highlighting the uncertainties pivotal to decision-making within a POMDP framework.

In the domain of autonomous driving, vehicles frequently navigate through dynamic environments, where the complexities are intensified by limited visibility and unpredictable changes. It is in these challenging terrains that POMDPs have demonstrated their advantages and benefits, providing an effective method to manage uncertainties.

[Haklidir and Temeltaş \(2022\)](#) examined autonomous driving by defining it as a POMDP problem, focusing on decisions in uncertain conditions. Their approach, termed the Guided Soft Actor-Critic (Guided SAC), demonstrated efficacy, especially in pedestrian crossing scenarios. The results closely matched those expected in environments with complete state knowledge, highlighting its effectiveness. [Pouya and Madni \(2020\)](#) extended the research horizon by crafting a probabilistic model aimed at decoding the behaviors of AVs. Their Expandable-POMDP models were notable for establishing a strong basis for strategic planning in environments with limited observability.

The complexity significantly increases at non-signalized intersections. Acknowledging this, [Quaglietta et al. \(2013\)](#) introduced a POMDP model based on the concept of Responsibility-Sensitive Safety (RSS). Their research showed positive outcomes, indicating improvements in traffic safety and driving smoothness.

Moreover, the application of POMDPs in the domain of robotics and autonomous systems has been explored extensively, with [Lauri et al. \(2022\)](#) survey outlining the integration of POMDP frameworks in various robot decision tasks. Lauri's work provides an exhaustive overview, emphasizing the efficacy of the POMDP framework in modeling and solving robot decision challenges under the issues of uncertainty. Further examining into algorithms based on POMDPs, [Lin et al. \(2019\)](#) categorizes these into approximate and exact algorithms. The study underscores the foundational role of exact algorithms, which subsequently inform and refine approximate methods.

Turning attention to autonomous systems dealing with non-Gaussian correlated uncertainty, [Chen and Zhang \(2019\)](#) introduced a new framework for chance-constrained stochastic model predictive control. The introduction of this method ([Cockburn et al., 2012](#)), established to manage non-Gaussian correlated uncertainties, is especially pertinent to autonomous vehicle control.

Fundamentally, the extensive research covering methods, applications, and challenges related to POMDPs, highlights their adaptability and effectiveness. In fields like robotics, autonomous driving, or any area dealing with uncertainty and limited observability, POMDPs have become a key methodology.

In conclusion, POMDPs is advantageous in decision-making under uncertainty, providing a comprehensive approach in environments where state observability remains uncertain. Their structure, which combines state transitions with probabilistic observations, makes them effective at investigating scenarios where applying other methods might be more complicated. For instance, while traditional MDPs requires an accurate view of the state space, POMDPs handle scenarios characterized by noisy, incomplete, or otherwise imperfect information. This capability is a significant benefit, particularly in real-world applications like autonomous driving, where environmental complexities and uncertainties are common.

- 3. Kalman Filters :** Kalman Filters (KFs) are advanced mathematical algorithms developed to offer accurate estimations of dynamic systems, even when faced with noisy measurements (Kalman, 1960). Historically used in aerospace applications, KFs have proven invaluable in a wide range of fields, including autonomous driving. At their foundation, KFs combine predictions from a system’s model with actual measurements to enhance state estimates. Conceptually, a KF relies on a two-step process: initially predicting the next state based on the system’s model and subsequently refining this prediction using the newest available measurements. By relying on Gaussian statistics, KFs effectively handle uncertainties both from system dynamics and sensor noise, optimizing accuracy in state estimations (Welch et al., 1995).

Autonomous driving has benefited from the integration of Kalman Filters. Manjunatha et al. (2023)’s work outlines this integration, where KFs are merged with neural structures, achieving improved performance when enriched with explicit vehicle models. The resilience of autonomous systems against security threats, as discussed by Yi and Chen (2023), underscores the potential of KFs.

Further extending the utility of KFs, Griebel et al. (2020) proposes a self-assessment strategy employing subjective logic. This strategy helps in gauging statistical uncertainties inherent in KFs, especially within the domain of autonomous driving. Taking a different approach, Nasir et al. (2017)’s research uses KFs for robot localization. This results in reduced errors, showing the advantages of KFs. In the same context, Khan et al. (2016) introduced a dynamic version of the KF, effective at predicting vehicle positions and velocities with precision. Finally, highlighting sensor fusion, Farag (2021)’s innovative approach combines LiDAR and Radar data, all coordinated by a KF, setting the stage for advancements in object detection and tracking.

To conclude, KFs have become essential tools in autonomous driving, bridging the gap between predictive models and real-world measurements. Their wide range of uses, from advanced neural networks to accurate localization and tracking, highlights their critical importance. Yet, recognizing their underlying assumptions and possible limitations is vital.

- 4. Reinforcement Learning (RL):** is an approach in which an agent learns to make decisions by interacting with its environment and receiving feedback in the form of rewards or penalties (Kaelbling et al., 1996). In fact, RL is a field that lies at the crossroads of computer science, intelligence, and control theory. Its main purpose is to develop algorithms that enable agents to make decisions by interacting with their environment to achieve goals. This approach has been extensively explained by Sutton and Barto (2018) who describe it as a process where agents adapt their actions based on feedback, in the form of rewards or penalties, to maximize long-term benefits. The essence of RL revolves around three primary entities: states, actions,

and rewards. As agents interact with the environment, they transition between different states, undertake actions, and accordingly receive rewards. The principal objective for an RL-based approach is to provide optimal policies, which is a defined mapping from states to actions, intending to maximize the expected rewards over time.

Over the years, RL has become increasingly popular in decision-making applications, especially for autonomous systems. In the automotive industry, there has been a significant increase in using RL to develop advanced driving strategies.

In the context of autonomous driving, [Hu et al. \(2022\)](#) thoroughly examined the complexities of motion planning in decision-making. Through their work, they proposed a novel deep Reinforcement Learning (RL) model that focuses on three key aspects: safety, efficiency, and smoothness. By integrating these factors, their approach provided a model for current research in developing reliable and safe AVs. In a concurrent vein, [Yang et al. \(2022\)](#) introduced the SMART (deciSion-Making frAmework based on ReinforcemenT learning) decision-making framework that prioritizes the robustness of RL. Their emphasis was on optimizing two crucial vehicular attributes: velocity and steering angle decisions. The SMART framework, by assimilating these parameters, aimed for a more dynamic and adaptive autonomous driving experience.

Notably, the SMART framework also caught the attention of [Xia et al. \(2023\)](#), who advanced this approach further. Expanding on the foundational principles set by Yang, Xia's version of the SMART framework used RL, specifically focusing on velocity and steering angle decisions, thus emphasizing the increasing recognition and potential of the SMART approach in the automotive domain. Overtaking, a complex vehicular maneuver filled with dynamics and potential hazards, became the focal point of research for [Zhang et al. \(2023a\)](#). In their work, a deep RL model was developed specifically for these overtaking scenarios in autonomous driving. By understanding the myriad challenges posed by overtaking, Zhang's model stood out as a pioneering solution to a long-standing vehicular challenge.

Another contribution came from [Cui et al. \(2023\)](#), who started the process of combining various aspects of decision-making for autonomous driving. Their proposition was a comprehensive model built on the tenets of deep RL. Examining specifics, they used the Deep Q Network (DQN) and its subsequent variants to create an integrated solution, catering to a spectrum of driving scenarios and challenges.

5. **Particle Filters** : also known as Sequential Monte Carlo (SMC) methods, have emerged as an influential tool in the domain of nonlinear and non-Gaussian estimation problems ([Del Moral, 1997](#)). Originating from the Monte Carlo methods, particle filters employ a set of random samples or '*particles*' to represent the posterior distribution of some stochastic process.

In conceptualizing particle filters, the process starts with the initial phase, known as *Initialization*. Here, particles are typically generated from a known distribution or are spread evenly across the state space. The objective is to cover the space adequately and anticipate where relevant observations might appear. After initialization, each particle begins by receiving a weight in the *Weighting* phase. This weight is directly associated with the probability that the observed data aligns with the particle's state. As the system progresses over time, the particles are selected in the *Resampling* phase. Those with higher weights continue, sometimes reproducing to replace those with lower weights, while those with very low weights are removed. Finally, in the *Propagation* phase, the particles move through the state space in a way determined by the system's dynamics ([Djuric et al., 2003](#)).

The field of autonomous systems has seen significant advancements in the application and optimization of particle filters, evidenced by various contributions in the literature. [Wu and Li \(2020\)](#) outlined the parameter estimation complexities of particle filters as applied to decision-making in autonomous driving, while works by [Song et al. \(2022\)](#) studies the topic from a quantum decision theory perspective, testing it against the Cumulative Prospect Theory model. In the domain of localization, an indispensable function in autonomous navigation, [Jonchery et al. \(2021\)](#) employed a particle filter approach using diverse landmark types and various sensor data fusion. Moreover, the relevance of particle filters in automotive applications goes beyond just theoretical frameworks; it extends to practical real-world applications, as reviewed by [Berntorp and Di Cairano \(2019\)](#). Notably, while [Iyer et al. \(2021\)](#) and [Katwe et al. \(2021\)](#) focuses on using particle filters for accurate localization of AVs, especially in GPS-compromised environments, [Yang et al. \(2022\)](#) pivot to proposing an adaptive self-driving tracking algorithm based on particle filter techniques.

The scope of research also covers freeway driving scenarios, wherein [Guan et al. \(2022\)](#) introduced a discrete decision-making strategy to improve efficiency and safety, leveraging the discrete Soft Actor-Critic (CAC) with a sample filter algorithm. Finally, in the context of autonomous robotics, [Ueda and Arai \(2007\)](#) use a particle filter combined with a real-time Quantitative Markov Decision Process (Q-MDP) value method, for refined state estimation and decision-making processes.

In the field of autonomous decision-making, many computational methods have been developed to assist systems in making informed decisions, especially in situations marked by uncertainty and constant changes. As outlined in Table 2.2, these methods are established to specific applications, making them effective for distinct scenarios. For instance, Markov Decision Processes (MDPs) are effective in environments with predictable state transitions, allowing for the creation of optimal strategies and policies for such conditions. On the other hand, Partially Observable MDPs (POMDPs) are designed for situations where complete state visibility is rare. Reinforcement Learning (RL) is notable for its flexibility, permitting agents to learn through interaction without a pre-established model. The Kalman Filter (KF) and Particle Filter (PF) focus on state estimation, with KF being preferred for its accuracy in environments with Gaussian noise, and PF valued for its ability to manage non-linear dynamics and non-Gaussian noises. Together, these methods represent a comprehensive set of tools for researchers and practitioners in the field of autonomous decision-making, with each technique offering distinct advantages for different scenarios and challenges.

Table 2.2: Summary of non-deterministic methods for Autonomous decision-making

Method	Advantages	Limitations	Use in autonomous decision-making
MDPs	Captures dynamic decision-making in stochastic environments. Facilitates derivation of optimal and safe policies. Versatile and adaptable.	Assumes the Markovian property: future state depends only on present state and action. May not always hold in real-world scenarios.	Modeling driving scenarios. Deriving driving policies. Intention-aware decision-making.
POMDPs	Handles partial observability and uncertainty. Extends MDPs to include observations. Offers a more realistic representation of many real-world problems.	Computationally intensive due to belief state space. Difficult to find optimal solutions for large problems.	Situations where the system state isn't fully observable. Decision-making under uncertainty.
RL	Enables learning optimal strategies from interaction. Doesn't require a model of the environment. Highly adaptive to changing environments.	Requires a lot of data/samples for learning. Exploration vs. exploitation trade-off.	Learning driving strategies. Adapting to changing driving environments.
KFs	Provides accurate state estimations in noisy environments. Merges predictions with measurements. Handles uncertainties from dynamics and noise.	Assumes system noise and measurement noise are Gaussian. Not suitable for non-linear systems without extensions.	State estimation in autonomous driving. Object tracking and detection. Sensor fusion.
PFs	Can handle non-linear systems and non-Gaussian noise. More flexible than KF. Can represent multi-modal beliefs.	Requires a large number of particles for accuracy. Resampling can introduce degeneracy issues.	Non-linear state estimation. Handling non-Gaussian uncertainties. Complex environments with multi-modal uncertainty.

2.6 Risk assessment in railways

2.6.1 Risk assessment for conventional trains

Safety and risk assessment in railways is an essential process, ensuring that the operations and functionalities of rail systems remain within an acceptable safety level. This assessment is ongoing processes that extend from the design stage to the operational stages, ensuring that the railways function efficiently without compromising safety.

The EN 50126, EN 50128 and EN 50129 series of European standards, for instance, provides a framework for the specification and demonstration of the Reliability, Availability, Maintainability, and Safety (RAMS) of railway systems. The suite, comprising various parts, delineates the requirements concerning system safety (EN 50126), software (EN 50128), and safety cases (EN 50129) aspects. Among the basic principles behind these standards is to enable a structured approach to risk assessment, making sure that potential hazards are identified, assessed, and appropriately addressed throughout the entire life-cycle of the rail system.

Parallely, and in line with these standards, the Common Safety Method for Risk Evaluation and Assessment (CSM-RA), mandated by the European Union, provides a harmonized approach to evaluating and ensuring the safety of rail operations. This method underscores the significance of a rigorous risk assessment and management process, promoting a holistic view of the rail system. It necessitates the identification of all possible hazards and the quantification of associated risks. Subsequently, the CSM-RA establishes requirements for demonstrating that these risks have been reduced to a tolerable level, given the current state of the art and societal expectations.

Throughout the life-cycle (see Figure 2.5), these standards and methodologies advocate for:

1. **Hazard identification** : recognizing potential sources of harm or adverse events;
2. **Risk analysis**: estimating the risks, considering the frequency and severity of potential consequences;
3. **Risk evaluation**: (which include risk analysis) determining whether the identified risks are acceptable or require mitigation;
4. **Risk mitigation**: implementing measures to eliminate or reduce the severity and likelihood of the risk to an acceptable level;
5. **Continuous monitoring and review**: to ensure that risk controls remain effective during operations and in light of new information or changing conditions.

Fundamentally, the risk assessment of traditional railway systems is an iterative and thorough process that requires the systematic application of standardized methods throughout the life-cycle of the system. These methods are driven by a focus on safety, while taking into account the complex connections between different elements of the rail system.

As railway technology advances towards increased autonomy (GoA4), the approaches of risk assessment and safety must be adapted to address the complexities of autonomous functions. This includes the integration of comprehensive algorithms, sophisticated sensors, and ML components. The following section represents the challenges and methods for ensuring the safety and risk mitigation of autonomous trains.

2.6.2 Risk assessment methods

Risk assessment is an essential process in any safety-critical domain, providing a systematic approach to identify, evaluate, and mitigate potential hazards. Over the years, many

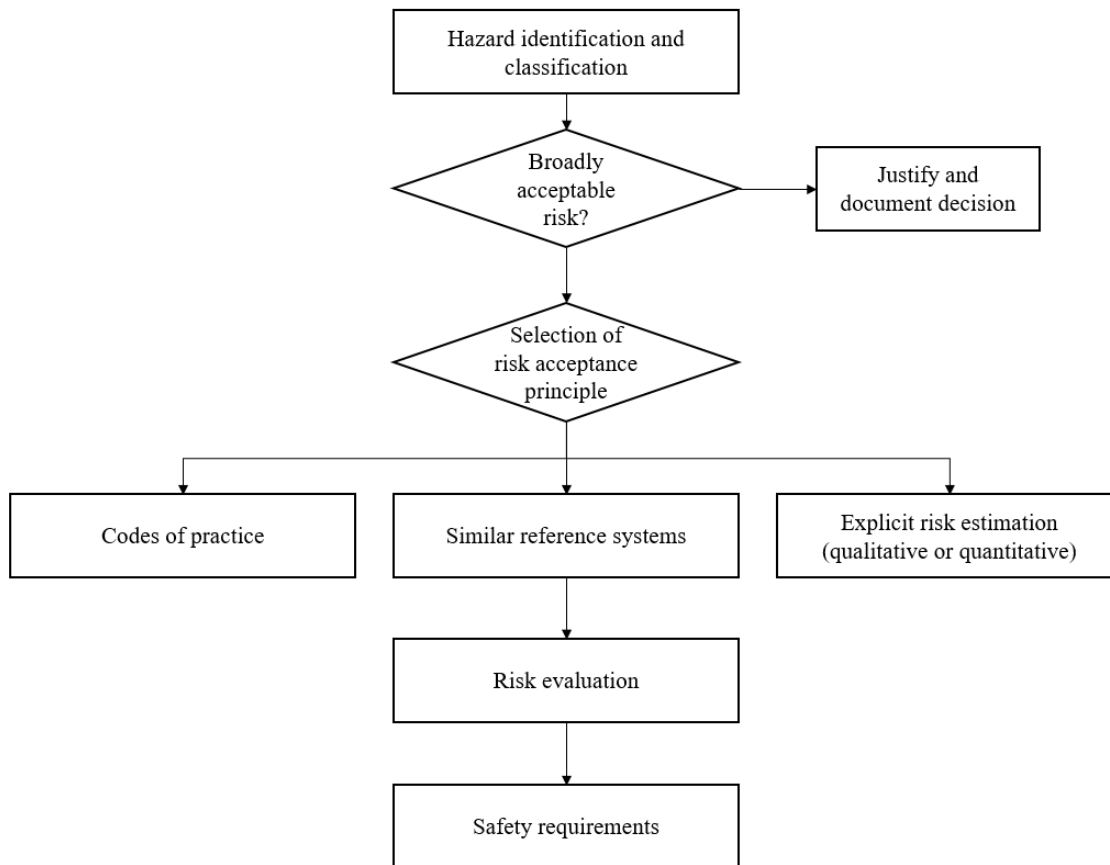


Figure 2.5: Process of risk assessment related to phases 3 and 4 of the life-cycle ([EN-50126, 2017](#))

different methods have been developed to adapt to the changing challenges of different industries, including railways. These methods, when applied properly, not only ensure operational safety but also ensure system efficiency and reliability. Selecting an appropriate assessment technique often revolves around the specifics of the system in focus. Here, we categorize prominent risk assessment methods into static and dynamic approaches:

1. Static approaches

- **Fault Tree Analysis (FTA):** using a top-down approach, FTA provides a graphical representation of events leading to a failure, enabling a clear identification of system vulnerabilities ([Vesely et al., 1981](#));
- **Event Tree Analysis (ETA):** as a complementary bottom-up method, ETA starts with an initial event and traces its potential subsequent outcomes, aiding in the visualization of the repercussions of a system disruption ([Kenarangui, 1991](#));
- **Hazard and Operability Study (HAZOP):** a systematic tool that delves into existing or proposed processes to spot and evaluate potential risks, ensuring that industries remain cognizant of and equipped to tackle threats ([Dunjó et al., 2010](#));
- **Failure Mode and Effects Analysis (FMEA):** a systematic method used to identify potential failure modes within a system, product, or process, FMEA also evaluates the associated consequences of these failures. By ranking the severity of the effects and their likelihood of occurrence, FMEA assists engineers

and designers in prioritizing the most critical failures so that mitigation actions can be implemented (Gilchrist, 1993).

2. Dynamic approaches

- **Monte Carlo simulation:** using randomness to solve problems that might be deterministic in principle, this method is widely adopted in risk analysis to model the probability of different outcomes in processes that cannot easily be predicted due to the intervention of random variables (Mooney, 1997);
- **Bayesian Networks:** this graphical model represents variables and their conditional dependencies. It is a flexible tool for modeling complex dependencies in risk assessment, and especially effective when dealing with uncertainty and when updating risks based on new information Friedman et al. (1997);
- **Markov chains:** mathematical models employed to represent systems that transition from one state to another over time. Characterized by the “memoryless” Markov Property, the probability of the system transitioning to a subsequent state is dependent only on its current state, not on the sequence that preceded it. In the context of risk assessment, Markov Chains are used to model and predict system behaviors, particularly in scenarios with probabilistic transitions and uncertainties (Norris, 1998).

While traditional risk assessment methods, both deterministic and probabilistic, have provided essential knowledge into the potential risks and failures of railway systems, the emergence of autonomous trains and real-time operations necessitates a shift in perspective. In fact, static risk models, which mainly depend on predefined scenarios and historical data, may not fully understand the changing environment and dynamic interactions. Consequently, as railway systems advance technologically, and as the demand for immediate responsiveness to unexpected situations increases, the need for dynamic risk assessment becomes crucial. Such assessment consider real-time data and situational changes, ensuring that safety evaluations remain relevant and adaptive to current conditions.

2.6.3 Dynamic risk assessment for the autonomous train

In the context of railways, especially with the introduction of autonomous trains, DRA becomes particularly essential. Here, the system would need to assess risks continuously, updating its operational strategy based on factors like track conditions, weather conditions, equipment’s state of health, obstacles, and other trains’ presence and movements. Such a continuous and adaptive approach ensures that the system is always operating within respect to safety requirements, and any potential risks are identified and addressed appropriately (Park, 2014).

In railway operations, risk assessment is essential, comprising various methodologies and approaches. For instance, deterministic risk models use specific parameters and scenarios to forecast outcomes, focusing on consistency. On the other hand, probabilistic risk models consider the uncertainties and variations that naturally occur in railway operations. Among these models, evaluating the risk of collisions is particularly important, underlining the need to prevent collisions in railway systems. The following discussions explore these methods, highlighting their characteristics, uses, and the challenges they introduce.

In order to adequately address these complexities, a focused approach to safety and risk assessment is necessary. This method goes further than traditional frameworks to address the unique challenges of autonomous operations.

1. **Perception:** At the core of autonomous train operation lies the crucial process of perception, which is the system’s ability to accurately detect, recognize, and interpret its environment. This includes identifying obstacles, comprehensive track conditions, and even interpreting signals (Wall et al., 2014). Ensuring the reliability of perception systems, which often rely on a fusion of sensors like LIDAR, radar, and cameras, becomes a cornerstone of safety assessment;
2. **Dynamic environment interaction:** Autonomous trains continuously engage with dynamic environments, utilizing sophisticated sensor systems and decision-making algorithms (Quaglietta et al., 2013). The constant validation and monitoring of these algorithms and sensors ensure that emerging risks are promptly identified and addressed;
3. **Advanced computational elements:** Leveraging contemporary computational methodologies, such as machine learning and artificial intelligence, autonomous trains can make decisions in real-time (Tazoniero et al., 2007). Validating these non-deterministic algorithms’ safety and reliability is imperative, requiring innovative techniques that factor in the probabilistic nature of their outputs;
4. **Risk-based approaches:** Due to the dynamic nature of autonomous operations, conventional deterministic safety assessments may have limitations. Probabilistic risk-based approaches that concentrate on recognizing potential hazardous scenarios while assessing their likelihood and severity provide a more comprehensive perspective (Johnsen et al., 2018). This transition from a strictly reactive standpoint to a proactive, scenario-driven evaluation amplifies the system’s capacity to foresee and mitigate potential safety issues;
5. **Decision-making:** Integral to autonomous train operations is the decision-making process—how the system reacts to perceived inputs. The transparency and predictability of this process are paramount (Stopka et al., 2020). Ensuring that decision-making pathways are robust, logical, and free from unforeseen biases or errors becomes a core component of safety assessments.

Existing standards and regulations, such as the EN5012x series and the CSM-RA, may require enhancements and improvements or supplementary directives to fully address the nuances of autonomous systems and ensure their holistic risk evaluation remains relevant.

In conclusion, safety and risk assessment for autonomous trains demand sophisticated strategies that integrate traditional principles with the unique challenges of autonomy. This ensures the continued effectiveness of safety requirements in the context of autonomous transportation.

2.7 Safety assurance of the autonomous train

Ensuring adherence to railway safety standards requires the collation of evidence supporting established safety requirements and objectives (Council et al., 2007). While standards provide guidelines for demonstrating safety, their practical application is critical because their descriptive nature leaves gaps for various interpretations (Nair et al., 2014). Typically, a *Safety Case* includes evidence and arguments demonstrating the safety of a system in specific operational contexts. However, as system complexity increases, establishing coherent and credible safety arguments becomes challenging for designers and developers.

In the next sections, we examine the importance of safety cases in ensuring safety and the argumentation process in railway systems. Then, we review various graphical argumentation methods used in the literature.

2.7.1 Safety cases

Safety cases have played an increasingly critical role in the railway industry's evolution towards a safer and more reliable mode of transportation (Leveson, 2011). These structured arguments have provided the approach for assessing, documenting, and ensuring safety measures are in place, significantly reducing the risk of accidents and incidents (Graydon, 2013). The history of safety cases in railways is marked by milestones that reflect the industry's commitment to safety assurance. In the pre-modern era of the rail domain, safety was a substantial concern. However, formal safety cases, as they are known today, did not exist. Safety measures were often developed in response to specific accidents and incidents (Rolt). For example, the boiler explosions that plagued early steam locomotives led to the creation of the Boiler Explosions Act of 1882 in the United Kingdom (Bartrip, 1980), which mandated boiler inspections and maintenance (Parliament of the United Kingdom, 1882). The end of World War II marked a significant turning point for railway safety (Divall, 2016). The increasing complexity of railway systems, the advent of electric and diesel locomotives, and the growing speed and density of rail traffic necessitated more formalized safety approaches. It was during this period that the British Railways Board¹ (BRB) introduced Safety Management Systems (SMS) to improve safety and reduce risks (U.K. Health and Safety Executive, 1999).

The introduction of safety cases in the railway industry can be traced back to the 1970s. The Haddon-Cave Report of 1971 (Haddon-Cave, 1971) is often cited as one of the earliest safety cases in railways, primarily focusing on safety analysis for rail equipment. The concept of the 'safety case' gained attention in the United Kingdom in the 1990s, thanks to the Health and Safety Executive² (HSE) and the U.K. Railway Inspectorate³. The idea of structured safety arguments, formally documented in safety cases, began to take hold.

The railway industry recognized the importance of formalized safety arguments and began standardizing safety practices. The introduction of ISO 9000 (ISO-9000, 2015) and IEC 61508 (IEC-61508, 2010) standards set the stage for international safety management practices in the industry. The systematic approach to safety, grounded in formal safety cases and safety arguments, became a cornerstone of railway safety assurance.

In recent years, the railway domain has seen a significant shift towards the development of safety cases that comply with international standards, such as CENELEC EN 50126 (EN-50126, 2017), EN 50129 (EN-50129, 2018), and ISO 15026 (ISO-15026, 2020). These standards require railways to develop comprehensive safety cases that demonstrate the achievement of safety goals and compliance with safety requirements. Modern railway safety management systems are increasingly dependent on structured safety argumentation to ensure the safety and reliability of railway systems, especially with the introduction of autonomous train operations and innovative technologies.

This brief history reflects the ongoing transformation of the railway industry in its relentless pursuit of safety and reliability. Today, structured safety argumentation, in combination with international safety standards, is instrumental in meeting the challenges of modern railway systems.

In examining textual safety argumentation within railway systems, it becomes clear that the complexities of autonomous trains often surpass the capabilities of textual representation alone. This highlights the need for a more comprehensive and graphical approach to manage the complexity of safety cases. Consequently, the focus now is shifted towards graphical safety argumentation, which uses graphical tools and techniques to complement textual argumentation.

¹<https://discovery.nationalarchives.gov.uk>

²<https://www.hse.gov.uk/>

³<https://www.orr.gov.uk>

2.7.2 Graphical safety argumentation

As mentioned before, safety argumentation plays a pivotal role in ensuring safe operation of complex systems, particularly in safety-critical industries such as railways. Here, we provide a comprehensive overview of safety argumentation frameworks, with a specific focus on their applications in the railway industry. These frameworks assist in systematically developing structured safety arguments, which are essential for demonstrating the satisfaction of safety goals and requirements, thereby ensuring safe railway operations.

1. **Goal Structuring Notation (GSN)**: a graphical notation, is a cornerstone of structured safety argumentation in railways ([Kelly and Weaver, 2004](#)). This framework provides a systematic approach for structuring safety arguments, arranging them hierarchically with goals, sub-goals, strategies, and evidence. The GSN methodology simplifies complex arguments by breaking them down into comprehensible components, facilitating communication and understanding within the railway safety domain;
2. **Claims-Argument-Evidence (CAE)**: The CAE framework focuses on creating structured arguments with clear connections between claims and supporting evidence. While it may not provide the same hierarchical structure as GSN, CAE simplifies argument structure by emphasizing the relationship between claims and their justifications. This makes it suitable for applications that require clarity and direct links between evidence and safety claims ([Bishop and Bloomfield, 2000a](#));
3. **Structured Assurance Case Metamodel (SACM)**: SACM, an Object Management Group standard, is an increasingly adopted framework that defines a metamodel for structuring safety cases and assurance arguments ([Wei et al., 2019](#)). Its primary advantage lies in the standardization of information exchange within safety-critical railway systems. SACM helps in maintaining a systematic approach to developing and sharing safety cases across different stakeholders;
4. **Knowledge Acquisition in Automated Specification (KAOS)**: is a graphical method used for goal-oriented requirement engineering, often applied in safety-critical systems to model system goals, agents, and their relationships. It revolves around modeling requirements as goals and agent commitments, where goals represent desired system behavior. These goals can be organized hierarchically, providing a structured way to represent system requirements. Agents and actors are introduced to assign responsibilities in the system, while obstacles represent conditions hindering goal achievement. The refinement process breaks down high-level goals into subgoals, helping create a goal hierarchy for thorough requirement analysis ([Gruber, 1990](#)).
5. **Safety Specification Graph (SSG)** : is a graphical representation used for safety analysis in complex systems, particularly in safety assurance and risk assessment ([Boulinier et al., 2004](#)). It is structured as a directed graph where nodes represent components, requirements, and constraints. Edges between nodes signify various safety-related relationships, such as dependencies, hierarchies, and constraints. SSGs play an important role in visualizing and specifying safety requirements, facilitating hazard analysis and risk assessment. They include traceability links for connecting safety requirements to design elements, ensuring a systematic and transparent approach to safety assurance.

Table 2.3 provides a comprehensive overview of each graphical safety argumentation method along with their advantages, limitations and real-life applications in different industries :

Table 2.3: Graphical safety argumentation methods

Method	Advantages	Limitations	Applications
SSG	Clear representation and systematic hazard identification.	Complexity for simple systems; might not capture all aspects.	<ul style="list-style-type: none"> • Aerospace • Automotive
CAE	Clear distinction and easy communication.	Over-simplification for complex systems; blurred distinctions.	<ul style="list-style-type: none"> • Nuclear • Railway • Aerospace
KAOS	Detailed representation and traceability.	Modeling complexity; might be excessive for small projects.	<ul style="list-style-type: none"> • Software engineering for complex projects.
GSN	Intuitive visual representation; widely adopted.	Requires training; challenging for large-scale systems.	<ul style="list-style-type: none"> • Automotive • Aerospace • Nuclear • Railway
SACM	Standardization of information exchange and systematic safety case development.	May require extensive training and familiarity with the metamodel.	<ul style="list-style-type: none"> • Safety-critical railway systems

In the domain of safety argumentation, GSN presents an array of features that differentiate it from other graphical methods. At its core, GSN is structured to delineate safety goals, the strategies to achieve them, and the evidence that underscores their realization (Kelly, 1999b). Unlike the SSG, which is narrowly centered on specification-based safety criteria (Heimdahl and Leveson, 1996), or KAOS, anchored primarily in knowledge acquisition protocols (Van Lamsweerde, 2001), GSN offers a more expansive canvas, emphasizing the logical articulation of safety arguments.

When compared with the Claims Argument Evidence framework, Goal Structuring Notation demonstrates a more dynamic approach (Bishop and Bloomfield, 2000b). The Claims Argument Evidence framework follows a direct route from formulating claims to collecting evidence. In contrast, Goal Structuring Notation allows for the representation of various strategies aimed at achieving a single goal (Kelly and Weaver, 2004). This adaptability is important, especially in situations where safety concerns are complex and require a comprehensive perspective.

The industry's shift towards Goal Structuring Notation is highlighted by its practical use, shown through its adoption in various fields, from aerospace engineering to automotive systems (Hawkins et al., 2011). Additionally, the specialized toolsets developed for Goal Structuring Notation improve its effectiveness in practice, establishing it as a strong option for creating safety cases (Denney et al., 2019). Essentially, while each graphical safety argumentation approach offers unique benefits, Goal Structuring Notation's structured design and proven effectiveness emphasize its leading role in safety argumentation.

Considering the methodological strength and empirical evidence supporting Goal Structuring Notation, it is important to assess its practical application in real-world situations. The effectiveness of a safety argumentation framework depends not only on its theoretical basis but also on its usability in different operational environments. Therefore, the upcoming section will present real-life examples and case studies illustrating how Goal Structuring Notation is used in various industries. This will highlight its practical relevance and versatility. Furthermore, the methodology will be elaborated upon in Chapter 3.

2.7.3 Goal Structuring Notation (GSN) application examples

Since the introduction of GSN at York University, several safety-critical industries have begun to use it and explore it, particularly defense, transport, nuclear, and medical devices. Earlier industrial use of GSN was limited to trial and pilot projects and then, the use has been broadened to concrete industrial use. Hereafter, we provide a non-exhaustive list of some relevant uses:

- QinetiQ and BAE Systems have collaborated with York University to extend GSN to support the management of "modular" and compositional safety cases to support the cost-effective certification of modular avionics systems Bate and Kelly (2003);
- GSN is considered to be used for representing safety arguments within safety cases for the European Air Traffic Management Eurocontrol;
- U.K. NATS (National Air Transportation System) have used GSN for representing their safety management system (as mentioned in Leveson's white paper).
- GSN was chosen as a method for representing arguments and evidence in the Preliminary Safety Case development process for the flight control system of a helicopter with a fly-by-wire system. This application for Western Helicopters Ltd⁴ was made in the frame of HEAT/ACT Chinneck et al. (2004) project ⁵;
- GSN was used in the safety case of Nimrod MR2 XV230 Aircraft. A review report⁶ mentioned that GSN was used as technique allowing to structure and present complex safety arguments in order to demonstrate the fulfillment of main safety goals supported by strategies and evidence.
- Safety case patterns have been generated using GSN in order to help clinicians study problems in e-therapy for children Attention Deficit Hyperactivity Disorder (ADHD). The project was a contribution between Ge et al. (2012) and the Hospital of St. Andrew, Portugal.
- Extensions of GSN were used to generate safety case patterns and to build high-level safety arguments for a Patient Controlled Analgesic (PCA) infusion pump⁷ Ayoub et al. (2012),Feng et al. (2014);

⁴<https://westernhelicopters.com/>

⁵<https://www.icc.illinois.gov>

⁶https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/229037/1025.pdf

⁷<https://www.ncbi.nlm.nih.gov/books/NBK551610/>

- The GSN was cited by the international automotive safety standard ISO 26262 published in 2012. Additionally, a work of the members of the Motor Industry Software Reliability Association (MISRA) shows the use of GSN in automotive safety arguments in compliance with the ISO 26262 standard [Birch et al. \(2013\)](#). An industrial case study based on a typical electric vehicle architecture was presented as well in this work (specific details have been abstracted from the paper for reasons of commercial sensitivity);
- The International Rail Industry Engineering Safety Management Handbook [INTESM \(2013\)](#) mentioned that GSN is an useful technique for structuring and illustrating safety cases;
- NASA developed GSN-based toolset called AdvoCATE⁸ to help assurance cases development [Denney et al. \(2012\)](#);
- Recently, Aurora⁹ - an American self-driving vehicle technology company, has made public its GSN-based safety case framework to assure that Aurora's self-driving vehicles are acceptably safe to operate on public roads (see [https:// safetycaseframework.aurora.tech/gsn](https://safetycaseframework.aurora.tech/gsn)). The framework combines guidance from United States government organizations, practices from safety-critical industries, voluntary industry standards and consortia, and academic research.

To sum up, an interesting work published by REF 2014¹⁰ (*Research Excellence Framework 2014 - Impact case studies*) has gathered references and resources showing the impact and use of GSN in safety-critical industrial domains.

Within the scope of autonomous train operation, the safety assurance and argumentation frameworks act as essential components for validating predefined safety requirements. However, it is essential to acknowledge that inherent risks persist regardless of the system's technical sophistication. These risks, primarily due to operational uncertainties, variations in system behaviors, or unpredictable external influences, necessitate a methodical approach towards their recognition, quantification, and subsequent mitigation.

