Bayesian Validation Based Fault Tree Analysis for Enhancing Autonomous Vehicle Safety

by

Lansu Dai

A Thesis Submitted to the University of Ottawa
in Partial Fulfillment of the Requirements
for the Master's degree in Computer Science

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Examining Committee

The following served on the Examining Committee for this thesis.

External Member(s): Omair Shafiq
Associate Professor,
School of Information Technology,
Carleton University

Internal Member(s): Iluju Kiringa
Professor,
School of Electrical Engineering & Computer Science,
University of Ottawa

Supervisor(s): Burak Kantarci
University Research Chair, Professor,
School of Electrical Engineering & Computer Science,
University of Ottawa

Abstract

Ensuring the safety of autonomous vehicles in complex and uncertain environments remains a critical challenge. While ISO 26262 provides a robust framework for functional safety, it primarily focuses on hazards arising from hardware or software malfunctions. The Safety of the Intended Functionality (SOTIF), defined in ISO 21448, extends the traditional ISO 26262 safety framework by addressing hazards that arise from performance limitations and functional insufficiencies, even in the absence of system faults. Addressing such hazards requires methodologies that go beyond traditional fault-based analysis and are capable of representing uncertainty and causal dependencies. Fault Tree Analysis (FTA) is a widely used safety assessment method within ISO 26262 Functional Safety (FuSa) framework. However, traditional FTA is limited by its binary event structure, static assumptions and inability to capture probabilistic dependencies between variables. These limitations constrain its applicability in SOTIF-related safety analysis, where context-dependent and performance-driven hazards are critical. By integrating FTA with Bayesian Network (BN), these limitations can be addressed, that enabling probabilistic inference, representation of causal dependencies, and dynamic risk quantification. This thesis proposes an integrated methodology that combines FTA with BN to support probabilistic safety analysis in alignment with SOTIF principles. The approach remains the deductive, structured hierarchy of FTA while incorporating the probabilistic reasoning capabilities of BN. This integration enables the representation of multi-state variables, explicit modeling of interdependencies, and dynamic risk quantification through Bayesian inference. The proposed methodology is validated through two case studies. The first focuses on collision scenarios in autonomous driving to demonstrate the integration of FTA with BN, the result shows the perception system is the main contributor to collision risk in autonomous vehicles. The second case study examines object detection failure to evaluate the framework's alignment with SOTIF principles by modeling performance limitations and triggering conditions. Results show that adverse weather and occlusion are the most significant contributors to object detection failures, with posterior probabilities of 45.76% and 58.72%, respectively. The findings demonstrate the framework's ability to capture causal relationships and provide both qualitative and quantitative insights of SOTIF-related hazards. By aligning with ISO 26262 and ISO 21448, this research advances a comprehensive and extensible safety assessment approach that accounts for both fault-based and performance-based hazards in autonomous vehicle systems.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Burak Kantarci, for his continuous support, insightful guidance, and encouragement throughout my research and thesis work.

I would also like to extend my sincere thanks to ReasonX Labs for the opportunity to collaborate and for providing valuable resources and insights that significantly contributed to this research.

Special thanks to all the members of the NEXTCON-SCVIC lab for their technical discussions and camaraderie during this journey.

Finally, I extend my heartfelt appreciation to my family for their unwavering love, encouragement, and support, which have been essential to the completion of this work.

This work is supported in part by MITACS Accelerate Program project IT40981, NSERC CREATE TRAVERSAL program, Ontario Research Fund-Research Excellence (ORF-RE) program under RE012-026.

Publications

- Lansu Dai, and Burak Kantarci, "Advancing Autonomous Vehicle Safety: A Combined Fault Tree Analysis and Bayesian Network Approach", IEEE International Conference on Engineering Reliable Autonomous Systems (ERAS), May 2025. (Accepted)
- Lansu Dai, and Burak Kantarci, "A Survey on Enhancing Autonomous Vehicle Safety through Fault Tree Analysis and SOTIF Integration". (Submitted to ACM Journal on Autonomous Transportation Systems)
- Lansu Dai, and Burak Kantarci, "A SOTIF-Oriented Framework for Safety Assessment in Autonomous Vehicles Using Integrated Fault Tree and Bayesian Network Analysis". (Submitted to ACM Journal on Autonomous Transportation Systems)

Table of Contents

List of Tables

List of Figures

Abbreviations

ADAS Advanced Driver-Assistance Systems 12, 16

ADS Automated Driving System 31

AEV Autonomous Electric Vehicles 31

AI Artificial Intelligence 11, 30, 38, 52, 54, 60, 66

AIOps Artificial Intelligence for IT Operations 85

ASIL Automotive Safety Integrity Level 41, 43, 48, 52, 56

ATA Attack Tree Analysis 35, 36

AVP Automated Valet Parking 35

BDD Binary Decision Diagrams 32, 33

BMS Battery Management System 31

BN Bayesian Network iii, x, 2–6, 17, 18, 29–33, 37–39, 41–46, 50, 52, 54, 56–60, 63–65, 67–70, 74, 75, 77–80, 82–85

CCF Common Cause Failures 27

CPTs Conditional Probability Tables x, 47, 63, 68, 70, 74, 75, 80, 84

CTA Causal Tree Analysis 35, 36

DBNs Dynamic Bayesian Networks 2

DDT Dynamic Driving Task 12

DFTA Dynamic Fault Tree Analysis 32, 33, 37

ECU Electronic Control Unit 24

FFTA Fuzzy Fault Tree Analysis 30, 33

FMEA Failure Mode and Effects Analysis 8, 9, 16, 18, 31, 44

FOV Field of View 16

FTA Fault Tree Analysis iii, ix, x, 1–10, 16, 18, 22–39, 41–46, 50, 52, 54, 56–65, 67, 68, 70, 82–85

FuSa Functional Safety iii, 7, 21, 35–37, 39, 43, 48, 60

GPS Global Positioning Systems 12, 14–16

HARA Hazard Analysis and Risk Assessment 62

HAZOP Hazard and Operability 8, 16, 18, 29, 31

HMMs Hidden Markov Models 2

I2V Infrastructure-to-Vehicle 19

IMU Inertial Measurement Unit 12, 14

MCS Minimal Cut Sets 26, 27

MLE Maximum Likelihood Estimation 69

MLOps Machine Learning Operations 85

MPS Minimal Path Sets 26, 27

ODD Operational Design Domain 17, 31, 62, 66, 67, 71, 72, 79

OEDR Object and Event Detection and Response 11–13

RBD Reliability Block Diagrams 32, 33

RGB Red Green Blue 16

RGBD Red Green Blue-Depth 14

SAE Society of Automotive Engineers 11, 12

SOTIF Safety of the Intended Functionality iii, x, 2–12, 19–21, 32–39, 57–60, 63, 65–67, 70–72, 75, 78, 79, 81–84

SPA Sense-Plan-Act x, 20, 21, 35, 36

STPA System-Theoretic Process Analysis 8, 9, 17, 18, 29, 31

UAV Unmanned Aerial Vehicle 2, 30, 42, 44, 64

V2V Vehicle-to-Vehicle 19

Chapter 1

Introduction

1.1 Background and Motivation

The rapid development of autonomous vehicles has brought significant advancements in modern transportation. Autonomous vehicles have the potential to significantly improve road safety, reduce traffic congestion, and enhance mobility for various user groups. However, the deployment of autonomous vehicles in real-world scenarios introduces a range of complex safety challenges. Autonomous vehicle systems must consistently perceive, interpret, and respond to a variety of dynamic and unpredictable environmental conditions [73].

To ensure the autonomous vehicle safety, several automotive safety standards such as ISO 26262 [43] and ISO 21448 [44] has been established. ISO 26262 focus on identifying and mitigating hazards caused by hardware and software malfunctions in safety-critical electronic systems. Within this framework, Fault Tree Analysis (FTA) is a widely used technique for system-level safety assessment [110]. FTA provides a top-down, deductive method for identifying how combinations of component failures can lead to hazardous events. It offers a structured and hierarchical representation of failure propagation, enables traceability from system-level failures to root causes, and allows for qualitative and quantitative risk assessment. These characteristics make FTA particularly valuable in early system design and safety certification.

Despite its widespread use, FTA is inherently limited when applied to emerging challenges in autonomous vehicle safety. As autonomous systems increasingly rely on perception algorithms, sensor fusion, and learning-based components, safety hazards may arise not from malfunction, but from functional insufficiencies or performance limitations. These

include failures to detect objects due to occlusion, the inability to recognize rare scenarios not covered during training, or degraded performance under adverse weather conditions. These scenarios can lead to safety risks even when the system functions as intended.

To address these broader safety concerns, the ISO 21448 standard, also known as the Safety of the Intended Functionality (SOTIF) is introduced. SOTIF extends the scope of safety analysis beyond traditional fault-based approaches. It focuses on ensuring the intended performance of a system under all foreseeable operating conditions. However, the standard does not prescribe specific tools or quantitative methodologies for analyzing SOTIF-related hazards. This gap emphasizes the need for a more advanced methodology that can reason under uncertainty and capture the complex causal factors involved in SOTIF-related hazards.

1.2 Problem Statement

While FTA remains a valuable method for system-level hazard assessment within the ISO 26262 framework, it exhibits several limitations when applied in the context of SOTIF. One key limitation is its inability to model uncertainty and event interdependencies. FTA assumes static system behavior and binary state variables, which are insufficient for modeling the dynamic and probabilistic nature of autonomous vehicle operations. Furthermore, FTA lacks the flexibility to model interdependencies among events or to account for environmental and algorithmic factors that affect system performance.

In contrast, Bayesian Network (BN) offers a probabilistic graphical modeling approach that explicitly captures causal dependencies and conditional probabilities among variables. BN supports reasoning under uncertainty, updating beliefs based on new evidence, and performing both predictive and diagnostic inference. These capabilities make BN particularly well-suited for addressing the safety analysis needs outlined in the SOTIF, where performance limitations and contextual variability are central concerns. Although temporal extensions such as Dynamic Bayesian Networks (DBNs) and Hidden Markov Models (HMMs) offer additional capabilities for modeling sequential dependencies and evolving system behaviors, this thesis begins with classical BN to establish a solid and interpretable foundation. BN preserve FTA's hierarchical structure and allow integration with limited data, whereas DBNs and HMMs are identified as promising future directions once richer temporal datasets become available.

Although the FTA-BN integration has been successful in safety-critical domains, such as Unmanned Aerial Vehicle (UAV) [138], oil and gas domains [8]. However, its application

on autonomous vehicle systems remain underexplored, especially from SOTIF perspectives. There is currently no standardized methodology that support both qualitative and quantitative evaluation of performance limitations and environmental triggering conditions. This thesis aims to fill this gap by proposing an integrated framework that combines FTA and BN to support SOTIF-aligned safety analysis in autonomous vehicles.

1.3 Research Objectives

The primary objective of this thesis is to develop and validate an integrated safety analysis framework that combines FTA and BN to support SOTIF-aligned safety assessment for autonomous vehicles. The specific research objectives are as follows:

- To identify and characterize the limitations of traditional FTA in modeling SOTIF-relevant scenarios;
- To design a structured process for converting FTA modelings into BN representations that allow probabilistic reasoning;
- To implement and validate the proposed methodology through case studies;
- To conduct both qualitative and quantitative analysis to assess how individual risk factors contribute to system-level hazards.

1.4 Contributions

This thesis makes the following key contributions toward advancing the safety analysis of autonomous vehicles:

- An in-depth review of existing safety assessment methodologies is presented, with a particular focus on FTA, the SOTIF framework (ISO 21448), and their respective roles in the context of autonomous vehicle safety. The review highlights the limitations of traditional fault-based approaches and identifies a methodological gap in analyzing performance limitations and context-dependent hazards.
- A novel framework is proposed that integrates FTA with BN to support probabilistic reasoning under uncertainty that aligns with SOTIF safety analysis. This methodology enables the modeling of conditional dependencies, environmental triggering

conditions, and functional insufficiencies, which are the key aspects emphasized in the SOTIF framework. The proposed approach bridges the gap between deductive structural modeling and probabilistic causal inference.

- This paper develops a structured process for converting FTA with BN representations to support both qualitative and quantitative analysis in autonomous vehicle safety.
- This paper validated the proposed methodology with two case studies in autonomous driving to show the integration is not only feasible, but also practical for analyzing autonomous vehicle safety.

The proposed methodology is validated through two representative case studies in autonomous driving. The first case study validates the effectiveness of integrating FTA and BN in the field of autonomous vehicles. The results show that perception system failures, particularly failure to detect existing objects and object misclassification, are the most significant contributors to collision risk, accounting for 46.06 FIT. The second case study focuses on object detection failure, the framework is applied to model triggering conditions, functional insufficiencies, and failure modes. The results show that adverse weather and occlusion are dominant contributors to object detection failures, underscoring the importance of accounting for performance limitations even in the absence of faults. These studies illustrate the practical applicability of the framework in both qualitative and quantitative safety analysis.

1.5 Methodology

We present a novel methodology that integrates FTA and BN within the SOTIF framework to enable both qualitative structure-based analysis and quantitative probabilistic reasoning. The methodology begins by identifying SOTIF-related influence factors relevant to the vehicle's operational design domain. With input from domain experts, a fault tree is constructed to capture the causal structure of failures. During the qualitative analysis, minimal cut sets are extracted to highlight critical combinations of failures. Next, we convert the fault tree into a BN, enabling probabilistic reasoning and dynamic analysis. While FTA does provide a form of quantitative analysis, it is inherently limited by its binary event structure and assumption of independence among failures, making it difficult to model dependencies, common-cause failure, or multi-state conditions such as varying weather or occlusion. By contrast, BN allows explicit modeling of conditional dependencies, support both predictive and diagnostic inference, and can incorporate data-driven

learning to refine conditional probability tables. This makes BN more suitable for addressing the uncertainty and variability emphasized in SOTIF. Parameter learning is applied by combining expert knowledge with empirical data to refine the conditional probability tables. Iterative refinement ensures continuous improvement with new observations or test results. Finally, the analysis step identifies high-impact factors, which can directly inform safety-driven design decisions. This structured approach bridges FTA and BN under SOTIF principles, ensuring both qualitative and quantitative safety analysis.