2.8 Conclusion

In conclusion, our review of the latest developments in autonomous vehicles and autonomous trains has identified various important aspects. We started by reviewing the key components of the autonomous driving systems for autonomous transportation systems. Furthermore, we focused on the process of risk assessment for conventional vehicles (with driver) and dynamic risk assessment for autonomous vehicles. Moreover, we review the use of risk models and their important role in ensuring safety for autonomous vehicles. Subsequently, we examined the decision-making process for autonomous vehicles, by investigating both deterministic and non-deterministic approaches. This discussion extended to railway systems, comparing risk assessment methods for conventional and autonomous trains. Finally, we provide an examination of the safety assurance process for autonomous trains, starting with safety cases, the use of graphical safety argumentation methods, and some examples of some GSN(Goal Structuring Notation)-based applications.

This detailed examination of recent advancements underscores key challenges in autonomous train development, emphasizing the essential need for a rigorous safety assurance

⁸<https://www.nasa.gov/collection-asset>

⁹www.aurora.tech

¹⁰<https://impact.ref.ac.uk/casestudies/CaseStudy.aspx?Id=43445>

process, thorough safety argumentation, risk assessment, and robust decision-making. Collaboratively, these aspects establish the basis for the safe and efficient autonomous train operations.

Chapter 3

Graphical argumentation using GSN for autonomous trains

Contents

3.1	Introduction	51
3.2	GSN for graphical safety argumentation	51
3.2.1	Key concepts	51
3.2.2	Development processes	52
3.2.3	Issues resolved with GSN	54
3.3	GSN-based safety cases in transportation systems	55
3.3.1	GSN in the automotive domain	55
3.3.2	GSN in the aviation domain	56
3.3.3	GSN in the railway domain	56
3.3.4	Toward using GSN for autonomous systems	58
3.4	Safety assurance approach of autonomous train	58
3.4.1	Overall system level	59
3.4.2	AI-based component level	60
3.4.3	AI techniques level	60
3.5	Use case	63
3.5.1	Anti-collision function	63
3.5.2	Discussion	65
3.6	Conclusion	68

3.1 Introduction

As explained and discussed in Chapter 2, the Goal Structuring Notation (GSN) presents several of advantages, such as interconnected sources of evidence and pragmatic applicability in different domains.

In Chapter 3, we focus on using GSN for building credible and efficient safety argumentation for autonomous train safety demonstration. We chose GSN for this purpose due to its graphical notation to present safety arguments, growing interest in applications in different domains, and the ability to efficiently and clearly link evidence and arguments. Therefore, this method represents the first pillar of the framework that we propose for the safety assurance of the autonomous train. As will be explained during this chapter, it makes it possible to resolve issues related to traceability, complexity, and efficiency of evidence and arguments in safety cases.

Notice that the main results of this chapter have been published in the Reliability Engineering & Systems Safety (RESS) journal (Chelouati et al., 2023a).

The remainder of this chapter is structured as follows. In Section 3.2, we discuss the key concepts and development processes of GSN, along with issues that can be resolved using GSN. In Section 3.3, we review the use of GSN-based safety cases in the transportation domains (automotive, aviation, and railways) and then we discuss its investigation for autonomous systems and vehicles. In section 3.4, we address the opportunities and challenges of using the GSN approach for presenting and structuring safety cases for autonomous trains. Furthermore, we propose a graphical safety argumentation using GSN. Additionally, in Section 3.5, we present the creation of GSN structures for a safety function: the anti-collision function. Finally, we provide some concluded remarks and point out some future research directions in Section 3.6.

3.2 GSN for graphical safety argumentation

In this section, we present the safety argumentation through the GSN framework, especially the two options for developing the arguments.

3.2.1 Key concepts

A safety argument is a logical representation of a set of safety claims, goals, assumptions, justifications, and evidence. Contrarily to the textual representation, such as Trust case¹ (Cyra and Górski, 2007; Falessi et al., 2011), structured HTML (Brown, 1998) or free-text, graphical representations are a suitable tool to capture these elements in a graphical notation and provide a clear representation of complex arguments with their supporting evidence. Indeed, the graphic notations have been designed to facilitate the description of assurance cases in a manner that is easy for humans to understand and for machines to manipulate (Armstrong and Paynter, 2004; Graydon et al., 2007). As presented in Chapter 2, the main graphical approaches are GSN, CAE, SACM, KAOS, and SSG.

The use of GSN for safety argumentation has received a lot of attention in both industry and academia during the last two decades. Indeed, GSN has been adopted by a growing number of companies within safety-critical industries and is now recommended by many safety standards (GSN-WG, 2021). According to the Goal Structuring Notation Community Standard, the GSN is given as “*a graphical argument notation that may be used to formally describe the contents and structure of any argument, as well as the link*

¹Trust case evokes a structured textual form of safety claims, arguments, and evidence, presented as assumptions with references to documentation.

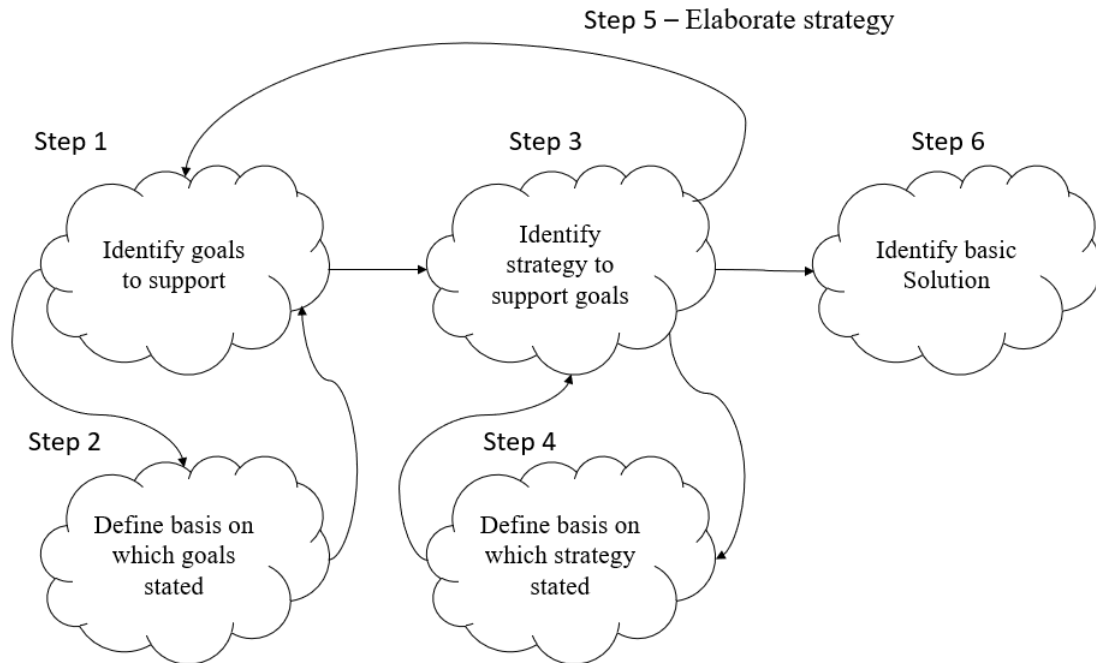


Figure 3.1: Six-step process for top-down developing goal structure

between the argument and evidence”. Furthermore, the GSN explicitly represents the individual elements of any safety argument (safety objectives or claims, evidence, and context) and the relationships that exist between these elements.

The key elements of the GSN are (1) the assurance *goal* or the claim, (2) the *evidence* that the goal has been satisfied, and (3) an *argument* linking the evidence to the goal in a way that leads one to believe that the evidence justifies the goal. When these elements of GSN are linked together in a network they are described as a ‘*goal structure*’. This basic structure is applied recursively to produce, for systems, a hierarchic structure with the overall goal for the system at the root (Graydon et al., 2007). Other (graphical) elements that can be used in GSN are strategies, assumptions, justifications, and context. The principal symbols of the notation are shown in Table 3.1.

3.2.2 Development processes

In GSN, goal structures are commonly implemented in a top-down manner (i.e., each claim is decomposed into sub-claims and so on). According to GSN Community Standard, the development of the arguments can be elaborated even in the bottom-up process (GSN-WG, 2021).

1. **The top-down development of goal structures:** To establish a top-down GSN structure, six steps should be followed (GSN-WG, 2021). The recursive process starts with the identification of a claim (step 1) and an explicit statement of the context in which the claim is valid (step 2). Then, a strategy to support it is identified (step 3) and justified (step 4). In particular cases, a claim needs to be supported immediately through reference to the associated evidence (step 6). Moreover, it is common to identify sub-claims to refine the argument to the needed level of detail at which the evidence used to argue the claim is considered sufficient (step 5). This process recursive process is illustrated in Figure 3.1.
2. **The bottom-up development of goal structures:** In some cases, it is useful or

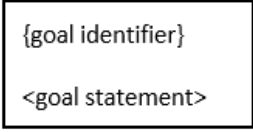

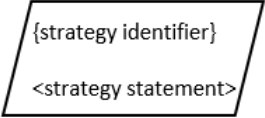
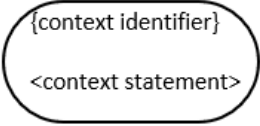
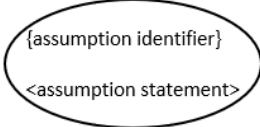
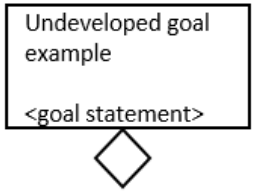
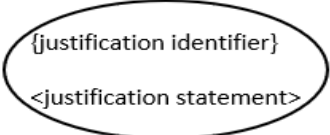
GSN elements	Definition
	A <i>goal</i> , rendered as a rectangle, presents a claim forming part of the argument.
	A <i>solution</i> , rendered as a circle, presents a reference to an evidence item.
	A <i>strategy</i> , rendered as a parallelogram, describes the inference that exists between a goal and its supporting goal(s).
	A <i>context</i> , rendered as shown left, presents a contextual artifact. This can be a reference to contextual information, or a statement.
	An <i>assumption</i> , rendered as an oval with the letter 'A' at the top- or bottom-right, presents an intentionally unsubstantiated statement.
	<i>Undeveloped element decorator</i> , rendered as a hollow diamond applied to the bottom center of an element, indicates that a line of argument has not been developed. It can apply to goals and strategies. For example, an undeveloped goal, rendered as a rectangle with the hollow-diamond undeveloped element decorator at the center-bottom, presents a claim that is intentionally left undeveloped in the argument.
	A <i>justification</i> , rendered as an oval with the letter 'J' at the top- or bottom-right, presents a statement of rationale.

Table 3.1: The main GSN elements. (GSN-WG, 2021)

necessary to establish a GSN bottom-up argument, beginning with the available evidence. In line with Kelly's prescribed methodology for top-down Goal Structuring Notation (GSN) development (Kelly, 1999a), the process of building a goal structure with a bottom-up process can be articulated as follows: Firstly, the identification of pertinent evidence to be presented as GSN solutions is necessary. Subsequently, '*evidence assertion*' claims are inferred to serve as direct underpinnings for these solutions, with each one serving as distinct GSN goals. Moving up the hierarchy,

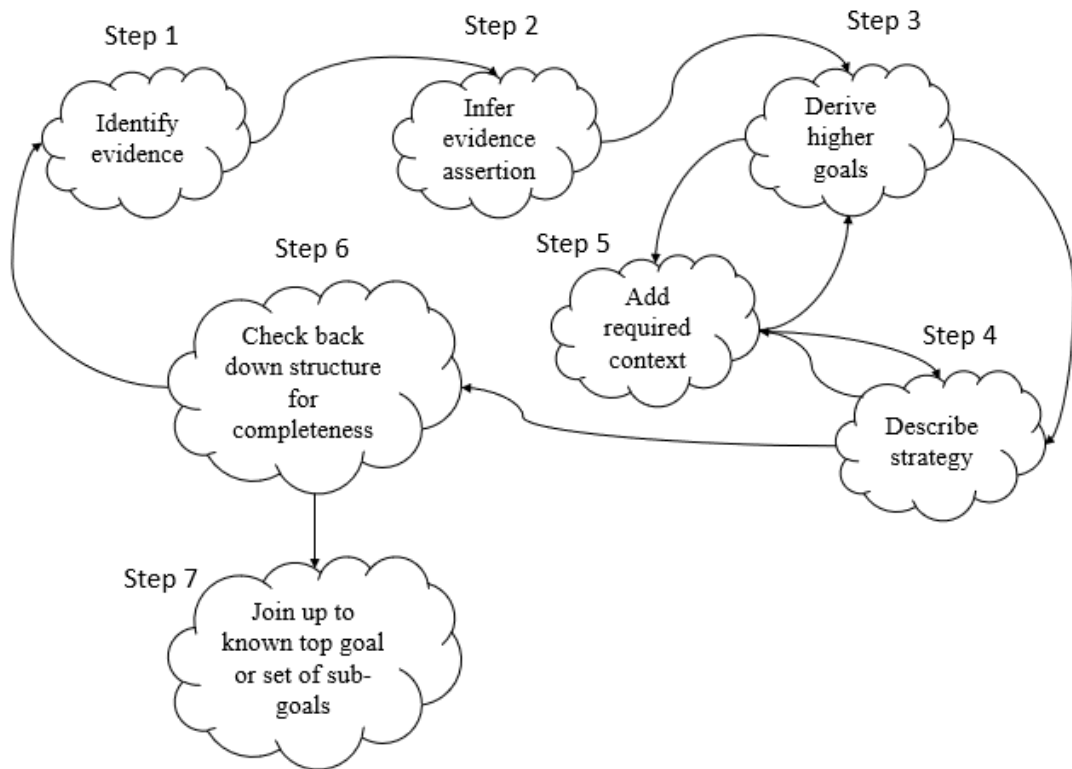


Figure 3.2: Bottom-up process for developing goal structure

higher-level goals are then derived, those which are substantiated by the initial evidence assertions. Furthermore, elucidation of how each stratum of goals aligns with their parent goal, essentially delineating the strategic relationship, is imperative. This is further underscored by the necessity to ensure the inclusion of all pertinent contextual information. An additional critical step involves comprehensive verification, by descending through the structure to ascertain completeness. Finally, the resulting goal structure is to be conjoined with an established top-level goal or an array of goals, thereby creating a coherent, hierarchically structured GSN framework. The seven steps of the recursive process are illustrated in Figure 3.2.

3.2.3 Issues resolved with GSN

A main advantage of the GSN is its contribution to the enhancement of traceability. This is particularly due to: (i) its directed/oriented structure (goal, argument, evidence) which allows to graphically explore the branches in up-down and bottom-up ways, (ii) the decomposition and refinement of high-level objectives until they meet the low-level evidence, and (iii) the modularity of the graphs, which offers several abilities, such as, partitioning the goal structures and having historical vision on the development of arguments.

The simplification of complexity through GSN is achieved by providing a clear, graphical representation that shows the direct relationships between safety goals, the arguments that support these goals, and the evidence that substantiates the arguments. This visual method transforms complex safety claims into more straightforward sections, enhancing stakeholders' ability to understand and evaluate the safety case effectively. In fact, GSN facilitates this by organizing safety information into a hierarchy that begins with high-level safety objectives and breaks these down into specific pieces of evidence. Each piece of evidence is linked to the associated safety objectives, ensuring a transparent and traceable path from general safety requirements to the detailed evidence that supports them.

Furthermore, GSN utilizes a set of standardized symbols (Table 3.1) to represent different elements of the safety argument, such as goals (what we aim to achieve for safety), strategies (how we plan to meet these goals), and evidence (the data or information that supports our claims). This structured approach allows for a systematic examination and validation of each safety claim, improving the safety argumentation process. The use of GSN ensures that all aspects of the safety case are thoroughly documented and easily accessible, making the overall argumentation not only more comprehensible but also simpler to manage and update as needed. Through the explicit connections made by GSN, stakeholders can efficiently manage the safety argumentation, ensuring that every claim is appropriately supported by appropriate evidence.

Finally, GSN enhances the efficiency of evidences and arguments by adopting the modular approach within the safety case development. This modularity allows for the decomposition of the argumentation into smaller sections, making it easier to update and refine individual parts without affecting the whole. The ability to track the development and modifications of arguments over time ensures that the safety case remains comprehensive and adaptable, capable of adding new evidence or responding to changes in system design or operational context. By enabling a more structured and coherent assembly of evidence, GSN not only improve the development of safety arguments, but also improves their credibility and comprehensibility, ensuring that safety claims are thoroughly substantiated and clearly communicated.

Finally, it is worth noticing that while the benefits of GSN in terms of expressiveness, clarity, and traceability are effective, its efficiency and added value to the quality of safety argumentation remain to be proved. Accordingly, several works and research have raised and discussed these claims of efficiency and added value (see (Leveson, 2020)). Interestingly, a NASA report (Rinehart et al., 2017) has provided a noteworthy overview of the safety assurance cases and GSN, and examined several claimed benefits of safety assurance, while considering the opinions and views of academic researchers and industrial practitioners.

3.3 GSN-based safety cases in transportation systems

In this section, we present a succinct survey regarding the use of GSN-based safety cases in transportation domains, and then, looking ahead, we discuss the research advances on the GSN use for safety cases of autonomous systems.

3.3.1 GSN in the automotive domain

Recently, GSN method has been incorporated into ISO 26262 to satisfy the critical safety assurance of automotive systems (Yang et al., 2017). From the literature review, the GSN approach has been used in the automotive domain mainly (i) to structure the content of safety cases (Luo et al., 2015; Martin et al., 2016) (ii) to establish reusable patterns and modules (particularly for software safety cases) (Wagner et al., 2010; Palin et al., 2011; Ruiz et al., 2017; Martin et al., 2020), and (iii) to automate the generation process of GSN modules with respect to the model-driven development and assessment (SysML for example) (Habli et al., 2010; Luo et al., 2019). Moreover, both process-based (Gallina, 2014; Martin et al., 2016) and product-based (Luo et al., 2015; Dardar et al., 2012) safety cases have been discussed, while focusing more on software and subsystems safety cases. Surprisingly, most of the research works sought compliance with the functional safety standard ISO 26262. Thus, various patterns and modular (process and product-based) extensions have been elaborated to cover all parts of the standard.

3.3.2 GSN in the aviation domain

In the aviation domain, the standard DO-178B requires documentation of safety cases. The GSN safety cases have been used to deal with some particular safety assurance issues, broadly related to critical software components, such as Verification & Validation process (Guarro et al., 2017), improving confidence (Clothier et al., 2017; Nešić et al., 2021), safety architecture (Denney et al., 2015), etc. While the GSN is mainly used in automotive for functional safety cases, in aviation, more “system” safety aspects have been investigated (Farnell et al., 2019); for instance, the operational safety (Williams et al., 2014), the Informed Risk (and Safety) (Guarro et al., 2017; Clothier et al., 2017), dynamic and real-time safety cases (Kurd et al., 2009; Denney et al., 2019; Asaadi et al., 2020; Javed et al., 2021). It is worth mentioning here that the latter is capturing increased attention when it comes to autonomous systems. Moreover, some papers were more concerned with the safety argument and its validity in the safety case. Indeed, Belle et al. (2019) addressed the uncertainties’ propagation through GSN-based safety argumentation, while Witulski et al. (2016) applied GSN to construct a radiation assurance case for spacecraft, highlighting its effectiveness in complex safety argumentation.

3.3.3 GSN in the railway domain

The investigation on GSN started later in the railway as compared to the automotive and aviation domains. One of the pioneer works, in European railway, was the INESS project² (for *Integrated European Signalling System* (Müller et al., 2009)), which aimed to reduce the time and the cost for the safety case process building by avoiding unnecessary and redundant procedures. Hence, a formal safety case process model was proposed following standard EN 50129, and a dedicated GSN-based tool was developed. Taguchi et al. (2014) have proposed a GSN-based reusable module in compliance with the railway safety standards to improve the traceability and thus the quality of the safety case. Interesting works have been developed in (Wang et al., 2018, 2019, 2016a; Idmessaoud et al., 2021), to assess the confidence level in the attributed arguments of a railway safety case. The main idea consists in investigating the Dempster-Shafer theory (Dempster, 2008; Sentz and Ferson, 2002).

Finally, it is worth noticing from the succinct survey above, that the adoption of GSN is not yet prevalent in the railways compared to the other transportation domains. This may be justified by the fact that railway standard EN 50129, which establishes the safety case guidance process, appeared early (first draft in 1998, first update in 2003, and last update in 2018) before the emergence of GSN framework; hence, the railway specialists and experts had already developed and adopted textual argumentation for safety cases. Contrarily, the automotive ISO 26262 standard appeared (first version in 2005, and then updated in 2011 and 2018) in the period where GSN-based safety cases began receiving attention and acceptance by the industries.

In Table 3.2, we provide a non-extensive summary of some relevant works in the automotive, aviation, and railway, while pointing out some features of the works.

²<http://www.iness.eu/>

Table 3.2: The use of the GSN method in automotive, aviation, and railway domains

References	Theoretical concepts	Process	Product	Compliance	System level	Sub-system level	Applications & features
Automotive domain							
Wagner et al. (2010)	-	-	✓	ISO 26262	-	✓	-
Rudolph et al. (2018)	✓	✓	-	ISO 26262	✓	-	-
Gallina (2014)	✓	✓	-	ISO 26262	✓	-	-
Luo et al. (2015)	-	-	✓	ISO 26262	-	✓	Power Window System
Schmid et al. (2019)	-	-	✓	ISO 26262	✓	-	Vehicle Motion Control
Luo et al. (2019)	✓	✓	-	ISO 26262	✓	-	-
Burton et al. (2017)	✓	✓	-	ISO 26262	-	✓	-
Martin et al. (2016)	-	✓	-	ISO 26262	-	✓	High Voltage System
Dardar (2014)	-	-	✓	ISO 26262	-	✓	Fuel Level Estimation
Palin et al. (2011)	✓	✓	✓	ISO 26262	✓	-	-
Habli et al. (2010)	-	✓	✓	ISO 26262	✓	✓	Air Suspension System
Aviation domain							
Kurd et al. (2009)	-	✓	✓	-	✓	✓	Gas Turbine Aero Engine Control
Williams et al. (2014)	-	✓	-	-	✓	-	Managing Mid-Air Collision Risk
Guarro et al. (2017)	-	✓	-	DO-178C	✓	-	Unmanned Aircraft System case study
Clothier et al. (2017)	-	✓	-	Subcommittee F38.01 ASTM International	✓	-	Small Unmanned Aircraft
Denney et al. (2015)	-	✓	-	-	✓	✓	Unmanned Aircraft System
McDermid et al. (2019)	-	✓	-	-	✓	-	-
Asaadi et al. (2020)	✓	✓	-	-	✓	-	Unmanned Aircraft System
Railway domain							
Taguchi et al. (2014)	-	✓	-	EN 50126	✓	-	Traceability Information Model
Stålhane and Myklebust (2016)	✓	-	-	EN 50129	✓	-	-
Gallina et al. (2017)	✓	-	-	ISO 26262	✓	-	MDSafeCer
Wang et al. (2017)	-	-	✓	EN 50129	✓	✓	Wheel Slide Protection System
Hirata and Nadjm-Tehrani (2019)	-	✓	✓	-	✓	-	Train Door Controller
Pissoort et al. (2019)	-	✓	✓	EMC Directive (2014/30/EU)	✓	✓	EMC Equipment/ Large Machines

3.3.4 Toward using GSN for autonomous systems

In the last few years, the GSN-based argumentation has received particular attention for guiding and structuring the safety assurance cases for autonomous systems. Indeed, in (Wardziński, 2008), the author has used the GSN to argue about the safety assurance of autonomous vehicles while investigating both traditional static (i.e., during the development phases) and dynamic (i.e., during the operation phase) risk assessment. Thus, two GSN safety argument patterns have been proposed. In (Alexander et al., 2009), the GSN is used to tackle a key safety activity for autonomous systems, which is the deriving and traceability of safety requirements. The work consists of specifying top-level requirements that are progressively decomposed until the lowest level of requirements is reached. The process is in fact an argument for the completeness and adequacy of requirements that can be expressed efficiently using GSN. The authors in (Heikkilä et al., 2017) have proposed a safety qualification process using goal-based safety case for an autonomous vessel prototype. They used a GSN pattern to instantiate the elaborated process with an illustration for the situational awareness system. The authors in (Cheng et al., 2020) have used GSN patterns to fill gaps at run-time phases to manage the self-adaptive operation of a robot operating system. Vierhauser et al. (2019) have discussed the concept of *pluggable* GSN-based safety cases to demonstrate compliance of unmanned aerial vehicles. More recently, (Schwalbe and Schels, 2020) have been interested in ensuring the safety of neural networks, and generic and modular GSN patterns have been proposed to build safety argumentation templates for safety cases of neural network software systems.

The Assuring Autonomy International Program (AAIP)³ provided a body of knowledge to support the development of safe autonomous systems by providing practical guidance on assurance and regulation. To illustrate how the proposed framework can be achieved, a structure for the assurance argument in the form of a safety assurance case pattern represented using GSN is proposed. By the same program, guidance on the Assurance of Machine Learning components used in Autonomous Systems (called AMLAS) is proposed (Hawkins et al., 2021; Picardi et al., 2020). In AMLAS, a set of GSN safety argument patterns and instances are used to illustrate and justify the safe design and deployment of machine learning components integrated into autonomous systems.

3.4 Safety assurance approach of autonomous train

The introduction of autonomy to railways requires a review of the previously established safety studies, safety cases, and their underlying safety principles/processes. The case of autonomous trains raises a fundamental issue regarding their definition, concept, and integration with the railway system; particularly due to the withdrawal of the human operators from the train control loop. With regard to the European railway safety standards, two methods can be considered when performing the autonomous train safety demonstration. The first one is to consider the autonomous train as an entirely *new* system that implies a from-scratch overall safety demonstration. Thus, the safety objectives (in terms of Safety Integrity Levels - SILs) already allocated to the safety functions (and their sub-systems) need to be reviewed and reapportioned. In contrast, the second manner consists in considering that the autonomous train only brings a *significant* change to the conventional trains, and thus the safety activities have to focus in autonomous/automated driving system (ADS) and its interaction with the existing systems (human agents or technical systems) and surroundings. This is mainly due to the fact that the ADS performs some safety and safety-related functions when it comes to operate in GoA 3/4.

³<https://www.york.ac.uk/assuring-autonomy>

The European railway regulation (CSM-RA, 2017) is applied to any significant (technical, operational, or organizational) change that may impact the railway system safety. Thus, considering the autonomous train as a modification (with a significant change) of conventional trains is more convenient to comply with the CSM-RA process. In this regard, the GSN is a suitable framework for building safety argumentation modules for the modified part of the train (i.e., the ADS) and efficiently incorporating it within the overall safety case of the autonomous train.

The safety assurance of autonomous trains requires a set of safety activities and processes at three (hierarchical) system levels (see Figure 3.3): (i) *overall system level* (i.e., train), (ii) *AI-based component level* (i.e., perception and decision-making components) and (iii) *AI techniques and technologies* (i.g., obstacle detection software) (Tonk et al., 2022). These safety activities have to be performed in parallel to the development life-cycle process of the overall system. Notice that several initiatives and white papers for safety assurance of autonomous systems have recommended and adopted such a hierarchical framework (Wozniak et al., 2021; Alexander et al., 2020).

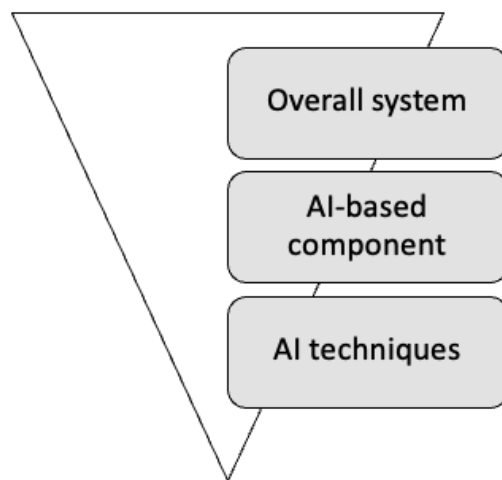


Figure 3.3: The three hierarchical system levels

3.4.1 Overall system level

At this level, the safety goal is to “*assure that the autonomous train remains safe when operating under its specified environment*”. Such an objective shall be achieved by assuring that all the hazards have been identified, assessed and controlled or reduced to an acceptable level. At this level, the railway standard EN 50126 specifies the safety activities and processes to be performed and evidence to be produced, starting from the definition of the system, the definition of its operational context, the risk analysis and evaluation, and finally the specification of the system safety requirements. However, in the case of autonomous train, these safety activities are not sufficient and thus need to be achieved by three activities: (1) specification of the autonomy aspects, (2) specification of the interactions between human operators and the autonomous train, and (3) definition of the operational design domain (ODD). Notice that the ODD should be specified starting from the detailed description of the operational context and the concept of operation (ConOps) of the system. This permits to take into account all the operational conditions related to the operational scenarios of the system. The European Union Aviation Safety Agency roadmap in (EASA, 2021) can be an interesting starting point for achieving this task. Moreover, the work of Hawkins (2019) provides relevant insights and guidance to achieve

these activities. Figure 3.4 proposes a high-level GSN argument pattern for the system level of the autonomous train. Notice that the same issues related to the overall system safety and risks have risen in the maritime domain with autonomous ships (Fan et al., 2022; Chang et al., 2021; Ramos et al., 2019b; Chen et al., 2021) and the provided studies can be helpful for the railway domain.

3.4.2 AI-based component level

At this level, the main safety goal is to "*assure that the AI-based component satisfies the allocated system safety requirements in its defined sub-ODD*". The safety activities at this stage concern (1) the specification of the (safety) architecture of the component (sensors, hardware, and software computing unit), and (2) the functional hazard analysis at the component level, i.e., identify the contribution of the component failures to the (potential) overall system hazards.

An additional safety activity to be considered in the case of autonomous trains is the analysis of *safety of the intended functionality* (known as SOTIF in the automotive domain (ISO-21448, 2022)). The SOTIF is concerned with managing risks due to inherent design limitations that are present even when the system is functioning as intended (e.g., the sensors' insufficiency due to the technological limitations and the severe environmental conditions). Another safety activity, to be handled at this stage, is the refinement of the safety requirements that shall be allocated to the AI/ML models. This activity is crucial in the sense that starting from this point, there is a switch from the textual requirements to data requirements, i.e., each requirement shall be implicitly represented by sets of data to be provided to AI model to learn.

3.4.3 AI techniques level

The output of this level is an AI/ML model to be deployed within an AI-based component. Similarly to the previous levels, the safety goal is to "*assure that the implemented AI model satisfies the allocated safety requirements*". Such a safety objective is a crucial challenge since the existing safety standards and safety engineering methods are no longer suitable for adaptive and learning software. Moreover, it is not only the safety activities that need to be carefully conducted but also the development process. Thus, the safety argumentation needs to argue the rigorous development process (i.e., data management, model learning, model verification and validation) and the associated safety activities. Besides the (quantitative) performance evaluation of the developed model, particular attention needs to be paid to AI/ML-specific activities, such as sufficiency of training and learning process, robustness and adversarial attacks verification, interpretability evaluation the completeness of the test with respect to the specified ODD. Notice that some standards and technical reports have been recently issued (or are under development) to deal with various aspects of AI development and assessment processes. For instance, ISO/IEC TR 29119-11 (ISO/IEC29119, 2020) presents some guidelines on the testing of AI systems, ISO/IEC TR 24028 (ISO/IEC24028, 2020) surveys topics related to trustworthiness in AI systems and approaches to assess its attributes, and ISO/IEC TR 5469⁴ which deals the functional safety related to AI systems.

Finally, in order to produce an efficient structured safety assurance case for the autonomous train, GSN-based safety argument patterns shall be established for the aforementioned safety activities and processes. Figure 3.5 depicts the main steps to establish the GSN patterns.

⁴ISO/IEC CD TR 5469 (2024) Artificial Intelligence — Functional Safety and AI systems.

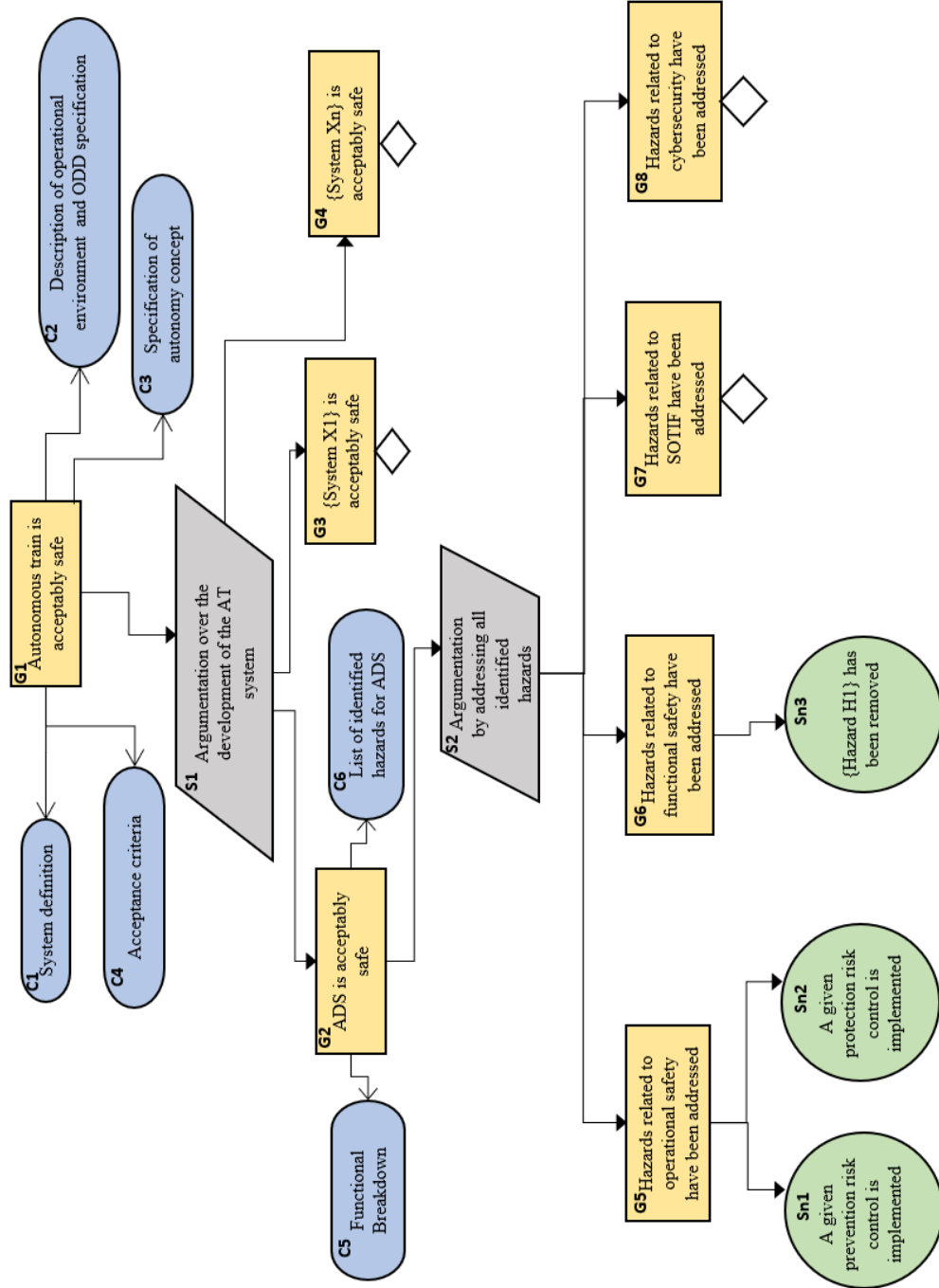


Figure 3.4: A system-level argument pattern for the autonomous train safety assurance

- (**Step 1**) Firstly, it is obvious that the autonomous train necessarily includes some system modules and components which have been already used for conventional trains and which are in interaction with the ADS system. Thus, safety *argument patterns* for these modules need to be firstly generated (from previous existing safety case documents) and transformed to GSN models;
- (**Step 2**) It involves building a *reusable template* for the safety case using the modular GSN argument patterns introduced in Step 1. This template should be created using GSN to justify the overall level of safety and should clearly argue the safe usage of components existing in conventional trains;
- (**Step 3**) It incorporates creating a GSN module to present the safety argument for the ADS system. To build the overall safety case of an autonomous train, the GSN module for the ADS must be established and incorporated into the template given in Step 2;
- (**Step 4**) This step aims to determine and evaluate *the confidence level* of the arguments to be used in safety assurance case. In fact, it allows modeling uncertainty and weighting the interrelationships between the various arguments and pieces of evidence presented in GSN structures.

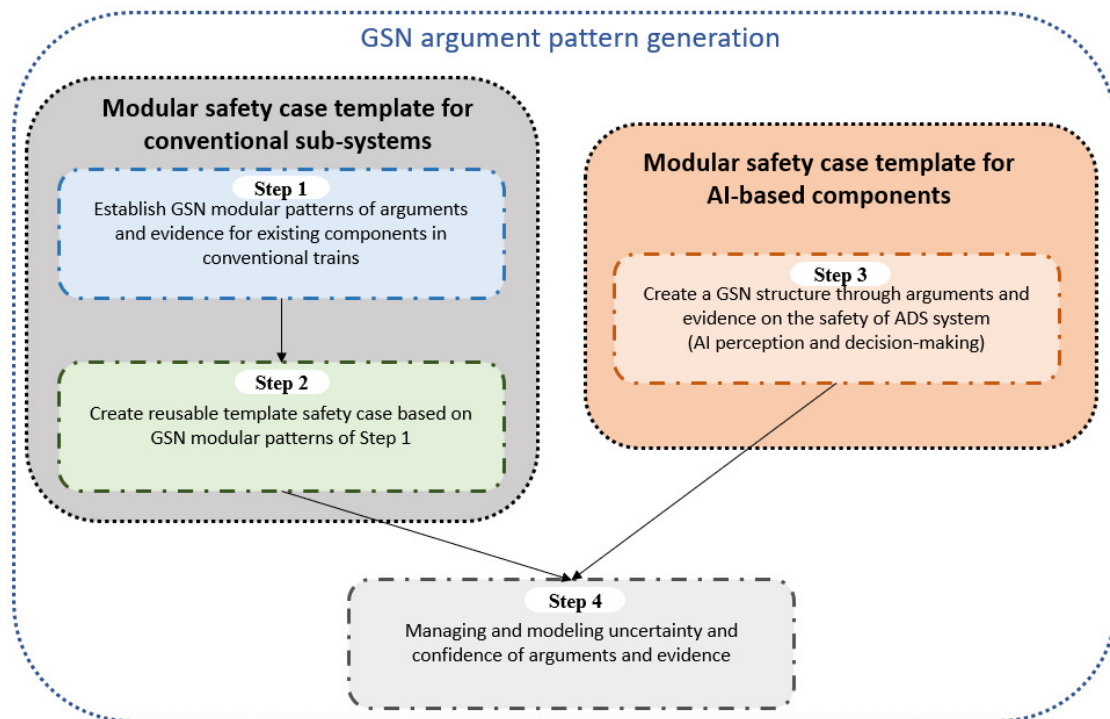


Figure 3.5: Main steps for building GSN safety argument patterns for autonomous trains

Finally, we notice that the safety assurance framework we present here is based on the assumption that all the identified hazards and their associated mitigation measures were identified during the system development phases. However, due to the transition of safety responsibility from human drivers to ADS, the autonomous train (and autonomous systems in general) shall be able to dynamically examine the risk associated with each control action performed by its ADS (as it is the case with human driver).

3.5 Use case

To illustrate the use of GSN-based safety argumentation within autonomous trains components, we take the anti-collision function as an example in this section.

3.5.1 Anti-collision function

The *anti-collision function* in autonomous train systems is engineered to proactively prevent collisions. This system integrates various sensors like LIDAR, radar, and cameras (Chouhan et al., 2014), which are essential in monitoring the train’s environment and surrounding (Abdazimov and Zuhridinov, 2023). These devices detect potential obstacles, ranging from static objects to moving entities. After the detection, advanced algorithms assess the collision risk based on the train’s speed, trajectory, braking capabilities, and distance to the obstacle. Depending on the assessed risk, the system autonomously implements appropriate measures, varying from minor speed adjustments to a full stop, thereby enhancing safety and reducing dependence on human operators.

GSN enables a precise decomposition of safety requirements (i.e., safety goals/objectives) into targeted sub-goals. These sub-goals are created to address the specific risks of railway operations, ensuring that each aspect of the system’s safety is defined and managed according to the associated standards.

The depicted GSN structure in Figure 3.6 provides a structured pattern for articulating the safety arguments of the autonomous train’s on-board ADS. For instance, the GSN structure starts with the top-level goal **G.1**, which asserts the necessity to ensure the ADS safety. Strategically, **S.1** underpins **G.1** by prescribing the implementation of integral safety functions. This strategy branches into specific objectives. For example, **G.2** focuses on mitigating collision risks, emphasizing the criticality of proactive safety measures within the system’s operation.

Context elements, such as **C.1**, specify the operational environment, including the railway track and weather conditions, which are essential factors for the ADS’s safety performance. Similarly, **C.3** encompasses the broader operational dynamics, such as the movements of other trains and dynamic obstacles, indicating the complex interplay of factors the ADS must operate.

Subsequent sub-goals, **G.3** to **G.8**, delineate the detailed safety targets that contribute to achieving **G.1**. Furthermore, **G.3** is tasked with the detection of light signals, while **G.4** is concerned with monitoring the train’s environment, ensuring that the system remains aware of its surroundings. Moreover, **G.5** ensures safe and effective stopping, while **G.6** introduce system redundancies and locational accuracy, which are crucial for maintaining operational safety under various conditions.

In addition, **G.7** and **G.8** address the train’s interaction with stationary and moving objects, respectively, underscoring the importance of the train’s ability to discern and react to both static and dynamic obstacles within its path. The inclusion of strategies, such as **S.3**, which details the detection and avoidance measures for static obstacles, further refines the safety argument by specifying the actions taken to prevent collisions. The contexts, **C.2** and **C.3**, lie in the GSN structure to detail the fixed infrastructure and the dynamic nature of the operational conditions. These elements frame the ADS’s functional requirements within the real-world context of the railway system.

Transitioning from the initial GSN structure, Figure 3.7 illustrates the system’s approach to address static obstacles. To achieve **G.7**, strategy **S.3** is supported by goal **G.9**, which aims to enhance sensor accuracy for static obstacle detection. Indeed, accurate sensor data is essential for reliable static obstacle detection and interpretation. Additionally, context **C.5** provides the basis for this goal, pointing to the need for precise sensor calibration to ensure the accuracy required for reliable detection.

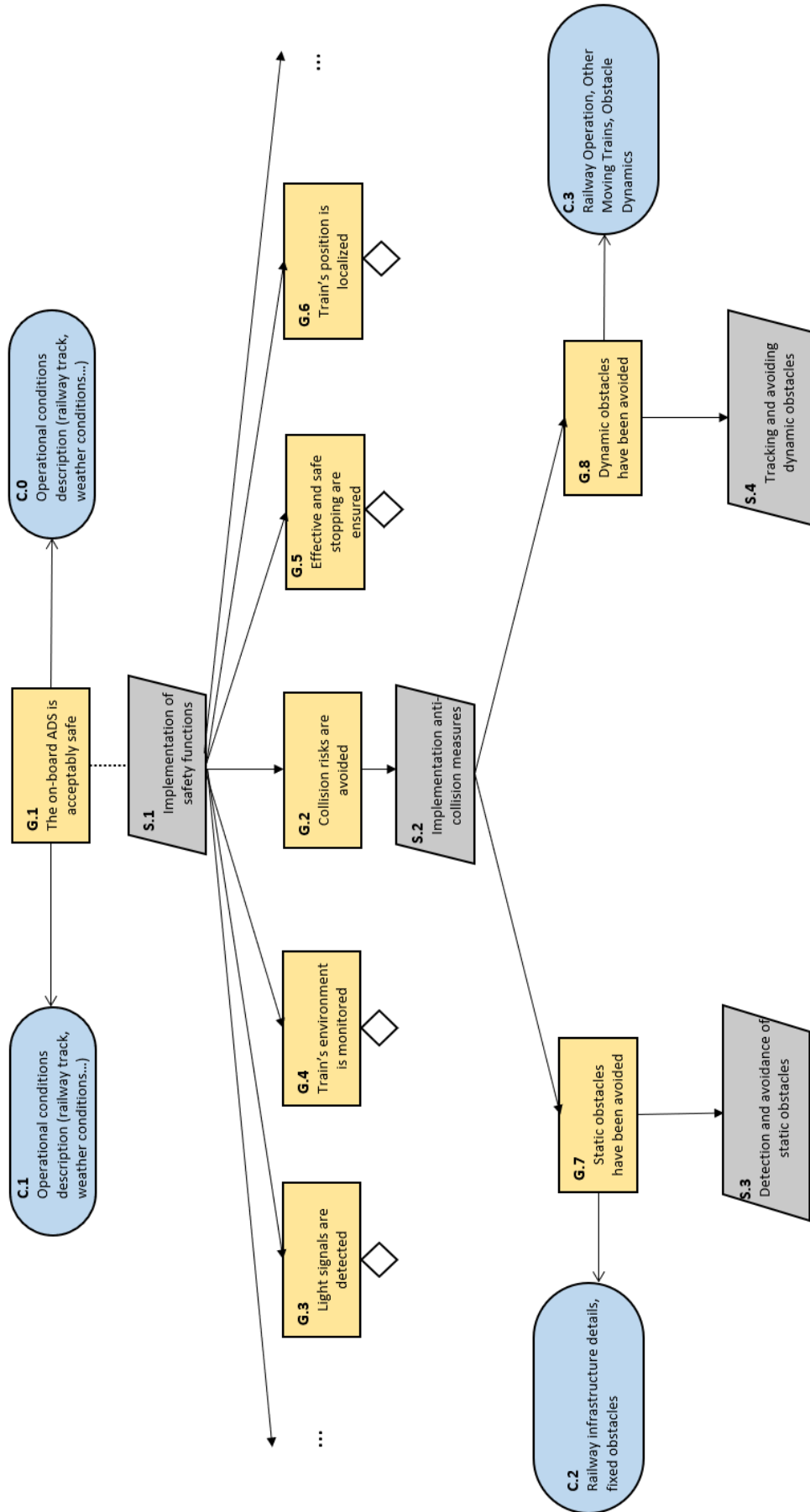


Figure 3.6: Decomposition of high-level goal: Ensure safety of the on-board ADS

Furthermore, the strategy **S.5** associated with goal **G.9** delves into the technical aspects of improving sensor data interpretation. This advancement is necessary to filter out noise and irrelevant information, ensuring that the sensors can reliably identify static obstacles. Another branch developed from **S.3** leads to goal **G.11** that aims to develop signal noise reduction algorithms. Here, context **C.7** provides noise sources and sensor calibration. Finally, following strategy **S.7** that consists of algorithms development and testing, solutions **s.1** and **s.2** respond to goal **G.11** with the design of signal noise reduction algorithms, and testing of these algorithms to evaluate the associated results.

In parallel, goal **G.12** addresses the need for systems robustness, implementing sensor redundancy and fail-safe mechanisms. Then, strategy **S.8** ensures that the system has multiple layers of safety (i.e., redundancy configuration and failure scenarios). Finally, solutions **s.3** and **s.4** outlines the actions to be taken in the case of failures. In this decomposition, context **C.8** highlights the importance of considering various redundancy configurations and potential failure scenarios to maintain safety integrity.

On the other hand, goal **G.10** aims to validate sensors' performances for static obstacle recognition, while strategy **S.6** is aimed at assessing the practical effectiveness of sensors and algorithms in real-world conditions. This process is guided by the considerations in context **C.6**, which specifies the test scenarios and assessment criteria.

Having established the GSN structures for static obstacles, the focus now shifts to dynamic obstacles. The next GSN, illustrated in Figure 3.8, extends the argumentation to include the detection and avoidance of moving hazards, a critical aspect of maintaining continuous operational safety. The central element of this structure is goal **G.8**, supported by strategy **S.4** detailing the system's approach to continuously monitor and react to moving obstacles. This strategy encompasses the development and refinement of sensor fusion techniques, as specified in goal **G.13**. The accurate tracking of dynamic obstacles is essential for real-time response, and context **C.10** emphasizes the need for precise sensor calibration and appropriate prediction models to achieve the goal. Furthermore, the prediction models, as highlighted in strategy **S.9**, are crucial for anticipating the orientations and trajectories of moving obstacles. In fact, solution **s.7** focuses on the prediction model design and parameters, laying the foundation for these predictive algorithms, while solution **s.8** ensures the model's accuracy and reliability through validation processes.

In the same manner, goal **G.14** reinforces the necessity of empirical evidence to support the system's capabilities. Additionally, solutions **s.9** and **s.10**, which provide the necessary testing and evaluation, along with the test scenarios and evaluation criteria defined in context **C.11**, provide a structured approach to testing the system's performance under a variety of conditions. This ensures that the autonomous train can safely operate in dynamic, changing, and unpredictable environments.

Collectively, the previous GSN diagrams present a structured safety argumentation for the anti-collision function of the autonomous train. The decomposition highlighted in each structure is essential in ensuring the reliable implementation of safety measures, particularly within the dynamic and unpredictable operational conditions of the autonomous train. In conclusion, through the application of GSN, the anti-collision function has been analyzed, showcasing the method's ability to structure and clarify the safety argument effectively. The GSN has demonstrated the different aspects of the function and highlighted its utility in the safety assurance process for autonomous trains.

3.5.2 Discussion

In the context of autonomous train systems, the role and significance of structured safety argumentation are carried out in this study through the use cases of the '*Anti-Collision Function*'. This function is essential in ensuring the safety and reliability of autonomous train operations, addressing safety-critical aspects such as collision risk mitigation, train

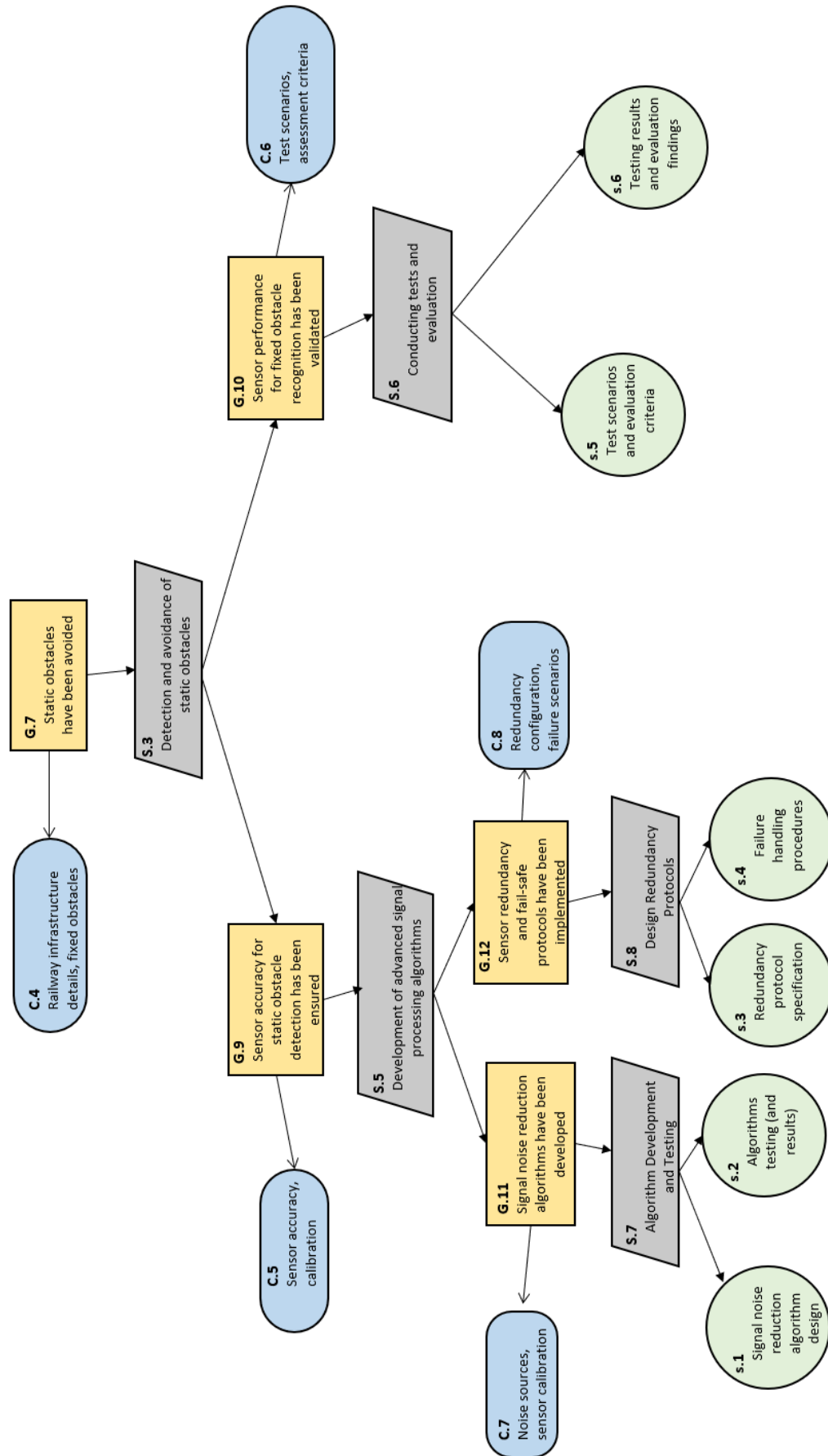


Figure 3.7: Safety argumentation structure for addressing static obstacles

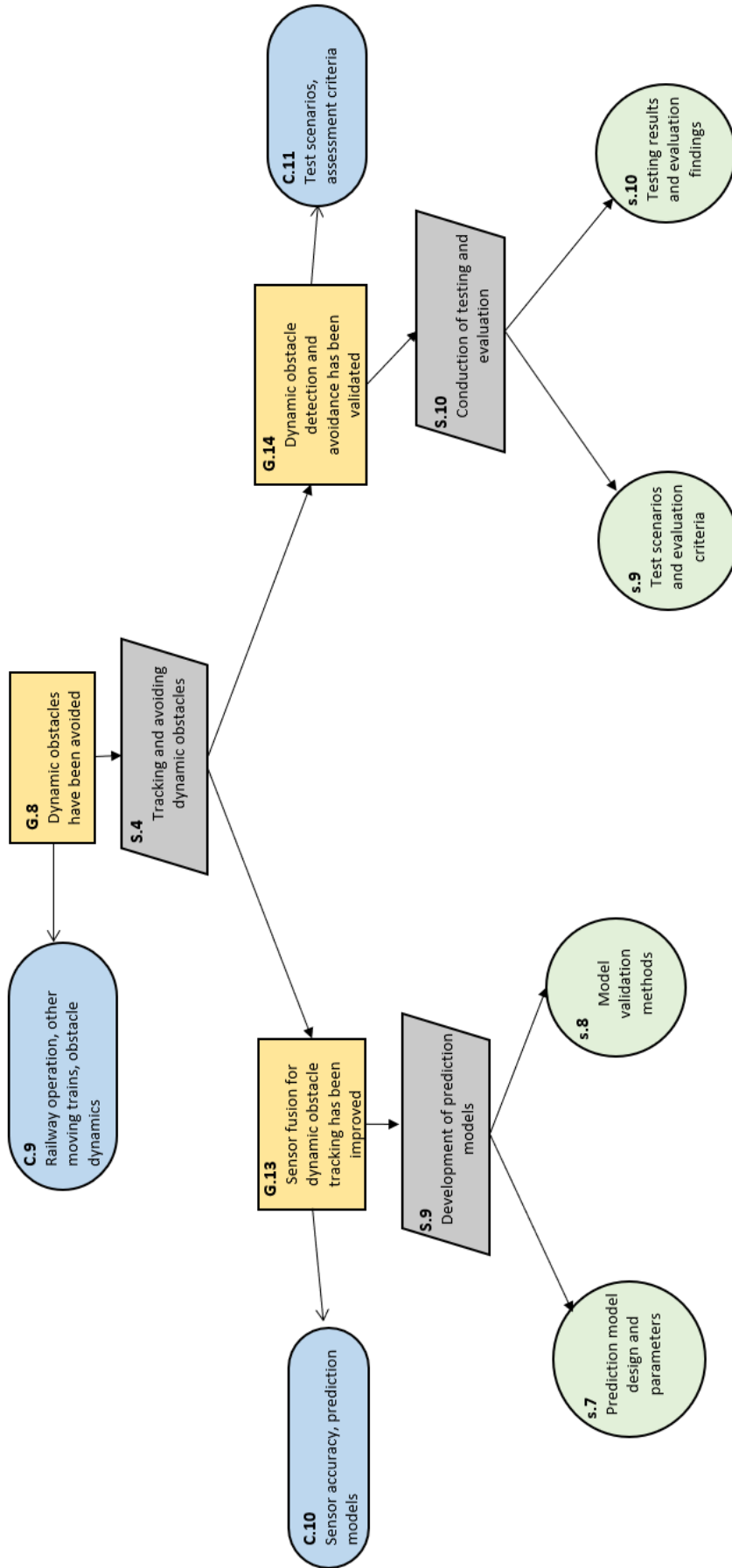


Figure 3.8: Safety argumentation structure for addressing dynamic obstacles

stopping capabilities, and the safety of railway crossings. Through the application of the GSN this use case have been systematically decomposed, allowing for a comprehensive approach to safety assurance.

The advantages of using GSN in this use case are highlighted. GSN provides a clearly defined framework to structure safety arguments, enabling the systematic breakdown and decomposition of complex safety claims/goals into more manageable sub-goals, contexts, strategies, and solutions. This decomposition enhances both the clarity and traceability of safety requirements, ensuring that each element within the safety argumentation is coherently justified and linked to the overall operational objectives. In addition, this structured approach offers a clear and rigorous method for justifying the safety of autonomous train systems, providing assurance to stakeholders and regulatory bodies.

Furthermore, the modular nature of GSN serves as a significant advantage in the context of autonomous train safety. This modularity allows for the development of distinct safety argument structures established for various functions of subsystems with the autonomous train overall system. For instance, creating separate GSN structures for ‘*Sensor Redundancy*’ and ‘*anti-collision*’ enables detailed and specific safety analyses for each area. The strength of this approach lies also in its ability to address the requirements and challenges of each subsystem while maintaining a coherent overall safety argumentation process. The development of these individual modules forms a collective and comprehensive safety argumentation while ensuring traceability throughout the process. Consequently, the modular approach of GSN is instrumental in building a detailed and integrated safety case for the overall autonomous train system, effectively validating its reliability and ensuring its safety at every level.

In conclusion, the use cases of the anti-collision function outline the significance of structured safety argumentation in autonomous train systems. GSN offers a robust and methodical approach to addressing safety argumentation challenges, ensuring that safety requirements are comprehensively justified and validated.

3.6 Conclusion

In this chapter, we discussed the use of graphical safety argumentation to build a safety assurance case for autonomous systems. Firstly, we presented a survey on the use of Goal Structuring Notation (GSN) for building safety cases in conventional transportation systems and then in automated and autonomous ones. Then, we elaborated an overall GSN-based framework for building a safety assurance argumentation for the autonomous trains. Finally, we provided GSN structures for a use case: the *anti-collision function*.

Chapter 4

SA & DRA framework for autonomous trains

Contents

4.1	Introduction	69
4.2	Context and concepts	70
4.2.1	Autonomous Driving System (ADS)	70
4.2.2	Situation Awareness (SA)	73
4.2.3	Complementarity between SA and DRA concepts	75
4.3	A DRA and SA framework for autonomous trains	76
4.3.1	Perception module	76
4.3.2	Understanding & prediction module	78
4.3.3	Decision-making module	79
4.4	Illustrative case: anti-collision function	79
4.4.1	Perception module	81
4.4.2	Understanding & prediction module	82
4.4.3	Decision-making module	84
4.5	Conclusion	84

4.1 Introduction

To ensure the operational safety of conventional trains, drivers carry out a real-time assessment of the risk associated with the operational environment. This task relies on the driver's SA to perform his/her driving task safely and enables him/her to manage potential risks dynamically according to the current operational conditions. The concept of SA is central to risk assessment in operations. In this chapter, its underlying principles will be adapted to the context of autonomous trains, to enable them to operate efficiently and safely in dynamic and unpredictable environments.

In autonomous systems, the situational awareness (SA) involves the perception, understanding, and projection of environmental elements, with a particular focus on the understanding of the current situation, interpreting its significance, and anticipating future scenarios (Chauvin et al., 2008). This concept is distinct from and complementary to the dynamic risk assessment (DRA) discussed in Chapter 2 (Conges et al., 2023). DRA is a proactive, adaptive process focused on identifying, assessing, and mitigating risks, particularly in environments prone to change or uncertainty (Patel and Liggesmeyer, 2021).

Whereas DRA is concentrated towards ongoing real-time risk management, SA emphasizes environmental awareness and projection. This makes both concepts essential to the decision-making of autonomous systems in dynamic environments.

In the context of the autonomous train (with a high level of autonomy - GoA3 and GoA4), the SA and the DRA processes will have to be integrated within the on-board ADS. Indeed, the on-board ADS should be able to perform its functions safely in all predictable and unpredictable situations/operational conditions. In order to achieve this goal, the ADS must integrate a DRA layer in its high-level control/decision-making architecture (Parhizkar et al., 2022). In fact, with strong interactions with the perception, planning, and control units, such a layer can continuously update the probability estimations for the occurrence of (hazardous) events.

In this chapter, we propose a framework allowing the on-board ADS to continuously perform the situational awareness process and provide run-time probability estimations for the occurrence of railway hazards while accounting for (internal and external) environment perception.

In this chapter, we firstly review the dynamic risk assessment and situational awareness key aspects for autonomous train operations. Then, we propose a framework allowing the on-board ADS to continuously perform the situational awareness process and provide run-time probability estimations for the occurrence of railway hazards while accounting for (internal and external) environment perception. The research work of this chapter has been published in the European Safety and Reliability Conference (ESREL) (Chelouati et al., 2022).

The remainder of this chapter is organized as follows. We first detail, in Section 4.2 the concept of SA and its complementarity with the concept of DRA after having highlighted how such concepts can intervene in a high-level architecture of the on-board ADS embedded in autonomous trains. Then, in Section 4.3 we present the situation-awareness and dynamic risk assessment framework as part of the ADS decision-making architecture. Finally, in Section 4.4, we illustrate it through an operational safety function (anti-collision function).

4.2 Context and concepts

Given the introduction of the autonomous driving systems in Chapter 2, this section concentrates on analyzing ADS's architectures within autonomous vehicles. Additionally, a high-level architecture of the on-board ADS is presented, outlining its design and operational features. In addition, concepts of situational awareness and dynamic risk assessment are discussed with a focus on the dynamic risk assessment components. Finally, we propose a dynamic risk assessment framework for the anti-collision function in the context of the autonomous train.

4.2.1 Autonomous Driving System (ADS)

This subsection examines detailed aspects of Autonomous Driving Systems (ADS) in the railway context. It discusses the main components of ADS, including sensor technologies, data processing algorithms, along with command and control systems. Particular focus is given to the aspects of autonomous railway systems compared to other ADS applications, highlighting the challenges in track management, signaling, and interaction with existing railway infrastructure.

In the rapidly evolving domain of autonomous transportation, various ADSs have been developed, each established to specific operational needs and technological capabilities. The variety in these systems reflects the broad spectrum of applications they serve, from

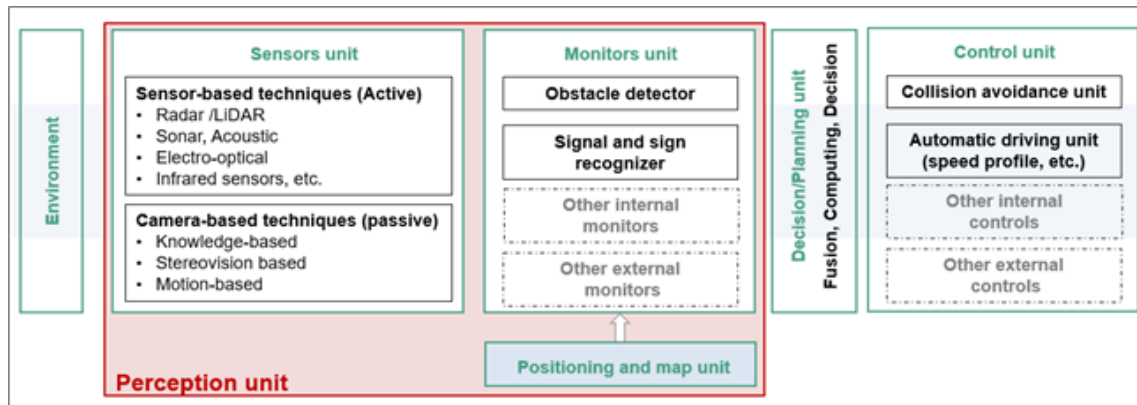


Figure 4.1: A high-level architecture of the on-board ADS of the autonomous train with a main focus on the decision-making process

automotive to aerial and maritime domains. To understand the extensive range and comprehensive details of ADS architectures, Table 4.1 provides a comparative overview of several key systems. This comparison includes well-known automotive ADS like the Audi zFAS System (Jung et al., 2018). This system integrates sensors, communication technologies, and AI to achieve SAE Level 3 Traffic Jam Pilot capabilities. It excels in providing high self-awareness in different traffic conditions but is constrained to specific scenarios like traffic jams and faces challenges in sensor fusion integration. Notice that the SAE levels of driving automation are presented in Chapter 1.

Moreover, Tesla’s Full Self Driving (FSD) Arbiter (Talpes et al., 2020) prioritizes sensors, cameras, and AI, the FSD Arbiter offers comprehensive environmental awareness and the ability to make sophisticated autonomous decisions. However, its performance may be affected in adverse weather due to a heavy reliance on camera systems, and it is still undergoing development and regulatory review. Furthermore, other notable systems such as Waymo’s ADS (Grigorescu et al., 2020), Mobileye’s EyeQ Chip (Ingle and Phute, 2016), and NVIDIA DRIVE Platform (Jagannadha et al., 2019) are featuring advanced LiDAR, radar, and camera systems to achieve high level of autonomy (SAE level 4 and 5). Additionally, architectures are included from other domains like Autonomous Aerial Vehicles (AAV) Systems (Koh and Wich, 2012), Maritime Autonomous Surface Ships (MASS) Systems, and Autonomous Train Operation Systems. Each system is evaluated in terms of its primary advantages and limitations, offering a comprehensive view of the current landscape in autonomous system architectures. Finally, each architecture addresses different aspects of autonomous operation, from obstacle detection and avoidance to navigation and safety management, highlighting the multiple aspects of the ADS development and the ongoing evolution of autonomous transportation technologies.

Table 4.1: Comparison of Various ADS Architectures

ADS Architecture	Advantages	Limitations
Audi zFAS System	<ul style="list-style-type: none"> - Combines sensors, communication technologies, and AI for SAE L3 Traffic Jam Pilot. - Offers high self-awareness in various traffic conditions. 	<ul style="list-style-type: none"> - Limited to specific traffic conditions (e.g, traffic jams). - Reliant on extensive sensor fusion, posing integration challenges.
Tesla Full Self Driving (FSD) Arbiter	<ul style="list-style-type: none"> - Utilizes sensors, cameras, and AI for comprehensive environmental awareness. - Capable of sophisticated autonomous decisions and reactions. 	<ul style="list-style-type: none"> - High reliance on camera systems may limit performance in adverse weather. - Ongoing development and regulatory approval process.
L2-automated driving system	<ul style="list-style-type: none"> - Allows hands-free operation on highways. - Capable of lane changes and overtaking. 	<ul style="list-style-type: none"> - Requires driver supervision and intervention. - Limited to highway driving conditions.
CEHSS	<ul style="list-style-type: none"> - Self-contained with independent sensors. - Does not rely on other driving automation systems. 	<ul style="list-style-type: none"> - May have limited capability in complex driving scenarios. - Integration with broader vehicle systems might be challenging.
MSS	<ul style="list-style-type: none"> - Monitors vehicle status and provides driver feedback. - Enhances driver awareness and safety. 	<ul style="list-style-type: none"> - Driver-centric, not fully autonomous. - Relies on driver's response to feedback for safety.
Waymo ADS	<ul style="list-style-type: none"> - Advanced LiDAR, radar, and camera systems. - High level of autonomy (SAE Level 4 and 5). 	<ul style="list-style-type: none"> - Requires highly detailed mapping. - Still under regulatory review for widespread use.
Mobileye's EyeQ Chip	<ul style="list-style-type: none"> - Specializes in visual sensor processing. - Powers advanced ADAS solutions. 	<ul style="list-style-type: none"> - Visual limitations under certain weather conditions. - Dependent on camera clarity and range.
NVIDIA DRIVE Platform	<ul style="list-style-type: none"> - High-performance AI computing. - Scalable across different autonomy levels. 	<ul style="list-style-type: none"> - Requires extensive training data. - High computational power demand.
AAV Systems	<ul style="list-style-type: none"> - Advanced navigation using GPS and IMUs. - Suitable for unmanned aerial vehicles. 	<ul style="list-style-type: none"> - Limited by battery life. - Sensitive to weather and environmental conditions.
MASS Systems	<ul style="list-style-type: none"> - Integrates radar, sonar, GPS for maritime navigation. - Advanced for autonomous ships. 	<ul style="list-style-type: none"> - Complex in terms of maritime traffic management. - Challenges in long-range communication.
Autonomous Train Operation Systems	<ul style="list-style-type: none"> - Utilizes ground-based and onboard sensors for safe train operation. - Advanced signaling and safety algorithms. 	<ul style="list-style-type: none"> - Dependent on track infrastructure. - Integration with existing rail systems.

Based on the general architectural concepts of ADS in autonomous vehicles presented in Chapter 2 (cf. Figure 2.2), a specific architecture of the on-board ADS is proposed for autonomous trains. Figure 4.1 illustrates the proposed high-level architecture with a main focus on the decision-making process. This architecture is designed to allow the flow of information going from the environment through the system (from left to right in Figure 4.1), ensuring real-time responsiveness and safe operations.

In details, the perception unit is the initial layer, where an array of active and passive sensors, including radar/LIDAR and camera-based technologies, collect environmental data. This unit functions as the empirical foundation, gathering crucial real-time inputs about the train’s external conditions. Furthermore, the decision/planning unit is the brain of the system, where it processes the sensory data into meaningful information. It uses advanced algorithms to detect obstacles and understand traffic signals, effectively converting raw data into a structured format for decision-making. Moreover, the control unit receives the intelligence from the decision/planning unit and acts on the decisions. It coordinates the collision avoidance unit to prevent possible dangers, and it controls the automatic driving unit to change the train’s speed based on the situation. In conclusion, this architecture represents a layered approach to autonomous operation, where the combination of data collection, analytical processing, and reactive action is essential for the safe movement of an autonomous train through its operational environment.

4.2.2 Situation Awareness (SA)

The concept of Situational Awareness (SA) was initially introduced in the aviation domain for the research in the human factors field (for aircraft pilots). The following definition has been formulated in (Endsley, 1988): “*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future*”. Then, this concept has been extended to almost all of the safety-critical domains that involve human operation, such as the nuclear power industry, automobile, air traffic control, medical systems, and railways. Achieving SA is one of the concerns of these communities for assuring relevant and efficient decision-making (Garland et al., 1996).

For conventional vehicles (with drivers), operators need to *recognize, understand, and predict* relevant information (for the system and its operational environment) to know what is happening and what is going to happen in the near future. This three-level model established by Endsley (1995) to characterize SA is the most widely used. Individual factors such as experience, bias and goals, as well as system factors such as interface design, complexity, and automation strongly influence the capacity of the SA. Figure 4.2 illustrates the three-level model of SA and its role in the decision-making process for human operators.