1.6 Thesis Organization

The remainder of the thesis is organized as follows: Chapter 2 provides a comprehensive literature review covering autonomous vehicle safety, traditional FTA techniques, the SOTIF framework, and recent integration approaches using probabilistic modeling. Chapter 3 introduces the integration of FTA and BN, detailing the foundational methodology developed to support probabilistic reasoning in safety analysis, and validating through a case study on collision risk in autonomous vehicles. Chapter 4 extends the proposed methodology by aligning it with the SOTIF framework. This chapter demonstrates how the combined FTA–BN approach can be applied to model functional insufficiencies, triggering conditions, and reasonably foreseeable misuse, which are the key concepts in ISO 21448, with validation provided through a case study at object detection failure in autonomous vehicle systems. Chapter 5 concludes the thesis with a summary of findings, limitations of the current work, and suggestions for future research directions.

Chapter 2

Literature Review

This chapter is the outcome of the following publication: Lansu Dai, and Burak Kantarci, "A Survey on Enhancing Autonomous Vehicle Safety through Fault Tree Analysis and SOTIF Integration", ACM Journal on Autonomous Transportation Systems. (Submitted)

As autonomous vehicles continue to evolve and integrate into modern transportation systems, ensuring their safety in uncertain and dynamic environments remains a critical challenge. FTA is widely applied in autonomous vehicle risk assessments, offering structured insights into system failure. However, traditional FTA mainly address hardware and software malfunctions, while hazards arising from functional insufficiencies remain less systematically explored. The SOTIF, outlined in ISO 21448, focuses on these non-fault-based hazards. There is a lack of surveys that thoroughly investigate of FTA within the SOTIF context to comprehensively address both malfunction-induced and functionality-induced risks. It is valuable to bridge the traditional risk analysis with SOTIF-specific hazard evaluation through a dedicated review. This survey reviews the fundamentals and key challenges of autonomous vehicle systems and FTA, and discusses how FTA can be adapted to support SOTIF-based safety assessments. Qualitative and quantitative FTA methods are summarized, along with advanced extensions like BN and Dynamic Fault Tree Analysis. The survey highlights the critical research trends in integrating FTA with SOTIF for autonomous vehicle safety assurance. Finally, current challenges, opportunities and future directions of use of FTA within the SOTIF framework are outlined to enhance risk analysis frameworks for next-generation autonomous vehicles.

2.1 Introduction

Autonomous vehicles have the potential to revolutionize transportation by enhancing safety, reducing congestion, and expanding mobility. As autonomous vehicles become more prevalent, it is crucial to ensure their safety in complex and dynamic environments. Autonomous vehicle systems must interpret their surroundings, predict the behavior of other road users, and respond appropriately under diverse scenarios [11, 130]. As a result, comprehensive and reliable safety assessments are essential to support the reliable deployment of autonomous vehicle technologies [135].

FTA is a widely used risk assessment methodology in engineering systems [145, 146]. FTA is a structured, top-down approach used to analyze how component-level failures can propagate to produce hazardous top-level outcomes. It represents causal relationships through graphical logic trees, allowing engineers to trace the root causes of undesired events. FTA supports both qualitative and quantitative analysis [110]. Qualitative FTA is used to identify minimal cut sets, which are the smallest combinations of events causing undesired events. Quantitative FTA enables probabilistic risk assessment by evaluating the likelihood of undesired events based on the failure probabilities of the basic events. In the automotive industry, FTA is a important tool in supporting of ISO 26262, the international standard for functional safety in road vehicles [115]. ISO 26262 focuses on the hardware and software malfunctions in autonomous vehicle systems [43]. However, this standard does not cover all safety concern. Some hazardous situation arise even when the system behaves as designed, but the intended functionality is insufficient to ensure safety in certain real-world conditions.

To address these non-fault-based hazards, the SOTIF is introduced and formalized under ISO 21448 [44]. SOTIF expands the safety perspective by focusing on risks that emerge from limitations in system behavior, not from faults. These risks include degraded perception in adverse weather, failure to recognize rare objects, or misinterpretation of complex road scenarios. SOTIF also considers reasonably foreseeable misuse, where users interact with the system in unintended and predictable ways. By addressing these aspects, SOTIF complements Functional Safety (FuSa) standards by expanding the definition of safety-relevant hazards.

Although SOTIF establishes a conceptual structure for identifying these hazards, it does not prescribe a specific method for systematically modeling or analyzing them. This leads to growing interest in using FTA within the SOTIF framework [65, 83]. By adapting FTA to represent functional insufficiencies and triggering conditions, engineers can visualize and evaluate how non-malfunctioning components may still contribute to hazardous outcomes.

Recent enhancements to traditional FTA, such as the use of Dynamic Fault Trees [7] and Bayesian Networks [138], enable the modeling of uncertain and adaptive system behaviors. These extensions are especially well-suited to autonomous vehicles, which operate in uncertain environments, rely on probabilistic perception, and increasingly employ learning based control systems.

The integration of FTA with SOTIF provides a more complete foundation for safety assessment, as it can evaluate both malfunction-induced failures and insufficient system performance hazards in a unified manner. Despite this emerging interest, there is currently no survey that systematically examines the use of FTA within the SOTIF framework. A summary of existing survey papers related to SOTIF is provided in Table 2.1. As shown in the comparison, previous surveys primarily addressed SOTIF from a broad perspective or focused on specific subdomains. For instance, Wang et al. [56] provide a broad overview of SOTIF in autonomous driving systems, outlining its theoretical foundations and general safety assessment approaches. Their comprehensive work briefly mentions FTA as a potential analysis tool, but it can be complemented by a survey or tutorial that additionally explores the methodological integration or practical application of FTA within the SOTIF framework. Tang et al. [129] present an extensive review of validation strategies for SOTIF-related safety in autonomous driving. Their work focuses on scenario-based and accelerated testing techniques, which are essential for evaluating system behavior under rare and potentially hazardous conditions. The authors examine frameworks for scenario generation, risk assessment, and testing strategies, including statistical, knowledge-based, and game-theoretic approaches. It emphasizes the importance of simulation-based evaluation for uncovering rare or unseen hazards. Xu et al. [141] focus on the safety challenges related to transitions between manual and automated driving in Level 3 autonomous vehicles. It highlights SOTIF-related risks, such as the limitations of the automated system and driver misuse, and reviews existing safety analysis methods (e.g., Failure Mode and Effects Analysis (FMEA), Hazard and Operability (HAZOP), and System-Theoretic Process Analysis (STPA)), as well as current approaches in human-machine interface design and shared control strategies. Meanwhile, several studies apply FTA to assess safety in autonomous vehicle systems [6,14,23,34,55,74,79,114,118,125,150]. However, these efforts primarily focus on malfunction-related hazards as defined by ISO 26262 and do not consider SOTIF-specific concerns such as functional insufficiency or reasonably foreseeable misuse. This survey seeks to address this critical gap by examining how FTA can be adapted to model and analyze hazards arising from the SOTIF perspective.

This survey explores the combined use of FTA and the SOTIF framework in the context of autonomous vehicle safety. It discusses how this integration can enhance hazard identification and risk analysis by effectively addressing both fault-induced hazards and

Table 2.1: Comparison of SOTIF-Related Survey Papers

Survey Paper	**Focus Area**	**Key Contribution**	**Fault Tree Analysis**	**Human-Machine Interaction**	**SOTIF analysis**
Xu et al. [141] (2022)	Safety concerns in Level 3 autonomous vehicles handover scenarios	Reviews SOTIF hazards and safety analysis methods during mode transitions (e.g., FMEA, STPA)	✗	✓	✓
Wang et al. [56] (2024)	SOTIF principles and safety assessment	Provides a broad overview of SOTIF principles and safety challenges	✗(brief)	✓	✓
Tang et al. [129] (2025)	Scenario-based testing for SOTIF	Reviews accelerated testing techniques using simulation, statistics, and game theory	✗	✗	✓
This Survey	Integration of FTA in SOTIF framework	Systematically explores how FTA models functional insufficiencies and triggering conditions in SOTIF	✓	✓	✓

hazardous scenarios that arise without system malfunctions. Furthermore, the paper identifies critical safety challenges in autonomous vehicle systems, emphasizing the limitations of traditional safety frameworks when dealing with probabilistic behaviors and adaptive systems. By reviewing recent literature on the integration of FTA with SOTIF, this survey highlights existing methodological gaps and proposes future research directions. These recommendations are intended to enhance the applicability and scalability of safety analysis methods for autonomous vehicles, contributing toward more reliable autonomous vehicle deployments.

The structure of this survey is shown in Fig. 2.1. The rest of the paper is organized into the following five sections. Section 2.2 provides an overview of safety in autonomous vehicles, including the definition and operational scope of autonomous vehicle systems, remaining key safety challenges, and an introduction to the SOTIF standard. Section 2.3

presents the principles and applications of FTA, with a focus on its role and relevance in autonomous vehicle systems. Section 2.4 discusses the integration of FTA and SOTIF, highlighting how the combination can support more comprehensive safety assessments. This section also reviews recent research efforts that apply FTA within SOTIF-based analyses. Section 2.5 summarizes key lessons learned, outlines current challenges, and identifies open issues related to the integration of FTA with the SOTIF framework. Finally, section 2.6 concludes and summarizes the survey.

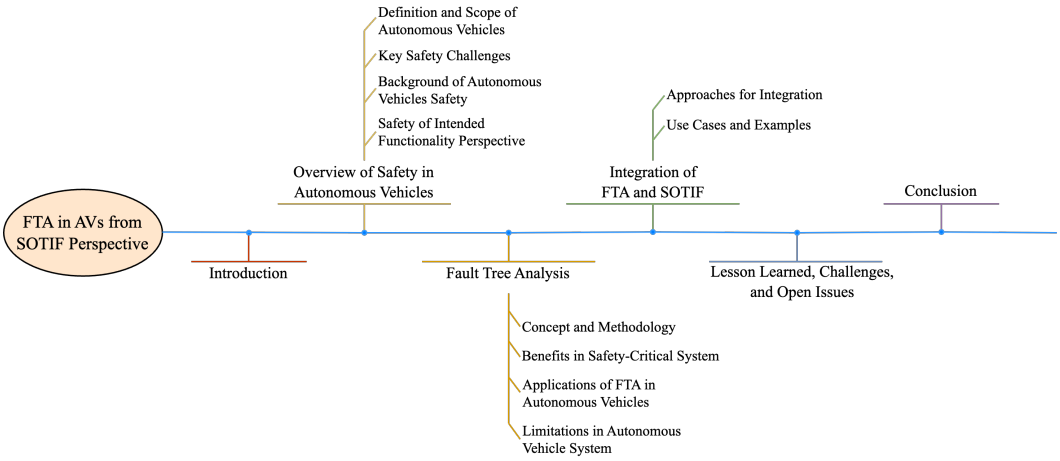

Figure 2.1: The Structure of Survey

2.2 Overview of Safety in Autonomous Vehicles

Autonomous vehicles are expected to transform transportation by enhancing mobility, reducing congestion, and improving road safety. However, their operation in open and unpredictable environments introduces unique safety challenges. This section provides an overview of safety in autonomous vehicle systems, beginning with a definition of autonomous vehicles and the scope of their functional capabilities. It then outlines key safety concerns, particularly those not adequately addressed by ISO 26262, which primarily addresses hazards caused by system malfunction. To address these limitations, the SOTIF, as defined by ISO 21448, is introduced as a complementary framework. SOTIF expands the safety perspective to include hazards resulting from functional insufficiencies or foreseeable misuse, even when the system operates without fault. This broader view is essential for assessing the real-world safety performance of autonomous vehicles.

2.2.1 Definition and Scope of Autonomous Vehicles

Autonomous vehicles are expected to significantly improve road safety by eliminating human errors, which contribute to 94% of traffic accidents [30]. Through the integration of sophisticated sensors, advanced perception, and real-time decision-making algorithms, autonomous vehicles are designed to reduce driver-related mistakes and improve overall traffic safety. Despite their potential benefits, autonomous driving technologies pose new safety challenges. These challenges arise from the complexity of autonomous vehicle software, the unpredictability of environmental conditions, and the need for autonomous vehicles to interact seamlessly with human drivers, pedestrians, and other road users [134]. Ensuring the safety of autonomous vehicles requires a comprehensive approach that addresses not only traditional functional safety concerns, such as hardware and software failures, but also emerging risks associated related to the system's intended functionality. This section provides a comprehensive overview of autonomous vehicle safety. It introduces the definition and classification of autonomous vehicle systems, examines key safety challenges, and presents the SOTIF framework, which complements traditional safety standards such as ISO 26262.

Level of Driving Automation

Autonomous vehicles are advanced transportation systems designed to navigate and operate with little or no human intervention. They rely on a combination of sensors, Artificial Intelligence (AI) algorithms, and control mechanisms to perceive their surroundings, interpret environmental data, and execute driving decisions. To establish a common understanding of automation capabilities, the Society of Automotive Engineers (SAE) defines 6 levels of driving automation ranging from 0 (fully manual) to 5 (fully autonomous) [95]. This classification framework provides a structured approach to categorize the capabilities and limitations of automated driving technologies.

At Levels 0 through 2, the driver must be actively engaged and constantly monitor the vehicle. Level 0 provides only warning systems and momentary assistance, such as emergency braking or lane departure warning, and does not provide sustained vehicle control. Level 1 introduces basic driver assistance, including either lane centering or adaptive cruise control, but not both simultaneously. Level 2 enhances these capabilities by allowing the vehicle to perform both lane centering and adaptive cruise control data at the same time.

A significant transition occurs at Level 3, where the vehicle is capable of handling all driving tasks under specific, predefined conditions. At this level, the system assumes responsibility for Object and Event Detection and Response (OEDR), which refers to

the tasks within the Dynamic Driving Task (DDT) that involve monitoring the driving environment and executing appropriate responses to objects and events. These responses are necessary for completing the DDT and/or managing a fallback scenario when the system requires human intervention. While the system performs OEDR, the driver must be prepared to respond to a takeover request when the system encounters a situation beyond its capabilities [44].

Levels 4 and 5 eliminate the need for driver intervention, with Level 4 functioning autonomously within operational conditions, while Level 5 achieves full automation in all environments without any human involvement. Currently, Level 2 and 3 automation are the most commonly implemented in commercially available vehicles, with features such as Advanced Driver-Assistance Systems (ADAS) and automated highway driving becoming increasingly common [112]. Table 2.2 summarizes the SAE levels of driving automation, highlighting key responsibilities, operational use cases, and potential SOTIF hazards associated with each level.

Architecture of Autonomous Vehicles

The architecture of autonomous vehicles consists of multiple interconnected subsystems which are sensors, perception, decision-making, and vehicle control systems. These systems work collaboratively to interpret the vehicle's environment, make driving decisions, and execute control actions. Fig. 2.2 shows the subsystem architecture of autonomous vehicle systems.

The sensors are responsible for collecting real-time environmental data, providing the foundation of perception and navigation. Autonomous vehicles integrate multiple sensor types to ensure accurate and redundant environmental perception [132,137]. Cameras provide visual information that allows for lane detection, traffic sign recognition, and object classification. Radar is used to measure the distance and velocity of surrounding objects, particularly in adverse weather conditions where optical sensors may be less effective. LiDARs create high-resolution 3D maps of the surroundings by measuring the time-of-flight of laser pulses, which helps in precise obstacle detection. Additionally, Global Positioning Systems (GPS) data support vehicle localization by providing global positioning information, while an Inertial Measurement Unit (IMU) measures acceleration and angular velocity to enable precise motion tracking and state estimation.

The perception system is responsible for interpreting the vehicle's environment using sensor data [108]. Sensor fusion combines data from multiple sensors to improve the accuracy and reliability of environmental understanding. In general, there are two types of fu-

Table 2.2: Levels of Driving Automation

SAE Level	**System Example**	**Who Controls Lateral & Longitudinal Motion?**	**Who Handles OEDR?**	**Who is Responsible for Fallback?**	**Operational Use Cases**	**Key SOTIF Hazards**
Level 0	No Automation (Manual Driving)	Driver	Driver	Driver	All driving situations	Human error (e.g., distracted driving, slow reaction times)
Level 1	Adaptive Cruise Control	Driver & System	Driver	Driver	Maintain headway, adjust speed	Misinterpreting static objects (e.g., bridge perceived as an obstacle)
Level 2	Adaptive Cruise Control + Lane Keeping	System	Driver	Driver	Follow lead vehicle in lane, maintain speed & headway	Failure to detect a merging lead vehicle
Level 3	Traffic Jam Assist	System	System	Fallback-ready User	Following a vehicle below a certain speed, taking control if needed	Heavy fog reduces perception, and driver does not take control in time
Level 4	Highway Co-Pilot, Robo-Taxi	System	System	System	Highway-related and defined geo-fenced urban cases	Misclassification of objects due to lighting/color issues
Level 5	Fully Autonomous Vehicle	System	System	System	Any driving scenario, no human intervention needed	Extreme edge cases (e.g., rare road hazards, unexpected pedestrian behavior)

sion: object-level fusion and raw data fusion, which process sensor inputs differently [155]. In the object-level fusion approach, each sensor independently processes the perception tasks and then combines the results to create a more comprehensive understanding of the environment. In contrast, the raw data fusion approach integrates the raw data from multiple sensors before performing perception tasks. This method generates a dense and precise 3D environmental Red Green Blue-Depth (RGBD) model, allowing for more detailed and accurate scene reconstruction and improving the autonomous vehicle's ability to detect and track objects in complex environments. Both fusion techniques improve the perception capabilities of autonomous vehicles, contributing to more robust environmental awareness and improving overall driving safety.

Beyond sensor fusion, object detection and tracking are essential components of the perception system, enabling the identification and monitoring of objects in the vehicle's surroundings [10, 96]. Object detection involves identifying and classifying objects such as vehicles, pedestrians, traffic signs, and obstacles. The output information supports decision-making and motion planning tasks, enabling autonomous vehicles to anticipate potential hazards and respond accordingly. Once an object is detected, object tracking ensures continuous monitoring of its movement over time. It involves associating the detected object across multiple frames, estimating its trajectory, and predicting its future position. Effective tracking enhances safety and smoothness in autonomous vehicle operations by ensuring reliable decision-making in dynamic traffic environments. In parallel, the localization system determines the vehicle's precise position using a combination of GPS, IMU, and offline maps [76]. The perception system forms the foundation of autonomous driving technology by incorporating sensor fusion, object detection, object tracking, and localization.

The decision-making system generates safe and collision-free driving decisions that guide the autonomous vehicle toward the destination. This system has three main components: route planning, behavioral planning, and motion planning [86, 117, 120]. Route planning determines the optimal path from the vehicle's current location to its final destination. It takes into account several factors, including road networks, traffic conditions, and predefined constraints. Once the route is established, behavioral planning refines this process by selecting appropriate driving behaviors based on perceived agents, obstacles, and signage in the surroundings. It makes high-level driving decisions such as changing lanes, stopping at intersections, and adjusting speed based on surrounding traffic. Finally, motion planning translates these high-level decisions into detailed and executable trajectories. It involves estimating future vehicle poses and identifying collision-free spaces within the environment to ensure safe and smooth navigation. Combining three components enables autonomous vehicles to navigate complex traffic scenarios and maintain safety and

compliance with road regulations.

Once driving decisions are formulated, the vehicle control system executes the planned actions by transmitting control signals to various actuators. The accelerator pedal motor controls the vehicle's speed, while the brake pedal motor ensures smooth deceleration and stopping. The steering wheel motor controls lateral movements and direction changes. Moreover, the gear motor controls automatic transmission shifts based on the driving scenarios. These control mechanisms ensure that planned vehicle actions are executed safely. Through the interaction between subsystems, the autonomous vehicle system can operate efficiently and safely across a wide range of driving scenarios.

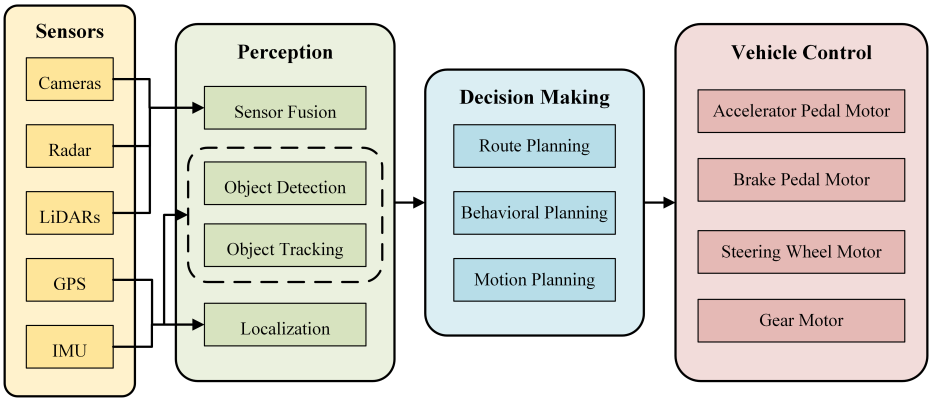

Figure 2.2: Subsystem Architecture of Autonomous Vehicle Systems [103, 156]

2.2.2 Key Safety Challenges

One of the primary safety challenges for autonomous vehicles is accurate perception and decision-making in diverse environments. Perception systems in autonomous vehicles integrate data from multiple sensors, including but not limited to LiDAR, radar, cameras and GPS. Each of these sensors contributes distinct advantages to environmental perception and situational awareness [40, 108, 131, 147]. However, inherent limitations in sensor technologies frequently introduce systematic uncertainties, leading to errors in object identification, classification accuracy, and spatial localization.

LiDAR provides high-resolution 3D point clouds for precise object detection, but it is vulnerable to performance degradation in adverse weather conditions, such as snow, rain, or fog [70, 105]. Radar has the advantage of maintaining reliable detection performance in adverse weather conditions. However, it also has some limitations, including low angular

resolution [15], a restricted Field of View (FOV) [152], and a tendency to produce false positives due to signal reflections and multipath effects [108]. Cameras are essential for classifying objects because they can detect Red Green Blue (RGB) information. They are highly sensitive to lighting variations (such as glare and shadows) and fail in extreme weather [21]. GPS is frequently used in autonomous vehicles for navigation and providing positioning data. These systems use data received from GPS satellites. It faces signal obstructions in urban canyons or tunnels [25].

To address the limitations of individual sensors, autonomous vehicles employ sensor fusion, which combines data from multiple sources to enhance perception accuracy. However, sensor fusion introduces additional challenges, such as sensor misalignment, and timing inconsistencies [137]. Combining data from sensors with different physical positions and orientations requires precise calibration. Even small misalignment can lead to errors when merging high-resolution data (e.g. LiDAR point clouds) with lower-resolution or differently oriented sensor outputs.

Moreover, autonomous vehicles must make real-time decisions while navigating complex and dynamic environments, including unstructured roads, dense traffic, and unpredictable interactions with pedestrians and other road users [27]. Learning-based algorithms, such as deep learning and reinforcement learning, are widely used in autonomous driving to enable continuous learning from extensive datasets. However, these algorithms still face limitations in their adaptive capabilities, particularly when faced with novel or highly variable driving conditions that were not well represented in their training data [33]. Autonomous vehicles may struggle to generalize across different environments, such as transitioning from structured urban streets to unpaved roads. Moreover, learning-based models often function as black boxes, making it difficult to interpret their decision-making processes and ensure safety in all scenarios [86]. This lack of transparency raises concerns about accountability and trust. Addressing these challenges in both perception and decision-making systems is crucial for improving the safety and reliability of autonomous vehicles.

2.2.3 Background of Autonomous Vehicles Safety

The increasing complexity of autonomous vehicle systems requires the development of diverse risk analysis and safety assurance methodologies. As autonomous vehicles transition from ADAS to higher levels of autonomy, the safety assurance process must address a broader range of potential hazards [68, 90, 102]. Traditional risk analysis techniques such as FTA [78], FMEA [82], and HAZOP [36], provide a structured method for identifying failure modes and their effects. More recent probabilistic and dynamic approaches, such

as BN [26], Markov Chains [28], Monte Carlo simulation [52], and STPA [99], provide enhanced capabilities for modeling interdependencies, system dynamics, and emergent behavior in safety-critical systems. Table 2.3 shows the comparison of risk analysis methods in autonomous vehicle safety. These methods collectively support the evaluation of both component-level failures and system-level hazards, resulting in a more comprehensive understanding of risk in autonomous driving contexts.

Despite the availability of these tools, a key challenge in autonomous vehicle safety remains which is ensuring reliable operation not only in the presence of faults, but also under normal operating conditions, when performance limitations or design insufficiencies may still result in unsafe outcomes [153]. Traditional safety standards, such as ISO 26262, are primarily focused on functional safety, addressing hazards caused by hardware or software malfunctions. However, they do not sufficiently address scenarios in which the system performs as intended but fails to ensure safety due to limitations in perception, decision-making, actuation, or interactions with complex environments.

The Operational Design Domain (ODD) plays a foundational concept in autonomous vehicle safety. ODD defines the specific operating conditions under which an autonomous vehicle system is expected to function safely [148]. These conditions include environmental factors such as weather, lighting, and road types, as well as operational constraints like speed limits, traffic scenarios, and geographic regions [69, 151]. As the level of automation increases, the scope of the ODD gradually expands, requiring systems to maintain safe performance under a wider range of operational and environmental conditions [75]. The precise definition of ODD makes it clear to understand the system's intended capabilities and boundaries, allowing engineers to tailor their safety assessments and validation strategies accordingly. Furthermore, the ODD helps to develop scenario-based testing strategies and ensures that the system is only deployed in environments where it has been validated. From a regulatory standpoint, a well-defined ODD enhances traceability and compliance by linking system capabilities to their intended use.

However, as autonomous vehicles are increasingly deployed in dynamic and unpredictable environments, it becomes apparent that fault-based safety approaches are insufficient [133]. Recent advancements in trajectory prediction and intelligent route planning [120] underscore the need for autonomous vehicle systems to adapt to real-time traffic and intelligently update routes in response to environmental variability. Nonetheless, unsafe behavior may still occur due to inherent system limitations or edge-case scenarios that are not fully captured during design and validation [72]. These challenges highlight the importance of looking beyond traditional assumptions and examining how performance constraints, design assumptions, and real-world variability contribute to residual safety risks. This recognition motivates a shift in the safety assurance paradigm toward a broader

Table 2.3: The Comparison of Risk Analysis Methods in Autonomous Vehicle Safety

Method	Reasoning		Nature of Analysis		Dynamic	Typical Use	Ref.
	Inductive	Deductive	Qualitative	Quantitative			
FTA		✓	✓	✓	✗	Identifying failure combinations leading to a top-level hazard	[14, 23, 34]
FMEA	✓		✓	✓	✗	Identifying and prioritizing failure modes at the component level	[24, 66, 97]
HAZOP	✓		✓		✗	Identifying process deviations and their consequences	[81, 93, 127]
BN	✓	✓		✓	✓	Modeling uncertain dependencies and probabilistic reasoning	[50, 57, 121, 124]
Markov Chains	✓	✓		✓	✓	Modeling system behavior over time with state transitions	[5, 107, 140]
Monte Carlo Simulation	✓			✓	✓	Estimating probabilities through repeated simulations	[4, 91, 149]
STPA	✓		✓		✗	Analyzing unsafe control actions and system-level interactions	[2, 85, 111, 139]

perspective that considers failure modes and the conditions under which a system may behave unsafely even when operating properly [116]. These insights provide the foundation for more comprehensive safety frameworks beyond functional safety. The following section introduces a framework that addresses the safety of intended functionality and explores its role in determining the safety of autonomous vehicles in complex and uncertain operational domains.