The SA has already been adopted to both manual and automated railway operations, such as train driving (Brandenburger and Naumann, 2019; Rose et al., 2018), rail maintenance (Golightly et al., 2013), or comprehending signalling and control in rail operations (Sharples et al., 2011). From the standpoint of human information processing, a considerable number of automated systems and procedures currently exist in the railway domain. Moreover, depending on the degree of automation, the range of tasks changes, and SA changes accordingly (ERA, 2021)¹. In railway, automation can be categorized into different levels, from manual operation by the human to self-driving without human intervention (GoA1 to GoA4). With the increasing level of automation, the SA responsibility is gradually transferred from the human driver to the ADS. For example, in GoA2, obstacle detection is a part of the driver functions, while in GoA3 or 4, it becomes a part

¹Automation Myth Busting Paper #1 (https://www.era.europa.eu/content/automation-myth-busting-paper-1-situation-awareness-remains-same_en)

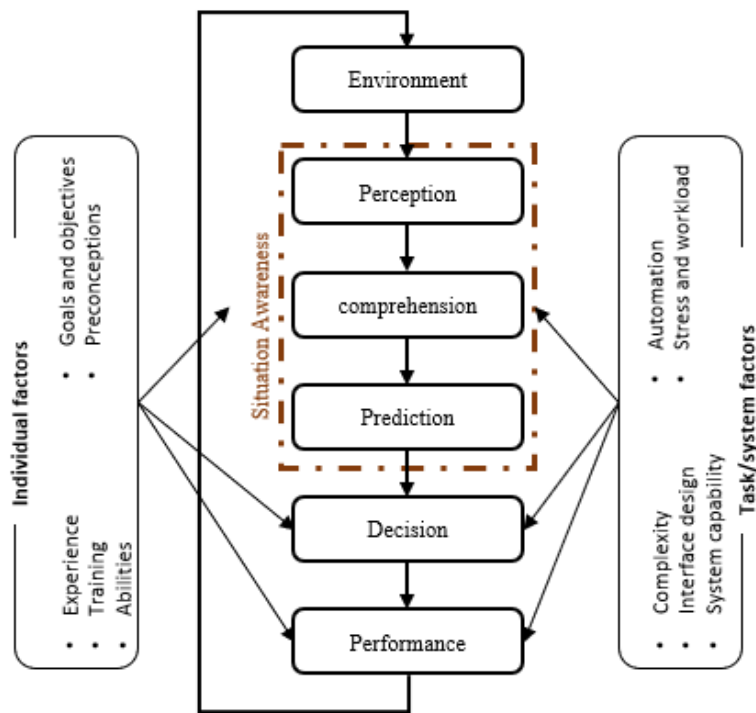


Figure 4.2: Three-level model of Situation Awareness [Endsley \(1995\)](#)

of the ADS functionalities. When the system is fully automated, i.e. GoA4, issues regarding the impact of automation on SA become increasingly challenging to address. In this case, the system should be self-aware and capable of making safe decisions in its dynamic operational conditions.

In autonomous systems, the convergence of SA and automation represents a significant research focus. For instance, [Vallikannu et al. \(2023\)](#) highlights the significance of SA within military contexts, suggesting the integration of AutoML (Automatic Machine Learning) systems to enhance situational prediction and risks mitigation. This aspect is carried out in other safety-critical domains, where SA's role in healthcare is also highlighted as a fundamental element for patient safety ([Parush et al., 2011](#)). Furthermore, the complexity of incorporating SA in vehicle automation is analyzed by [Salmon et al. \(2020\)](#), who presents a distributed situational model as essential for the design and safety of automated vehicles. This concept is further examined in the context of autonomous driving by [Laugier \(2019\)](#), who uses a combination of Bayesian methods and ML techniques to advance the understanding of SA in autonomous navigation systems.

In the domain of robotics, more specifically, robotic surgery, [Ginesi et al. \(2020\)](#) proposed a framework that establishes a SA module to ensure safety in surgical automation. Additionally, in the maritime domain, [Zhou et al. \(2019\)](#) proposed a quantitative SA model specifically designed for autonomous ship navigation, addressing the complexities of SA in such dynamic environments.

Moreover, the development and integration of SA in various systems have led to significant advancements in decision-making processes. In the context of human-machine interactions, a recent paper ([D'Aniello and Gaeta, 2023](#)) highlights the necessity of SA for making accurate and timely decisions. Meanwhile, a comparative study by ([Costa et al., 2023](#)) evaluates the applications of RL for adaptive automation under evolving conditions,

demonstrating the effectiveness of SA-based decision-making models. [Srivastava et al. \(2022\)](#) focuses on shared SA in AI-advised decision-making, leading to enhanced team performance. In the other hand, ([Insaurralde and Blasch, 2022](#)) introduces a decision support system for air traffic management, showing the benefits of ontological reasoning. This review outlines the evolution of SA techniques, from traditional applications to their integration with AI and ML in enhancing the efficacy of autonomous systems across diverse domains.

4.2.3 Complementarity between SA and DRA concepts

For autonomous railway systems, a main challenge is transitioning the human driver's intuitive capability (i.e., Situational Awareness) to *dynamically* assess risks to an Autonomous Driving System (ADS). A human driver, leveraging years of experience and training, naturally adapts to the changing railway environment, quickly making decisions using a variety of sensory information. In turn, the ADS must perform this complex (human) function to ensure that the overarching safety level does not decrease. The concept of Dynamic Risk Assessment (DRA) emerged from this imperative, highlighting the need for real-time, responsive, and adaptive risk analysis in autonomous train systems.

The complementarity between SA and DRA is essential for the development and operation of autonomous systems, playing a crucial role in ensuring their safety and reliability. SA involves the perception of environmental elements within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future. It provides the foundational layer for DRA by ensuring that the system has a continuous, updated understanding of its operating environment. DRA, on the other hand, is the process of analyzing and evaluating potential risks in real-time, adapting to changes in the environment, and implementing strategies to mitigate identified risks. It relies on SA to provide the necessary context for risk identification and analysis, making these two concepts interlinked and complementary. Together, they form an integrated approach that enables autonomous systems to operate safely within their dynamic and potentially unpredictable environments. In the context of autonomous systems, such as autonomous cars or autonomous trains, SA ensures that the system maintains an accurate understanding of its surroundings, including other vehicles, pedestrians, and changing road or track conditions. This awareness is critical for identifying potential hazards that might not have been anticipated during the planning or programming phases. DRA builds on this awareness, using algorithms and models to evaluate the risk associated with these hazards in real-time and to determine the most appropriate set of action to mitigate the risk. This might involve adjusting the vehicle's speed, changing its route, or taking evasive maneuvers to avoid an obstacle. The complementarity between these aspects is not just beneficial but essential for the successful operation of autonomous systems. SA provides the data and context needed for DRA, while DRA offers the mechanisms to use this information effectively to ensure safety. This complementarity enables autonomous systems to operate in complex environments, make safe decisions, and respond adequately to unpredictable challenges, thereby enhancing their overall safety and effectiveness.

The review concerning use of DRA for autonomous vehicles and systems is presented in Chapter 3. It explains how these systems identify and manage risks in changing conditions. Moving forward, in this chapter, we cover the SA aspect. This part shows how SA helps autonomous systems understand and react to their surroundings. Moreover, it links SA's role to the safety and effectiveness of AVs, showing how it supports DRA in making safe decisions in complex environments.

Recent advancements in SA for AVs highlight a variety of methodologies and technologies aimed at ensuring safety and operational efficiency. Examining Vehicle-to-Vehicle communications, [Metzner and Wickramaratne \(2019\)](#) explores improving SA through

additional sensor information in Vehicle-to-Vehicle communications. Moreover, [Hu \(2023\)](#) introduces the Doppler principle SA supplementing on-board radar for autonomous driving. In addition, [Nine \(2020\)](#) focuses on automating the process for SA, as a crucial process for autonomous vehicle operation. On the other hand, [Islam et al. \(2016\)](#) implemented a system-on-chip concept of the comprehension level of SA using an expert system, while [Nine et al. \(2021\)](#) proposed frameworks aimed at data fusion and decision-making based on sensing data. Lastly, [Dahn et al. \(2018\)](#) offers an application-agnostic definition of SA, integrating it into the perception component. Together, these studies illustrate diverse approaches to enhance SA in autonomous systems, from Vehicle-to-Vehicle communications and Doppler principles to System-on-chip implementations and expert systems, aiming to improve the ability of autonomous systems in dynamic changing environments.

Moving forward, the focus shifts to the application of these principles within the autonomous train. The next subsection examine how these advanced DRA and SA approaches are adapted and implemented in the context of railway operations, particularly in autonomous train systems, where safety and reliability are crucial.

4.3 A DRA and SA framework for autonomous trains

In this section, we describe how the DRA process need to be considered and handled by the on-board ADS to assess the risks of autonomous trains.

In fact, the on-board ADS should be able to perform its functions safely, in run-time and in all operational conditions. For this purpose, the on-board ADS have to consider an online DRA layer with a high-level decision-making architecture. Consequently, we propose a framework allowing the ADS to continuously manage the SA process and provide a run-time evaluation and prediction of the potential railway hazard risks (particularly, the estimation of occurrence probabilities). A high-level presentation of the framework is depicted in Figure 4.3.

This framework is applied at run-time while using environment perceived sensor data (in addition to historical data) as input. Thus, a risk-based approach is required to process and understand relevant information from the perception module such that the decision-making module efficiently takes the adequate and safe actions to avoid or reduced that impact of hazards. The modules involve on the DRA process are briefly discussed hereafter.

4.3.1 Perception module

The perception function is responsible for interpreting sensory data, comprehending its essential meaning, resolving uncertainty and imprecision from complex inputs, and producing relevant information. In fact, environmental perception is the ability of an autonomous system to acquire, analyze and interpret the raw data from its surroundings, such as images, sounds, or signals. By doing so, the autonomous system extracts meaningful useful information from the data, such as the location, shape, motion, or identity of objects or agents in the environment. Environmental perception is essential for an autonomous system to perform tasks that require interaction, navigation, or decision-making in complex and dynamic scenarios.

The objective of the perception is to establish a sufficiently accurate view of the real world appropriate to the function of the autonomous train (e.g., discern the difference between an animal and a person, discern the difference between a track worker and a trespasser etc.). Generally, this task consists of environment perception and localization/positioning.

In fact, the perception module is mainly composed of the *sensor unit*, which contains

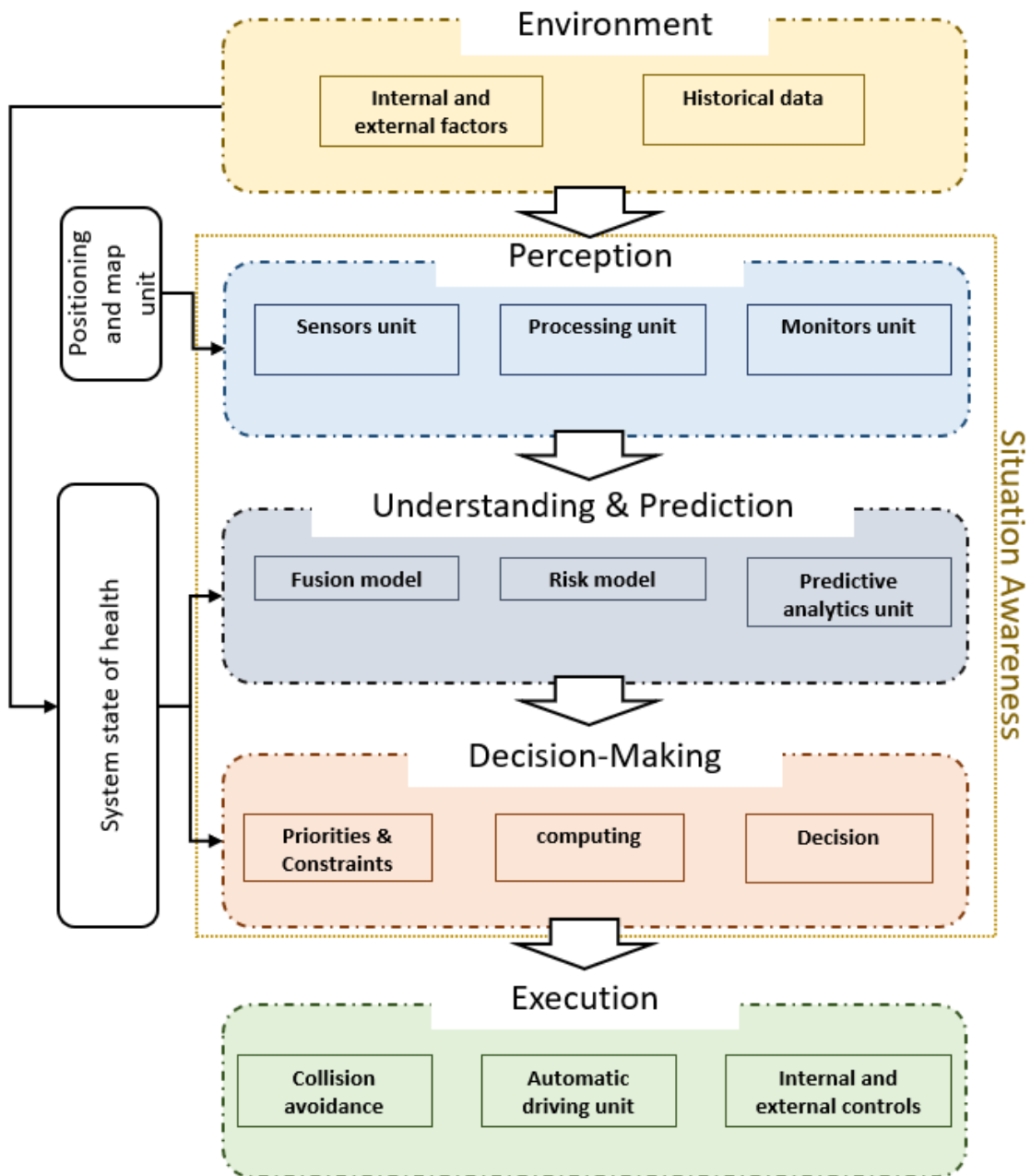


Figure 4.3: The autonomous train situational awareness framework

all the physical sensor devices that are used to capture and collect information and signals from the external or internal environment, and the *processing and monitoring units*, which contains the set of perception functions (and associated processing algorithms) that are used to perform the perception tasks, i.g., detection, recognition, classification, etc. (See Figure 4.1)

4.3.2 Understanding & prediction module

This module is the foundation of the SA process, since it uses the information provided by the perception module to create and constantly update an integrated run-time model representing the system's environment and its states. This model is then used for ongoing decision-making. From the DRA perspective, the module permits to compute an estimation of the current level of risk and to predict the potential railways hazards. Such a task is performed through the combination/fusion of all the information provided by the perception module with the historical data about the system and its operational conditions.

Within the Understanding & Prediction module, the fusion model plays an important role. This model integrates all sensor data collected by the perception module, using advanced algorithms to create a comprehensive view of the train's state and surrounding. By combining data from various sources like radar, LIDAR, and cameras, the fusion model compensates for the limitations of individual sensors, enhancing the accuracy and reliability of environmental perception. The basis of this process is algorithms such as the Kalman Filters or Particle Filters, which combines data but also filter noises associated to sensors' information, leading to a more precise understanding of the environment.

Furthermore, the evaluation of risk within the understanding & prediction module is conducted through a risk model. Initially established offline, this model integrates historical data and qualitative insights, such as expert judgements, to establish a risk assessment framework. Once operational, the risk model incorporates, dynamically, real-time data and observed risk factors from the train's surroundings, enabling the system to quantify and infer potential risks effectively. In fact, the selection of risk model should be established based on multiple factors, including the nature and volume of incoming data streams, computational demands, and the inherent uncertainties present in the autonomous train's environment. Among the model employed in autonomous systems are Partially Observable Markov Decision Processes (POMDPs) (Spaan, 2012) and Dynamic Bayesian Networks (DBN) (Junyung et al., 2021). These models provide robust frameworks for managing the complexities of risk assessment in dynamic and partially observable scenarios.

Ultimately, the prediction aspect of the understanding & prediction module is performed by the predictive analytics unit. This unit enhances the train's ability to estimate and predict future conditions and potential obstacles on its path. Using advanced algorithms, this unit interprets patterns and trends from historical and real-time data provided by the perception module. Moreover, it effectively predicts the behavior of other entities, estimates changes in environmental conditions, and anticipates potential system states. By doing so, it offers insight into the immediate future, enabling the decision-making module to proactively adjust operational parameters. This capability is crucial for mitigating risks that are not yet apparent but could result in hazardous events if not addressed.

Finally, the estimation process must comprehensively consider the associated uncertainties within each module, providing the decision-making module with three critical pieces of information: (i) estimation of current residual risk, (ii) prediction of future level of risk, and (iii) the level of confidence assigned to these estimations. These elements are essential for a robust decision-making framework that ensures informed and reliable autonomous operation.

4.3.3 Decision-making module

The decision-making module stands as the final brick in the autonomous train's DRA framework, synthesizing information from the perception and understanding & prediction modules to formulate actionable policies.

In addition, the decision-making module reflects the ability of the ADS to apply policies and make decisions to achieve higher order goals in response to the current operational conditions. This is achieved by combining the processed information about the environment (perceived view of the real world) with established policies, domain knowledge and learning regarding how to respond to the presented environment. In fact, the decision-making module is segmented into several integral units, each with a defined purpose in the module.

First, the priorities & constraints unit is tasked with defining the operational envelope (i.e., ODD of the autonomous train). It assimilates regulatory standards, safety requirements and guidelines, and operational objectives to establish the criteria against which decisions are evaluated. This unit ensures that all decisions align with the predefined safety requirements, regulatory compliance, and operational efficiency while considering the constraints of the train's current state and environmental conditions.

Secondly, the computing unit employs algorithms that interpret the information received from the estimation of residual risk and predictive analytics units. It processes the estimation of the current risk level, projections of future risk levels, and the associated confidence levels. By prioritizing the safe decisions and advanced computational models, this unit quantitatively assesses potential actions and their outcomes.

Finally, the decision unit integrates the insights provided by the computing unit to deliver final decisions. This unit is responsible for selecting the most appropriate course of action from a range of possible responses, from initiating collision avoidance maneuvers to maintaining current operational parameters. The decisions made are contingent upon achieving an optimal balance between risk mitigation and operational progress, guided by the defined priorities and constraints.

Once a decision has been reached, the information is conveyed to the execution module, which acts upon the directives with precision and timeliness. The execution module comprises units such as collision avoidance, automatic driving, and internal and external controls, each executing the necessary adjustments to the train's behavior in response to the decision-making module's outputs.

4.4 Illustrative case: anti-collision function

In railway, any external object on the track that can result in a collision with the train is considered as an obstacle. The obstacle detection function shall then detect and locate all elements that may represent a risk (of collision) for the train, and identify all parameters associated with these elements, such as distance, dimensions of the object, and velocity. The identification of these parameters allows for characterizing and estimating the objects that may represent a danger for the train movement or not. The anti-collision function in the autonomous train is performed by the *perception module*. In such systems, advanced sensor technologies, such as 2D cameras, 3D cameras and/or Radar and LiDAR are used to perceive the environment (Ristić-Durrant et al., 2020; Pavlović et al., 2018).

This function operates by employing a combination of advanced onboard sensors and complex processing algorithms to accurately detect and assess any potential obstacles that may block the train's path. Figure 4.4 serves as an illustration of the decision-making behavior diagram of obstacle detection and avoidance system (i.e., the anti-collision function) integrated to autonomous train operation. It represents a systematic and iterative

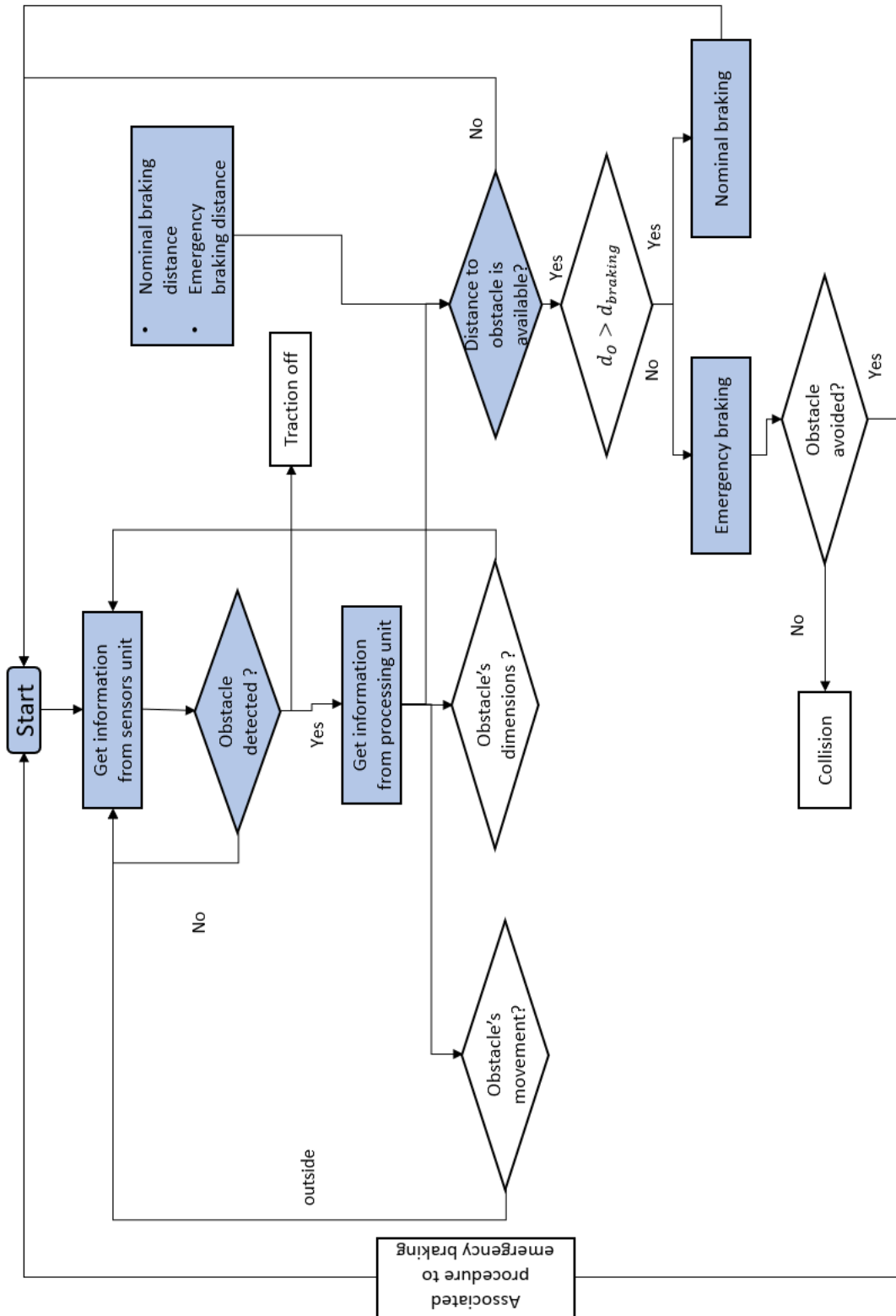


Figure 4.4: Decision-making flowchart of the anti-collision function for autonomous trains

approach where the train continuously monitors its environment through an array of sensors. Upon detection of an obstacle, the system initiates a series of steps to determine the appropriate set of actions (i.e., policies). This process is both essential and critical for ensuring the safety of the autonomous train as it operates through varying operational conditions. The figure breaks down this process into logical steps, from initial detection to the final decision-making, highlighting the train’s ability to respond to obstacles autonomously. The following description offers a clear and simplified overview of each step within this crucial safety system.

Starting with the initial data acquisition, the train’s sensors collect environmental information, a process depicted by the ‘*Get information from sensors*’ unit. This data is then scrutinized for potential hazards within the train’s operational envelope at the ‘*Obstacle detected?*’ decision node. After identifying a potential hazard, the process advances to the ‘*Get information from processing*’ unit, where the data is analyzed further. This stage is crucial for determining the obstacle’s dimensions and velocity (i.e., key factors that influence the train’s response policy). When the obstacle’s size surpasses predetermined safety parameters, the system evaluates its proximity to the train. When the obstacle is situated within the nominal braking range, a controlled deceleration is initiated, as depicted by the ‘*Nominal braking*’ route. However, if the obstacle presents an immediate danger, the ‘*Emergency braking*’ pathway is triggered, implementing rapid deceleration measures to mitigate a potential collision. Traction control may be disengaged during this phase to maximize braking effectiveness.

Moreover, the flowchart node ‘Distance to obstacle is available?’ assesses the proximity of an identified obstacle relative to the autonomous train. At this decision point, if the distance to the obstacle (d_o) is greater than the train’s braking distance ($d_{braking}$), nominal braking is initiated, allowing the train to decelerate under regular conditions. Conversely, if $d_o < d_{braking}$, the situation necessitates an emergency braking, leading the train to apply maximal braking forces in order to stop as quickly as possible, aiming to prevent a collision. Following each action, the process checks if the obstacle has been successfully avoided, thus determining the next steps in the operational procedure. After executing the braking maneuver, the system assesses the outcome. If the obstacle has been successfully performed, normal operations are resumed. In the event of a collision, the system transitions into an emergency state, activating additional safety measures.

The next sections detail the process of the obstacle detection and avoidance within the SA & DRA framework (which is also depicted in Figure 4.5, providing an understanding of each component’s role in identifying and responding to obstacles to maintain seamless and safe train operation. This examination clarifies how the system’s various elements function collectively to ensure the autonomous train’s continual safety and efficiency.

4.4.1 Perception module

In fact, environmental perception is the core of the perception module, allowing the autonomous train to identify and classify the critical elements of its surrounding. This process involves the detection and analysis of various signals, whether visual, auditory, or otherwise, to discern the dimensions and dynamics of objects or entities around the train. For example, the module should differentiate between static and dynamic obstacles, such as distinguishing a stationary platform from a moving vehicle.

The system’s ability to manage uncertainties and imprecision from multiple complex sensors inputs is crucial for establishing a reliable informational picture. The position tuple P_i (Equation 4.1) represent the distance to each obstacle i (d_o^i), its dimensions (\mathcal{D}_o^i), velocity of each obstacle (v_o^i), and its orientation (θ_o^i).

$$P^i = \langle d_o^i, \mathcal{D}_o^i, v_o^i, \theta_o^i \rangle \quad (4.1)$$

The position tuple ensures that the perception module provides a robust and accurate representation of the real world that is essential for the autonomous train's functions.

For each obstacle i detected at the time t , the perception module at least provides the following information: the kind of the obstacle (human, animal, etc.), the size of the obstacle, its position with respect to the train (coordinates) or the distance to the train d_o^i and the speed v_o^i of the obstacle (if it is in motion). Based on the distance to the obstacle i , the Time To Collision (see Equation 4.2) can be then estimated.

$$TTC_i(t) = \frac{d_o^i(t)}{v_T(t)} \quad (4.2)$$

Where $v_T(t)$ is the train speed, and $d_o^i(t)$ is the distance to obstacle i at time t .

The information and parameters related to the obstacle will be then shared with understanding & prediction module. Notice that these information are provided with some uncertainties due to the performance limitations of the sensors and the processing algorithm. Based on these data flows (with their corresponding uncertainties), the risk model determines and estimate (and predict) the current (and the future) level (or probability) of the collision risk.

4.4.2 Understanding & prediction module

In the understanding & prediction module, the system uses the processed sensory information to predict the future state of the environment. This involves estimating the current and future probability of hazardous events, such as potential collisions. This module calculates the predicted position vector \hat{P}^i of an obstacle i (Equation 4.3), which consists of its distance to the train (\hat{d}_o^i), the obstacle's dimensions ($\hat{\mathcal{D}}_o^i$), the obstacle's velocity (\hat{v}_o^i), and its orientation ($\hat{\theta}_o^i$).

$$\hat{P}^i = \langle \hat{d}_o^i, \hat{\mathcal{D}}_o^i, \hat{v}_o^i, \hat{\theta}_o^i \rangle \quad (4.3)$$

The system also computes an uncertainty vector δP^i (Equation 4.4, which accounts for measurement errors or uncertainties in the obstacle's position (δd_o^i), the obstacle's dimensions ($\delta \mathcal{D}_o^i$), the obstacle's speed (δv_o^i), and its orientation ($\delta \theta_o^i$). The prediction is based on whether the obstacle is moving in the direction of the rails and its orientation, which could influence the decision-making process. This vector enables the system to consider and reduce the natural variations and errors in sensor data, like the possible inaccuracy in an object's position or the speed at which it moves.

$$\delta P^i = \langle \delta d_o^i, \delta \mathcal{D}_o^i, \delta v_o^i, \delta \theta_o^i \rangle \quad (4.4)$$

Furthermore, the fusion model within this module integrates the diverse data flow to ensure a unified and consistent interpretation of the train's surroundings. Moreover, the module reconciles the inputs from various sensors to provide a reliable representation of the environment, which is essential for accurate risk assessment. Following data fusion, the risk model is the next unit in the process. The risk model employs algorithms to analyze the current surrounding and predict potential hazards. This proactive approach leverages the processed data to anticipate the trajectory of moving objects and the emergence of new obstacles, assessing how these elements may affect the train's path and safety.

Together, these components of the understanding & prediction module work to support the decision-making module. They provide a comprehensive risk profile that is instrumental in guiding the autonomous train's responses to a complex and dynamic operational conditions, ensuring that every decision is informed by a thorough understanding and anticipation of potential risks.

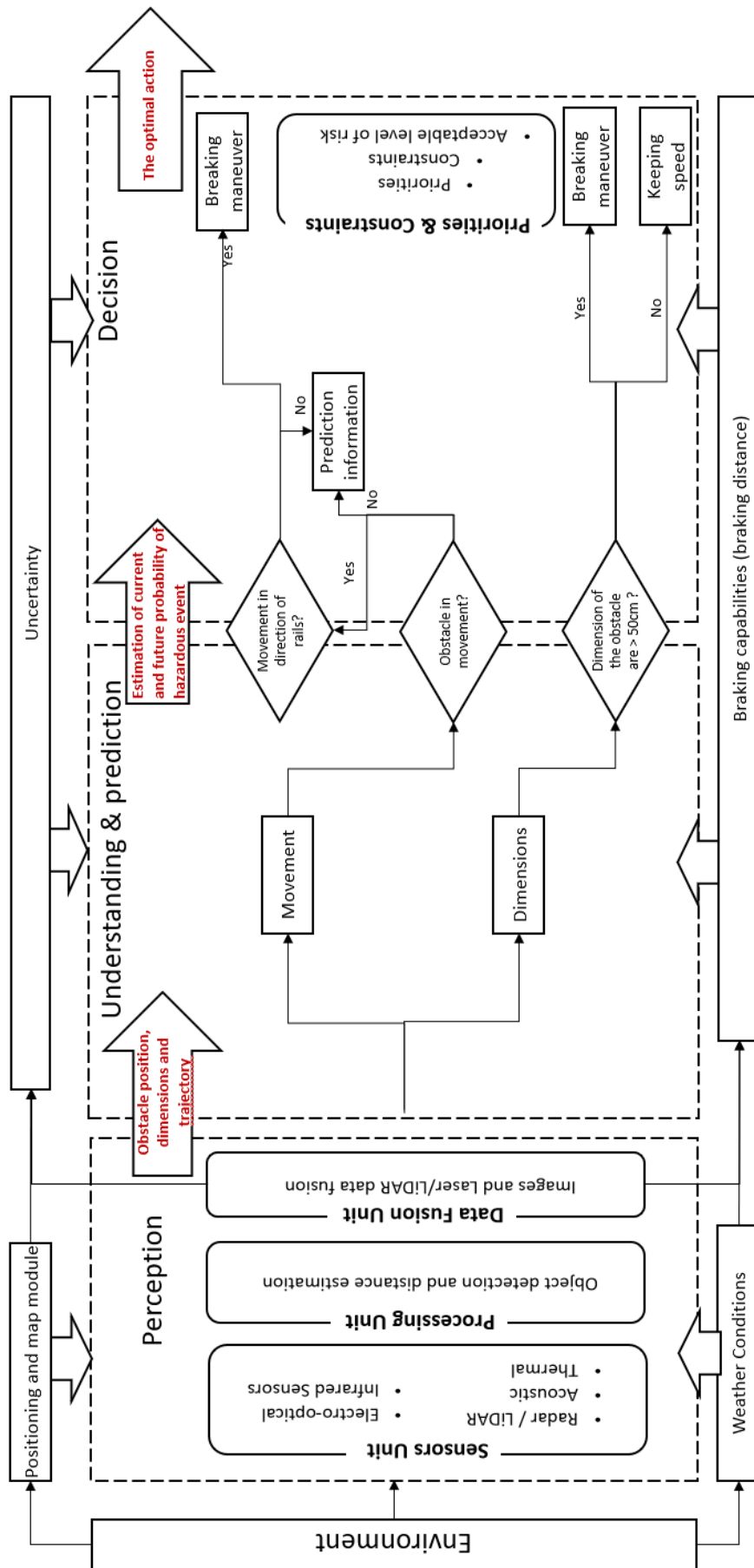


Figure 4.5: Framework for obstacle detection and avoidance (anti-collision) of autonomous trains

4.4.3 Decision-making module

The decision module is the executive function of the system. Based on the input from the understanding & prediction module, it decides whether to execute a braking maneuver, accelerate or to maintain speed. This decision is informed by various priorities, constraints, and what is considered an acceptable level of risk. The decision-making process is likely managed by algorithms designed to optimize safety and efficiency. The decision module receives the current and the future level/probability of collision with the obstacle. Taking into consideration the train braking capability, the module decides the kind of brake to be applied (Emergency brake, full service brake or holding brake).

In addition, the decision module leverages the braking capabilities vector P_b (see Equation 4.5). Here, ν stands for the train's velocity, η is the deceleration rate available, Δ is the brake delay time, τ indicates the braking state, and m represents the mass distribution of the train. These variables are crucial for determining the train's braking strategy in response to assessed risks

$$P_b = \langle \nu, \eta, \Delta, \tau, m \rangle \quad (4.5)$$

The final step within this module is the decision unit, which synthesizes all the processed data and prioritization to conclude the most appropriate action. It takes the output from the Computing Unit (i.e., the calculated risk levels and operational constraints) and determines whether to initiate a braking maneuver, accelerate or to continue at the current speed.

4.5 Conclusion

The responses of autonomous trains towards dynamic uncertainties in their external environment, that have not been anticipated during design time, shall be safe. In this chapter, we have proposed a Dynamic Risk Assessment (DRA) framework based on the Situation Awareness (SA). The proposed framework allows the on-board Autonomous Driving System to stay aware of its surrounding environment and entities, intending to provide run-time probability estimations for the occurrence of railway hazards while accounting for (internal and external) environment perception. In this chapter, we outline a strategy for testing the decision-making proposed framework for autonomous trains, as an essential step in assessing its practical applicability and effectiveness. However, it is important to note that, within the scope of this chapter, the framework has not yet been subjected to empirical testing or simulation. In fact, the focus here is on highlighting a prospective approach for future testing and validation of the framework. Nevertheless, this study lays the foundation for the model proposed in the next chapter, as well as the simulations and testing associated with it.

Chapter 5

POMDP-based decision-making process of autonomous trains

Contents

5.1	Introduction	85
5.2	Toward the use of POMDPs in ADS	86
5.2.1	Handling uncertainties in decision-making processes	86
5.2.2	Benefits of POMDP in decision-making processes	87
5.3	Decision-making related to the train’s anti-collision function	88
5.3.1	DRA of the anti-collision function	89
5.3.2	Structuring risk profiles with the DRA framework	91
5.4	Methodology	92
5.4.1	POMDP definition	92
5.4.2	POMDP modeling of the train anti-collision system	93
5.5	Simulation and results	100
5.5.1	Perceived state	100
5.5.2	Obstacle generation function	100
5.5.3	Belief updater	101
5.5.4	Solver choice	101
5.5.5	Variables initialization	102
5.5.6	Risk formulation	102
5.5.7	Results	103
5.6	Conclusion	108

5.1 Introduction

One of the challenges for the ADS to effectively implement safety functions lies in the existence of potential uncertainties associated with the perception system (including sensors and AI algorithms) and environmental conditions (Rosique et al., 2019; Nair and Bhat, 2021). Indeed, the unreliable received information could lead to missed detections and, at worst, to catastrophic consequences. Arising from this challenge is the need for a comprehensive and robust decision-making process capable of taking into account and handling uncertainties. This process should be designed to examine sensors’ information, taking into account the potential for inaccuracies, and react accordingly.

In this chapter, we propose a risk-based decision-making framework using Partially Observable Markov Decision Processes (POMDPs) for run-time monitoring of the environment during the train operations. This approach aims to assure the safe operation of the train with regard to collision hazards. Concretely, this consists of maintaining an acceptable risk level by estimating and updating the risk associated with the surroundings. The run-time risk estimation allows the system to make safe decisions while considering the inherent uncertainties in the train's state and the perceived environment. The approach is established and illustrated for the anti-collision function of the autonomous train.

Notice that the main research results of this chapter have been published in IEEE Access journal (Chelouati et al., 2023b).

The remainder of this chapter is organized as follows. Section 5.2 presents the related works addressing uncertainties in decision-making processes of autonomous systems. Additionally, the benefits of integrating POMDPs in such processes for risk control are discussed. In Section 5.3, the problem statement related to the anti-collision function for the autonomous train is detailed, along with the way to structure the associated risk data needed to complete the DRA task. Furthermore, the methodology of the proposed solution is described in Section 5.4, including the definition of the POMDP model, POMDP solvers, and the proposed risk model. The simulation results are presented in Section 5.5. Finally, Section 5.6 provides some concluding remarks and highlights some perspectives for future research.