2.2.4 Safety of Intended Functionality Perspective (SOTIF)

Overview of SOTIF Framework

The SOTIF framework, as outlined in the ISO 21448 standard [44], addresses potential hazards arising from the intended function of a system, particularly in automated and autonomous vehicle systems. Unlike traditional functional safety focuses on mitigating risks caused by system failures, SOTIF examines whether the required safety functionalities can be maintained under unknown or unforeseen conditions, even in the absence of a system failure [98]. It specifically considers risks arising from functional insufficiency and reasonably foreseeable misuse. Functional insufficiencies refer to limitations in the system's design, implementation, or performance that result in unsafe behavior [41]. These limitations may include sensor misinterpretations, incorrect decision-making algorithms, or failure to detect obstacles. On the other hand, reasonably foreseeable misuse involves unintended user interactions with the system that could compromise safety [119]. For example, when a driver misuses an autonomous function due to over-reliance on automation. Both categories are critical considerations in the assessment and mitigation of risks associated with autonomous technologies.

As shown in Fig. 2.3, SOTIF categorizes scenarios into four quadrants based on their hazard potential and awareness level: known not hazardous scenarios (Area 1), known hazardous scenarios (Area 2), unknown hazardous scenarios (Area 3), and unknown not hazardous scenarios (Area 4) [44]. These scenarios involve both inherent and external risks that impact autonomous vehicle safety. Inherent risks originate in the vehicle itself, including sensor limitations, perception errors, decision-making errors, and motion control inaccuracies. External risks are caused by environmental factors such as extreme weather, poor lighting, temporary obstacles, road defects, and influences from Infrastructure-to-Vehicle (I2V) and Vehicle-to-Vehicle (V2V) communication. The goals of SOTIF are to evaluate and mitigate risks by ensuring that known hazardous scenarios (Area 2) are minimized through functional modifications and to minimize the unknown hazardous scenarios by identifying and mitigating risks through validation and refinements. This classification

helps to systematically evaluate risk scenarios and implement appropriate safety measures to improve autonomous vehicle reliability.

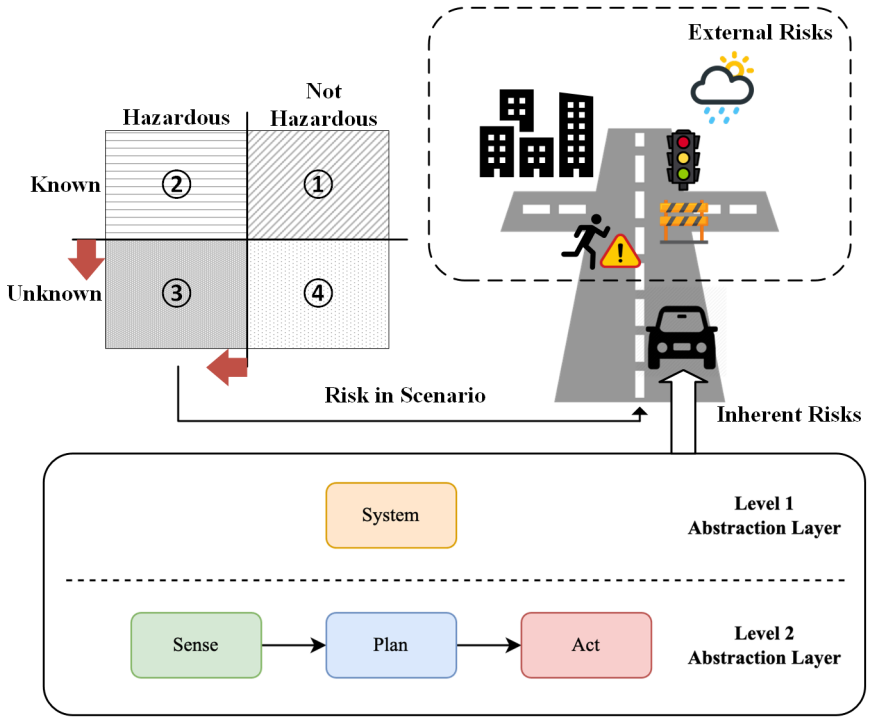

Figure 2.3: Integration of SOTIF Risk Scenarios with the Sense-Plan-Act (SPA) Model in Autonomous Vehicle Systems [44, 144]

Sense-Plan-Act Model

ISO 21448 emphasizes the use of the SPA model as a conceptual representation of key functional components and interactions with an autonomous driving system. According to Clause 4.2.3 of the SOTIF standard [44], possible causes of hazardous behavior are closely related to the system's ability to accurately perceive its environment, make appropriate decisions, and execute those decisions reliably. As shown in Fig. 2.3, the SPA model decomposes the system into three functional elements: Sense, Plan, and Act.

The Sense element is responsible for the perception process, including both external surroundings and internal vehicle states. This includes sensor data collection, sensor fusion, localization, and generation of an environmental model. The Plan element processes

the environmental model provided by the Sense element. It applies predefined goals and decision logic to evaluate the current context and determine appropriate control actions. This includes route planning, behavioral planning, and motion planning. Finally, the Act element executes the planned control actions by interfacing with the vehicle's actuators. These actuators manage essential vehicle functions such as acceleration, braking, and steering. The effectiveness of the Act element is critical to ensuring that the intended behavior is carried out safely and accurately, thereby maintaining overall system reliability and performance.

Moreover, ISO 21448 highlights that decision algorithms are embedded across all three elements of the SPA model. In other words, each element of the SPA model has its own set of decision algorithms that contribute to the overall behavior of the system. For example, in the Sense element, decision algorithms are applied to tasks such as object detection, object classification, and sensor fusion. These processes involve identifying relevant objects in the environment and integrating data from multiple sources to form a coherent environmental model. ISO 21448 further emphasizes the importance of selecting a capable and adaptable system architecture that aligns with the SPA structure. The system architecture should support the identification and mitigation of hazards throughout the entire development lifecycle, beginning from the early design stages. The standard also recommends that the system architecture should be regularly reviewed and refined as development progresses. This iterative approach ensures that the architecture remains robust and appropriate for addressing SOTIF-related risks as operational conditions and system functionalities change over time.

Key Differences from ISO 26262

ISO 21448, SOTIF, complements ISO 26262, FuSa, by addressing hazards that arise from intended functionality rather than hardware or software failures. Table 4.1 presents a comparison between the two standards. While FuSa focuses on preventing system malfunctions, SOTIF ensures that even a correctly functioning system does not lead to unsafe situations due to inherent design limitations or unforeseen environmental conditions [88, 104].

Fig. 2.4 illustrates the distinct causal pathways leading to hazards as defined by ISO 26262 and ISO 21448, respectively. ISO 26262 emphasizes hazards originating from system malfunctions [89]. In this context, the causal pathway begins with a fault, which is an abnormal condition that can cause an element to fail. This fault can lead to a failure, defined as the termination of an intended behavior due to fault manifestation. As a result, the system may exhibit malfunctioning behavior, referring to a deviation from the design intent, either through failure or unintended operation. Malfunctioning behavior has the

potential to give rise to a hazard, which is the potential source of harm including physical injury or health damage. In contrast, ISO 21448 addresses hazards that occur even in the absence of system faults or failures. It focuses on the hazardous behavior that may emerge from the system functioning as designed but with insufficient capabilities [71,84]. The key initiating element is the triggering condition, which is a specific situation factor within a scenario that initiates a system response. This may lead to functional insufficiency, referring to the system's inability to perform its intended task or inability to prevent a reasonably foreseeable misuse. Functional insufficiencies can result in hazardous behavior at the vehicle level, ultimately leading to a hazard. The figures highlight the importance of addressing both hazards arising from system malfunctions, as covered by ISO 26262, and hazards resulting from functional limitations in the absence of faults, as outlined in ISO 21448, to achieve a comprehensive and robust automotive safety assessment.

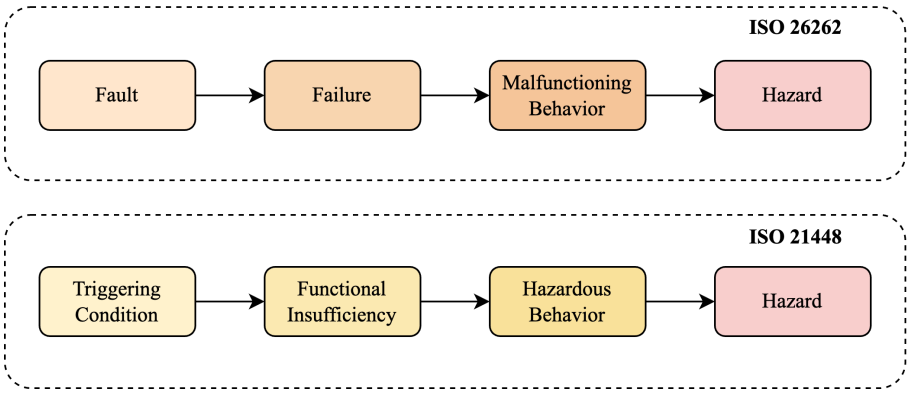

Figure 2.4: Root Cause Identification for the Existence of Hazard identified in ISO 26262 and ISO 21448

2.3 Fault Tree Analysis

Ensuring the safety of autonomous vehicles requires systematic methods for identifying and analyzing potential failures. FTA is a well-established deductive technique that models how combinations of lower-level component failures can lead to hazardous outcomes at the system level. This section introduces the foundational principles and structural elements of FTA, covering both qualitative and quantitative approaches. It emphasizes how FTA supports a structured understanding of system vulnerabilities and informs the prioritization of safety measures. The section also reviews representative applications of

Table 2.4: Comparison of ISO 26262 Functional Safety and ISO 21448 Safety of Intended Functionality Perspective

Standard	Focus on	Scope	Applicability	Example
ISO 26262	Prevent hazards caused by malfunction of electrical/electronic systems	Address failures: Hardware fault, systematic design failures	Any safety-critical system	Prevent unintended acceleration due to a software glitch or faulty sensor
ISO 21448	Mitigate risks from functional insufficiencies and performance limitations of the intended functionality	Address limitations: Sensor performance insufficiency, perception limitation, decision-making limitation, actuator accuracy, human misuse	Primarily ADAS and autonomous vehicles	Prevent unintended acceleration due to sensor weakness or a driver misuse a feature

FTA in the autonomous vehicle systems, highlights their limitations, and examines recent advancements aimed at extending FTA to accommodate dynamic, uncertain, and adaptive system behaviors.

2.3.1 Concept and Methodology

FTA is a top-down, deductive approach used to evaluate the reliability and safety of the system. It begins by identifying an undesired event as a top event and systematically breaks it down into its contributing factors, which are visually represented in a graphical model called a fault tree. This method helps to clarify complex interactions between system components and makes it easier to identify potential failure points.

Fault Tree Structure

FTA employs a tree structure to illustrate the logical relationships between component failures and system-level faults. The fault tree consists of two main types of nodes: events, which denote different kinds of failure conditions, and gates, which describe how multiple events combine logically to produce higher-level failures [110]. Table 2.5 summarizes the symbols commonly used for these event and gate types in the fault tree.

An example of a simple fault tree for an autonomous vehicle steering system is shown in Fig. 2.5 [55]. In this example, the top event is defined as steering system failure. The basic events AS2, AS3, AS11, and AS12 represent failures of the Electronic Control Unit (ECU), steering electrical motor, torque sensor, and angle sensor, respectively. By systematically breaking down complex systems into discrete events and analyzing their interdependencies, FTA provides critical insights into the underlying causes of potential failures.

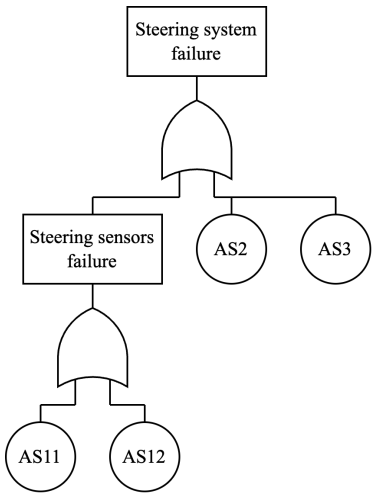

Figure 2.5: Example of a Fault Tree of Steering System in Autonomous Vehicle [55]

General Procedure for Fault Tree Analysis

The process of conducting FTA follows a structured and systematic approach to identifying the root causes of an undesired event within a system, the following steps outline the general procedure of performing an effective FTA [9, 39]. The first step is to clearly define the undesired event, also known as the top event. This step should specify the problem of interest that will be addressed, such as a system shutdown, or a safety issue. A well-defined top event determines the scope and focus of the entire analysis.

Once the top event is established, the next step involves identifying the contributing events that may directly lead to its occurrence. These include basic events, which are primary failure events with no further decomposition, and intermediate events, which result from combinations of other events. Then the fault tree is developed by logically connecting events using Boolean logic gates to represent their relationships. The construction begins with the top event and proceeds systematically downward, identifying the events and conditions until all branches terminate in the basic events. This process ensures the fault tree

Table 2.5: FTA Event and Logic Gate Symbols with Definitions

Type	**Name**	**Symbol**	**Definition**
Event	Basic Event		Represents a primary failure at the component level, which cannot be further decomposed.
	Intermediate Event		Represents a fault that occurs as a result of one or more basic events.
	Top Event		Represents the overall system-level failure that the fault tree aims to analyze and prevent.
	Undeveloped Event		Represents an event that is not further developed because of lack of information.
	Conditioning Event		Represents a specific condition or restriction that applies to a logic gate, often used in conjunction with Inhibit gates.
Gate	AND Gate		Indicates that all input events must occur simultaneously for the output event to happen.
	OR Gate		Indicates that at least one of the events must occur for the output event to occur.
	XOR Gate		Produces an output when exactly one of its input events occurs.
	K/N Gate		Produces an output when at least K out of N input events occur.
	Inhibit Gate		Allows the input event to propagate only when a conditional event is also present.

provides a complete and logical representation of the failure pathways. After constructing the fault tree, relevant data are gathered to support the analysis. The data typically comes from historical data or expert opinions.

The next step is to perform a qualitative analysis by identifying the minimal cut sets, which are the smallest combinations of basic events that can lead to the occurrence of the top event. This helps in understanding the structure of the system's vulnerabilities and potential failure scenarios. Then, the likelihood of the top event is calculated based on the failure probabilities of basic events. Quantitative analysis enables risk quantification and supports informed decision-making. Finally, the results of both analyses are interpreted to identify the most critical vulnerabilities in the system. Based on these findings, recommendations for corrective actions or design improvements are proposed to reduce the associated risks. The construction of the fault tree requires a thorough understanding of the system's operation. This ensures the accuracy and effectiveness of the analysis. The following section expands on the qualitative and quantitative analysis used in FTA.

Qualitative and Quantitative FTA

FTA can be categorized into qualitative and quantitative approaches, each providing unique insights into system safety and reliability [9, 59, 110]. These complementary methods offer a comprehensive understanding of potential failure modes and their impacts on overall system performance.

Qualitative FTA focuses on understanding the logical relationships and hierarchical pathways leading to system failure without relying on probabilistic data. By mapping these pathways through logic gates, qualitative FTA identifies Minimal Cut Sets (MCS) or Minimal Path Sets (MPS) [42]. This process identifies critical vulnerabilities within the system, such as single points of failure and dependencies between components. Mathematically, let T represents the top event, and let $E_1, E_2, \ldots, E_n$ denote the set of basic events. A cut set C satisfies the condition

$$\phi((E_1, E_2, \ldots, E_n) = 1$$

when all $E_i \in C$ are true (i.e. $E_i = 1$), where structure function ϕ defines the logical relationships in the fault tree. A minimal cut set C_{min} cannot be further reduced without losing its cut set status. Formally, for any subsets $C' \subsetneq C_{min}$, C' does not satisfy $\phi(C') = 1$. Minimal cut sets thus represent the smallest set of basic events that can result in the top event. Identifying MCS allows engineers to prioritize risks and develop effective redundancy strategies. By systematically analyzing the failure structure, qualitative FTA

provides valuable guidance on risk mitigation and system design improvements. While MCS focus on system failure, MPS focus on system success. A MPS is the smallest set of basic events that, if all these events function properly, will result in system success [110]. Understanding MPS is important for reliability analysis, as it helps engineers understand which components must be functional to ensure system success. Another important consideration in qualitative FTA is the Common Cause Failures (CCF). A common cause failure refers to a single failure event that simultaneously affects multiple system components [13]. CCF introduce dependencies among basic events in FTA, which affect overall system reliability and risk assessment. Identifying and mitigating CCF is crucial in designing robust systems, as they can significantly impact overall system safety and effectiveness.

On the other hand, the quantitative FTA provides a numerical assessment of system reliability and risk [35]. This approach assigns probabilities to each basic event, typically based on historical data and expert judgment [110]. These probabilities are then propagated these values through the fault tree to estimate the likelihood of the top event. The probabilities of the intermediate events are determined based on the basic events and the logical gate connecting them. The two most common gates are the AND gate and the OR gate [63]. For an AND gate, all input events must occur for the output event to occur. The probability of the output event $P(T_{AND})$ is the product of the probabilities of all input events:

$$P(T_{AND}) = \prod_{i=1}^{n} P(E_i) \tag{2.1}$$

For an OR gate, the output event occurs if at least one input event occurs. The probability of the output event is calculated using the complement of the product of the complements of the input probabilities:

$$P(T_{OR}) = 1 - \prod_{i=1}^{n} (1 - P(E_i)) \tag{2.2}$$

The probability of the top event $P(T_{OR})$ is further calculated by the probability of intermediate events. For more complex gates or combinations, such as NOT gates, additional calculations are required. This probabilistic approach enables a detailed numerical evaluation of system risk, quantifies the failure probability of critical events, and assesses the effectiveness of mitigation strategies. Both qualitative and quantitative FTA provide a comprehensive understanding of system risks and failure pathways, enabling for more informed risk management decisions.

2.3.2 Benefits in Safety-Critical Systems

FTA offers several significant benefits when applied to safety-critical systems, particularly in industries such as nuclear power [53], oil and gas [8], and automotive [14, 17]. By systematically identifying potential hazards and failure modes, FTA enables engineers to proactively address safety concerns and design more resilient systems. The effectiveness of FTA is demonstrated across multiple domains, where it plays a crucial role in ensuring regulatory compliance and improving system reliability. Hassan et al. [53] apply FTA to analyze the time-dependent reliability of a safety injection system in an advanced pressurized water reactor. The study demonstrates the critical role of safety injection pumps in both short-term and long-term system reliability.

FTA provides a structured approach to identifying potential hazards and failure modes within complex systems [38, 39]. This systematic method helps engineers and safety analysts to thoroughly examine all potential paths that could lead to a critical failure. By enabling a comprehensive analysis of system vulnerabilities, FTA contributes to a more robust system design and enhanced safety measures. One of the key advantages of FTA is its ability to support quantitative risk assessment [1, 58]. By assigning probabilities to basic events, engineers can calculate the likelihood of top-level system failures, allowing for more informed safety decisions. This quantitative aspect is especially useful in industries where accurate risk estimation is critical for regulatory compliance and operational safety.

FTA also enables the prioritization of safety measures through the analysis of minimal cut sets and importance measures. Bhavsar et al. [14] perform a fault tree analysis to assess the risks associated with autonomous vehicle failure in mixed traffic streams. They identify critical failure probabilities and minimal cut sets that impact road safety. In addition, Chen et al. [23] conduct a fault tree analysis to identify potential failure sources in automated driving, focusing on the interactions between human drivers and automated driving systems. Their study highlights the importance of control transitions and the need for improved system interfaces to mitigate risks. This prioritization allows for more efficient allocation of resources in implementing safety mechanisms and redundancies. By leveraging these benefits, organizations can significantly enhance the safety and reliability of critical systems. Therefore, this will reduce the risk of catastrophic failures and improve overall operational performance in safety-critical domains.

2.3.3 Application of FTA in Autonomous Vehicles

FTA is extensively used in the safety assessment of autonomous vehicles due to its systematic, deductive approach to identifying the root causes of hazardous system-level failure.

Table 2.6 provides a summary of representative studies that apply FTA in various autonomous vehicle contexts, ranging from traditional static fault modeling to more advanced probabilistic and dynamic techniques.

Early works, such as Bhavsar et al. [14], focus on modeling failures in vehicular and infrastructure components using traditional FTA. Schonemann et al. [118] use FTA to systematically derive safety requirements for automated valet parking. While these efforts provide foundational insights, they are primarily focused on hardware failures and do not account for dynamic environmental factors or human factors. Later studies expand the analytical scope of FTA by incorporating complementary methods. For instance, Kramer et al. [74] combine FTA with HAZOP and STPA to capture a broader range of hazards in automated driving systems. However, the method remains constrained by its dependence on predefined scenarios, limiting its scalability across broader operational contexts.

Several studies explore the quantitative extension of FTA. Hedal et al. [55] employ Monte Carlo simulations to evaluate subsystem reliability in autonomous electric vehicles, providing probabilistic estimates under component degradation. Sreeraj et al. [125] apply probabilistic FTA to evaluate failure probabilities in battery management systems, aligning their analysis with ISO 26262 requirements. These quantitative approaches offer actionable insights but often assume static event independence, which limits their ability to capture system interdependencies or adapt to real-time data.

To address these limitations, recent research explores the integration of FTA with dynamic and probabilistic modeling techniques. Samadi et al. [114] integrate statistical model checking with FTA to assess failure probabilities over time, offering a temporal dimension to fault analysis. The most recent work focuses on integrating FTA with probabilistic graphical models to improve adaptability. Dai et al. [34] combine FTA with BN to evaluate collision risk in autonomous vehicle systems. This hybrid model supports dynamic updating of failure probabilities and captures interdependencies between events, addressing core limitations of traditional FTA. However, integrating BN introduces added modeling complexity and computational overhead, which may affect its applicability in real-time systems.

Overall, these studies demonstrate the versatility and limitations of FTA in autonomous vehicle safety analysis. While FTA effectively supports the identification of failure modes and hazardous pathways, many studies do not capture dependencies between basic events, which could lead to oversimplification of complex system interaction [6, 14, 74, 79, 118, 125, 150]. Moreover, the lack of quantitative FTA in several studies limits the ability to prioritize risks based on their probability and severity [23, 79, 150]. These limitations underscore the need for more comprehensive approaches that can capture the complex interdependency

in autonomous vehicle systems and provide quantitative risk assessments to guide safety improvements.

2.3.4 Limitations in Autonomous Vehicle Systems

Traditional FTA remains a valuable tool for analyzing system failures across various industries [39, 63]. However, it faces several significant limitations when applied to complex systems like autonomous vehicles. The dynamic nature of autonomous vehicles, which operate in highly uncertain and continuously evolving environments, introduces complexities that traditional FTA struggles to address [46]. This section highlights the key limitations of traditional FTA in autonomous driving scenarios and presents existing solutions to address these limitations.

FTA is a static and deductive methodology that models failures based on predefined logic gates and fixed failure probabilities [32]. However, autonomous driving systems operate through continuous and dynamic interactions between sensors, software algorithms, and external environmental conditions. Traditional FTA struggles to capture how these dynamic interactions propagate through the system and influence the top event in real-time [63]. As a result, safety assessments based on FTA alone may be incomplete or may not fully reflect the complexity of real-world operational environments. Furthermore, the adaptive nature of autonomous systems adds another layer of complexity. Autonomous vehicles incorporate machine learning algorithms that allow them to learn from new data and adjust their behavior over time. As these systems evolve, their potential failure modes and associated risks also change. Traditional FTA, with its static structure, limits its effectiveness in assessing the reliability of AI-based decision-making processes.

To address these challenges, researchers propose several enhanced approaches that combine FTA with other reliability analysis techniques to better model uncertainty, dynamic behavior, and adaptive systems. Table 2.7 summarizes advanced methodologies that extend traditional FTA, highlighting their mechanisms and specific advantages in dealing with dynamic, uncertain, and adaptive nature of autonomous vehicle systems. BN are the most popular method for dealing with uncertainty. BN provides a probabilistic and dynamic framework for modeling dependencies between system components and updating failure probabilities as new evidence is introduced [87]. This capability makes BN especially suitable for autonomous vehicle systems that constantly learn and adapt over time. Xiao et al. [138] integrate FTA with BN to improve UAV safety analysis, demonstrating the effectiveness of BN in capturing statistical dependencies among components and allowing dynamic risk assessment based on real-time flight data. Fuzzy Fault Tree Analysis (FFTA)

Table 2.6: Comparison of FTA Applications in Autonomous Vehicle Systems

Ref.	Year	Method	Remark	Limitation
[14]	2017	FTA	Develop fault trees for vehicular and infrastructure components	Limit modeling of unpredictable human behavior and dynamic traffic events
[118]	2019	FTA	Systematically derive safety requirements for automated valet parking	Limit applicability for all safety goals
[74]	2020	FTA with HAZOP and STPA	Develop an integrated safety assessment method for Automated Driving System (ADS), analyze hazardous scenarios and derive risk mitigation measures	Limit to scenario-based hazard identification and lack general-purpose applicability beyond ADS
[6]	2020	FTA	Analyze dataset-related failures in machine learning systems	Does not account for dataset quality issues
[114]	2021	FTA with Statistical Model Checking	Analyze failure probabilities over time in autonomous systems	Traditional FTA does not effectively analyze time-dependent failures
[150]	2021	Extend Traditional FTA	Model failures due to both natural and adversarial perturbations in neural network-based perception systems	Challenges in quantifying neural network-based failures due to their black-box nature
[23]	2022	FTA	Construct FTA considering ODD structure and human processing stages	Lack quantitative analysis and event dependencies
[55]	2023	FTA with Monte Carlo simulation	Evaluate reliability of autonomous electric vehicle subsystems	Focus primarily on electric components in autonomous vehicles
[125]	2023	FTA	Use quantitative FTA to evaluate failure probabilities in the battery management system of an Autonomous Electric Vehicles (AEV)	FTA does not fully capture dynamic failures or time-dependent risks in Battery Management System (BMS)
[79]	2024	FMEA with FTA	Develop a fault database and demonstrate fault injection benefits	Lack quantitative FTA analysis
[34]	2025	FTA with BN	Combine FTA and BN to assess autonomous vehicle collision risk and dynamically update failure probabilities	BN integration improves adaptability but increases modeling complexity

represents another advanced extension of traditional FTA. It integrates fuzzy logic to address the uncertainty and imprecision in failure data, especially when precise statistical information is unavailable. Zhao et al. [154] propose a hybrid method that combines it with fuzzy logic and a Noisy-OR gate BN to overcome the static limitations of traditional FTA. Their approach enables more flexible and realistic risk assessments, as demonstrated through a case study on maritime navigation accidents at Qinzhou Port.

Further enhancements to FTA involve computational enhancements through logic-based representations. Combining FTA with Binary Decision Diagrams (BDD) enables efficient quantitative analysis and identification of the critical components. Marquez et al. [48] apply this integrated approach to the reliability assessment of wind turbine systems. By transforming the fault tree into a BDD structure, it can efficiently evaluate system failure probabilities and reduce the computational complexity associated with traditional FTA. Reliability Block Diagrams (RBD) also offer a complementary perspective. RBD model how overall system reliability depends on the reliability of individual components [19]. Jakkula et al. [60] conduct a case study of Load-Haul-Dumper systems in underground mining, integrating RBD with FTA. This method provides a broader perspective on system reliability by using RBD to assess system reliability and FTA to identify critical failure modes.