5.2 Toward the use of POMDPs in ADS

Figure 5.1 recalls the essential components of the ADS in an autonomous train. In fact, the decision-making unit receives all the necessary information from the perception unit, computes main (operational and safety) indicators, and takes the adequate actions. The dynamic risk assessment task has to form the safety basis (via risk model) of the train decision-making process. Depending on the evaluated risk level, the ADS should then decide on an action plan. It could choose, for example, to accelerate or maintain speed to meet the speed profile of the train when no obstacle is present on the horizon, decelerate if a potential obstacle is detected at a safe distance, or initiate an emergency braking procedure if an immediate collision risk is identified.

Notice that, in railway standards, particularly as outlined in EN 50126, a '*risk model*', as discussed in Chapter 2, is the comprehensive framework designed for the systematic identification, assessment, and management of risks in railway operations. The risk model's main objective is quantifying the likelihood and severity of potential hazardous events, evaluating the effectiveness of existing safety measures, and determining the need for additional risk mitigations. The model typically (for conventional railway systems) encompasses the identification of hazards, the risk analysis (including frequency and consequences of hazardous events), and the evaluation of risk against predefined acceptability criteria. On the other hand, risk models for autonomous trains should incorporate real-time information. This allows for an adaptive response to changing environmental conditions and operational scenarios. Using advanced algorithms, the model evaluates risk levels continuously, considering both historical data and real-time sensory inputs.

5.2.1 Handling uncertainties in decision-making processes

Addressing uncertainties in decision-making for autonomous systems had emerged as a central research focus, identifying key problematic such as sensor fusion (Gupta and Snigdh, 2022; Shao et al., 2023), perception under varying environmental conditions (Molloy and McDermid, 2022; Gu et al., 2023), and dynamic system state evaluation (Yang et al.,

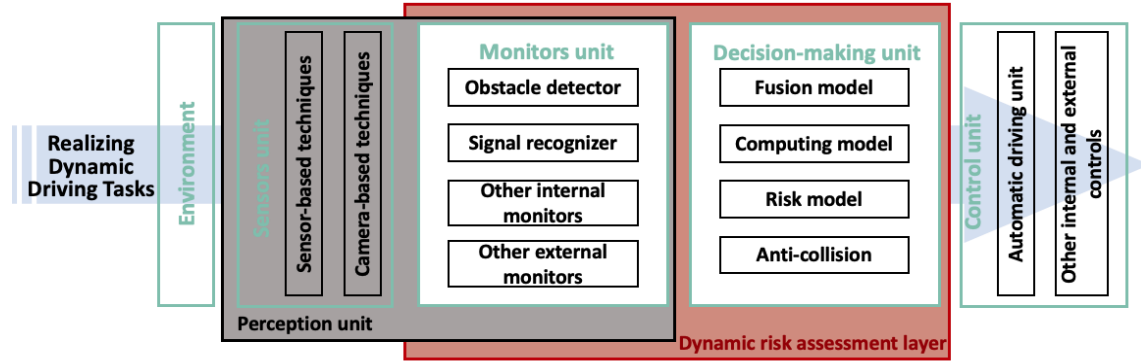


Figure 5.1: A simplified architecture of the ADS with a main focus on the DRA layer, strengthening the decision-making task

2023a,b). These challenges are critical as they directly impact the safety and reliability of autonomous operations. Sensor fusion is particularly essential for ensuring comprehensive perception (Zhang et al., 2023b), as it integrates data from multiple sensors to form a coherent understanding of the environment, compensating for the limitations of individual sensors (Lobato et al., 2023). The literature reveals that environmental conditions significantly affect the perception accuracy (Johansen et al., 2023), where factors such as lighting, weather, and obstructions can lead to uncertainties in detecting and classifying objects (Bolbot et al., 2023). Moreover, maintaining an accurate system state is imperative, as it forms the basis for all subsequent decisions (Sarker et al., 2023). Variability in operational conditions and the need for real-time responsiveness necessitate robust frameworks and methodologies capable of adapting to sudden changes and predicting future states.

In fact, several research works in the literature focus on robust decision-making methodologies capable of taking into account various types of uncertainties. For instance, Bayesian Networks (BN) provides a graphical model to comprehend the probabilistic relationship among a set of variables and manage uncertain information (Hegde et al., 2018; Junyung et al., 2021), while Dynamic Bayesian Networks (DBN) extend this capability by handling temporal dependencies between variables (Weber et al., 2012; Weber and Simon, 2016). Moreover, decision trees offer a simple and intuitive method to model decisions and their possible consequences, including outcomes, resource costs, and utility (Abaei et al., 2021). Lastly, Reinforcement Learning (RL) offers an interactive approach to learning an optimal policy for direct trial-and-error interaction with a dynamic environment (Kiran et al., 2021; Morato et al., 2023; Plissonneau et al., 2021). However, among these methodologies, Partially Observable Markov Decision Processes (POMDPs) have gained significant attention in the context of autonomous systems as described below.

5.2.2 Benefits of POMDP in decision-making processes

POMDPs have proven several advantages when dealing with the decision-making process. Firstly, POMDPs explicitly account for the uncertainty in both the system's state and the observations. This feature is essential in autonomous systems where sensor readings may not always be reliable or complete, and the actual state of the environment is hard-to-specify and hard-to-predict. Secondly, unlike methodologies such as decision trees that operate on discrete models, POMDPs can handle continuous states, actions, and observation spaces. This is particularly useful for autonomous systems where the environment is often better represented as a continuous space, such as the relative positions and speeds of

vehicles (Tran and Bae, 2021). Finally, while the RL is also a powerful tool for decision-making under uncertainty, it typically requires a large number of trials to learn the optimal policy, which may not always be feasible or safe in critical applications like autonomous trains. On the other hand, POMDPs offer a model-based approach that allows efficient policy computation based on the system's model.

In addition to their capability to address uncertainty, POMDPs can also model both the stochasticity in environment transitions and imperfect sensory information (Pouya and Madni, 2020). This dual capability becomes vital when dealing with real-time sensor data that inherently contains observational noise and varying environmental states.

A number of studies have focused on the use of POMDPs for various tasks related to the decision-making process, including dynamic probabilistic risk assessment (Maidana et al., 2022), cruise control of high-speed trains (Xu et al., 2019), collision avoidance in uncertain environments (Ragi and Chong, 2013), and behavior planning for autonomous vehicles (Pouya and Madni, 2020). In the field of robotics, POMDPs have also been applied for fault management in autonomous underwater vehicles (Svendensen and Seto, 2020). A survey by Lauri et al. (2022) provided a comprehensive overview of the use of POMDPs in robotics.

The literature also provides a range of algorithms and techniques for solving POMDPs, including online solvers (Ross et al., 2008), Monte-Carlo planning (Silver and Veness, 2010), and regularization methods (Somani et al., 2013). In addition, various tools and frameworks have been developed to aid in the modeling and analysis of autonomous system behavior using POMDPs, such as TAPIR (Klimenko et al., 2014), an online approximating and adapting software toolkit (Sunberg and Kochenderfer, 2017), and the Expandable-Partially Observable Markov Decision-Process Framework (Pouya and Madni, 2020). Equivalently, the use of Deep Reinforcement Learning (DRL) in combination with POMDPs has been gaining popularity in recent years. For example, Xiang and Foo (2021) explored the recent advances in DRL applications for solving POMDP problems in various fields, including transportation, industries, communication, and networking.

The above-mentioned papers highlight the various methods and techniques that have been developed to solve POMDPs in real-time and address the challenges of uncertain environments and dynamic parameters. Therefore, by using POMDPs, autonomous systems can make informed decisions that balance the trade-off between safety and efficiency (or even comfort), providing an important step toward the widespread adoption of autonomous systems.

5.3 Decision-making related to the train's anti-collision function

The anti-collision function represents the train capacity to detect and react appropriately and safely to any potential obstacles that could instigate a collision. Notice that the obstacles to be considered are physical entities, such as other trains, vehicles, individuals, trees, and so on. It is essential for an autonomous train to be outfitted with the necessary sensors and algorithms to accurately identify the nature of an obstacle, and estimate its distance from the train and its trajectory, in order to compute and evaluate the associated risk. To realize the anti-collision function, the ADS monitors the operational state of the train and its surrounding environment, constantly scanning for potential obstacles.

Figure 5.2¹ illustrates a scenario where an autonomous train, depicted in green, is approaching an intersection point in its track where the rail of another train merges. This is a potential area of conflict that the train's ADS recognizes and reacts to in a safer

¹This figure was generated using an AI-based image generation tool

manner. Furthermore, on the horizon, a car intersects the railway track, indicating a level-crossing scenario. A few individuals, along with their animals, are seen near the crossing, preparing to cross or possibly cross the railway track. This adds another layer of complexity to the scene, and the train's ADS must be capable of reacting to any potential obstacle and making decisions ensuring an acceptable safety level.

Moreover, the presence of trees alongside the rails is not merely an environmental feature in the figure. It signifies another set of potential risks, such as the danger of a fire, or the possibility of animals wandering onto the tracks from the forested areas. In such complex and unpredictable scenarios, the train's anti-collision function serves as the backbone, ensuring the safety of the autonomous train. It needs to efficiently process the potential risks arising from different aspects of the scenario (e.g., another train, humans and animals near the level crossing, cars, potential forest threats; etc.). The anti-collision function objective is not just to detect and identify these threats but also to measure the level of risk associated with each one so that the decision can be made based on the most updated and accurate risk information.

Figure 5.2 serves as a reminder of the vast array of potential risks that an autonomous train might face, and how a robust, dynamic, and real-time risk assessment based on the anti-collision function can play a critical role in ensuring the safe operation of the train.



Figure 5.2: Generic illustration of the anti-collision function

5.3.1 DRA of the anti-collision function

Given the uncertainties associated with real-world environments and sensor information, the observations help to form an *uncertainty estimation*. This estimation is a probabilistic representation of the current situation of the train, summarizing possible states of the train and its surroundings. Once the uncertainty estimation is established, the risk assessment

inherent to the anti-collision function should be carried out. This refers to the DRA task, which has to be performed by the ADS to evaluate and update the level of risks associated with the current state of the train, the environment, and the available actions the train might take. This assessment is based on uncertainty estimation, considering both the likelihood and potential consequences of a collision. In addition, the uncertainty estimation plays an important role in establishing the risk profile, as it provides the probabilistic basis from which potential hazardous scenarios and their associated risks are assessed, and classified within the risk profile.

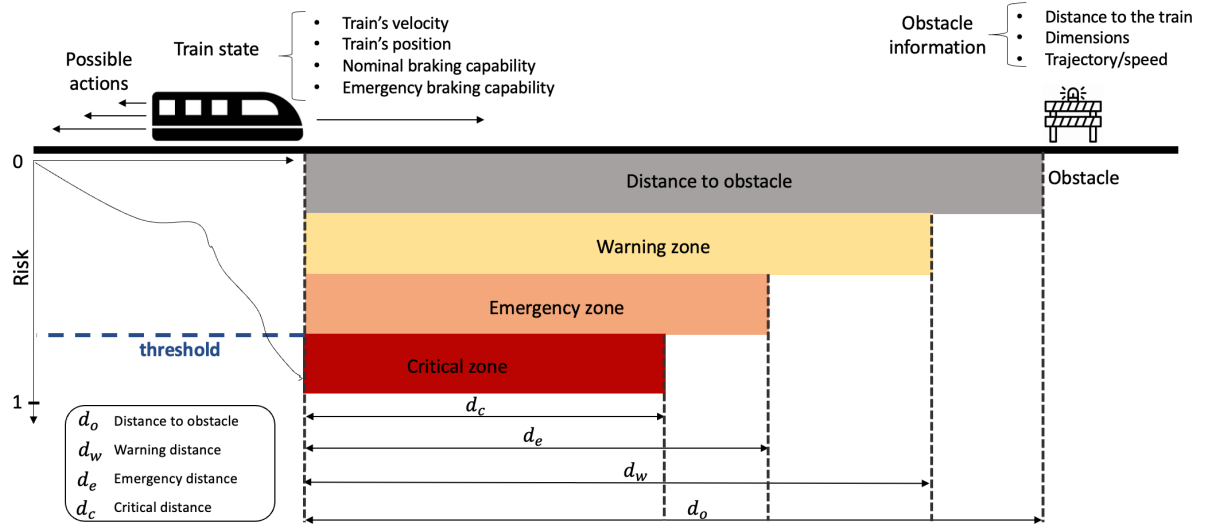


Figure 5.3: Illustrative representation of the anti-collision function

Figure 5.3 shows an illustrative scenario involving an autonomous train and a potential obstacle in its track. Different control actions are available for the train, in response to the surroundings and with respect to the criticality of the evaluated risks, namely, accelerating, maintaining the current speed, and various types of braking. In Figure 5.3, the obstacle is located at a certain distance on the track of the train. With respect to the distance from the train, three zones are considered: warning, emergency, and critical zones. The *warning* zone (in yellow color) indicates a distance from where no immediate action is needed (i.e., the obstacle is so far or not detected yet). The *emergency* zone (in orange color) signifies a cautionary distance from where the train may need to adjust its speed or brake in order to avoid a collision. Finally, the *critical* zone (in red color) signifies that the presence of an obstacle can lead to a collision (i.e., in this zone, the obstacle is considered close to the train, and even with an emergency braking the risk of collision is high). The associated risk level, represented on the vertical axis with a scale between 0 and 1, is estimated according to the distance to the obstacle. Obviously, the closer the obstacle is to the train, the higher the risk level is. The threshold to reach the unacceptable risk level (visualized in the figure by the intersection between the blue dashed line with the vertical axis) is crossed when the train crosses the *critical* zone.

Note that, the DRA task must not only lead to a safe reaction of the ADS when a collision risk is identified, but also learn from every decision made. The consequences of each decision have to be monitored and analyzed to understand the effectiveness of the actions taken. This feedback loop allows the system to continuously adapt and evolve, improving its performance over time. Therefore, the anti-collision function, performed by the DRA, acts as a dynamic learning and protection layer, ensuring a higher level of safety

in the operation of autonomous trains.

5.3.2 Structuring risk profiles with the DRA framework

The proposed framework, presented in the previous chapter, provides a structured approach to decision-making, taking into account the train state uncertainty and the perception of the environment. The DRA framework is designed to take into account the various factors that influence the decision-making process (cf. Figure 5.4). This includes the train’s speed, the distance to the obstacle, and the perception of the environment, among other internal and external factors.

This framework enables the ADS to perform a real-time evaluation and prediction of potentially hazardous situations by estimating their occurrence probabilities and severity. It exploits not only the information collected from the perception module but also translates this information into an actionable risk profile². This profile then guides the decision-making process to efficiently determine the appropriate and safe actions necessary to avoid or mitigate the impact of hazards. Therefore, the integration of risk profiles into a DRA framework for autonomous trains allows for the real-time management and mitigation of risks, thereby improving overall safety in autonomous train operations.

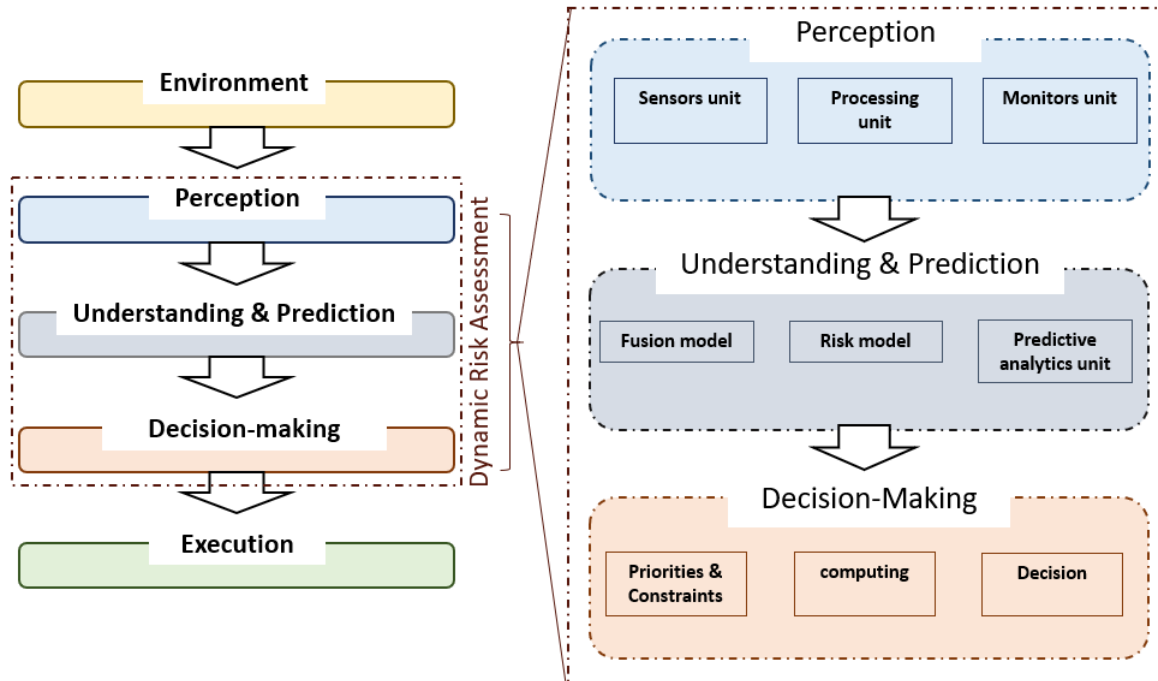


Figure 5.4: The autonomous train dynamic risk assessment framework

In the present chapter, our focus is on the Understanding & Prediction and Decision-making modules in the case of the anti-collision function. The Understanding & Prediction module utilizes the information provided by the Perception module to create and continually update an integrated real-time model that represents the system environment and its states. This model is subsequently utilized for run-time decision-making. From the perspective of DRA, this module enables the computation of a current risk estimate and

²According to (Kumamoto and Henley, 1996), a risk profile is defined as an “outcome, likelihood, significance, causal scenario, and population affected [are factors that] determine the risk profile”. The purpose of the risk profiles is to understand better which scenarios are relatively riskier (i.e., how the risks are compare to each other.)

the prediction of potential railway hazards. Subsequently, this risk estimate is evaluated through the risk model. This risk model for anti-collision purposes integrates both historical data, which reflects past system performance and incidents, with real-time sensor information to enhance the accuracy of potential collision predictions. Moreover, it evaluates several parameters, such as the train's current speed, position, and braking capabilities as well as the positions and velocities of detected obstacles. By continuously updating these parameters in real-time the model is able to adjust and update the risk estimates associated with each potential action and thus assists in selecting the safest action for the autonomous train.

5.4 Methodology

In this section, we first recall the preliminary definitions and notions of POMDP, and then, we describe the different components of the POMDP model for the train's anti-collision decision-making process.

5.4.1 POMDP definition

A POMDP is a probabilistic method that models the sequential process of a system under uncertainty. It is a generalization of Markov Decision Process to situations where the system state is partially unknown. Formally, a POMDP is a tuple $\langle S, A, O, T, Z, R, \gamma \rangle$, where S and A are the sets of states and actions, T is the transition function that defines the conditional probability P of moving from one state $s \in S$ to another state $s' \in S$ as a result of executing an action $a \in A$, i.e., $T(s, a, s') = P(s' | s, a)$. O is the observation space that defines the information received (from sensors) after the execution of an action. Z is the corresponding observation function that defines the conditional probability of observing a particular outcome $o \in O$ after executing an action $a \in A$ to reach to state $s' \in S$, i.e., $Z(o, a, s') = P(o | s', a)$. R is the reward function $R(s, a)$ that defines the immediate reward received for being in a particular state $s \in S$ and taking a particular action $a \in A$. Finally, $\gamma \in [0, 1]$ is the discount factor that determines the relevance (or not) of future rewards.

In a POMDP, only partial and noisy knowledge of the system and its environment is considered; thus, a belief about the model states, known as a belief state $b(s)$, is continually inferred. The belief state is a probability distribution over the state space that reflects the degree of certainty maintained by the POMDP model about the current state of the system. Accordingly, a policy $\pi : B \rightarrow A$ is used as a mapping from the set of possible belief states to the set of actions, in order to determine the adequate action that should be taken.

Solving a POMDP involves finding the optimal policy π^* in terms of current action or finite sequence of actions to be executed in order to maximize (or optimize) the expected cumulative reward over time, taking into account the belief state. Formally,

$$\pi^*(b) = \operatorname{argmax}_{a \in A} \left\{ \sum_{s'} P(s' | b, a) [R(b, a, s') + \gamma \cdot E[V^*(b')]] \right\} \quad (5.1)$$

To evaluate the potential reward of taking an action a and transitioning to state s' , equation 5.1 considers the probability $P(s' | b, a)$ of transitioning to state s' given the current belief state b and the action a taken. It also accounts for the immediate reward $R(b, a, s')$ obtained from the action a in the belief state b and transitioning to state s' . Moreover, the equation considers the expected value (i.e., expected reward) of the optimal value function, $E[V^*(b')]$ for the next belief state b' resulting from the transition to state s' . This component accounts for the potential future rewards and outcomes taking action a .

The optimal policy in a POMDP can be computed using two main categories of solvers: *online* and *offline* solvers. These solvers differ in the way they find the optimal policy and the computational resources they require. Online solvers are designed to run in real-time and make decisions based on the current state of the system, while offline solvers are designed to run offline and make decisions based on historical data. The choice of the solver depends on the specific use case and the computational resources available.

5.4.2 POMDP modeling of the train anti-collision system

Train anti-collision system modeling

The anti-collision system takes as input internal information regarding the train state, and external information about the environment. As explained in subsection 5.3.2 presenting the DRA framework, the internal inputs encompass sensor information about the train position and velocity (generally, provided by the localization and the speed measuring modules), as well as nominal and emergency braking (i.e., deceleration) capabilities, which can be transformed into the nominal and emergency distances to stop the train. On the other hand, external inputs refer to information about the surrounding obstacles (coming from the perception module), including their positions, dimensions, velocity, and intentions (for moving obstacles). The output of the system is the adequate control action (or sequence of actions) to be taken in order to avoid (when possible) any collision with the detected obstacle. Figure 5.5 presents a general view of the POMDP input-output structure used to implement the anti-collision function.

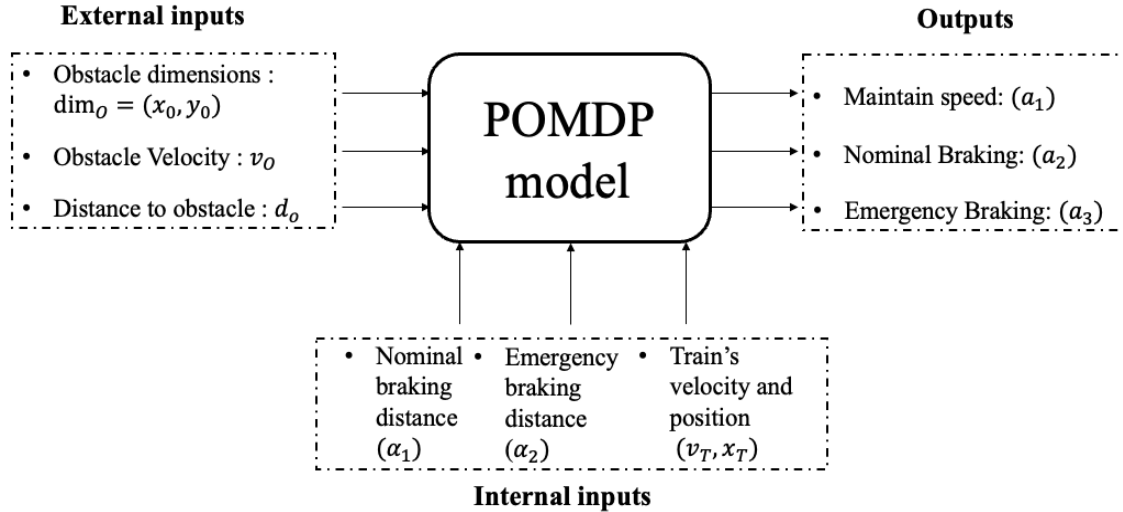


Figure 5.5: A generic illustration of the POMDP model.

The continuous state-space of the POMDP model includes the state of the train and the states of the (possibly) surrounding obstacles. The state of the train s_T contains its position (x_T, y_T) , its velocity v_T , and its orientation θ_T . Similarly, the state of each obstacle s_i is composed of its position (x_i, y_i) , its dimension \mathcal{D}_i , its instantaneous speed (v_{xi}, v_{yi}) , and its orientation θ_i . It is worth noticing that such a formulation of the state space is performed on a global (or earth) coordinate system. An arbitrary point on the track is chosen as the origin of the coordinate system. Notice that several coordinate systems can be considered, as local and relative systems (See [Temizer et al. \(2010\)](#); [Leurent \(2018\)](#) for more details).

While the continuous formulation of the state space is a faithful representation of the real system, it remains a very high-dimensional continuous space, which requires significant

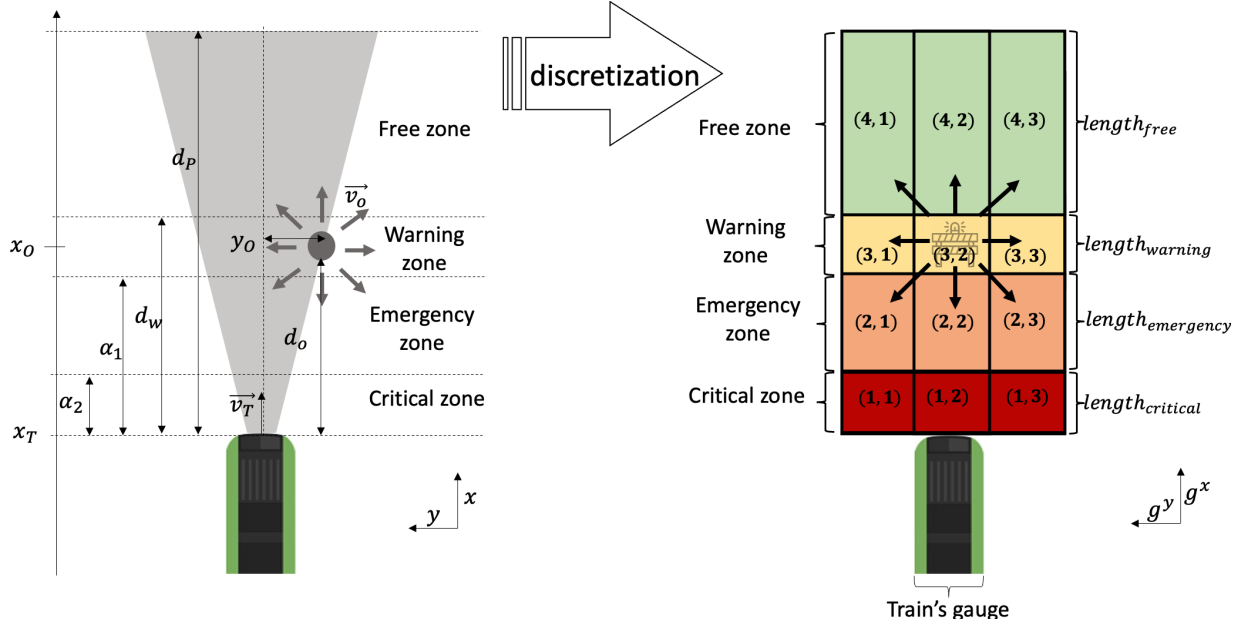


Figure 5.6: A generic spatial discretization of Cartesian plan into adaptive grid map for autonomous train navigation

computation time and space to solve the model and find the adequate policy. Moreover, the existing algorithms to solve the continuous POMDP do not scale well when it comes to high-dimension continuous models. In order to remedy this issue, we consider a discrete POMDP with a discrete representation of state space, action space, and observation space.

Modeling the discrete state space

The discretization of the state space is performed using a two-dimensional adaptive grid fixed to the head of the train. Thus, a local coordinate (egocentric) system with the head of the train as system origin is considered. This means that instead of explicitly representing the positions of the obstacles as continuous variables within the model states, they are represented implicitly through several variables indicating the occupancy or not of the grid cells. The positive x -axis is in the direction of train driving, and the positive y -axis is directed to the left of the train head. Notice that the adaptive grid cell size is dependent on the tangible braking capabilities of the train, the presence of obstacles in (or alongside) the track, and the gauge of the train.

Figure 5.6 presents a two-part illustration from a real-world scenario of the adaptive grid map. The first part (*on the left of the figure*) shows a train moving along its track with an obstacle appearing in its path, visualized using a global coordinate system. The second part of the illustration (*on the right of the figure*) depicts the adaptive grid map resulting from this discretization process. Furthermore, On the right side, the concept of discretization is shown. This is represented by a grid overlay on the track, with the grid cells numbered in parentheses. The cells are color-coded consistent with the zones described on the left: green for the free zone, yellow for the warning zone, orange for the emergency zone, and red for the critical zone. This grid represents a method for discretizing the continuous space around the train into manageable sections for the anti-collision system to evaluate risk more effectively. This discretization allows the transfer from the global coordinate system to an adaptive grid map. The lengths of each zone in this adaptive grid map are indicated on the right side of the grid as $length_{free}$, $length_{warning}$, $length_{emergency}$, and $length_{critical}$. The train's gauge, which is the width of the train or the tracks, is also noted at the bottom of the grid. In fact, figure 5.6 illustrates our

approach to risk quantification, which, at first glance, emphasizes proximity and braking distances. However, the model's architecture inherently accommodates additional critical parameters. Lateral position is factored into the discretized grid map, where each cell corresponds to a specific lateral and longitudinal zone relative to the train, allowing us to account for the lateral positioning of obstacles. Moreover, obstacle velocity is incorporated into the risk assessment through dynamic cell updates that reflect the changing positions of obstacles over time. This enables the system to anticipate and react to moving obstacles, with a higher risk attribution for those with significant relative velocity towards the train.

The adaptive grid map is structured as a 12-cells grid, where each cell is defined based on the relative position of the obstacle (g^x, g^y) , and its relative discrete orientation θ^d . Notice that the orientation of the obstacle is determined based on its velocity projections (v_x, v_y) (or its angular velocity ω_o), and represents the possible transitions to the eight surrounding grid cells, i.e., $\theta^d \in \{0, \frac{2\pi}{8}, \frac{4\pi}{8}, \frac{6\pi}{8}, \pi, \frac{10\pi}{8}, \frac{12\pi}{8}, \frac{14\pi}{8}\}$.

Thus, the state set \mathcal{S} can be expressed as follows:

$$\mathcal{S} = \begin{cases} g^x, & \text{with } g^x \in \{1, 2, 3, 4\} \\ g^y, & \text{with } g^y \in \{1, 2, 3\} \\ \theta^d, & \text{with } \theta^d \in \{0, \frac{2\pi}{8}, \frac{4\pi}{8}, \frac{6\pi}{8}, \pi, \frac{10\pi}{8}, \frac{12\pi}{8}, \frac{14\pi}{8}\} \end{cases} \quad (5.2)$$

The variable g^x represents the discretization of the obstacle's position in the x -axis and can take four values $\{1, 2, 3, 4\}$, corresponding to the number of lines in the grid. The variable g^y represents the discretization of the obstacle's position in the y -axis and can take three values $\{1, 2, 3\}$, corresponding to the number of columns in the grid. Additionally, the variable θ^d represents the orientation of the obstacle and is discretized from a continuous space (from 0 to 2π) to eight discrete values $\{0, \frac{2\pi}{8}, \frac{4\pi}{8}, \frac{6\pi}{8}, \pi, \frac{10\pi}{8}, \frac{12\pi}{8}, \frac{14\pi}{8}\}$, representing the possible transitions to the eight surrounding cells. In fact, each unique combination of g^x , g^y , and θ^d represents a distinct state in the adaptive grid map, indicating the position and orientation of the obstacle (see Figure 5.6). With four possible values for g^x , three possible values for g^y , and nine possible values for θ^d , the total number of possible states in the adaptive grid map is $N_{\mathcal{S}} = 4 \times 3 \times 8 = 96$. These 96 states capture all the possible configurations of an obstacle within the adaptive grid map, enabling the POMDP model to effectively reason about its movement and potential interactions with the train in real-world scenarios.

In order to establish the sizes of each cell in the adaptive grid map, the next step of the discretization process is the definition of different zones (*Free*, *Warning*, *Emergency* and *Critical* zones). The boundaries of each zone are determined as functions of the nominal and emergency braking distances α_1 and α_2 . In fact, the length of the cells in *Critical*, *Emergency*, and *Warning* zones are respectively equal to the emergency braking distance ($length_{critical} = \alpha_2$), the nominal braking distance ($length_{emergency} = \alpha_1 - \alpha_2$), the distance to the obstacle ($length_{warning} = d_o - \alpha_1$) with d_o equivalent to g_o^x (i.e., $d_o = x_o - x_T$ in the global coordinate system). Additionally, the length of the *free* zone cells is determined by the maximal perception distance (or the perception range) d_p of the autonomous train ($length_{free} = d_p - (\alpha_1 + d_m)$). On the other hand, the width of all cells in the adaptive grid map is equal to the gauge of the train. Equation 5.3 shows the boundaries of each zone :

$$\begin{cases} Freezone & = \{(g^x, g^y, \theta^d) \in \mathcal{S} \mid \text{for } g^x = 4; g^y = 1, 2, 3\} \\ Warningzone & = \{(g^x, g^y, \theta^d) \in \mathcal{S} \mid \text{for } g^x = 3; g^y = 1, 2, 3\} \\ Emergencyzone & = \{(g^x, g^y, \theta^d) \in \mathcal{S} \mid \text{for } g^x = 2; g^y = 1, 2, 3\} \\ Criticalzone & = \{(g^x, g^y, \theta^d) \in \mathcal{S} \mid \text{for } g^x = 1; g^y = 1, 2, 3\} \end{cases} \quad (5.3)$$

These zones include the *Free* zone where no obstacle is detected, the *Warning* zone

where an obstacle is present but can be avoided by a nominal braking, the *Emergency* zone where an obstacle can only be avoided by an emergency braking, and the *Critical* zone where an obstacle cannot be avoided and a collision is imminent. In fact, in the adaptive grid map, each zone consists of three cells, resulting in a total of 12 cells.

From a safety perspective, if an obstacle is in one of the three cells within each zone, whatever the speed of the obstacle compared to the speed of the train, and knowing that its orientation is toward a lateral direction (i.e., $\theta^d = 0$ or π , meaning that the next obstacle state will remain in the same zone), the associated level of risks can be considered to be similar for the autonomous operation. If the obstacle orientation is forward (i.e., $\theta^d = \frac{2\pi}{8}$ or $\frac{4\pi}{8}$ or $\frac{6\pi}{8}$) or backward (i.e., $\theta^d = \frac{10\pi}{8}$ or $\frac{12\pi}{8}$ or $\frac{14\pi}{8}$), the risk will respectively decrease (only if $v_o \geq v_T$) or increase (only if $v_o > 0$). In order to define POMDP states with comprehensible risk levels, we adopt the following assumptions.

Assumptions for defining risk levels

It can be observed that most of the 96 states from the adaptive grid map can exhibit similar safety implications. In particular, the three cells within each zone can be related to a similar level of risk. In other words, multiple states might present an analogous level of risk for autonomous train operations. Such similarities across various states can be attributed to factors such as the immediate threat an obstacle can raise, the available reaction time for the train, and the potential consequences of inaction. Rather than distinguishing among these numerous states, which might only offer marginal differences in the actual risk, it appears to be more pragmatic and efficient to aggregate them based on their overall risk level. This not only streamlines the decision-making process but also ensures clarity in defining distinct risk levels.

Moreover, in the initial simulation setup described herein, it is assumed that obstacles detected by the autonomous train's perception unit are static (i.e., $v_o = 0$) in the immediate environment. This assumption simplifies the predictive aspect of obstacle movement and trajectory, allowing the decision-making process to forgo consideration of these dynamics. Consequently, the orientation (θ^d) of the obstacles is not taken into account when transitioning to discrete safety states. The focus is primarily on identifying obstacles and gauging their proximity to the train (i.e., the distance to obstacle d_o). In contrast, the second simulation setup advances this model by integrating the velocity of obstacles and their nature (i.e., static or dynamic). This not only reflects a more realistic operational scenario but also challenges the system to account for the additional complexity in its risk assessment and decision-making algorithms. Furthermore, developing two simulation setups highlights the adaptability of the approach, showcasing its capacity to integrate multiple factors, whether they are external factors related to obstacles or internal factors associated with the train itself.