Dynamic Fault Tree Analysis (DFTA) extends traditional FTA by incorporating dynamic logic gates, such as sequence-enforcing, standby, and priority gates, to represent time-dependent interactions and failure orderings in complex systems [7]. This extension enables the modeling of failure behaviors that change over time, which static fault trees cannot adequately capture. Moreover, Monte Carlo Simulation is a stochastic analysis technique that relies on repeated random sampling to estimate the probabilistic behavior of systems, particularly under uncertainty [106]. It is especially effective for analyzing systems with many interdependent variables and complex operational conditions. Rao et al. [37] integrate DFTA with Monte Carlo Simulation to assess the reliability of nuclear power plant systems, enabling more realistic modeling of dynamic failure behaviors and overcoming the computational limitations. These advanced methods collectively enhance the applicability of FTA in autonomous vehicle systems by enabling more flexible modeling of uncertainty, dynamic interactions, and adaptive behaviors.

2.4 Integration of FTA and SOTIF

FTA mostly applies to assess malfunction-induced hazards in safety-critical systems. The SOTIF framework introduces new challenges that arise from functional insufficiencies and

Table 2.7: Enhanced FTA Methods and Their Advantages for Autonomous Vehicle Systems

Method	Description	Advantage over Traditional FTA	Ref.
FTA + BN	Integrates FTA structure with probabilistic graphical models to capture conditional dependencies between components	Enables dynamic risk assessment and belief updating as new data becomes available	[138]
FFTA	Applies fuzzy logic to model vague or incomplete failure data when precise probabilities are unavailable	Accommodates uncertainty and expert judgment in the absence of statistical failure rates	[154]
FTA + BDD	Transforms fault trees into BDD for efficient logical evaluation and probability computation	Reduces computational cost in large-scale or combinatorial complex systems	[48]
FTA + RBD	Combines fault-based and reliability-based system models to capture both failure logic and structural dependencies	Provides a broader view of system reliability and identifies both critical paths and failure modes	[19, 60]
DFTA	Introduces time-sequencing logic gates to capture order-dependent failures and dynamic behavior	Models failures that depend on timing, such as standby and priority-based behaviors	[7]
DFTA + Monte Carlo Simulation	Combines dynamic fault trees with stochastic simulation for realistic modeling of system evolution	Handles complex interactions and time-varying failure probabilities in uncertain environments	[37]

reasonably foreseeable misuse. This section explores how FTA can be extended and adapted to support safety analysis within the SOTIF framework. It examines methodological strategies for incorporating SOTIF-specific risk factors into fault tree structures. By bridging the deductive modeling capabilities of FTA with the hazard categories defined in SOTIF, this integration enables a more comprehensive assessment of safety risks in autonomous vehicle systems. This section also highlights recent studies that demonstrate the feasibility of this combined approach.

2.4.1 Approaches for Integration

Integrating FTA into the SOTIF framework provides a structured approach to extending traditional risk assessment techniques to cover hazards related to intended functionality. While FTA traditionally evaluates functional safety by identifying failure paths and quantifying the likelihood of system malfunctions, its application within the SOTIF context enables a comprehensive assessment of potential risks. SOTIF focuses on hazards arising from the system's intended functionality, such as sensor limitations or environmental conditions [3]. By applying FTA to these scenarios, it becomes possible to model how functional insufficiencies and triggering conditions lead to hazardous outcomes, even in the absence of component failures.

Clause 7.3 of ISO 21448 states that inductive, deductive, or exploratory methodologies can be used to analyze potential functional insufficiencies and triggering conditions. The analysis can be qualitative, quantitative, or a combination of both. Moreover, quantitative targets can be defined at the element level and derived from acceptance criteria or validation targets at the vehicle level [44]. FTA is a deductive, top-down safety analysis technique that closely matches the analysis expectations. It can also be adapted to serve the goal of SOTIF by addressing hazards caused by performance limitations and reasonably foreseeable misuse.

FTA supports the SOTIF framework through both qualitative and quantitative methodologies [18, 34]. Qualitative FTA systematically identifies and illustrates logical relationships among functional insufficiencies, triggering conditions, and hazardous events. This approach effectively identifies root causes, allowing safety analysts to trace hazardous outcomes back to their initial contributing factors. In addition, qualitative FTA supports the identification of minimal cut sets, revealing the smallest combinations of conditions that cause the hazardous event. By highlighting these critical vulnerabilities, the qualitative approach supports risk prioritization and targeted mitigation strategies [109, 123]. Quantitative FTA extends the qualitative approach by incorporating numerical probabilities so that enabling precise probabilistic risk assessments [45, 122]. By assigning likelihood values to basic events, quantitative FTA facilitates the establishment of explicit safety performance targets at both component and system-element levels.

2.4.2 Use Cases and Examples

Several studies demonstrate the effective integration of FTA within the SOTIF framework, offering both methodological advancements and practical application. Table 2.8 presents

a comparative overview of representative studies that integrate FTA within the SOTIF framework, detailing their main focus areas, the role of FTA, addressed SOTIF aspects, and key contributions to safety assessment in autonomous vehicle systems. Schönemann et. al [118] present a structured methodology employing FTA to derive functional safety requirements for cooperative Automated Valet Parking (AVP) systems. Their approach is based on the ISO 26262 framework and extends to address key aspects of SOTIF. In particular, it considers performance limitations and triggering conditions that are not caused by system malfunctions. The authors analyze the AVP system using a SPA model. This model provides a conceptual framework for breaking down the system's functional behavior into sequential operational phases. This decomposition makes it easier to identify and evaluate safety-critical events at each stage of the vehicle's operation. A notable contribution of their study is the ability to model safety goal violations that may arise in the absence of hardware or software failures. For instance, in a narrow parking environment, the system may operate as intended but fail to detect or classify pedestrians due to sensor occlusion. This scenario represents a potential hazard caused by a functional insufficiency rather than a traditional malfunction. By incorporating these non-failure causes into the fault tree, the proposed methodology enables the derivation of safety requirements that address both malfunction-based and SOTIF-related hazards.

Liu et al. [83] propose a Tri-Safety Integration Analysis Strategy which unifies FuSa, SOTIF, and Cybersecurity within a single analytical framework. Their approach is centered on a collaborative tree structure that integrates FTA with Causal Tree Analysis (CTA) and Attack Tree Analysis (ATA). Their study decomposes the overall safety goals into contributing events related to functional failures, performance insufficiencies, and cybersecurity threats. In the context of SOTIF, this study applies FTA and CTA to systematically identify root causes of performance limitations and evaluate their potential to result in hazardous behavior under specific triggering conditions. The combination of FTA with ATA is used to model cybersecurity threats, tracing how malicious actions could result in safety goal violations. The proposed method improves system safety and reliability and obtains more comprehensive safety requirements.

Kaiser [65] provides a complementary perspective, which focuses on the practical integration of FuSa and SOTIF using model-based tools such as medini analyze. His methodology employs FTA to conduct causal analysis of hazardous events, extending traditional malfunction scenarios to include performance limitations and environmental factors. In this context, SOTIF-related hazards, such as degraded object detection in low-light conditions, are explicitly modeled as part of the fault tree. Moreover, Kaiser suggests that SOTIF analysis and simulation should be more closely integrated. Since simulation can reveal real-world triggering conditions that might not be detectable through analysis alone.

By incorporating such scenarios into the FTA, the methodology supports iterative validation and refinement of the SOTIF concept. The study demonstrates how triggering conditions, functional weakness, and misuse scenarios can be causally linked to hazardous events, highlighting the importance of FTA as a deductive tool for comprehensive SOTIF analysis.

Integrating SOTIF with FTA aligns closely with this guideline mentioned in ISO 21448. This integration offers a unified and systematic approach to hazard identification, supporting a comprehensive assessment of potential risks throughout the system. This approach effectively captures and analyzes all hazards resulting from functional limitations and system malfunctions so that enhancing the robustness of risk assessment. It also enables more accurate prioritizing of risks and facilitates the development of safety measures.

Table 2.8: Comparison of FTA-SOTIF Use Cases

Ref.	**Main Focus**	**Role of FTA**	**SOTIF Aspect Addressed**	**Key Contributions**
[65]	Practical integration using model-based tools; emphasis on SOTIF and simulations	Use for causal analysis including performance limitations and environmental factors	Focus on low-light object detection and misuse scenarios revealed by simulation	Iterative simulation-informed validation of SOTIF scenarios
[83]	Unifies analysis of FuSa, SOTIF, and Cybersecurity	Combine with CTA and ATA to model functional failures and cyber threats	Model performance insufficiencies and triggering conditions under security threats	Tri-safety integration enhances completeness of safety goals
[118]	Automated Valet Parking; SPA model-based FuSa and SOTIF	Use to derive functional and SOTIF-based safety requirements	Consider performance limitations and sensor occlusions	Captures non-fault hazards in narrow parking contexts

2.5 Lesson Learned, Challenges, and Open Issues

This section summarizes the key insights derived from integrating the FTA and SOTIF frameworks for autonomous vehicle safety. It also discusses current challenges and highlights potential directions for future research in the context of autonomous vehicle safety assurance.

2.5.1 Lessons Learned

A key insight emerging from this study is that FTA offers a structured methodology that aligns well with the objectives of the SOTIF framework. FTA is traditionally used within the ISO 26262 FuSa to evaluate hazards arising from system malfunctions or component faults. SOTIF, as defined by ISO 21448, extends the safety scope to include hazardous behaviors that may occur even when the system operates correctly [98]. These include sensor limitations, insufficient system performance in edge-case scenarios, or reasonably foreseeable misuse. The distinct focuses of FTA and SOTIF are not contradictory but complementary. Their integration allows for a more comprehensive safety assessment that addresses both fault-induced failures and risks associated with the intended functionality under uncertain operating conditions.

A key requirement in SOTIF is to identify triggering conditions and functional insufficiencies that could lead to hazardous outcomes [31]. This objective aligns well with the FTA's deductive structure, which traces causal pathways from basic events to top-level hazards. When applied in a SOTIF context, FTA provides the structured visualization of these causal relationships. It can represent performance limitations, environmental conditions, and user interactions as basic or intermediate events within a fault tree. This approach provides a visual and logical framework for understanding how various factors might interact to produce unsafe behavior, even in the absence of system malfunctions. In addition to its qualitative capabilities, FTA also enables quantitative reasoning. While SOTIF currently lacks standardized risk quantification methods [142], integrating FTA into the SOTIF analysis allows safety engineers to estimate the probability of specific hazardous scenarios. This supports comparative analysis of the related risk contributions associated with different insufficiencies and informs risk acceptance decisions based on severity and likelihood. As a result, FTA provides a valuable extension to the largely qualitative nature of current SOTIF assessments. Application studies, such as cooperative valet parking systems, battery management, and perception-based hazard detection, demonstrate the practical relevance of integrating FTA and SOTIF in real-world scenarios.

Recent advancements further extend the analytical capabilities of FTA. Techniques such as DFTA, probabilistic modeling with BN, and simulation-based approaches like Monte-Carlo analysis are integrated to address limitations in classical fault tree methods. These techniques support the modeling of time-dependent behaviors, accommodate uncertainty, and enable the evaluation of adaptive systems that evolve in response to environmental changes or learning algorithms.

While FTA provides a deductive and structured foundation, its binary and static assumptions limit its ability to fully represent uncertainty, multi-state variables, and in-

terdependencies among events. These constraints become particularly evident in SOTIF contexts, where hazards arise from functional insufficiencies and performance limitations. BN address these gaps by modeling probabilistic dependencies through conditional probability tables, allowing event likelihoods to be updated dynamically as new evidence becomes available. This capability enables BN extend FTA beyond static risk estimation toward a more adaptive and evidence-driven form of safety analysis.

The advantages of BN are especially relevant to autonomous vehicle safety. Their ability to support both predictive and diagnostic inference provides benefits for proactive risk assessment and post-incident investigation. In the context of SOTIF, BN can represent triggering conditions, functional insufficiencies, and failure modes in a single probabilistic framework, quantifying their combined impact on system-level hazards. As a results, BN is more than just alternatives to FTA. They are natural extensions that enhances their applicability in scenarios involving uncertainty and contextual variability.

2.5.2 Challenges

Despite these promising insights, there are several challenges remain. One key limitation is the absence of a standardized methodology for integrating FTA with the SOTIF framework. Current approaches are often developed for specific case studies or system architectures, which limits their scalability, reusability, and regulatory compliance across different autonomous vehicle platforms or development teams. Additionally, modern autonomous vehicle systems increasingly rely on AI-based perception and decision-making systems that exhibit stochastic and adaptive behavior. Traditional FTA rely on well-defined and deterministic event structures, which struggle to capture this complexity, limiting their applicability in modeling SOTIF-relevant hazards introduced by learning-based components.

Another limitation lies in the absence of quantitative safety thresholds in SOTIF, unlike ISO 26262, which provides numerical failure rate targets. This makes it challenging to apply probabilistic FTA in a consistent and traceable manner. Moreover, traditional FTA is not inherently scenario-based. While SOTIF emphasizes the importance of identifying unknown and context-sensitive scenarios, fault trees do not naturally capture the variability of real-world environments. Without support for dynamic scenario generation, critical edge cases may be overlooked.

2.5.3 Open Issues and Future Research Directions

To overcome these limitations, future research should focus on developing hybrid safety analysis frameworks that integrate FTA with probabilistic and causal modeling techniques, such as BN or structural causal models. This type of integration can capture both deterministic and probabilistic dependencies while also accommodating dynamic and uncertain behaviors. There is also a need to define quantitative safety metrics tailored to SOTIF-relevant hazards. These metrics would support more consistent risk assessment and strengthen regulatory alignment. Additionally, the advancement of automated scenario generation techniques would improve coverage of edge cases and enable more realistic validation of autonomous vehicle behavior under uncertain conditions. Finally, incorporating human factors modeling into fault and hazard modeling frameworks would support a more comprehensive evaluation of misuse-related risks in autonomous systems.

2.6 Conclusion

This survey has presented a comprehensive review of current literature on FTA and SOTIF in the context of autonomous vehicle safety. In particular, it has explored the feasibility of integrating FTA, a well-established deductive method for analyzing failure propagation, with the SOTIF framework, which focuses on addressing risks that may arise even when a system operates as intended. The paper begins by introducing the overall architecture of autonomous vehicle systems and highlighting key safety concerns. It then outlines the scope of SOTIF, as defined in ISO 21448, and compares it to ISO 26262 FuSa, highlighting the importance of assessing both malfunction-induced and functionality-related risks.

The survey further investigates how FTA can be adapted to model SOTIF-relevant factors such as functional insufficiencies and triggering conditions. Recent methodological advancements, including dynamic modeling and probabilistic reasoning, are reviewed for their potential to enhance FTA's applicability in SOTIF contexts. In addition, practical case studies demonstrate the feasibility of combining FTA with SOTIF frameworks in real-world autonomous vehicle scenarios. The paper also highlights several ongoing challenges. These include the lack of standardized integration methods, the absence of quantitative metrics for SOTIF, and the limited capability of current approaches to support scenario-based validation. Addressing these challenges is essential for developing adaptive safety assurance frameworks for autonomous vehicle systems. At the same time, these challenges give rise to meaningful opportunities for future research. Enhancing FTA through probabilistic modeling, formalizing scenario-based analysis techniques, and systematically

incorporating human factors into safety assessments can greatly strengthen the ability to capture real-world complexity. Collectively, these directions offer a promising foundation for the development of more comprehensive and context-aware safety frameworks aligned with the advanced autonomous vehicle technologies.

Chapter 3

Advancing Autonomous Vehicle Safety: A Combined Fault Tree Analysis and Bayesian Network Approach

This chapter is the outcome of the following publication: Lansu Dai, and Burak Kantarci, "Advancing Autonomous Vehicle Safety: A Combined Fault Tree Analysis and Bayesian Network Approach", IEEE International Conference on Engineering Reliable Autonomous Systems (ERAS), May 2025. (Accepted)

This chapter integrates FTA and BN to assess collision risk and establish Automotive Safety Integrity Level (ASIL) B failure rate targets for critical autonomous vehicle components. The FTA-BN integration combines the systematic decomposition of failure events provided by FTA with the probabilistic reasoning capabilities of BN, which allow for dynamic updates in failure probabilities, enhancing the adaptability of risk assessment. A fault tree is constructed based on autonomous vehicle subsystem architecture, with collision as the top event, and failure rates are assigned while ensuring the total remains within 100 FIT. Bayesian inference is applied to update posterior probabilities, and the results indicate that perception system failures (46.06 FIT) are the most significant contributor, particularly failures to detect existing objects (PF5) and misclassification (PF6). Mitigation strategies are proposed for sensors, perception, decision-making, and motion control to reduce the collision risk. The FTA-BN integration approach provides dynamic risk quantification, offering system designers refined failure rate targets to improve autonomous

vehicles safety.

3.1 Introduction

With the increased use of autonomous vehicles, it has become critical to ensure the safety of autonomous vehicles in complex and dynamic environments so that they can accurately perceive, predict, and respond to diverse scenarios to mitigate risks [73]. According to ISO 26262, risk is defined as the combination of the probability of harm occurring and the severity of its consequences [43]. In the context of autonomous vehicles, risk arises from system limitations, component failure, and external environmental conditions that may lead to undesirable outcomes, such as collisions. Mitigating risks are important to ensure the safe operation of autonomous vehicles. This highlights the importance of risk assessment frameworks to address the uncertainty and complexity of the real-world environment.

FTA is a widely used technique in safety analysis, particularly in complex systems [110]. It provides a structured approach to identify potential hazards and assessing the likelihood of failure. Autonomous vehicles rely on multiple interconnected systems, they cooperate to ensure the safe and reliable operation. FTA provides a systematic approach to analyze failures across these systems. However, traditional FTA has limitations in handling complex interactions and dependencies in systems [63]. It assumes that the system components fail independently and with static probabilities, making dynamic relationships difficult to model. This challenge becomes evident in autonomous vehicles, where subsystem interactions continuously evolve.

BN excel at modeling uncertain events and capturing dependencies between components, enhancing the overall accuracy and flexibility of risk assessment [100]. BN enable probabilistic inference and dynamic updates based on new evidence. IntegratingFTA with BN provides a comprehensive risk assessment framework, combining FTA's ability to decompose failure events systematically with BN's ability to model complex dependencies. While FTA-BN integration has been successfully applied to safety-critical domains, such as UAV [138] and radio altimeter systems [47], its application to autonomous vehicles risk assessment at the subsystem level remains underexplored. Bhavsar et al. [14] use traditional FTA to analyze autonomous vehicles systems operating in mixed traffic streams. However, existing analyses rely on static failure probabilities modeling the autonomous vehicles system as a whole. Therefore, state of the art remains the open issue of capturing the intricate interdependencies between its subsystems. This gap leaves a critical aspect of autonomous vehicles safety unexplored, that how failures in individual subsystems contribute to overall system risk and how these risks evolve over time.

This study addresses the limitations of traditional FTA by incorporating Bayesian inference to derive target failure rates for basic events, which provides valuable guidance for system designers in ensuring autonomous vehicles safety compliance. This chapter contributes to risk-based autonomous vehicles safety assessment by focusing on subsystem-level analysis and utilizing BN to dynamically update failure probabilities. The main contributions of this chapter can be summarized as follows:

1. We define target failure rates for autonomous vehicles subsystems that align with ASIL B under ISO 26262 FuSa standards [43], ensuring that autonomous vehicles system designers can use these failure rate thresholds as guidelines for safety compliance.

2. We decompose the autonomous vehicles system into subsystems based on its architecture, which includes sensors, perception, decision-making, and motion control systems. This structured approach allows for more targeted risk mitigation strategies and system improvements.

3. We identify the most significant contributors to collision risk and map critical failure pathways, offering insights for prioritizing safety enhancements and resource allocation in autonomous vehicles development.

Our quantitative analysis demonstrates that perception system failures are the primary contributors to autonomous vehicles collision risk, accounting for 46.06 FIT, almost half of the total 100 FIT failure rate. Specifically, failures in detecting existing objects (PF5), misclassification (PF6), and delayed response to obstacles (DMF2) are identified as critical factors. The rest of the chapter is organized as follows. Section 3.2 reviews the related work and highlights the novelty of this study. Section 3.3 details the methodologies, including FTA and BN integration. The experimental results and discussion are presented in Section 3.4. Finally, Section 3.5 concludes with future directions.

3.2 Related Work

FTA has been widely used to evaluate autonomous vehicles safety by identifying potential failure points and assessing associated risks. Bhavsar et al. [14] employ FTA to analyze vehicular and infrastructure components in mixed-traffic environments, identifying critical failure points in autonomous vehicles. While promising, their study can further be improved by incorporating dynamic failure probabilities and accounting for real-time updates

based on dynamic traffic conditions. Chen et al. [23] use FTA to examine control transitions in Level 2 and 3 autonomous vehicles, identifying key failure sources in operational design domains and human-machine interactions. Their study highlights the significant role of human factors in autonomous vehicles safety, particularly in takeover scenarios. Expanding their study to include broader system failures and incorporating in-depth quantitative methods for handling dependencies could further strengthen the findings. Li et al. [79] integrate FTA with FMEA to build a fault database. This structured method improves failure diagnosis in electric autonomous vehicles. The work could be further improved by incorporating real-time adaptability and capturing dependencies between events to provide a more comprehensive and dynamic risk assessment framework.

These studies demonstrate the effectiveness of FTA in autonomous vehicles risk assessment but also highlight its limitations, particularly its static nature and inability to model interdependencies dynamically. To address these challenges, integrating FTA with BN has been explored in other domains, such as radio altimeter systems [47] and unmanned aerial vehicles [138]. Gao et al. [47] used an FTA-BN approach to fault diagnosis in radio altimeter systems. They highlight the advantages of using BN to capture dependencies between system components and update failure probabilities in real-world scenarios. Similarly, Xiao et al. [138] demonstrate the effectiveness of BN in UAV safety analysis, by modeling statistical dependencies among UAV components and dynamically updating risk assessments based on real-time flight data.

Despite its success in these fields, FTA-BN integration remains underdeveloped in autonomous vehicles safety research. Most existing studies, including Bhavsar et al. [14] could benefit from incorporating Bayesian inference. This gap provides an opportunity to apply the FTA-BN framework to autonomous vehicles risk assessment, particularly in modeling subsystem interactions and dynamic risk probabilities. This study aims to bridge that gap, using FTA-BN integration to develop a more adaptive framework for improving autonomous vehicles safety.

3.3 Methodology

3.3.1 Fault Tree Analysis (FTA)

The fault tree is a top-down approach for modeling the pathways leading to a specific system failure or undesired event. It has two main types of nodes: events (basic, intermediate, or top events) and gates (AND/OR) to define logical relationships.

FTA supports both qualitative and quantitative analysis [110]. The qualitative FTA uses minimal cut sets to identify a system's vulnerabilities without requiring numerical data. The minimal cut sets represent the smallest combination of basic events that can cause the top event. Formally, a minimal cut set C_{min} is a set of basic events $E_1, E_2, \ldots, E_n$: $\phi(C_{min}) = 1$ and $\forall C' \subsetneq C_{min}$, $\phi(C') = 0$, where ϕ is the structure function representing the logical relationships in the fault tree.

Quantitative FTA calculates the failure probability of the undesired event occurring by propagating the failure probabilities of the basic events through gates. For an AND gate, the output probability is calculated as the product of the probabilities of all input events. For an OR gate, the output probability is computed as the complement of the product of the complements of the input probabilities. The failure probability of the top event is calculated by iteratively applying these formulas from the bottom to the top of the fault tree. Conducting both qualitative and quantitative FTA provides a comprehensive understanding of system risks and failure pathways, enabling informed risk management decisions.

3.3.2 Bayesian Network (BN)

Bayesian Networks have both forward and backward analysis. The forward analysis calculates the probability of occurrence of any node in the network based on the prior probability of the parent node and the conditional dependence of each node. This provides the prediction of outcomes given known input. The backward analysis focuses on the computation of the posterior probability of any given set of variables given evidence. This allows for reasoning and diagnostics based on known outcomes [67].

BN is widely used to model uncertainty, make inferences, and predict outcomes based on partial information. Therefore, they are useful for modeling uncertainty in autonomous vehicles, as dependencies between variables can be probabilistic and dynamic. BN can incorporate evidence, such as the failure rate of specific nodes, and then use Bayes' theorem to calculate the posterior failure probability. This makes them a valuable tool for risk assessment.

3.3.3 Integrating FTA with BN

Integrating BN with FTA enhances risk assessment by capturing probabilistic dependencies between system components. Any fault tree can be converted into a corresponding BN by creating a binary BN node for each event in the fault tree [113]. According to Xiao

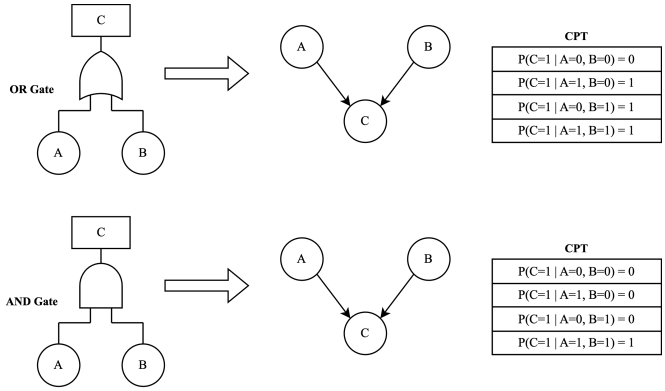

Figure 3.1: Transformation of AND and OR gates from Fault Tree to Bayesian Network

et al. [138], the conversion rule from fault tree to BN can be considered as two parts: graphical and numerical mapping.

Let $B = \{B_1, B_2, \ldots, B_n\}$ and $I = \{I_1, I_2, \ldots, I_m\}$ represent the set of basic events and intermediate events in FTA, respectively. Let T represent the top event in FTA. The graphical mapping to BN is:

$$B \xrightarrow{\mathcal{M}_G} Pa = \{Pa_1, Pa_2, \ldots, Pa_n\} \tag{3.1}$$

$$I \xrightarrow{\mathcal{M}_G} N = \{N_1, N_2, \ldots, N_m\} \tag{3.2}$$

$$T \xrightarrow{\mathcal{M}_G} C \tag{3.3}$$

where Pa is the set of parent nodes, N is the intermediate nodes, and C is the child node in BN. For the numerical mapping, let $P(B_i)$ represent the occurrence probability of a basic event B_i and G represent a logic gate (e.g., AND, OR) in the fault tree.

$$P(B_i) \xrightarrow{\mathcal{M}_N} \text{Prior}(Pa_i) \tag{3.4}$$

$$G(B_1, B_2, \ldots, B_k) \xrightarrow{\mathcal{M}_N} \text{CPT}(C|N_1, N_2, \ldots, N_k) \tag{3.5}$$

where $\text{Prior}(P_i)$ is the prior probability in BN and $\text{CPT}(C|N_1, N_2, \ldots, N_k)$ is the conditional probability table in the BN corresponding to the logic of gate G. The translation rule can be summarized as:

$$\text{FTA}(B, I, T, P(B), G) \xrightarrow{\mathcal{M}} \text{BN}(Pa, N, C, \text{Prior}(Pa), \text{CPT}) \tag{3.6}$$

where $\mathcal{M} = \mathcal{M}_G \cup \mathcal{M}_N$ represents the combined graphical and numerical mapping functions. Fig. 3.1 shows the transformation of the two-states AND and OR gates rule. These Conditional Probability Tables (CPTs) form the foundation for probabilistic reasoning in BN derived from fault trees, allowing us to calculate the likelihood of events under uncertainty and update probabilities when new evidence is introduced. Bayes' Theorem is used to incorporate evidence updates.

$$P(E_i \mid \text{Evidence}) = \frac{P(\text{Evidence} \mid E_i) \cdot P(E_i)}{P(\text{Evidence})} \tag{3.7}$$

where $P(E_i \mid \text{Evidence})$ is the updated probability of event E_i after considering the evidence, $P(\text{Evidence} \mid E_i)$ is the probability of observing the evidence, given the event E_i is true, and $P(E_i)$ is the occurrence probability before considering the evidence. $P(\text{Evidence}$ is the marginal probability of the evidence, which can be computed as:

$$P(\text{Evidence}) = \sum_j P(\text{Evidence} \mid E_j) \cdot P(E_j) \tag{3.8}$$

where E_j represents all possible events in the network.