Based on the outlined considerations, we have identified four discrete states. This delineation is not just a reduction, but a methodical classification and categorization based on the risk levels that several states in the adaptive grid map might be associated with. This structured approach provides a clear representation of collision risks, facilitating an efficient response by the autonomous train system to safety-critical situations. The specifics of these four states are detailed in Equation 5.4.

Finally, the state *Safe* indicates that no obstacle is detected in the train's surroundings. This situation applies to the *Free* zone, where the distance to obstacle $d_O \rightarrow \infty$. Conversely, the *ObstacleDetected* state signifies that an obstacle is located in the *Warning* zone. In this zone, the obstacle can be avoided by nominal braking. However, if the obstacle breaches the *Emergency* zone, the state switches to *AboutToCrash*. This state represents a significant risk that necessitates the immediate application of emergency braking

$$\mathcal{S} = \begin{cases} s_1 \\ s_2 \\ s_3 \\ s_4 \end{cases} = \begin{cases} \textit{Safe} & = \{(g^x, g^y, \theta^d) \mid g^x = 4, \forall g^y, \theta^d\}, \\ \textit{ObstacleDetected} & = \{(g^x, g^y, \theta^d) \mid g^x = 3, \forall g^y, \theta^d\}, \\ \textit{AboutToCrash} & = \{(g^x, g^y, \theta^d) \mid g^x = 2, \forall g^y, \theta^d\}, \\ \textit{Crash} & = \{(g^x, g^y, \theta^d) \mid g^x = 1, \forall g^y, \theta^d\} \end{cases} \quad (5.4)$$

to prevent a collision. Finally, the *Crash* state denotes the situation where the obstacle is located in the *Critical* zone, and a collision is inevitable despite any measures.

Modeling the action space

The dynamic behavior of the train is mainly controlled by the continuous action of acceleration (and intrinsic deceleration and braking). To simplify the model, we consider a discretization of the acceleration space into three discrete values $\mathcal{A} = \{a_1, a_2, a_3\}$, which represent respectively: maintaining the speed, nominal braking, and emergency braking.

It is worthwhile noticing that, in the context of obstacle avoidance, the (positive) acceleration action can also be considered. This action is generally taken in the case of hazardous situations related to fires in the track or the presence of smoke in tunnels. In this study, such a kind of situation is not considered.

Modeling the observation space

The observation space, denoted \mathcal{O} , is defined as the set of possible observations that the autonomous train can make at each time step. In fact, all observable variables constructing the observation space, such as train position and velocity, can be updated directly from sensor measurements. Noise in these sensor measurements can also be taken into account during observation and belief updates. In our case, the observation space comprises the obstacle's position in the adaptive grid map, represented by the variables g^x and g^y . This representation captures the relative location of the obstacle with respect to the train's position and enables the assessment of potential collision risks. Thus, two observations are defined in the following set :

$$\mathcal{O} = \begin{cases} g^x, & \text{with } g^x \in \{1, 2, 3, 4\} \\ g^y, & \text{with } g^y \in \{1, 2, 3\} \end{cases} \quad (5.5)$$

Modeling the transition function

Based on the probability distribution of the initial (or current) state of the model, at each step time δt , an action is taken and probability distribution over the state space is updated according to the transition function model $T(s, a, s') = P(s' \mid s, a)$. The transition function depicts the dynamic behavior of the train and obstacles after each step time δt . We consider v_T , x_T , and a_{cc}^T being the train velocity, position, and acceleration respectively, with the time sample δt . The following equation shows the train's transition model (i.e., train's dynamics) in the global (or earth) coordinate system:

$$\begin{bmatrix} v_T(t + \delta t) \\ x_T(t + \delta t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \delta t & 1 \end{bmatrix} \cdot \begin{bmatrix} v_T(t) \\ x_T(t) \end{bmatrix} + \begin{bmatrix} \delta t \\ \frac{\delta t^2}{2} \end{bmatrix} \cdot a_{cc}^T(t) \quad (5.6)$$

Similarly, the obstacle's transition model (i.e., obstacle's dynamics) in the global coordinate system is described as follows:

$$\begin{bmatrix} v_o(t + \delta t) \\ x_o(t + \delta t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \delta t & 1 \end{bmatrix} \cdot \begin{bmatrix} v_o(t) \\ x_o(t) \end{bmatrix} \quad (5.7)$$

Notice that in the case of the obstacle's transition model, the acceleration is not considered. In addition, we assume that the transitions are deterministic and the obstacle remains static in time. The new distance to obstacle d_o after a time step (i.e., the obstacle's transition model) in the global coordinate system is represented by the following equation:

$$d_o(t + \delta t) = d_o(t) - v_T(t) \cdot \delta t - a_{cc}^T(t) \cdot \frac{\delta t^2}{2} \quad (5.8)$$

However, the distance to the obstacle in the local coordinate system (adaptive grid map) is defined as follows:

$$d_o(t) \approx g_o^x(t) \quad (5.9)$$

Modeling Observation function

The main objective of the observation function $Z(o|s, a)$, in this case, is to calculate the distance traveled by the train after a time step, in the global coordinate system. This distance allows keeping track of the new distance to the obstacle in each action selected from the action space.

$$d_T^{traveled}(t + \delta t) = v_T(t) \cdot \delta t + a_{cc}^T(t) \cdot \frac{\delta t^2}{2} \quad (5.10)$$

The new distance to the obstacle, after a time step, becomes :

$$d_0(t + \delta t) = d_0(t) - (v_T(t) \cdot \delta t + a_{cc}^T(t) \cdot \frac{\delta t^2}{2}) \quad (5.11)$$

Similarly, the orientation of the obstacle can be updated, at each time step δt , based on the obstacle's orientation at the previous time step and the obstacle's angular velocity ω_o . Thus:

$$\theta_{t+\delta t}^d = \theta^d(t) + \omega_o \cdot \delta t \quad (5.12)$$

Notice that equation 5.12 assumes that the obstacle's angular velocity (ω_o) remains constant over the time step δt . This fits the assumption made previously that the obstacle remains static in the global coordinate system. In fact, the static obstacle's position in the global coordinate system corresponds to a constant angular velocity in the local (or relative) coordinate system (i.e., adaptive grid map).

Reward function

The reward function is in the form of costs (or negative rewards), assigned to each decision (action) made by the model within a specified state (Temizer et al., 2010). The role of the reward function is to encourage decisions that advance the system's goals while imposing penalties on those that do not. Whilst the primary objective of the anti-collision system is to prevent train collisions, it remains desirable to consider other secondary objectives, such as respecting the timetable schedule, maintaining a smooth velocity, etc.

For the primary objective, negative rewards (i.e., penalties) are assigned to states that are considered unsafe, such as those that have a high probability of collision with an obstacle (e.g., the *Crash* state). By assigning higher negative rewards to riskier states, the

ADS can be incentivized to take safer actions and avoid collisions. This reward adaptation according to risk embodies the risk model mentioned in section 5.3. It can be updated in real-time as new information about the environment becomes available, allowing the system to continuously adapt to changing conditions and maintain a safe operation. Moreover, the reward function assigns numerical values to each state-action pair to simulate the desired behavior. In our case, the main objective of the system is to avoid when possible (i.e., minimize the risk of) the train collisions. Hence, we define an important penalty to the train to be in the *Crash* state (s_4), another penalty for the *AboutToCrash* state (s_3), and a reward for being in the *Safe* state (s_1). The reward function is represented by equation 5.13 :

$$\mathcal{R}(s) = \begin{cases} 10, & \text{if state } s = s_1 \text{ (Positive Reward)} \\ -10, & \text{if state } s = s_2 \text{ (Minor Penalty)} \\ -100, & \text{if state } s = s_3 \text{ (Moderate Penalty)} \\ -1000, & \text{if state } s = s_4 \text{ (Severe Penalty)} \end{cases} \quad (5.13)$$

One important consideration when designing the risk model for the ADS is the trade-off between safety and efficiency. In particular, for states such as *ObstacleDetected* and *AboutToCrash* (i.e., s_2 and s_3), the reward function should balance the desire to avoid collisions with the need to maintain efficient driving behavior. Assigning overly negative rewards/penalties to these states may cause the system to become overly cautious and overly slow, which can lead to inefficient or impractical driving behavior. On the other hand, assigning insufficiently negative rewards (i.e., penalties) may lead to unsafe driving behavior, where the system takes risky actions in order to maintain high efficiency. Finding the right balance between safety and efficiency is a key challenge in designing the risk model for the autonomous driving system. For instance, we established the reward function as follows:

The method described in this section serves as the basic framework for conducting simulations and presenting the results in Section 5.5.

Choice of POMDP solver

The POMDP problem can be implemented using a dynamic programming algorithm, such as value iteration (Zhang and Zhang, 2001). The algorithm takes into account, as discussed above, the current state of the train and the observed distance to the obstacle, and generates outputs as optimal action to take. The algorithm operates by updating a value function that represents the expected long-term reward from each state. At each iteration, the value function is updated using a Bellman equation (Jaakkola et al., 1994) (see equation 5.14) that takes into account the transition probabilities, rewards, and discount factor.

$$V^*(s) = R(s) + \gamma \sum_{s' \in S} P(s' | a) \sum_{O \in O} P(o | a, s') V^*(s') \quad (5.14)$$

where V^* is the value function of the selected policy.

In order to effectively solve the anti-collision problem for the autonomous train, it is necessary to select a suitable POMDP solver. In this study, we provide a comprehensive overview of the most commonly used POMDP solvers. We characterize each solver in terms of the nature of its state space, action space, and observation space, as well as whether the solvers operate in an online or offline manner. These key characteristics are presented in Table 5.1.

In this application, the state space, action space, and observation space are discrete. Moreover, an online approach is needed to evaluate the train environment in real-time to

solver	abbreviation	offline	online	state space	action space	observation space
Q-Markov Decision Process	QMDP	x	-	Discrete	Discrete	Discrete
Fast Informed Bound	FIB	x	-	Discrete	Discrete	Discrete
Belief Grid Value Iteration	BGVI	x	-	Discrete	Discrete	Discrete
Successive Approximations of the Reachable Space under Optimal Policies	SARSOP	x	-	Discrete	Discrete	Discrete
Basic Partially Observable Monte Carlo Planning	Basic POMCP	x	x	Continuous	Discrete	Discrete
Point Based Value Iteration	PBVI	-	x	Discrete	Discrete	Discrete
Anytime Error Minimization Search	AEMS	-	x	Discrete	Discrete	Discrete
Anytime Regularized-DEterminized Space Partially Observable Tree	AR-DESPOT	x	x	Continuous	Discrete	Discrete
Partially Observable Monte Carlo Planning with Observation Widening	POMCPOW	-	x	Continuous	Continuous	Continuous
Monte Carlo Value Iteration	MCVI	x	-	Continuous	Discrete	Continuous

Table 5.1: POMDP solvers choice

avoid collisions with obstacles. After evaluating the various POMDP solvers in terms of their state space, action space, and observation space, as well as their online/offline nature, the study found that Anytime Error Minimization Search (AEMS) and Point-Based Value Iteration (PBVI) solvers are the most suitable options for the obstacle detection and avoidance problem (Pineau et al., 2003). Of these, PBVI was selected as the solver of choice for the simulation in this study due to its efficient and effective representation of the POMDP model.

5.5 Simulation and results

In this section, we provide a detailed description of the experimental set-up, elaborate on the process of variable initialization, and present the simulation results.

The simulations established in this chapter provide insights into the decision-making processes of the autonomous train, with a particular emphasis on ensuring safety and an effective anti-collision function. We present two simulation scenarios: the original, based on the POMDP model that takes only the distance to an obstacle as input, and an advanced setup that integrates the velocities and the nature of obstacles (i.e., static or dynamic obstacles). These simulations collectively offer a way to evaluate the system’s performance under controlled yet realistic conditions, negating the risk and financial implications associated with real-world testing. This process’ practical use includes essential components each with an important role in the simulation process:

5.5.1 Perceived state

The perceived state is crucial for connecting the real and simulated environments. In fact, observed distance and perceived obstacles in this simulation are subject to Gaussian noise, emulating uncertainties inherent to real-world sensing. The train’s next action is decided based on the perceived state, derived from these noisy observations and not from the actual state.

5.5.2 Obstacle generation function

In the simulation model used in this chapter, obstacles are generated stochastically in the train’s path. The appearance of an obstacle is determined by a random function, occurring approximately 20% of the time, with the distance to a new obstacle drawn from a uniform distribution. This obstacle-generation process introduces diversity into the simulation

and allows testing of the reliability of the train’s decision-making in various situations. Moreover, obstacles are generated, following a uniform distribution, between the mean of the nominal and emergency braking distances (α_1 and α_2) and 50 meters beyond this mean respectively. This ensures that the obstacles are generated within a reasonable range of distances where the autonomous train could have a fair chance to detect them and react appropriately.

This choice of obstacle generation provides a balance between the extremes of having all obstacles too close, which might not provide sufficient reaction time for the train, and having them too far, which might not pose any real danger or challenge to the train’s ADS.

5.5.3 Belief updater

The belief updater is a critical component of the model. It retains a distribution over potential states the autonomous train may occupy, integrating the actual state, perceived state, and actions taken. The belief state is generated for each time step, playing an essential role in handling uncertainties in the system and enabling more robust decision-making. The belief update equation is given by:

$$b'(s') = \eta \cdot P(o|s', a) \cdot \sum_{s \in \mathcal{S}} P(s'|s, a) \cdot b(s) \quad (5.15)$$

In equation 5.15, η is the normalization constant to ensure that the updated belief state b' is a valid probability distribution (i.e., sums to 1 over all states). $b(s)$ and $b'(s)$ are the probability of being, respectively, in the current state belief state s and the updated belief state s' .

This equation updates the belief about the current state after taking an action a and observing an outcome o . The new belief $b'(s')$ is proportional to the likelihood of the observation o given that we end up in state s' , times the sum of the probabilities of reaching s' from all possible states s under an action a , weighted by the current belief about being in the state s .

5.5.4 Solver choice

For this problem, a Point-Based Value Iteration (PBVI) algorithm (Pineau et al., 2003; Spaan and Vlassis, 2005) is employed as the solver due to its efficiency and compatibility with problems possessing small, finite discrete state and action spaces. The PBVI solver iteratively optimizes the value function, updating the maximum expected reward for each state-action pair over a number of iterations. The resulting policy, which assigns actions to states, is extracted from this optimal value function. Equation 5.16 shows how the PBVI works:

$$V_{n+1}(b) = \max_{a \in \mathcal{A}} \left[R(s) + \gamma \sum_{o \in \mathcal{O}} P(o|b, a) \max_{\alpha \in \Gamma_n} \sum_{s \in \mathcal{S}} \alpha(s) b'(s) \right] \quad (5.16)$$

In this equation, $V_{n+1}(b)$ represents the value of belief state b at the $n + 1$ iteration. $R(b, a)$ is the expected immediate reward for taking an action a in belief state b . In addition, $\alpha(s)$ represents the value of state s for α -vector (defined below). Finally, the $\max_{\alpha \in \Gamma_n}$ operation selects the α -vector that yields the highest value for the updated belief state b' .

The aim of PBVI is to find an approximate solution of the POMDP by computing a set of α -vectors. Each α -vector corresponds to a specific action and provides a mapping from the state space to real numbers. In each iteration, the α -vectors are updated according to the equation 5.16 to improve the value function approximation. The algorithm continues until a termination condition is met, such as a maximum number of iterations or a minimal

improvement threshold. In the simulation established in this chapter, the condition is related to the maximum number of iterations.

5.5.5 Variables initialization

Before the simulation is run, all necessary variables associated with the states, actions, and policy are initialized. Initial settings for the train's position, speed, and distance from the obstacle are also established. As the simulation progresses, the position and speed are continuously updated according to the chosen action and the train's current state. These initial values provide a baseline from which the train learns to make optimal decisions (see Table 5.2).

Variable	Initial Value	Unit
Initial train speed	40	$m.s^{-1}$
Initial train position	0	m
Nominal braking distance (α_1)	300	m
Emergency braking distance (α_2)	100	m
Time sample	0.1	s
Rewards $[r_{s_1}, r_{s_2}, r_{s_3}, r_{s_4}]$	$[10, -10, -100, -1000]$	-
Actions forces $[a_1, a_2, a_3]$	$[0, -1, -3]$	$m.s^{-2}$
Discount factor (γ)	0.95	-

Table 5.2: Variables initialization

5.5.6 Risk formulation

Once the environment is perceived, the next step is the risk estimation. Here, possible scenarios that can lead to unsafe conditions/collisions are identified and their probability is estimated based on current and predicted states. This involves the identification of potential hazards, assessment of their possible impact, and the calculation of the risk associated with each hazard. To this end, the risk is calculated in two manners, as described in the following equations :

$$R_1 = 1 - \frac{1}{1 + \exp(-5 \cdot \frac{d_o}{\alpha_1})} \quad (5.17)$$

$$R_2 = \frac{\alpha_1 - d_o}{\alpha_1 - \alpha_2} \quad (5.18)$$

Equation 5.17 utilizes a logistic function to present the scenario where risk is relatively low when the train is far from the obstacle ($d_o > \alpha_1$). The use of the logistic function offers a smooth and sigmoidal transition from a low-risk state to a high-risk state. This feature is ideal for representing scenarios where risk is initially low but increases as the train approaches the obstacle, and eventually saturates as the obstacle gets very close. Additionally, this characteristic caters to the fact that when the obstacle is far enough, the train has enough time to react, and the risk is low. On the other hand, when the obstacle is very close ($d_o < \alpha_2$) the train could have already engaged its emergency braking, implying that it has already acknowledged the risk and is attempting to mitigate it.

Equation 5.18 linearly increases the risk as the train gets closer to the obstacle, from the nominal braking distance (α_1) to the emergency braking distance (α_2). This is logical as when the train is within its nominal braking distance, it should ideally start decelerating to avoid a collision, and failure to do so progressively increases the risk. The risk reaches

its peak when the train is at its emergency braking distance, signifying that if the train does not stop immediately, the collision is inevitable.

In summary, both equations are established as a probability ($R_1, R_2 \in [0,1]^2$) to collectively encapsulate the two critical regions of autonomous train operations from a safety perspective: the proactive safety measures (equation 5.17) and the reactive safety measures (equation 5.18).

5.5.7 Results

The following figures illustrate the system's performance in a dynamic railway environment, providing valuable insights into its ability to detect and respond to obstacles, estimate risk levels, and ensure safe and efficient operations. In concluding our discussion on the simulation setups, it is important to note that by presenting two distinct scenarios, we demonstrate the inherent advantages of our approach in terms of adaptability and the ease with which new elements or factors can be integrated. The original scenario establishes a baseline, while the enhanced simulation scenario takes a leap forward by incorporating dynamic elements such as obstacle velocities and behaviors.

Actual state, perceived state, and chosen action

Figure 5.7 shows the evolution of the actual state (*in blue color*), perceived state (*in red color*), and the chosen action (*in green color*) over time. The actual state represents the ground truth state of the train, while the perceived state is based on the observations made by the train's sensors. The chosen action is the decision made by the POMDP model based on the perceived state. The plot provides valuable insights into how the perception process impacts decision-making, and it showcases the effectiveness of the model in adapting to the dynamic environment. Moreover, the x-axis in the figure represents the different time steps during the simulation, capturing the sequential evolution of the system's decision-making process. On the y-axis (*on the left*), the values $s_1, s_2, s_3,$ and s_4 correspond to the different states the system can be in. On the other hand, the y-axis (*on the right*) represents also the available actions that the system can take in response to its perceived state. These actions are depicted as $a_1, a_2,$ and a_3 .

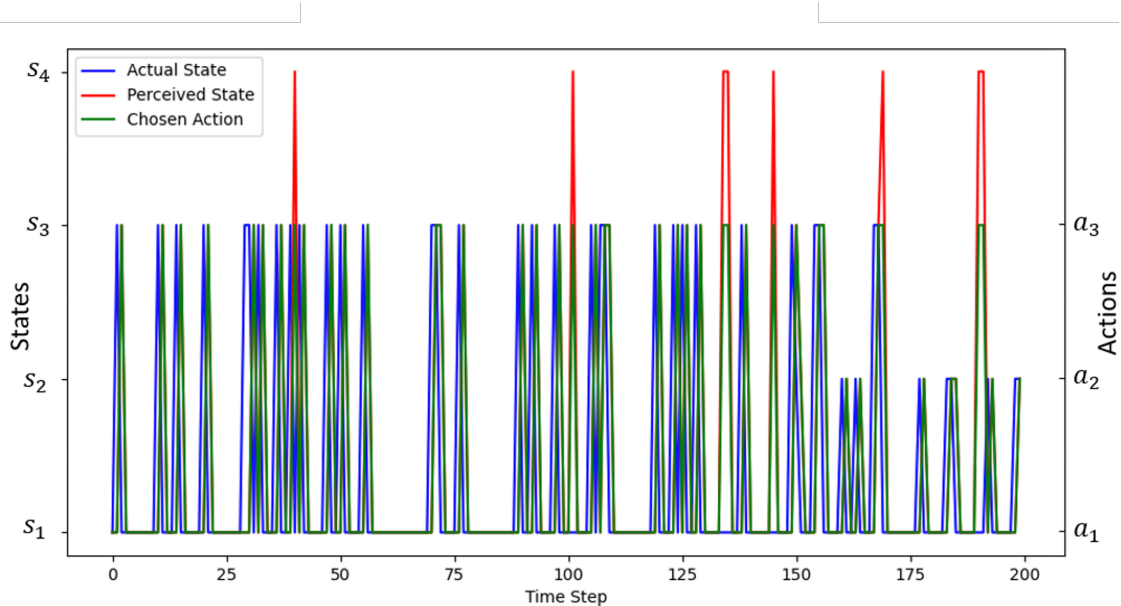


Figure 5.7: The evolution of the actual state, perceived state, and the chosen action (setup 1)

The perceived state follows the trajectory of the actual state, underscoring the system’s ability to accurately perceive its environment. However, some occasional divergences between the two trajectories (perceived and actual state) are present at specific time steps. These divergences are interpreted as false positives (perceiving an obstacle that is not present/false alert) and false negatives (falling to detect an obstacle/missed detection).

Similarly, Figure 5.8 provides a visualization of the autonomous train’s state transitions alongside the corresponding actions taken over the simulation period. The graph displays perceived states in red, actual states in blue, and chosen actions are highlighted in green for clear differentiation and easy interpretation. The plot shows the model’s responsiveness to changes in risk levels, transitioning to more conservative actions as the perceived risk increases (i.e., state s_4). Notably, the shift from s_1 to s_4 prompts an immediate action change to a_3 , demonstrating the system’s capacity for rapid reaction to imminent collision risks.

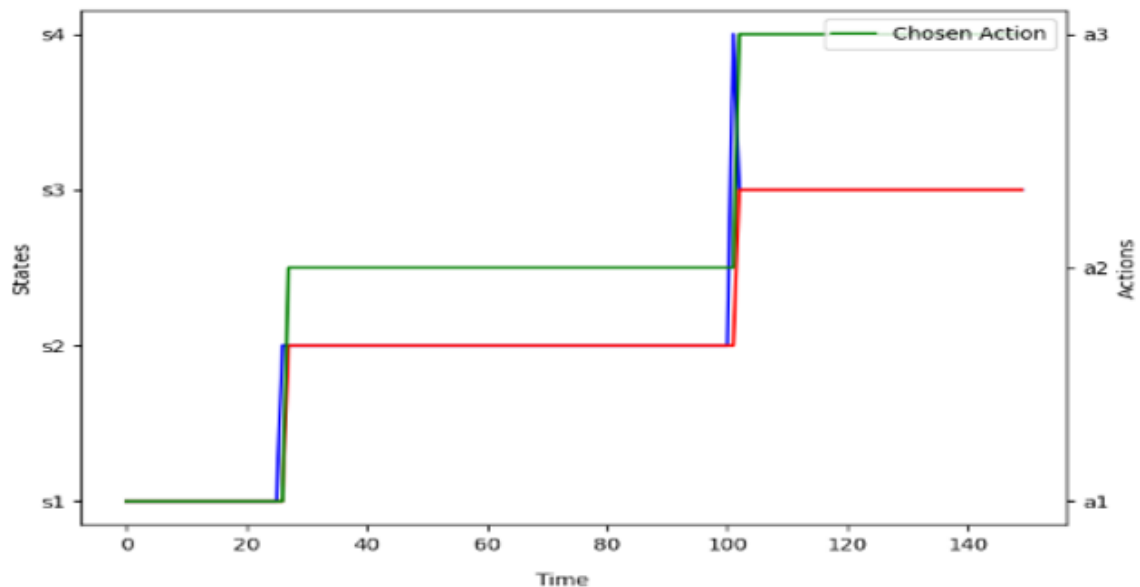


Figure 5.8: The evolution of the actual state, perceived state, and the chosen action (setup 2)

Rewards over time

Figure 5.9 displays the immediate rewards (and penalties) obtained by the system over time. The rewards are directly linked to the perceived state and the chosen action. Positive rewards indicate safety (*Safe* state), while negative rewards represent potential risks (*ObstacleDetected*, *AboutToCrash*, and *Crash* states). The scatter plots in the figure also highlight false positives (*in green points*) and false negatives (*in red points*) in the decision-making process, showing instances where the model’s perception deviates from the actual state. Notable false positives occur at times 40, 101, 134, 135, 145, 169, 190, 191, while false negatives occur at times 20, 30, 119, 149.

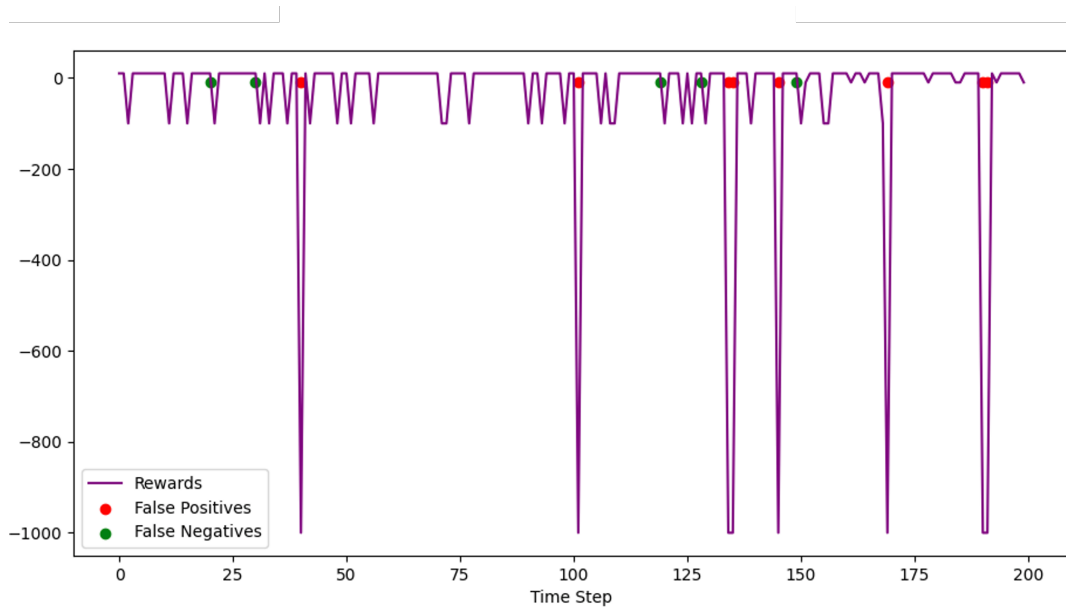


Figure 5.9: The evolution of rewards over time (setup 1)

Correspondingly, Figure 5.10 shows the dynamics of the rewards function for the second setup of simulation. The figure clearly denotes the penalty incurred as the system approaches a high-risk state, highlighting the impact of strategic decision-making on the train's overall safety. In the rewards function of the second simulation setup, the concentration is oriented towards the model's ability to integrate dynamic properties of obstacles, such as their velocities and nature. As such, the delineation of false positives and negatives was deemed less pertinent for this particular analysis, given that the primary interest was to observe how the integration of obstacle dynamics affects the overall reward structure and safety performance of the autonomous system.

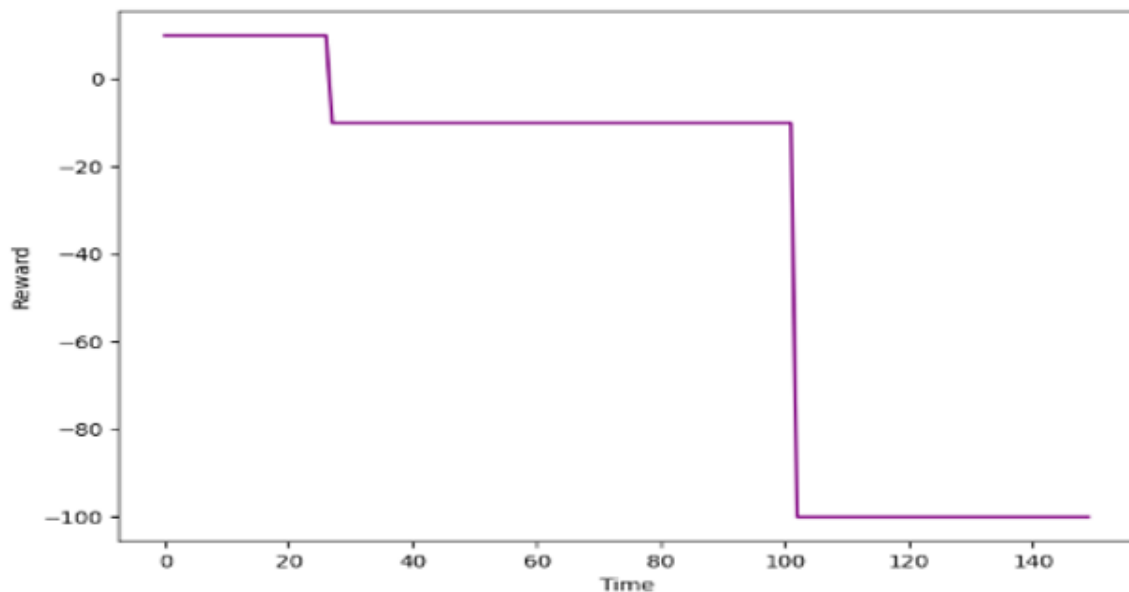


Figure 5.10: The evolution of rewards over time (setup 2)

Risk estimation over time

Figure 5.11 illustrates two risk estimation methods: *risk estimation 1* (equation 5.17) and *risk estimation 2* (equation 5.18) employed in the model. These estimations assess the risk level associated with the observed distance to the obstacle. Higher risk values indicate a higher likelihood of collision. The plot enables a comprehensive understanding of the risk assessment process and its role in determining appropriate actions.

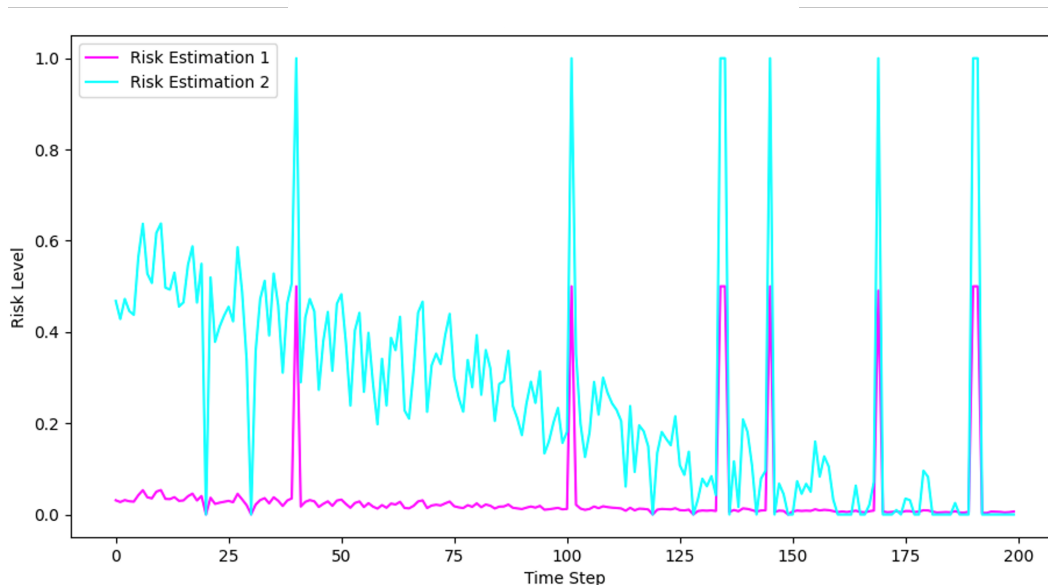


Figure 5.11: The risk estimation over time (setup 1)

Risk estimation 1 (depicted in *magenta color*) mainly describes low-risk scenarios across most states (i.e., states s_1 , s_2 and s_3 , except for the *Crash* state (i.e., state s_4), where risk is high. This approach seems cautious, as it maintains a conservative risk assessment. In contrast, *risk estimation 2* (illustrated in *cyan color*) describes a more dynamic risk evaluation. As the model navigates from the *Safe* state to *AboutToCrash* state, risk steadily increases, reaching approximately 0.5, indicating a heightened state of caution. However, once the model enters the *Crash* state, risk reaches its maximum value of 1, underscoring the severe consequences of this state. These differing risk estimation strategies shed light on the adaptability of the model, reveal the ability to respond to different levels of risk, and provide valuable insights into decision-making process.

Equally, Figure 5.12 illustrates the fluctuating risk levels as perceived by each method over time, with *Risk Estimation 1* and *Risk Estimation 2* plotted on the same graph for direct comparison. The divergence of two methods underscores the variability in risk perception and the importance of selecting a robust model that accurately reflects the operational condition's inherent uncertainties.

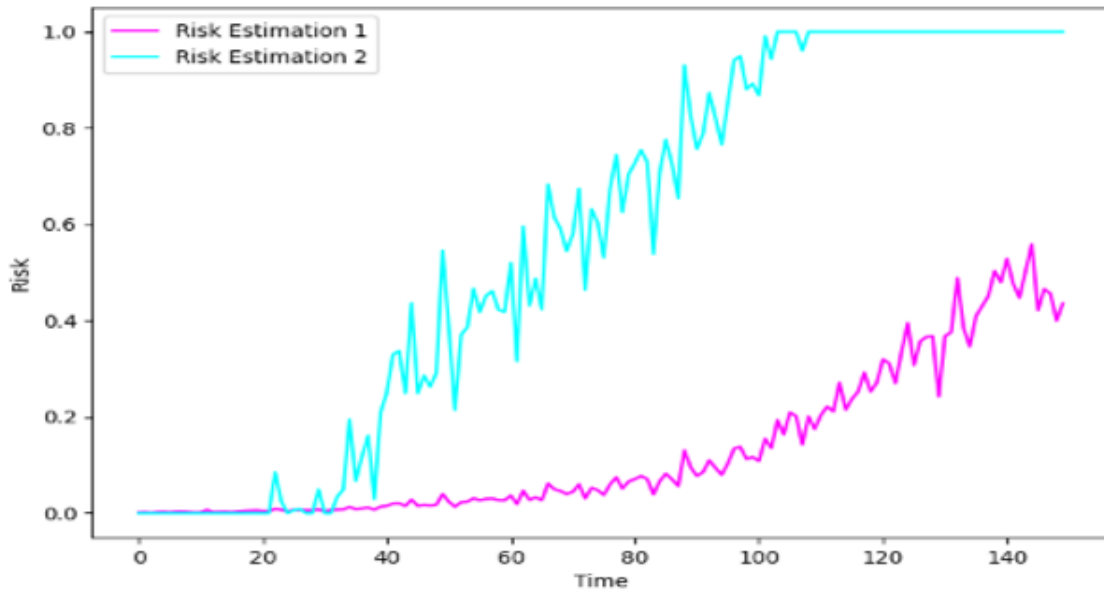


Figure 5.12: The risk estimation over time (setup 2)

Observed distance to obstacle

Figure 5.13 depicts the observed distance to the obstacle over time. It tracks how the perceived distance fluctuates as the train's sensors detect and interact with the environment. The red and blue dashed lines represent the thresholds for the nominal and emergency braking distances (α_1 and α_2 , respectively). When the observed distance crosses these thresholds, the model may initiate braking actions accordingly to prevent potential collisions. On the other hand, Figure 5.14 showcases the observed distance to the nearest obstacle throughout the simulation timeline. In this second simulation setup, the model considers multiple obstacles, both static and dynamic, and calculates the distance to the nearest obstacle (i.e., the *distance to obstacle* variable). The plot is a testament to the system's ability to maintain situational awareness and adapt its responses based on real-time assessments.

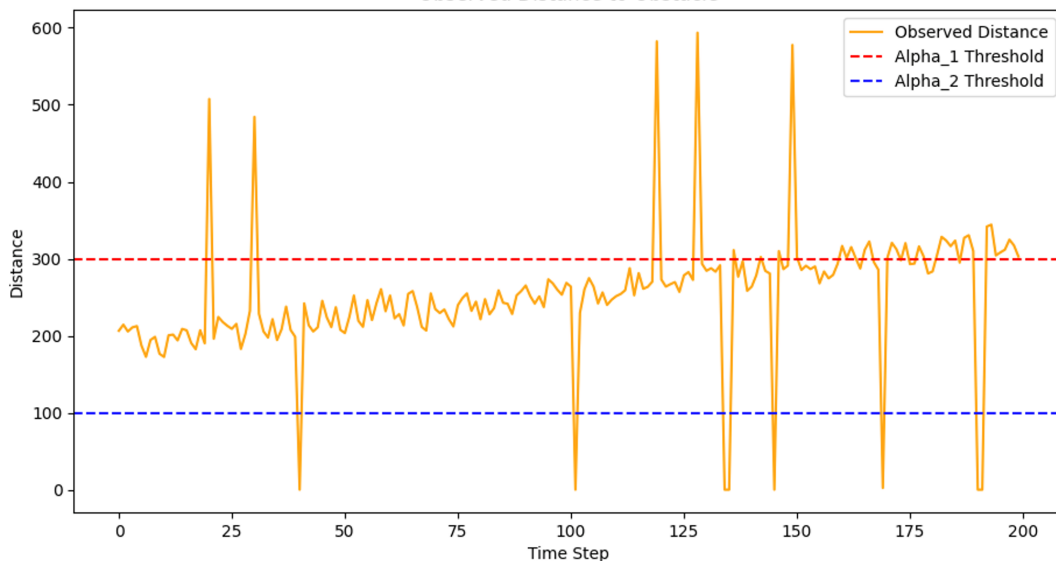


Figure 5.13: The evolution of the observed distance to obstacle over time (setup 1)

The results of the simulation demonstrate the effectiveness of the proposed risk-based POMDP process for the autonomous train anti-collision function. The results show that the proposed model is able to provide a safe and efficient solution for the anti-collision function, which takes into account the uncertainties related to the train's state and its perception of the environment. Moreover, this highlights the potential of the proposed process to be applied to real-world scenarios and provides a basis for further research to improve and extend the process to handle more complex environments. Finally, the dual-scenario structure not only showcases the robustness of our model but also represents the initial steps towards a more generic and comprehensive approach. In future iterations, the model could evolve to include additional complexities such as the precise dimensions of obstacles, their predicted trajectories, and other environmental factors. These advancements will allow for a more detailed and far-reaching application of the POMDP model, pushing the boundaries of autonomous train safety and operational efficiency.

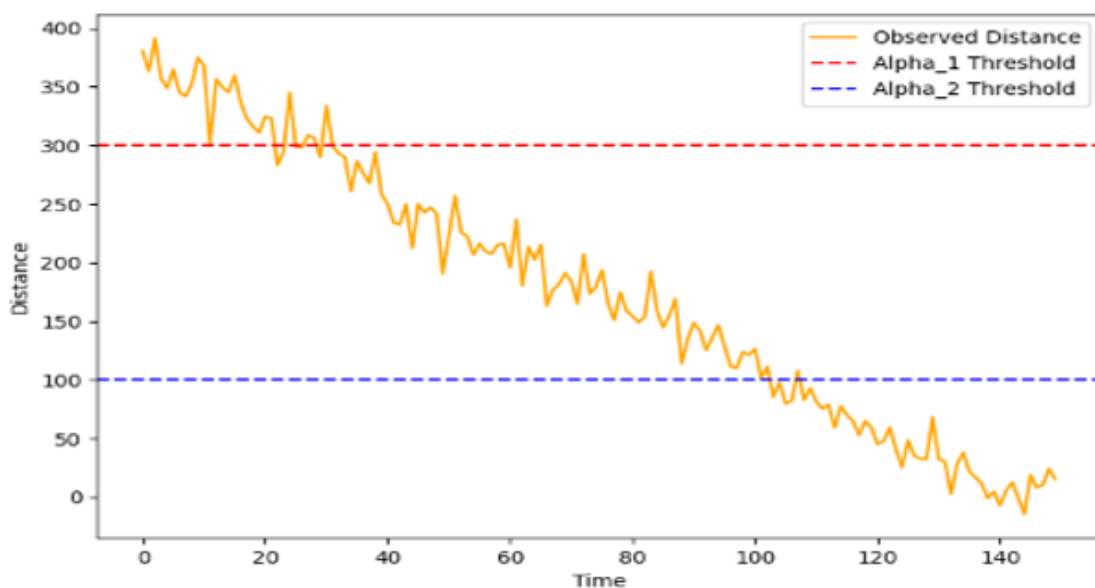


Figure 5.14: The evolution of the observed distance to obstacle over time (setup 2)

5.6 Conclusion

In this chapter, we proposed a risk-based decision-making approach for autonomous trains, leveraging the capabilities of Partially Observable Markov Decision Processes (POMDPs) to facilitate effective and real-time environmental monitoring of trains. The core contribution of this study lies in the ongoing monitoring and risk estimation, which is crucial for ensuring the safe operation of autonomous trains. This approach integrates dynamic risk assessment into the process of decision-making, enabling the train to proactively manage potential collision hazards. It effectively addresses uncertainties in both the train's operational state and its interaction with the environment. By doing so, the approach enhances the autonomous train's ability to make informed and safe decisions.

Chapter 6

Conclusion and perspectives

Contents

6.1	Conclusions	109
6.2	Perspectives	110

6.1 Conclusions

This dissertation has dealt with several challenges related to the safety assurance of autonomous trains, namely the safety argumentation, the situational awareness and dynamic risk assessment aspects, and the decision-making process for the autonomous driving system. The main contributions of this PhD thesis are presented in Chapters 3, 4, and 5.

In Chapter 3, we have proposed a high-level safety argumentation framework for the autonomous driving system of the autonomous train. The main objective of this framework is to show how to present and structure safety arguments for autonomous trains by using Goal Structuring Notation (GSN) graphical models. The proposed framework is illustrated through the use case of a safety function (anti-collision function).

In Chapter 4, we have proposed a framework allowing the autonomous train to continuously perform a situational awareness of its surrounding environment and provide a run-time probability estimation for the occurrence of railway hazards. To achieve this, the framework integrates a dynamic risk assessment layer in its high-level decision-making architecture. Furthermore, we illustrated the proposed framework through an operational safety function: collision detection and avoidance.

Finally, in Chapter 5, we have established a risk-based decision-making approach using Partially Observable Markov Decision Processes (POMDPs) for run-time monitoring of the autonomous train's environment during its operation. The main objective of the approach is to ensure the safe operations of the train, with respect to the collision hazards, by maintaining an acceptable risk level. Indeed, this level should be maintained by estimating and updating the risk associated with the operational and environmental conditions of the train. Moreover, this approach allows the system to make safe and informed decisions while considering the inherent uncertainties related to the train's state and its perceived environment. The approach is established and illustrated for the anti-collision function of the autonomous train.

These approaches collectively contribute to advancing the safety assurance of autonomous trains by addressing the complexities of autonomous systems, enhancing situational awareness, and establishing a robust decision-making framework based on dynamic risk assessment.

6.2 Perspectives

While this dissertation provides substantial contributions to the field of autonomous train safety assurance, it also opens several directions for future research :

1. **Integration of confidence assessment in safety arguments:** Future research can further examine integrating confidence aspects into safety argumentation for autonomous trains. This involves using approaches such as belief functions to quantitatively and qualitatively assess the confidence in safety claims and the evidence supporting them. Such an approach would allow for a more comprehensive understanding of the reliability of different components of the autonomous train system and their impact on overall safety. By systematically evaluating the confidence level of each safety argument, researchers can identify areas requiring additional evidence or more robust safety measures (i.e., dynamic safety cases). This methodology not only improves the explainability and traceability of safety cases but also contributes to a deeper trust in autonomous train technologies among stakeholders and the public. Finally, as autonomous trains move closer to widespread implementation, research into regulatory frameworks and standards will be essential to ensure safety, interoperability, and public acceptance.
2. **Enhancement of situational awareness models:** Advancing situational awareness models to account for a wider array of environmental variables (i.e., weather conditions, obstacle trajectory prediction, signalling, etc.) and operational scenarios is another important area for future investigation. This improvement requires enhancing models to better understand and adapt to changing complex environments. By using sophisticated sensors and data analysis methods, future models could significantly improve the autonomous driving system's ability to anticipate and respond to unexpected events, thereby maintaining an acceptable level of safety for autonomous trains.
3. **Empirical validation and testing:** To confirm the real-world applicability and effectiveness of the proposed models and approaches, conducting thorough field tests and simulations with real data is essential. This empirical validation involves assessing the models under real conditions to evaluate their performance and identify any necessary adjustments. Such detailed testing is critical to ensure the reliability of safety assurance methods for autonomous trains, aiming for the highest standards of safety and operational performance.

Bibliography

- Abaei, M.M., Hekkenberg, R., BahooToroody, A., 2021. A multinomial process tree for reliability assessment of machinery in autonomous ships. *Reliability Engineering & System Safety* 210. doi:[10.1016/j.ress.2021.107484](https://doi.org/10.1016/j.ress.2021.107484).
- Abdazimov, S., Zuhridinov, H., 2023. Analysis of monitoring and forecasting of emergency situations in railway transport. *Theoretical aspects in the formation of pedagogical sciences* 2, 85–88.
- Agrawal, V., Achuthan, B., Ansari, A., Tiwari, V., Pandey, V., 2021. Threat/hazard analysis and risk assessment: A framework to align the functional safety and security process in automotive domain. *SAE International Journal of Transportation Cybersecurity and Privacy* 4, 83–96.
- Alawad, H., An, M., Kaewunruen, S., 2020a. Utilizing an adaptive neuro-fuzzy inference system (anfis) for overcrowding level risk assessment in railway stations. *Applied Sciences* 10, 5156.
- Alawad, H., Kaewunruen, S., An, M., 2020b. A deep learning approach towards railway safety risk assessment. *IEEE Access* 8, 102811–102832.
- Alexander, R., Asgari, H., Ashmore, R., Banks, A., Bongirwar, R., Bradshaw, B., Bragg, J., Clegg, J., Fenn, J., Harper, C., et al., 2020. Safety assurance objectives for autonomous systems. *Safety Critical Systems Club (SCSC)* .
- Alexander, R., Herbert, N., Kelly, T., 2009. Deriving safety requirements for autonomous systems, in: *4th SEAS DTC Technical Conference*, pp. 1–8.
- Appoh, F., Yunusa-Kaltungo, A., 2022. Dynamic hybrid model for comprehensive risk assessment: a case study of train derailment due to coupler failure. *IEEE Access* 10, 24587–24600.
- Armstrong, J.M., Paynter, S.E., 2004. The deconstruction of safety arguments through adversarial counter-argument, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 3–16.
- Asaadi, E., Denney, E., Menzies, J., Pai, G.J., Petroff, D., 2020. Dynamic Assurance Cases: A Pathway to Trusted Autonomy. *Computer* 53, 35–46.
- Aven, T., 2016. Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research* 253, 1–13.
- Avilés, H., Negrete, M., Machucho, R., Rivera, K., Trejo, D., Vargas, H., 2022. Probabilistic logic markov decision processes for modeling driving behaviors in self-driving cars, in: *Ibero-American Conference on Artificial Intelligence*, Springer. pp. 366–377.
- Ayoub, A., Kim, B., Lee, I., Sokolsky, O., 2012. A safety case pattern for model-based development approach, in: *NASA Formal Methods Symposium*, Springer. pp. 141–146.

- Bagloee, S.A., Tavana, M., Asadi, M., Oliver, T., 2016. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. *Journal of modern transportation* 24, 284–303.
- Bartrip, P.W., 1980. The state and the steam-boiler in nineteenth-century Britain. *International review of social history* 25, 77–105.
- Bate, I., Kelly, T., 2003. Architectural considerations in the certification of modular systems. *Reliability Engineering & System Safety* 81, 303–324.
- Belle, A.B., Lethbridge, T.C., Kpodjedo, S., Adesina, O.O., Garzón, M.A., 2019. A novel approach to measure confidence and uncertainty in assurance cases, in: *IEEE 27th International Requirements Engineering Conference Workshops (REW)*, IEEE. pp. 24–33.
- Bellman, R., 1957. A Markovian decision process. *Journal of mathematics and mechanics* , 679–684.
- Berntorp, K., Di Cairano, S., 2019. Particle filtering for automotive: a survey, in: *22th International Conference on Information Fusion (FUSION)*, IEEE. pp. 1–8.
- Birch, J., Rivett, R., Habli, I., Bradshaw, B., Botham, J., Higham, D., Jesty, P., Monkhouse, H., Palin, R., 2013. Safety cases and their role in ISO 26262 functional safety assessment, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 154–165.
- Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer.
- Bishop, P., Bloomfield, R., 2000a. A methodology for safety case development. *Safety and Reliability* 20, 34–42.
- Bishop, P., Bloomfield, R., 2000b. A methodology for safety case development, in: *Safety and Reliability*, Taylor & Francis. pp. 34–42.
- Bogdoll, D., Uhlemeyer, S., Kowol, K., Zöllner, J.M., 2023. Perception datasets for anomaly detection in autonomous driving: A survey. *IEEE Intelligent Vehicles Symposium (IV)* doi:[10.1109/iv55152.2023.10186609](https://doi.org/10.1109/iv55152.2023.10186609).
- Bolbot, V., Theotokatos, G., Wennersberg, L.A., Faivre, J., Vassalos, D., Boulougouris, E., Jan Rødseth, Ø., Andersen, P., Pauwelyn, A.S., Van Coillie, A., 2023. A novel risk assessment process: Application to an autonomous inland waterways ship. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 237, 436–458.
- Boudardara, F., Boussif, A., Meyer, P.J., Ghazel, M., 2022. Interval weight-based abstraction for neural network verification, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 330–342.
- Boudardara, F., Boussif, A., Meyer, P.J., Ghazel, M., 2023. A review of abstraction methods towards verifying neural networks. *ACM Transactions on Embedded Computing Systems* .
- Boulinier, C., Petit, F., Villain, V., 2004. When graph theory helps self-stabilization, in: *Proceedings of the 23th annual ACM symposium on Principles of distributed computing*, pp. 150–159.

- Boussif, A., Tonk, A., Beugin, J., Collart Dutilleul, S., 2023. Operational risk assessment of railway remote driving system, in: *Safety and Reliability*, Taylor & Francis. pp. 1–22.
- Brandenburger, N., Naumann, A., 2019. On track: a series of research about the effects of increasing railway automation on the train driver. *IFAC-PapersOnLine* 52, 288–293.
- Brandenburger, N., Naumann, A., Jipp, M., 2021. Task-induced fatigue when implementing high grades of railway automation. *Cognition, Technology & Work* 23, 273–283.
- Brown, R., 1998. Improving the production and presentation of safety cases through the use of intranet technology, in: *Industrial Perspectives of Safety-critical Systems*. Springer, pp. 184–193.
- Buehler, M., Iagnemma, K., Singh, S., 2009. The DARPA urban challenge: autonomous vehicles in city traffic. volume 56. Springer.
- Burton, S., Gauerhof, L., Heinzemann, C., 2017. Making the case for safety of machine learning in highly automated driving, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 5–16.
- Chalvatzaras, A., Pratikakis, I., Amanatiadis, A.A., 2022. A survey on map-based localization techniques for autonomous vehicles. *IEEE Transactions on Intelligent Vehicles* 8, 1574–1596.
- Chang, C.H., Kontovas, C., Yu, Q., Yang, Z., 2021. Risk assessment of the operations of maritime autonomous surface ships. *Reliability Engineering & System Safety* 207, 107324.
- Chauvin, C., Clostermann, J., Hoc, J.M., 2008. Situation awareness and the decision-making process in a dynamic situation: avoiding collisions at sea. *Journal of cognitive engineering and decision making* 2, 1–23.
- Chelouati, M., Boussif, A., Beugin, J., El-Koursi, E.M., 2022. A framework for risk-awareness and dynamic risk assessment for autonomous trains, in: *32nd European Safety and Reliability Conference (ESREL)*, pp. 2128–2135.
- Chelouati, M., Boussif, A., Beugin, J., El Koursi, E.M., 2023a. Graphical safety assurance case using goal structuring notation (gsn) — challenges, opportunities and a framework for autonomous trains. *Reliability Engineering & System Safety* 230, 108933. doi:<https://doi.org/10.1016/j.ress.2022.108933>.
- Chelouati, M., Boussif, A., Beugin, J., El Koursi, E.M., 2023b. A Risk-Based Decision-Making Process for Autonomous Trains Using POMDP: Case of the Anti-Collision Function. *IEEE Access* 12, 5630 – 5647. doi:[10.1109/ACCESS.2023.3347500](https://doi.org/10.1109/ACCESS.2023.3347500).
- Chen, C., Seff, A., Kornhauser, A., Xiao, J., 2015. Deepdriving: Learning affordance for direct perception in autonomous driving, in: *Proceedings of the IEEE international conference on computer vision*, pp. 2722–2730.
- Chen, H., Zhang, Z., 2019. Stochastic model predictive control of autonomous systems with non-gaussian correlated uncertainty.
- Chen, X., Bose, N., Brito, M., Khan, F., Thanyamanta, B., Zou, T., 2021. A review of risk analysis research for the operations of autonomous underwater vehicles. *Reliability Engineering & System Safety* 216, 108011.

- Cheng, B.H., Clark, R.J., Fleck, J.E., Langford, M.A., McKinley, P.K., 2020. AC-ROS: Assurance case driven adaptation for the robot operating system, in: Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, pp. 102–113.
- Cheng, R., Cheng, Y., Chen, D., Song, H., 2021. Online quantitative safety monitoring approach for unattended train operation system considering stochastic factors. *Reliability Engineering & System Safety* 216, 107933.
- Chia, W.M.D., Keoh, S.L., Goh, C., Johnson, C., 2022. Risk assessment methodologies for autonomous driving: A survey. *IEEE transactions on intelligent transportation systems* 23, 16923–16939.
- Chia, W.M.D., Keoh, S.L., Michala, A.L., Goh, C., 2021. Real-time recursive risk assessment framework for autonomous vehicle operations, in: 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), IEEE. pp. 1–7.
- Chinneck, P., Pumfrey, D., Kelly, T., 2004. Turning up the HEAT on safety case construction, in: *Practical Elements of Safety*. Springer, pp. 223–240.
- Chouchani, N., Zinkunegi, I.L., 2022. Enhance railway digital map for slam: Feasibility study. *Computers in Railways XVIII: Railway Engineering Design and Operation* 213, 233.
- Chouhan, S., Pradesh, M., et al., 2014. Railway Anti-Collision System using DSLR Sensor. *International Journal of Engineering Sciences & Research* 3, 1199–1202.
- Clothier, R., Denney, E., Pai, G.J., 2017. Making a Risk Informed Safety Case for Small Unmanned Aircraft System Operations, in: 17th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations Conference, pp. 1–19.
- Cockburn, B., Karniadakis, G.E., Shu, C.W., 2012. *Discontinuous Galerkin methods: theory, computation and applications*. volume 11. Springer Science & Business Media.
- Collart-Dutilleul, S., Lecomte, T., Romanovsky, A., 2019. Reliability, safety, and security of railway systems. modelling, analysis, verification, and certification, in: 3th International Conference, RSSRail, Springer. p. 11495.
- Conges, A., Breard, L., Patruno, W., Ouro-Sao, A., Salatge, N., Fertier, A., Lauras, M., Graham, J., Benaben, F., 2023. Situational awareness and decision-making in a crisis situation: A crisis management cell in virtual reality. *International Journal of Disaster Risk Reduction* 97, 104002.
- Corso, A., Kochenderfer, M.J., 2020. Interpretable safety validation for autonomous vehicles, in: *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE. pp. 1–6.
- Coskun, S., Langari, R., 2018. Predictive fuzzy Markov decision strategy for autonomous driving in highways, in: *IEEE Conference on Control Technology and Applications (CCTA)*, IEEE. pp. 1032–1039.
- Coskun, S., Zhang, Q., Langari, R., 2019. Receding horizon Markov game autonomous driving strategy, in: *American Control Conference (ACC)*, IEEE. pp. 1367–1374.

- Costa, R.D., Hirata, C.M., Pugliese, V.U., 2023. A comparative study of situation awareness-based decision-making model reinforcement learning adaptive automation in evolving conditions. *IEEE Access* 11, 16166–16182.
- Costantini, F., Thomopoulos, N., Steibel, F., Curl, A., Lugano, G., Kováčiková, T., 2020. Autonomous vehicles in a GDPR era: An international comparison, in: *Advances in transport policy and planning*. Elsevier. volume 5, pp. 191–213.
- Council, N.R., et al., 2007. *Software for dependable systems: Sufficient evidence?* National Academies Press.
- CSM-RA, 2017. *Guide for the application of the commission regulation on the adoption of a Common Safety Method on Risk evaluation and Assessment*.
- Cui, J., Zhao, B., Qu, M., et al., 2023. An integrated lateral and longitudinal decision-making model for autonomous driving based on deep reinforcement learning. *Journal of Advanced Transportation* 2023.
- Cyra, Ł., Górski, J., 2007. Supporting compliance with safety standards by trust case templates, in: *Risk, Reliability and Societal Safety : Proceedings of the European Safety and Reliability Conference (ESREL)*, pp. –.
- Dahn, N., Fuchs, S., Gross, H.M., 2018. Situation awareness for autonomous agents, in: *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*, IEEE. pp. 666–671.
- D’Aniello, G., Gaeta, M., 2023. Situation awareness in human-machine systems. *Handbook of Human-Machine Systems* , 451–461.
- Dardar, R., 2014. *Building a Safety Case in Compliance with ISO 26262 for Fuel Level Estimation and Display System*. Ph.D. thesis. Malardalen University.
- Dardar, R., Gallina, B., Johnsen, A., Lundqvist, K., Nyberg, M., 2012. *Industrial Experiences of Building a Safety Case in Compliance with ISO 26262*. Master’s thesis. Mälardalen University, School of Innovation, Design and Engineering.
- Del Moral, P., 1997. Nonlinear filtering: Interacting particle resolution. *Comptes Rendus de l’Académie des Sciences-Series I-Mathematics* 325, 653–658.
- Dempster, A.P., 2008. The Dempster–Shafer calculus for statisticians. *International Journal of approximate reasoning* 48, 365–377.
- Denney, E., Pai, G., Habli, I., 2015. Dynamic Safety Cases for Through-Life Safety Assurance, in: *37th IEEE International Conference on Software Engineering*, IEEE. pp. 587–590.
- Denney, E., Pai, G., Pohl, J., 2012. Advocate: An assurance case automation toolset, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 8–21.
- Denney, E., Pai, G., Whiteside, I., 2019. The role of safety architectures in aviation safety cases. *Reliability Engineering & System Safety* 191, 106502. doi:[10.1016/j.ress.2019.106502](https://doi.org/10.1016/j.ress.2019.106502).
- Dickmanns, E.D., Behringer, R., Dickmanns, D., Hildebrandt, T., Maurer, M., Thomanek, F., Schiehlen, J., 1994. The seeing passenger car ‘VaMoRs-P’, in: *Proceedings of the Intelligent Vehicles’ Symposium*, IEEE. pp. 68–73.

- Divall, C., 2016. Railway Safety, Reliability, and Security: Technologies and Systems Engineering. IGI Global.
- Djuric, P.M., Kotecha, J.H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M.F., Miguez, J., 2003. Particle filtering. *IEEE signal processing magazine* 20, 19–38.
- Dunjó, J., Fthenakis, V., Vílchez, J.A., Arnaldos, J., 2010. Hazard and operability (HA-ZOP) analysis. a literature review. *Journal of hazardous materials* 173, 19–32.
- EASA, 2021. EASA Concept Paper First usable guidance for Level 1 machine learning applications - Proposed Issue.
- Eggert, J., 2018. Risk estimation for driving support and behavior planning in intelligent vehicles. *at-Automatisierungstechnik* 66, 119–131.
- EN-50126, 2017. Railway Applications - The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process.
- EN-50128, 2011. Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems.
- EN-50129, 2018. Railway applications - Communication, signalling and processing systems - Safety related electronic systems for signalling.
- Endsley, M.R., 1988. Situation awareness in aircraft systems: Symposium abstract, in: *Proceedings of the Human Factors Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA. pp. 96–96.
- Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. *Human factors* 37, 32–64.
- ERA, 2021. Automated Railway - Operation as Usual: Best Practice to Achieve Situational Awareness. URL: https://www.era.europa.eu/sites/default/files/events-news/docs/mythbustingpaper1_muehl_en.pdf.
- EU-Commission, 2017. High level group on the competitiveness and sustainable growth of the automotive industry in the european union (GEAR 2030). European Commission Brussels, Belgium .
- Eurocontrol, . Eurocontrol Safety Case Development Manual Ed 2.1 13-Oct-2006 | Argument | Safety.
- European-Commission, 2014. Directive 2014/45/eu of the european parliament and of the council. Official Journal of the European Union URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0045>.
- Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77, 167–181.
- Falessi, D., Nejati, S., Sabetzadeh, M., Briand, L., Messina, A., 2011. Safeslice: a model slicing and design safety inspection tool for sysml, in: *Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering*, pp. 460–463.
- Fan, C., Montewka, J., Zhang, D., 2022. A risk comparison framework for autonomous ships navigation. *Reliability Engineering & System Safety* 226, 108709.

- Farag, W., 2021. Kalman-filter-based sensor fusion applied to road-objects detection and tracking for autonomous vehicles. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 235, 1125–1138.
- Farnell, G., Saddington, A., Lacey, L., 2019. A new systems engineering structured assurance methodology for complex systems. *Reliability Engineering & System Safety* 183, 298–310.
- Feng, L., King, A.L., Chen, S., Ayoub, A., Park, J., Bezzo, N., Sokolsky, O., Lee, I., 2014. A safety argument strategy for PCA closed-loop systems: A preliminary proposal, in: *5th Workshop on Medical Cyber-Physical Systems, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik*. pp. 1–9.
- Ferber, J., Weiss, G., 1999. *Multi-agent systems: an introduction to distributed artificial intelligence*. volume 1. Addison-wesley Reading.
- Feth, P., Akram, M.N., Schuster, R., Wasenmüller, O., 2018. Dynamic risk assessment for vehicles of higher automation levels by deep learning, in: *Computer Safety, Reliability, and Security: SAFECOMP Workshops, ASSURE, DECSoS, SASSUR, STRIVE, and WAISE, Västerås, Sweden, Proceedings* 37, Springer. pp. 535–547.
- Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. *Machine learning* 29, 131–163.
- Frigerio, A., Vermeulen, B., Goossens, K.G., 2021. Automotive architecture topologies: Analysis for safety-critical autonomous vehicle applications. *IEEE Access* 9, 62837–62846.
- Gadmer, Q., Richard, P., Popieul, J.C., Sentouh, C., 2022. Railway automation: A framework for authority transfers in a remote environment. *IFAC-PapersOnLine* 55, 85–90.
- Gallina, B., 2014. A Model-Driven Safety Certification Method for Process Compliance, in: *IEEE International Symposium on Software Reliability Engineering Workshops*, pp. 204–209.
- Gallina, B., Gómez-Martínez, E., Benac-Earle, C., 2017. Promoting MBA in the rail sector by deriving process-related evidence via MDSafeCer. *Computer Standards & Interfaces* 54, 119–128.
- Garland, D.J., Endsley, M.R., Andre, A.D., Hancock, P.A., Selcon, S.J., Vidulich, M.A., 1996. Assessment and measurement of situation awareness, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA. pp. 1170–1173.
- Ge, X., Rijo, R., Paige, R.F., Kelly, T.P., McDermid, J.A., 2012. Introducing Goal Structuring Notation to explain decisions in clinical practice. *Procedia Technology* 5, 686–695.
- Gilchrist, W., 1993. Modelling failure modes and effects analysis. *International Journal of Quality & Reliability Management* 10.
- Ginesi, M., Meli, D., Roberti, A., Sansonetto, N., Fiorini, P., 2020. Autonomous task planning and situation awareness in robotic surgery, in: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE. pp. 3144–3150.

- Golightly, D., Ryan, B., Dadashi, N., Pickup, L., Wilson, J., 2013. Use of scenarios and function analyses to understand the impact of situation awareness on safe and effective work on rail tracks. *Safety science* 56, 52–62.
- Graydon, P., 2013. A perspective on safety argumentation: Aims, achievements, challenges, and opportunities, in: AAA2013, First International Workshop on Argument for Agreement and Assurance, Kanagawa, Japan, pp. 1–8.
- Graydon, P.J., Knight, J.C., Strunk, E.A., 2007. Assurance based development of critical systems, in: 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07), pp. 347–357.
- Griebel, T., Müller, J., Buchholz, M., Dietmayer, K., 2020. Kalman filter meets subjective logic: A self-assessing Kalman filter using subjective logic, in: IEEE 23rd International Conference on Information Fusion (FUSION), IEEE. pp. 1–8.
- Grigorescu, S., Trasnea, B., Cocias, T., Macesanu, G., 2020. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics* 37, 362–386.
- Gruber, T.R., 1990. Automated knowledge acquisition for strategic knowledge. Springer.
- GSN-WG, 2021. GSN community standard version 3.
- Gu, Z., Gao, L., Ma, H., Li, S.E., Zheng, S., Jing, W., Chen, J., 2023. Safe-state enhancement method for autonomous driving via direct hierarchical reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems* .
- Guan, J., Chen, G., Huang, J., Li, Z., Xiong, L., Hou, J., Knoll, A., 2022. A discrete soft actor-critic decision-making strategy with sample filter for freeway autonomous driving. *IEEE Transactions on Vehicular Technology* 72, 2593–2598.
- Guan, Y., Li, S.E., Duan, J., Wang, W., Cheng, B., 2018. Markov probabilistic decision making of self-driving cars in highway with random traffic flow: a simulation study. *Journal of Intelligent and Connected Vehicles* 1, 77–84.
- 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: American Institute of Aeronautics and Astronautics (AIAA) Information Systems-AIAA Infotech@ Aerospace, p. 0910.
- Gupta, S., Snigdh, I., 2022. Multi-sensor fusion in autonomous heavy vehicles, in: Autonomous and Connected Heavy Vehicle Technology. Elsevier, pp. 375–389.
- Habli, I., Ibarra, I., Rivett, R., Kelly, T., 2010. Model-based assurance for justifying automotive functional safety, in: Proc. SAE World Congress, pp. 1–16.
- Haddon-Cave, C., 1971. Investigation of railway rolling stock.
- Hagen, I., Kufoalor, D., Johansen, T., Brekke, E., 2022. Scenario-based model predictive control with several steps for colregs compliant ship collision avoidance. *IFAC-PapersOnLine* 55, 307–312.
- Haklidir, M., Temeltaş, H., 2022. Autonomous driving systems for decision-making under uncertainty using deep reinforcement learning, in: 30th Signal Processing and Communications Applications Conference (SIU), IEEE. pp. 1–4.

- Hartsell, C., Ramakrishna, S., Dubey, A., Stojcsics, D., Mahadevan, N., Karsai, G., 2021. Resonate: A runtime risk assessment framework for autonomous systems, in: International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), IEEE. pp. 118–129.
- Hathat, Y., Samai, D., Benlamoudi, A., Bensid, K., Taleb-Ahmed, A., 2022. SNCF workers detection in the railway environment based on improved YOLO v5, in: 7th International Conference on Image and Signal Processing and their Applications (ISPA), IEEE. pp. 1–7.
- Hawkins, R., 2019. Body of knowledge for assurance of Robotic and Autonomous Systems (RAS). Assuring Autonomy International Programme (AAIP) - technical report .
- Hawkins, R., Kelly, T., Knight, J., Graydon, P., 2011. A new approach to creating clear safety arguments, in: Advances in Systems Safety: Proceedings of the 19th Safety-Critical Systems Symposium, Southampton, UK, Springer. pp. 3–23.
- Hawkins, R., Paterson, C., Picardi, C., Jia, Y., Calinescu, R., Habli, I., 2021. Guidance on the Assurance of Machine Learning in Autonomous Systems (AMLAS). Assuring Autonomy International Programme (AAIP), University of York .
- Hegde, J., Utne, I.B., Schjølberg, I., Thorkildsen, B., 2018. A Bayesian approach to risk modeling of autonomous subsea intervention operations. Reliability Engineering & System Safety 175, 142–159. doi:[10.1016/j.ress.2018.03.019](https://doi.org/10.1016/j.ress.2018.03.019).
- Heikkilä, E., Tuominen, R., Tiusanen, R., Montewka, J., Kujala, P., 2017. Safety Qualification Process for an Autonomous Ship Prototype—a Goal-based Safety Case Approach, in: Marine Navigation. CRC Press, pp. 365–370.
- Heimdahl, M.P.E., Leveson, N.G., 1996. Completeness and consistency in hierarchical state-based requirements. IEEE transactions on Software Engineering 22, 363–377.
- Hirata, C., Nadjm-Tehrani, S., 2019. Combining GSN and STPA for Safety Arguments, in: International Conference on Computer Safety, Reliability, and Security, Springer. pp. 5–15.
- Hou, W., Li, W., Li, P., 2023. Fault diagnosis of the autonomous driving perception system based on information fusion. Sensors 23, 5110.
- Howard, R.A., 1960. Dynamic programming and Markov processes.
- Hu, J., Kong, H., Liu, T., Meng, Y., 2022. Autonomous motion decision-making based on deep reinforcement learning for autonomous driving, in: 6th CAA International Conference on Vehicular Control and Intelligence (CVCI), IEEE. pp. 1–6.
- Hu, Y., 2023. Doppler principle based situational awareness method for autonomous driving, in: 5th International Conference on Communications, Information System and Computer Engineering (CISCE), IEEE. pp. 84–89.
- Idmessaoud, Y., Dubois, D., Guiochet, J., 2021. Quantifying confidence of safety cases with belief functions, in: International Conference on Belief Functions, Springer. pp. 269–278.
- IEC-61508, 2010. Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems - Part 1: General Requirements.

- IEC-62267, 2009. Railway applications - Automated Urban Guided Transport (AUGT) . Standard. International Electrotechnical Commission. Geneva, CH.
- Ingle, S., Phute, M., 2016. Tesla autopilot: semi autonomous driving, an uptick for future autonomy. *International Research Journal of Engineering and Technology* 3, 369–372.
- Insaurralde, C.C., Blasch, E., 2022. Situation awareness decision support system for air traffic management using ontological reasoning. *Journal of Aerospace Information Systems* 19, 224–245.
- INTESM, 2013. International Rail Industry’s Engineering Safety Management Handbook. volume 2. International Rail Industry by Technical Programme Delivery Ltd., Issue 1.
- Islam, M.M., Lautenbach, A., Sandberg, C., Olovsson, T., 2016. A risk assessment framework for automotive embedded systems, in: *Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security*, pp. 3–14.
- ISO-15026, 2020. Systems and Software Engineering - Systems and Software Assurance - Part 1: Concepts and Vocabulary.
- ISO-21448, 2022. Road Vehicles - Safety Of The Intended Functionality. International Organization for Standardization .
- ISO-26262, 2018. Road Vehicles — Functional safety. URL: <https://www.iso.org/fr/standard/43464.html>.
- ISO-31000, 2018. Risk Management — Guidelines.
- ISO-34501, 2022. Road Vehicles - The Test Scenarios of Automated Driving Systems.
- ISO-9000, 2015. Quality management systems - fundamentals and vocabulary.
- ISO/IEC24028, 2020. ISO/IEC TR 24028 Information technology — Artificial Intelligence — Overview of trustworthiness in artificial intelligence. International Organization for Standardization .
- ISO/IEC29119, 2020. ISO/IEC TR 29119 Software and systems engineering - Software testing - Part 11 : Guidelines on testing of AI-based systems. International Organization for Standardization .
- Iyer, N.C., Kulkarni, A., Shet, R., Keerthan, U., 2021. Localization of self-driving car using particle filter, in: *Advances in Computing and Network Communications: Proceedings of CoCoNet, Volume 1*, Springer. pp. 147–155.
- Jaakkola, T., Singh, S., Jordan, M., 1994. Reinforcement learning algorithm for partially observable Markov decision problems. *Advances in neural information processing systems* 7.
- Jagannadha, P.K.D., Yilmaz, M., Sonawane, M., Chadalavada, S., Sarangi, S., Bhaskaran, B., Bajpai, S., Reddy, V.A., Pandey, J., Jiang, S., 2019. Special session: in-system-test (ist) architecture for nvidia drive-agx platforms, in: *IEEE 37th VLSI Test Symposium (VTS)*, IEEE. pp. 1–8.
- Javed, M.A., Muram, F.U., Hansson, H., Punnekkat, S., Thane, H., 2021. Towards dynamic safety assurance for Industry 4.0. *Journal of Systems Architecture* 114, 101914. Publisher: Elsevier.

- Johansen, T., Blindheim, S., Torben, T.R., Utne, I.B., Johansen, T.A., Sørensen, A.J., 2023. Development and testing of a risk-based control system for autonomous ships. *Reliability Engineering & System Safety* 234, 109195.
- Johnsen, S.O., Hoem, Å.S., Stålhane, T., Jenssen, G., Moen, T., 2018. Risk based regulation and certification of autonomous transport systems, in: *Proceedings of the 28th International European Safety and Reliability Conference (ESREL)*, pp. 17–21.
- Jonchery, S., Revilloud, M., Ouerhani, Y., 2021. Trajectory based particle filter: Asynchronous observation fusion for autonomous driving localization, in: *IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE. pp. 114–121.
- Jourdan, L., Sallak, M., Schön, W., Quost, B., Bouvet, Y., 2022. Validated autonomous train perception using interpretable machine learning, in: *Lambda Mu 22-Congrès de maîtrise des risques et de sûreté de fonctionnement*.
- Jung, M., McKee, S.A., Sudarshan, C., Dropmann, C., Weis, C., Wehn, N., 2018. Driving into the memory wall: The role of memory for advanced driver assistance systems and autonomous driving, in: *Proceedings of the International Symposium on Memory Systems*, pp. 377–386.
- Junyung, K., Xingang, Z., Shah, A.U.A., Kang, H.G., 2021. System risk quantification and decision-making support using functional modeling and dynamic Bayesian network. *Reliability Engineering & System Safety* 215. doi:[10.1016/j.ress.2021.107880](https://doi.org/10.1016/j.ress.2021.107880).
- Kaelbling, L.P., Littman, M.L., Cassandra, A.R., 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101, 99–134.
- Kaelbling, L.P., Littman, M.L., Moore, A.W., 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4, 237–285.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems.
- Katrakazas, C., Quddus, M., Chen, W.H., 2019. A new integrated collision risk assessment methodology for autonomous vehicles. *Accident Analysis & Prevention* 127, 61–79.
- Katwe, S., Iyer, N., Khan, M., Peters, M., Mahale, M., 2021. Particle filter based localization of autonomous vehicle, in: *2nd Global Conference for Advancement in Technology (GCAT)*, IEEE. pp. 1–6.
- Kelly, T., Weaver, R., 2004. The Goal Structuring Notation—a safety argument notation. *Proc Dependable Syst Networks Workshop Assurance Cases* .
- Kelly, T.P., 1999a. Arguing safety: a systematic approach to managing safety cases. Ph.D. thesis. University of York York, UK.
- Kelly, T.P., 1999b. Arguing safety—a systematic approach to safety case management. DPhil Thesis York University, Department of Computer Science Report YCST .
- Kenarangui, R., 1991. Event-tree analysis by fuzzy probability. *IEEE transactions on reliability* 40, 120–124.
- Khan, F., Hashemi, S.J., Paltrinieri, N., Amyotte, P., Cozzani, V., Reniers, G., 2016. Dynamic risk management: a contemporary approach to process safety management. *Current opinion in chemical engineering* 14, 9–17. doi:[10.1016/j.coche.2016.07.006](https://doi.org/10.1016/j.coche.2016.07.006).
- Khurana, N., Das, K., 2009. Investigation of human factors in railway accidents and safety on indian railways. *International Journal of Human and Social Sciences* 4, 315–324.

- Khurshid, T., Faisal, M.N., 2012. Human errors in railway accidents: A case study of multan city. *World Applied Sciences Journal* 17, 46–51.
- Kiran, B.R., Sobh, I., Talpaert, V., Mannion, P., Al Sallab, A.A., Yogamani, S., Pérez, P., 2021. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems* 23, 4909–4926. doi:[10.1109/TITS.2021.3054625](https://doi.org/10.1109/TITS.2021.3054625).
- Klimenko, D., Song, J., Kurniawati, H., 2014. TAPIR: A software toolkit for approximating and adapting POMDP solutions online, in: *Proceedings of the Australasian Conference on Robotics and Automation, Melbourne, Australia, Proceedings of Australasian Conference on Robotics and Automation*. pp. 1–9.
- Kocić, J., Jovičić, N., Drndarević, V., 2018. Sensors and sensor fusion in autonomous vehicles, in: *26th Telecommunications Forum (TELFOR)*, IEEE. pp. 420–425.
- Koh, L.P., Wich, S.A., 2012. Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation. *Tropical conservation science* 5, 121–132.
- Kufoalor, D.K.M., Johansen, T.A., Brekke, E.F., Hepsø, A., Trnka, K., 2020. Autonomous maritime collision avoidance: Field verification of autonomous surface vehicle behavior in challenging scenarios. *Journal of Field Robotics* 37, 387–403.
- Kumamoto, H., Henley, E., 1996. *Probabilistic Risk Assessment and Management for Engineers and Scientists*. IEEE Press.
- Kurd, Z., Kelly, T., McDermid, J., Calinescu, R., Kwiatkowska, M., 2009. Establishing a framework for dynamic risk management in ‘intelligent’ aero-engine control, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 326–341.
- Laugier, C., 2019. Situation awareness & decision-making for autonomous driving, in: *IROS IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE. pp. 1–25.
- Lauri, M., Hsu, D., Pajarinen, J., 2022. Partially observable Markov decision processes in robotics: A survey. *IEEE Transactions on Robotics* 39, 21–40.
- Lemonnier, A., Adélé, S., Dionisio, C., 2023. Acceptability of autonomous trains with different grades of automation by potential users: A qualitative approach. *Travel behaviour and society* 33, 100641.
- Leurent, E., 2018. A survey of state-action representations for autonomous driving.
- Leveson, N., 2020. White Paper on Limitations of Safety Assurance and Goal Structuring Notation (GSN). *Aeronautics and Astronautics MIT* , 2.
- Leveson, N.G., 2011. The use of safety cases in certification and regulation.
- Li, G., Yang, Y., Li, S., Qu, X., Lyu, N., Li, S.E., 2022. Decision making of autonomous vehicles in lane change scenarios: Deep reinforcement learning approaches with risk awareness. *Transportation research part C: emerging technologies* 134, 103452.
- Li, L., Ota, K., Dong, M., 2018. Humanlike driving: Empirical decision-making system for autonomous vehicles. *IEEE Transactions on Vehicular Technology* 67, 6814–6823.

- Lin, X., Zhang, J., Shang, J., Wang, Y., Yu, H., Zhang, X., 2019. Decision making through occluded intersections for autonomous driving, in: IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. pp. 2449–2455.
- Liu, P., Zhang, R., Yin, Z., Li, Z., 2021a. Human errors and human reliability. *Handbook of Human Factors and Ergonomics*, 514–572.
- Liu, Q., Li, X., Yuan, S., Li, Z., 2021b. Decision-making technology for autonomous vehicles: Learning-based methods, applications and future outlook, in: IEEE International Intelligent Transportation Systems Conference (ITSC), IEEE. pp. 30–37.
- Lobato, W., Mendes, P., Rosário, D., Cerqueira, E., Villas, L.A., 2023. Redundancy mitigation mechanism for collective perception in connected and autonomous vehicles. *Future Internet* 15, 41.
- Luo, Y., van den Brand, M., Engelen, L., Klabbbers, M., 2015. A Modeling Approach to Support Safety Assurance in the Automotive Domain, in: *Progress in Systems Engineering*, Springer International Publishing, Cham. pp. 339–345.
- Luo, Y., Saberi, A.K., Den Brand, M.v., 2019. Safety-Driven Development and ISO 26262, in: *Automotive Systems and Software Engineering: State of the Art and Future Trends*. Springer International Publishing, Cham, pp. 225–254.
- Ma, M., Dong, W., Sun, X., Ji, X., 2019. A dynamic risk analysis method for high-speed railway catenary based on bayesian network, in: *CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, IEEE. pp. 547–554.
- Macher, G., Armengaud, E., Brenner, E., Kreiner, C., 2016. A review of threat analysis and risk assessment methods in the automotive context, in: *Computer Safety, Reliability, and Security: 35th International Conference, SAFECOMP, Trondheim, Norway, Proceedings 35*, Springer. pp. 130–141.
- Machin, M., Guiochet, J., Waeselynck, H., Blanquart, J.P., Roy, M., Masson, L., 2016. Smof: A safety monitoring framework for autonomous systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48, 702–715.
- Maidana, R.G., Parhizkar, T., Gomola, A., Utne, I.B., Mosleh, A., 2022. Supervised dynamic probabilistic risk assessment: Review and comparison of methods. *Reliability Engineering & System Safety* doi:[10.1016/j.ress.2022.108889](https://doi.org/10.1016/j.ress.2022.108889).
- Manjunatha, H., Pak, A., Filev, D., Tsiotras, P., 2023. Karnet: Kalman filter augmented recurrent neural network for learning world models in autonomous driving tasks. arXiv preprint arXiv:2305.14644 .
- Martin, H., Krammer, M., Bramberger, R., Armengaud, E., 2016. Process-and product-based lines of argument for automotive safety cases.
- Martin, H., Ma, Z., Schmittner, C., Winkler, B., Krammer, M., Schneider, D., Amorim, T., Macher, G., Kreiner, C., 2020. Combined automotive safety and security pattern engineering approach. *Reliability Engineering & System Safety* 198, 106773.
- Martínez-Díaz, M., Soriguera, F., 2018. Autonomous vehicles: theoretical and practical challenges. *Transportation Research Procedia* 33, 275–282.
- Masson, É., Richard, P., Garcia-Guillen, S., Adell, G.M., 2019. TC-Rail: Railways remote driving, in: *Proceedings of the 12th World Congress on Railway Research*, Tokyo, Japan, pp. 1 – 6.

- McDermid, J.A., Jia, Y., Habli, I., 2019. Towards a Framework for Safety Assurance of Autonomous Systems, in: Artificial Intelligence Safety, CEUR Workshop Proceedings. pp. 1–7.
- Metzner, A., Wickramaratne, T., 2019. Exploiting vehicle-to-vehicle communications for enhanced situational awareness, in: IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), IEEE. pp. 88–92.
- Molloy, J., McDermid, J., 2022. Safety assessment for autonomous system perception capabilities. arXiv preprint arXiv:2208.08237 .
- Monahan, G.E., 1982. State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms. *Management science* 28, 1–16.
- Mooney, C.Z., 1997. Monte carlo simulation. 116, Sage.
- Morato, P.G., Andriotis, C.P., Papakonstantinou, K.G., Rigo, P., 2023. Inference and dynamic decision-making for deteriorating systems with probabilistic dependencies through bayesian networks and deep reinforcement learning. *Reliability Engineering & System Safety* 235. doi:10.1016/j.ress.2023.109144.
- Müller, J.R., Drewes, J., May, J., Trog, C., 2009. The formal representation of the safety case processes described in the en 5012x norms, in: International Railway Safety Conference (IRSC 2009), p. 46.
- Müller, M., Ghasemi, G., Jazdi, N., Weyrich, M., 2022. Situational risk assessment design for autonomous mobile robots. *Procedia CIRP* 109, 72–77.
- Murphy, K.P., 2012. Machine learning: a probabilistic perspective. MIT press.
- Nair, G.S., Bhat, C.R., 2021. Sharing the road with autonomous vehicles: Perceived safety and regulatory preferences. *Transportation research part C: emerging technologies* 122. doi:10.1016/j.trc.2020.102885.
- Nair, S., de la Vara, J.L., Sabetzadeh, M., Briand, L., 2014. An extended systematic literature review on provision of evidence for safety certification. *Information and Software Technology* 56, 689–717.
- Nasir, N.Z.M., Zakaria, M.A., Razali, S., bin Abu, M.Y., 2017. Autonomous mobile robot localization using kalman filter, in: MATEC Web of conferences, EDP Sciences. p. 01069.
- National Transportation Safety Board, 2002. Railroad accident report: Collision of two washington metropolitan area transit authority metrorail trains near shady grove passenger station, gaithersburg, maryland. URL: <https://www.nts.gov/investigations/AccidentReports/Reports/RAR0201.pdf>.
- Nešić, D., Nyberg, M., Gallina, B., 2021. A probabilistic model of belief in safety cases. *Safety Science* 138, 105187.
- Niestadt, M., Debyser, A., Scordamaglia, D., Pape, M., 2019a. Artificial intelligence in transport: Current and future developments, opportunities and challenges. European Parliamentary Research Service .
- Niestadt, M., Debyser, A., Scordamaglia, D., Pape, M., 2019b. Artificial intelligence in transport: Current and future developments, opportunities and challenges. European Parliamentary Research Service .

- Nine, J., 2020. Towards robust situation awareness in autonomous vehicles. *Embedded Selforganising Systems* 7, 1–3.
- Nine, J., Manoharan, S., Hardt, W., 2021. Concept of the comprehension level of situation awareness using an expert system, in: *IOP Conference Series: Materials Science and Engineering*, IOP Publishing. p. 012103.
- Norris, J.R., 1998. *Markov chains*. 2, Cambridge university press.
- Palanisamy, P., Qiao, Z., Muelling, K., Dolan, J.M., Mudalige, U.P., 2020. Autonomous driving decisions at intersections using hierarchical options markov decision process. US Patent App. 16/039,579.
- 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*, pp. 1–6.
- Parhizkar, T., Utne, I.B., Vinnem, J.E., et al., 2022. Online probabilistic risk assessment of complex marine systems. *Springer Series in Reliability Engineering* doi:[10.1007/978-3-030-88098-9](https://doi.org/10.1007/978-3-030-88098-9).
- Park, M.G., 2014. *RAMS management of railway systems*. Ph.D. thesis. University of Birmingham.
- Parliament of the United Kingdom, 1882. Boiler explosions act 1882. <https://www.legislation.gov.uk/ukpga/Vict/45-46/22/contents>. Accessed: date-of-access.
- Parush, A., Campbell, C., Hunter, A., Ma, C., Calder, L., Worthington, J., Abbott, C., Frank, J., 2011. *Situational awareness and patient safety*. The Royal College of Physicians and Surgeons of Canada: Ottawa .
- Patel, A.R., Liggesmeyer, P., 2021. Machine learning based dynamic risk assessment for autonomous vehicles, in: *2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC)*, IEEE. pp. 73–77.
- Pavlović, M.G., Ćirić, I.T., Ristić-Durrant, D., Nikolić, V.D., Simonović, M.B., Ćirić, M.V., Banić, M.S., 2018. Advanced thermal camera based system for object detection on rail tracks. *Thermal Science* 22, 1551–1561.
- Pek, C., Althoff, M., 2019. Ensuring motion safety of autonomous vehicles through online fail-safe verification, in: *Robotics: Science and Systems–Pioneers Workshop*, pp. 1–4.
- Picardi, C., Paterson, C., Hawkins, R.D., Calinescu, R., Habli, I., 2020. Assurance argument patterns and processes for machine learning in safety-related systems, in: *Proceedings of the Workshop on Artificial Intelligence Safety (SafeAI 2020)*, pp. 23–30.
- Pineau, J., Gordon, G., Thrun, S., et al., 2003. Point-based value iteration: An anytime algorithm for pomdps, in: *18th International Joint Conference on Artificial Intelligence (IJCAI '03)*, pp. 1025–1032.
- 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: *International Symposium on Electromagnetic Compatibility - EMC EUROPE*, IEEE, Barcelona, Spain. pp. 465–469.
- Plissonneau, A., Trentesaux, D., Ben-Messaoud, W., Bekrar, A., 2021. AI-based speed control models for the autonomous train: a literature review, in: *3rd International Conference on Transportation and Smart Technologies (TST)*, IEEE. pp. 9–15.

- Plissonneau Duquene, A., 2023. Apprentissage machine pour la décision de conduite autonome de véhicules guidés: Application dans le domaine ferroviaire. Ph.D. thesis. Valenciennes, Université Polytechnique Hauts-de-France.
- Pomerleau, D.A., 1988. Alvin: An autonomous land vehicle in a neural network. *Advances in neural information processing systems* 1.
- Pouya, P., Madni, A.M., 2020. Expandable-partially observable Markov decision-process framework for modeling and analysis of autonomous vehicle behavior. *IEEE Systems Journal* 15, 3714–3725.
- Puterman, M.L., 1990. Markov decision processes. *Handbooks in operations research and management science* 2, 331–434.
- Quaglietta, E., Corman, F., Goverde, R.M., 2013. Stability analysis of railway dispatching plans in a stochastic and dynamic environment. *Journal of Rail Transport Planning & Management* 3, 137–149.
- Ragi, S., Chong, E.K., 2013. UAV path planning in a dynamic environment via partially observable Markov decision process. *IEEE Transactions on Aerospace and Electronic Systems* 49, 2397–2412. doi:[10.1109/TAES.2013.6621824](https://doi.org/10.1109/TAES.2013.6621824).
- Ramanathan, P., Kartik, 2021. Autonomous driving cars: Decision-making. *Internet of Vehicles and its Applications in Autonomous Driving*, 31–39.
- Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A., 2019a. Autonomous systems safety—state of the art and challenges, in: *Proceedings of the First International Workshop on Autonomous Systems Safety, NTNU*. pp. 18–32.
- Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A., 2019b. *Proceedings of the first international workshop on autonomous systems safety*.
- Read, G.J., Naweed, A., Salmon, P.M., 2019. Complexity on the rails: A systems-based approach to understanding safety management in rail transport. *Reliability Engineering & System Safety* 188, 352–365.
- Reich, J., Trapp, M., 2020. Sinadra: towards a framework for assurable situation-aware dynamic risk assessment of autonomous vehicles, in: *16th European dependable computing conference (EDCC), IEEE*. pp. 47–50.
- Reich, J., Wellstein, M., Sorokos, I., Oboril, F., Scholl, K.U., 2021. Towards a software component to perform situation-aware dynamic risk assessment for autonomous vehicles, in: *Dependable Computing-EDCC Workshops: DREAMS, DSOGR1, SERENE 2021, Munich, Germany Proceedings 17, Springer*. pp. 3–11.
- Rinehart, D.J., Knight, J.C., Rowanhill, J., 2017. Understanding what it means for assurance cases to “work”.
- Ristić-Durrant, D., Haseeb, M.A., Franke, M., Banić, M., Simonović, M., Stamenković, D., 2020. Artificial intelligence for obstacle detection in railways: Project smart and beyond, in: *Dependable Computing-EDCC Workshops: AI4RAILS, DREAMS, DSOGR1, SERENE 2020, Munich, Germany, Proceedings 16, Springer*. pp. 44–55.
- Ristić-Durrant, D., Haseen, M.A., Emami, D., Gräser, A., Ćirić, I., Simonović, M., Nikolić, V., Nikolić, D., Eßer, F.P., Schindler, C., 2018. *Reliable Obstacle Detection for Smart Automation of Rail Transport*. Universitätsbibliothek der RWTH Aachen.

- Rolt, L.T.C., . Red for danger: a history of railway accidents and railway safety precautions.
- Rose, J., Bearman, C., Dorrian, J., 2018. The Low-Event Task Subjective Situation Awareness (LETSSA) technique: Development and evaluation of a new subjective measure of situation awareness. *Applied Ergonomics* 68, 273–282.
- Rosique, F., Navarro, P.J., Fernández, C., Padilla, A., 2019. A systematic review of perception system and simulators for autonomous vehicles research. *Sensors* 19, 648. doi:[10.3390/s19030648](https://doi.org/10.3390/s19030648).
- Ross, S., Pineau, J., Paquet, S., Chaib-Draa, B., 2008. Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research* 32, 663–704. doi:[10.1613/jair.2567](https://doi.org/10.1613/jair.2567).
- Rudolph, A., Voget, S., Mottok, J., 2018. A consistent safety case argumentation for artificial intelligence in safety related automotive systems, in: *9th European Congress on Embedded Real Time Software and Systems, ERTS*, pp. 1–9.
- Ruiz, A., Juez, G., Espinoza, H., de La Vara, J.L., Larrucea, X., 2017. Reuse of safety certification artefacts across standards and domains: A systematic approach. *Reliability Engineering & System Safety* 158, 153–171.
- Russell, S., Norvig, P., et al., 2016. *Artificial intelligence: a modern approach*. Malaysia; Pearson.
- SAE, I., 2016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, J3016.
- Salmon, P.M., Stanton, N.A., Walker, G.H., 2020. 13 distributed situation awareness and vehicle automation. *Handbook of human factors for automated, connected, and intelligent vehicles* .
- Sarker, A., Fisher, P., Gaudio, J.E., Annaswamy, A.M., 2023. Accurate parameter estimation for safety-critical systems with unmodeled dynamics. *Artificial Intelligence* 316, 103857.
- Schmid, T., Schraufstetter, S., Wagner, S., Hellhake, D., 2019. A Safety Argumentation for Fail-Operational Automotive Systems in Compliance with ISO 26262, in: *4th International Conference on System Reliability and Safety (ICSRS)*, pp. 484–493.
- Schulman, P., 1984. Bayes' theorem—a review. *Cardiology clinics* 2, 319–328.
- Schwalbe, G., Schels, M., 2020. Concept enforcement and modularization as methods for the iso 26262 safety argumentation of neural networks, in: *10th European Congress on Embedded Real Time Software and Systems (ERTS 2020)*, pp. 1–10.
- Schwarting, W., Alonso-Mora, J., Rus, D., 2018. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems* 1, 187–210.
- Sentz, K., Ferson, S., 2002. Combination of evidence in Dempster-Shafer theory .
- Shafaei, S., Kugele, S., Osman, M.H., Knoll, A., 2018. Uncertainty in machine learning: A safety perspective on autonomous driving, in: *Computer Safety, Reliability, and Security: SAFECOMP Workshops, ASSURE, DECSoS, SASSUR, STRIVE, and WAISE, Västerås, Sweden, Proceedings 37*, Springer. pp. 458–464.

- Shao, H., Wang, L., Chen, R., Li, H., Liu, Y., 2023. Safety-enhanced autonomous driving using interpretable sensor fusion transformer, in: Conference on Robot Learning, PMLR. pp. 726–737.
- Sharples, S., Millen, L., Golightly, D., Balfe, N., 2011. The impact of automation in rail signalling operations. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 225, 179–191.
- Silver, D., Veness, J., 2010. Monte-carlo planning in large POMDPs. Advances in neural information processing systems 23.
- Singh, P., Dulebenets, M.A., Pasha, J., Gonzalez, E.D.S., Lau, Y.Y., Kampmann, R., 2021. Deployment of autonomous trains in rail transportation: Current trends and existing challenges. IEEE Access 9, 91427–91461.
- Smiles, S., 1904. The Locomotive.
- Somani, A., Ye, N., Hsu, D., Lee, W.S., 2013. DESPOT: Online POMDP planning with regularization. Advances in neural information processing systems 26. doi:[10.1613/jair.5328](https://doi.org/10.1613/jair.5328).
- Song, Q., Fu, W., Wang, W., Sun, Y., Wang, D., Zhou, J., 2022. Quantum decision making in automatic driving. Scientific reports 12, 11042.
- Spaan, M.T., 2012. Partially observable Markov decision processes, in: Reinforcement learning: State-of-the-art. Springer, pp. 387–414.
- Spaan, M.T., Vlassis, N., 2005. Perseus: Randomized point-based value iteration for POMDPs. Journal of artificial intelligence research 24, 195–220. doi:[10.1613/jair.1659](https://doi.org/10.1613/jair.1659).
- Srivastava, D.K., Lilly, J.M., Feigh, K.M., 2022. Improving human situation awareness in ai-advised decision making, in: IEEE 3rd International Conference on Human-Machine Systems (ICHMS), IEEE. pp. 1–6.
- Steele, H., Roberts, C., 2022. Towards a sustainable digital railway, in: Sustainable Railway Engineering and Operations. Emerald Publishing Limited. volume 14, pp. 239–263.
- Stickel, S., Schenker, M., Dittus, H., Unterhuber, P., Canesi, S., Riquier, V., Ayuso, F.P., Berbineau, M., Goikoetxea, J., 2022. Technical feasibility analysis and introduction strategy of the virtually coupled train set concept. Scientific Reports 12, 4248.
- Stopka, O., Stopková, M., Lupták, V., Krile, S., 2020. Application of the chosen multi-criteria decision-making methods to identify the autonomous train system supplier. Transport problems 15.
- Stålhane, T., Myklebust, T., 2016. The Agile Safety Case, in: Computer Safety, Reliability, and Security, Springer International Publishing. pp. 5–16.
- Sunberg, Z., Kochenderfer, M.J., 2017. POMCPOW: An online algorithm for POMDPs with continuous state, action, and observation spaces.
- Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.
- Svendsen, K.A., Seto, M.L., 2020. Partially Observable Markov Decision Processes for Fault Management in Autonomous Underwater Vehicles, in: IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE. pp. 1–7.

- Taguchi, K., Daisuke, S., Nishihara, H., Takai, T., 2014. Linking Traceability with GSN, in: International Symposium on Software Reliability Engineering Workshops, IEEE, pp. 192–197.
- Talpes, E., Sarma, D.D., Venkataramanan, G., Bannon, P., McGee, B., Floering, B., Jalote, A., Hsiung, C., Arora, S., Gorti, A., et al., 2020. Compute solution for tesla’s full self-driving computer. *IEEE Micro* 40, 25–35.
- Tazoniero, A., Gonçalves, R., Gomide, F., 2007. Decision making strategies for real-time train dispatch and control. *Analysis and Design of Intelligent Systems Using Soft Computing Techniques*, 195–204.
- Temizer, S., Kochenderfer, M., Kaelbling, L., Lozano-Pérez, T., Kuchar, J., 2010. Collision avoidance for unmanned aircraft using Markov decision processes, in: AIAA guidance, navigation, and control conference, pp. –. doi:[10.2514/6.2010-8040](https://doi.org/10.2514/6.2010-8040).
- Tonk, A., Boussif, A., 2022. Operational Design Domain or Operational Envelope: Seeking a suitable concept for autonomous railway systems, in: ESREL, In 32nd European Safety And Reliability Conference, pp. 1–8.
- Tonk, A., Chelouati, M., Boussif, A., Beugin, J., El Koursi, M., 2022. A safety assurance methodology for autonomous trains, in: TRA, Transport Research Arena, pp. 1–8.
- Torens, C., Juenger, F., Schirmer, S., Schopferer, S., Zhukov, D., Dauer, J.C., 2023. Ensuring safety of machine learning components using operational design domain, in: AIAA SCITECH 2023 Forum, p. 1124.
- Tran, D.Q., Bae, S.H., 2021. Improved responsibility-sensitive safety algorithm through a partially observable Markov decision process framework for automated driving behavior at non-signalized intersection. *International journal of automotive technology* 22, 301–314.
- Trentesaux, D., Dahyot, R., Ouedraogo, A., Arenas, D., Lefebvre, S., Schön, W., Lussier, B., Chéritel, H., 2018. The autonomous train, in: 13th Annual Conference on System of Systems Engineering (SoSE), IEEE. pp. 514–520.
- Ueda, R., Arai, T., 2007. Real-time decision making of autonomous robot under uncertainty of state estimation by using particle filter and q-mdp value method. *Journal of the Robotics Society of Japan* 25, 103–112.
- U.K. Health and Safety Executive, 1999. Railway Safety Principles and Guidance. Technical Report. HM Railway Inspectorate.
- Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., Maurer, M., 2015. Defining and substantiating the terms scene, situation, and scenario for automated driving, in: IEEE 18th international conference on intelligent transportation systems, IEEE. pp. 982–988.
- Vallikannu, R., Rani, V.K., Kavitha, B., Sankar, P., 2023. An analysis of situational intelligence for first responders in military, in: International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), IEEE. pp. 1–4.
- Van Lamsweerde, A., 2001. Goal-oriented requirements engineering: A guided tour, in: Proceedings 5th IEEE international symposium on requirements engineering, IEEE. pp. 249–262.

- De la Vara, J.L., Panesar-Walawege, R.K., 2013. Safetymet: A metamodel for safety standards, in: *Model-Driven Engineering Languages and Systems: 16th International Conference, MODELS, Miami, FL, USA. Proceedings 16*, Springer. pp. 69–86.
- Ventikos, N.P., Chmurski, A., Louzis, K., 2020. A systems-based application for autonomous vessels safety: Hazard identification as a function of increasing autonomy levels. *Safety science* 131, 104919.
- Vesely, W.E., Goldberg, F.F., Roberts, N.H., Haasl, D.F., et al., 1981. *Fault tree handbook*. Systems and Reliability Research, Office of Nuclear Regulatory Research, US.
- Vierhauser, M., Bayley, S., Wyngaard, J., Xiong, W., Cheng, J., Huseman, J., Lutz, R., Cleland-Huang, J., 2019. Interlocking Safety Cases for Unmanned Autonomous Systems in Shared Airspaces. *IEEE Transactions on Software Engineering* 47, 899–918.
- Wagner, S., Schätz, B., Puchner, S., Kock, P., 2010. A Case Study on Safety Cases in the Automotive Domain: Modules, Patterns, and Models, in: *21st International Symposium on Software Reliability Engineering*, IEEE, pp. 269–278.
- Wall, J., Wittemyer, G., Klinkenberg, B., Douglas-Hamilton, I., 2014. Novel opportunities for wildlife conservation and research with real-time monitoring. *Ecological Applications* 24, 593–601.
- Wang, R., Guiochet, J., Motet, G., 2017. Confidence Assessment Framework for Safety Arguments, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 55–68.
- Wang, R., Guiochet, J., Motet, G., Schön, W., 2016a. Dempster-Shafer theory for argument confidence assessment, in: *International Conference on Belief Functions*, Springer. pp. 190–200.
- Wang, R., Guiochet, J., Motet, G., Schön, W., 2019. Safety case confidence propagation based on Dempster–Shafer theory. *International Journal of Approximate Reasoning*, Elsevier 107, 46–64.
- Wang, R., Guiochet, J., Motet, G., Schön, W., 2018. Modelling Confidence in Railway Safety Case. *Safety Science* 110, 286–299.
- Wang, Y., Zhang, M., Ma, J., Zhou, X., 2016b. Survey on driverless train operation for urban rail transit systems. *Urban Rail Transit* 2, 106–113.
- Wardziński, A., 2008. Safety assurance strategies for autonomous vehicles, in: *International Conference on Computer Safety, Reliability, and Security*, Springer. pp. 277–290.
- Weber, P., Medina-Oliva, G., Simon, C., Iung, B., 2012. Overview on bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence* 25, 671–682.
- Weber, P., Simon, C., 2016. *Benefits of Bayesian network models*. John Wiley & Sons.
- Wei, R., Kelly, T.P., Dai, X., Zhao, S., Hawkins, R., 2019. Model based system assurance using the structured assurance case metamodel. *Journal of Systems and Software* 154, 211–233.
- Welch, G., Bishop, G., et al., 1995. *An introduction to the Kalman filter*.
- White, D.J., 1993. A survey of applications of Markov decision processes. *Journal of the operational research society* 44, 1073–1096.

- 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 American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations Conference, p. 2286.
- Witulski, A., Austin, R., Evans, J., Mahadevan, N., Karsai, G., Sierawski, B., LaBel, K., Reed, R., Schrimpf, R., 2016. Goal structuring notation in a radiation hardening assurance case for cots-based spacecraft, in: GOMAC Tech Government Microcircuits Applications & Critical Technologies Conference.
- Wozniak, E., Putzer, H.J., Cârlan, C., 2021. AI-Blueprint for Deep Neural Networks., in: SafeAI workshop, Association for the Advancement of Artificial Intelligence (AAAI) Conference, p. 6.
- Wu, M., Yu, F.R., Liu, P.X., He, Y., 2022. A hybrid driving decision-making system integrating Markov logic networks and connectionist AI. *IEEE Transactions on Intelligent Transportation Systems* 24, 3514–3527.
- Wu, Y., Li, J., 2020. Particle filter estimation method of parameters time-varying discrete dynamic Bayesian network with application to UGV decision-making, in: 4th CAA International Conference on Vehicular Control and Intelligence (CVCI), IEEE. pp. 497–501.
- Xia, Y., Liu, S., Hu, R., Yu, Q., Feng, X., Zheng, K., Su, H., 2023. Smart: A decision-making framework with multi-modality fusion for autonomous driving based on reinforcement learning, in: International Conference on Database Systems for Advanced Applications, Springer. pp. 447–462.
- Xiang, X., Foo, S., 2021. Recent advances in deep reinforcement learning applications for solving partially observable Markov decision processes (POMDP) problems: Part 1—fundamentals and applications in games, robotics and natural language processing. *Machine Learning and Knowledge Extraction* 3, 554–581. doi:[10.3390/make3030029](https://doi.org/10.3390/make3030029).
- Xu, X., Peng, J., Zhang, R., Chen, B., Zhou, F., Yang, Y., Gao, K., Huang, Z., 2019. Adaptive model predictive control for cruise control of high-speed trains with time-varying parameters. *Journal of Advanced Transportation* 19. doi:[10.1155/2019/7261726](https://doi.org/10.1155/2019/7261726).
- Xue, Y., Xiang, P., Jia, F., Liu, Z., 2020. Risk assessment of high-speed rail projects: A risk coupling model based on system dynamics. *International journal of environmental research and public health* 17, 5307.
- Yang, J., Ward, M., Akhtar, J., 2017. The development of safety cases for an autonomous vehicle: A comparative study on different methods. Technical Report. SAE Technical Paper.
- Yang, K., Li, B., Shao, W., Tang, X., Liu, X., Wang, H., 2023a. Prediction failure risk-aware decision-making for autonomous vehicles on signalized intersections. *IEEE Transactions on Intelligent Transportation Systems* .
- Yang, K., Tang, X., Li, J., Wang, H., Zhong, G., Chen, J., Cao, D., 2023b. Uncertainties in onboard algorithms for autonomous vehicles: Challenges, mitigation, and perspectives. *IEEE Transactions on Intelligent Transportation Systems* .
- Yang, Z., Pei, X., Xu, J., Zhang, X., Xi, W., 2022. Decision-making in autonomous driving by reinforcement learning combined with planning & control, in: 6th CAA International Conference on Vehicular Control and Intelligence (CVCI), IEEE. pp. 1–6.

- Yi, R., Chen, J., 2023. Kalman filter-based adversarial patch attack defense for autonomous driving multi-target tracking, in: *IEEE International Conference on Industrial Technology (ICIT)*, IEEE. pp. 1–6.
- Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., Gao, Z., 2017. Research and development of automatic train operation for railway transportation systems: A survey. *Transportation Research Part C: Emerging Technologies* 85, 548–572.
- Yu, Q., Teixeira, Â.P., Liu, K., Rong, H., Soares, C.G., 2021. An integrated dynamic ship risk model based on bayesian networks and evidential reasoning. *Reliability Engineering & System Safety* 216, 107993.
- Zadeh, L.A., Klir, G.J., Yuan, B., 1996. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers*. volume 6. World scientific.
- Zarei, E., Gholamizadeh, K., Khan, F., Khakzad, N., 2022. A dynamic domino effect risk analysis model for rail transport of hazardous material. *Journal of Loss Prevention in the Process Industries* 74, 104666.
- Zhang, C., Huang, Z., Wang, S., Hong, Y., 2023a. Decision-making for overtaking in specific unmanned driving scenarios based on deep reinforcement learning, in: *IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, IEEE. pp. 680–685.
- Zhang, N.L., Zhang, W., 2001. Speeding up the convergence of Value Iteration in Partially Observable Markov Decision Processes. *Journal of Artificial Intelligence Research* 14, 29–51.
- Zhang, Y., Carballo, A., Yang, H., Takeda, K., 2023b. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing* 196, 146–177.
- Zheng, S., Wang, J., Rizos, C., Ding, W., El-Mowafy, A., 2023. Simultaneous Localization And Mapping (SLAM) for autonomous driving: Concept and analysis. *Remote Sensing* 15, 1156.
- Zhou, X., Liu, Z., Wu, Z., Wang, F., 2019. Quantitative processing of situation awareness for autonomous ships navigation. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation* 13.
- Zio, E., 2018. The future of risk assessment. *Reliability Engineering & System Safety* 177, 176–190.