3.3.4 Construction of Fault Tree

The fault tree for autonomous vehicles collision risk is structured based on the core subsystems of an autonomous vehicle, which includes sensors, perception system, decision-making system, motion control system, and external interactions [101, 103]. The subsystem architecture of autonomous vehicles is illustrated in Fig. 3.2.

Each subsystem contributes significantly to the overall functionality and safety of the vehicle. The sensors are responsible for collecting real-time environmental data, providing the foundation of perception and navigation. However, failures in this subsystem can affect the vehicle's ability to interpret its surroundings accurately. Sensor failures occur when camera (SF1), LiDAR (SF2), radar (SF3), GPS (SF4), or IMU (SF5) fails due to hardware malfunctions, environmental interference, or signal loss.

The perception system is responsible for interpreting the environment using sensor data. Failures in this subsystem affect the vehicle's ability to detect and respond to its surroundings. The basic events in the perception system are labeled PF1 through PF14. Data misalignment (PF1), coordinate frame errors (PF2), and algorithm fusion errors (PF3) may cause inconsistencies in sensor fusion. Furthermore, object recognition failures,

such as detecting non-existent objects (PF4), failure to detect existing objects (PF5), and misclassification (PF6), result in inaccurate scene interpretation. Low confidence score (PF7) and edge case limitation (PF8) present additional challenges for object recognition algorithms. Object tracking is responsible for continuously monitoring detected objects in the environment and predicting their future position over time. Failures in object tracking, including data association errors (PF9), drift in tracking output (PF10), and tracking loss (PF11), degrade the reliability of autonomous vehicles perception, leading to unsafe driving decisions. Map matching errors (PF12), coordinate transformation failures (PF13), and localization drift (PF14) in localization introduce navigation inconsistencies, affecting autonomous vehicles' positional accuracy.

The decision-making system processes perception outputs and determines appropriate navigation actions. Failures within this subsystem can result in incorrect or delayed driving decisions, increasing the likelihood of collisions. Incorrect path planning (DMF1), delayed response to obstacles (DMF2), and obstacle avoidance failure (DMF3), can directly lead to unsafe driving behavior. The motion control system ensures that planned vehicle actions are executed safely. Failures in acceleration control (MCF1), braking mechanisms (MCF2), or steering functionality (MCF3) affect vehicle stability. In addition to internal system failures, external interaction factors also contribute to autonomous vehicles safety risks. Adverse weather conditions (E1), degraded road conditions (E2), communication failure (E3), and cyberattacks (E4) pose significant challenges. These external factors may reduce sensor accuracy, interfere with decision-making processes, or compromise system security, increasing the likelihood of a collision. Based on the architecture of the autonomous vehicle, we constructed the fault tree with the top event as collision in autonomous vehicles as Fig. 3.3 and Fig. 3.4.

3.3.5 Risk-Based Safety Assessment Methodology

This chapter presents a risk-based safety assessment methodology for collision risk analysis in autonomous vehicles, focusing on defining target failure rates for critical components to align with FuSa standards, particularly ASIL B under ISO 26262. While the ASIL B failure rate target is originally defined for random hardware failures, we use it notionally as a reference for the overall system, including both hardware and software failures, to ensure a comprehensive risk assessment. The methodology systematically analyzes autonomous vehicles architecture by breaking it down into onboard sensors, perception, decision-making, and motion control systems to identify failure components and establish target failure rates. The root causes of collision as the top event are investigated and fault tree is developed based on literature reviews and expert opinions. In the qualitative analysis, minimal

Table 3.1: List of Basic Events with their Corresponding Posterior Failure Rate

Event Node	**Name of Event**	**Failure Rate (FIT)**
SF1	Camera Failure	6.36 ± 1.33
SF2	LiDAR Failure	5.67 ± 1.36
SF3	Radar Failure	6.51 ± 1.42
SF4	GPS Failure	6.87 ± 1.74
SF5	IMU Failure	6.46 ± 1.14
PF1	Data Misalignment	5.45 ± 0.516
PF2	Coordinate Frame Errors	4.82 ± 0.632
PF3	Algorithm Fusion Error	5.52 ± 0.806
PF4	Detecting Non-Existent Objects	6.42 ± 0.843
PF5	Failure to Detect Existing Objects	6.46 ± 0.917
PF6	Misclassification	6.56 ± 0.454
PF7	Low Confidence Scores	5.01 ± 0.49
PF8	Edge Case Limitations	4.27 ± 1.05
PF9	Data Association Errors	5.43 ± 0.905
PF10	Drift in Tracking Output	5.17 ± 0.617
PF11	Tracking Loss	5.01 ± 0.568
PF12	Map Matching Errors	5.19 ± 1.11
PF13	Coordinate Transformation Faults	5.44 ± 0.444
PF14	Localization Drift	5.46 ± 0.466
DMF1	Incorrect Path Planning	5.98 ± 0.787
DMF2	Delayed Response to Obstacle	6.56 ± 1.08
DMF3	Obstacle Avoidance Failure	6.31 ± 0.923
MCF1	Accelerator Control System Failure	5.58 ± 0.775
MCF2	Brake Control System Failure	5.62 ± 0.750
MCF3	Steering System Failure	4.96 ± 0.812
E1	Weather	4.48 ± 0.713
E2	Road Conditions	4.41 ± 0.607
E3	Communication Failure	4.76 ± 0.719
E4	Cyberattack	5.28 ± 0.626

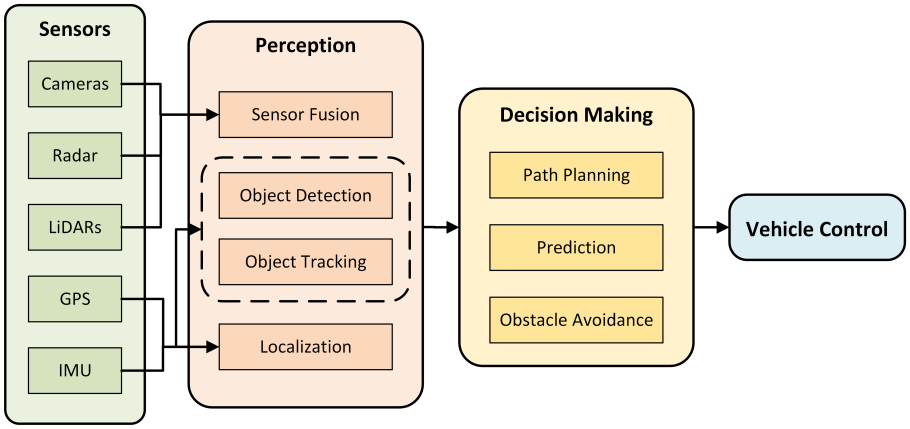

Figure 3.2: Subsystem Architecture of Autonomous Vehicles Systems

cut sets are generated using FTA to identify the critical failure combinations systematically. For the quantitative analysis, the fault tree is converted into a BN to perform the probabilistic risk assessment.

3.4 Experimental Results

3.4.1 Qualitative Analysis

The qualitative analysis of the fault tree structure provides insights into the system's minimal cut sets, which represent the simplest combinations of basic events that can cause system failure: collision. Analyzing these cut sets helps to identify which components are the most critical contributors to the risk associated with the system.

In this study, three order-2 minimal cut sets are identified: coordinate frame errors and data misalignment (PF1, PF2), low confidence scores and edge case limitation (PF7, PF8), and data association errors and drift in tracking outputs (PF9, PF10). These failure pairs indicate that a single failure alone is not sufficient to cause a system failure in these cases. However, the simultaneous occurrence of both failures in a set significantly increases collision risk. To mitigate the risks associated with these minimal cut sets, it needs targeted risk reduction strategies.

For PF1 and PF2, real-time sensor calibration should be implemented to continuously adjust sensor alignment and correct errors. Besides, sensor fusion consistency checks can

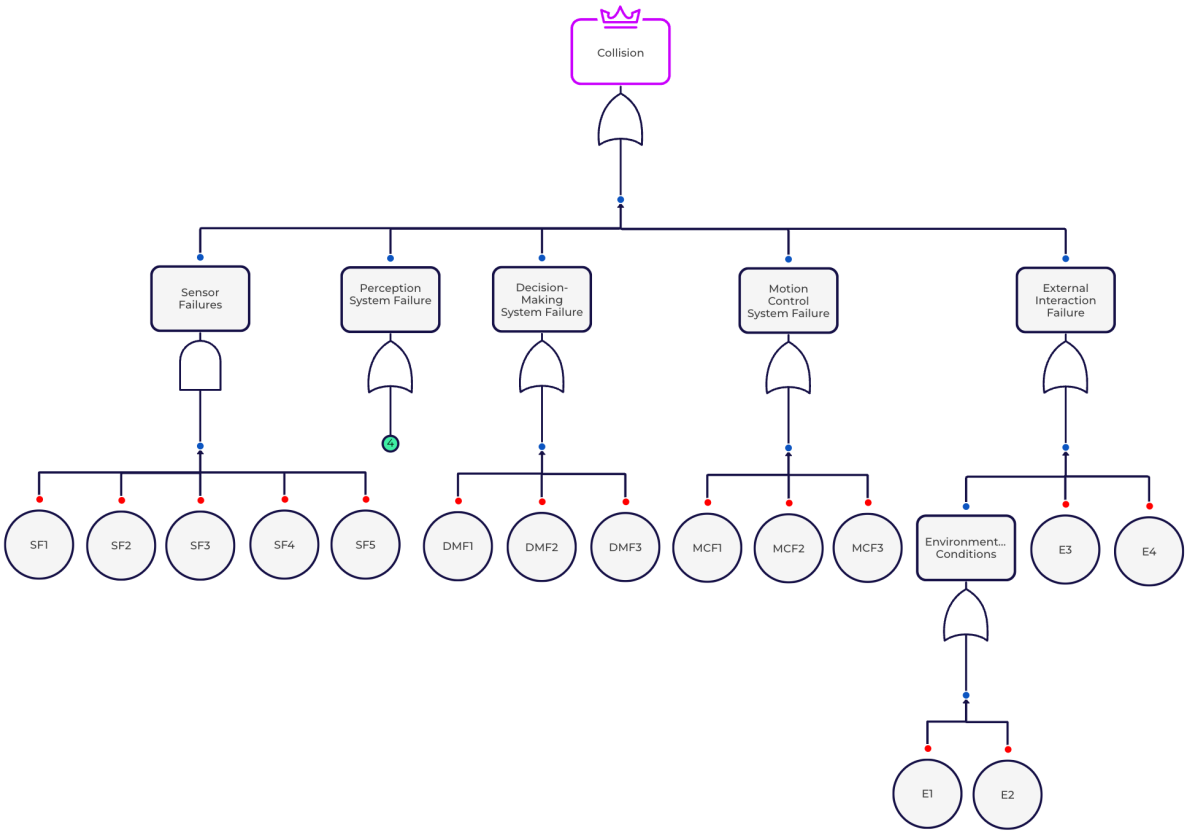

Figure 3.3: Fault Tree Diagram of Collision as Top Event

detect and mitigate misaligned data before it propagates. For PF7 and PF8, improvements in AI perception systems are necessary. Expanding training datasets with edge-case scenarios can enhance robustness. Adaptive AI models should be developed to handle low-confidence situations by triggering a fail-safe mode or a secondary decision validation process. PF9 and PF10 significantly impact the initialization of object tracking results. Incorporating appearance descriptors can help improve identification matching and reduce correspondence errors. Using probabilistic filters such as Kalman filters, can help reduce drift by estimating covariance and dynamically adjusting prediction confidence to reduce positional uncertainty. In addition, an AND gate connects multiple sensor failures (SF1-SF5). This structure indicates that individual sensor failures do not significantly influence collision risk unless multiple sensors fail simultaneously. This insight suggests that while individual sensor failures are less critical, maintaining overall sensor reliability remains essential to system robustness.

In addition to the minimal cut sets mentioned above, all other failures in the system are single points of failure, meaning that the failure of a single node directly impacts the probability of collision. These failure modes are particularly high-risk and require additional safety mechanisms to ensure system robustness. Mitigation strategies such as system redundancy (e.g. secondary braking mechanisms, alternative localization) to maintain backup functionality during failures, and AI-driven predictive maintenance to detect early component degradation through real-time operational data analysis. These mechanisms maintain operational safety and reduce catastrophic risk during critical failure.

3.4.2 Quantitative Analysis

To ensure compliance with ASIL B, we assume that the failure rate of a collision in autonomous vehicles systems does not exceed 100 FIT. FIT stands for failures in time which is a unit used to express the failure rate of a component or system. Following the fault tree structure, we randomly assign failure rates to basic events while ensuring the total failure rate remains on this predefined safety threshold. The fault tree is constructed and evaluated using Pathfinder, and failure probabilities for basic events are calculated using the formula: $P = 1 - e^{-\lambda t}$ where P denotes the failure probability, λ represents the failure rate, and t stands for the time. In our study, t is set to 10,000 operational hours since we are working on an autonomous system. Then these probabilities are mapped into a BN for posterior probability estimation.

Following the conversion rules from FTA to BN, we construct a Bayesian Network model using Python, as shown in Fig. 3.5. To validate the BN structure, we assume a

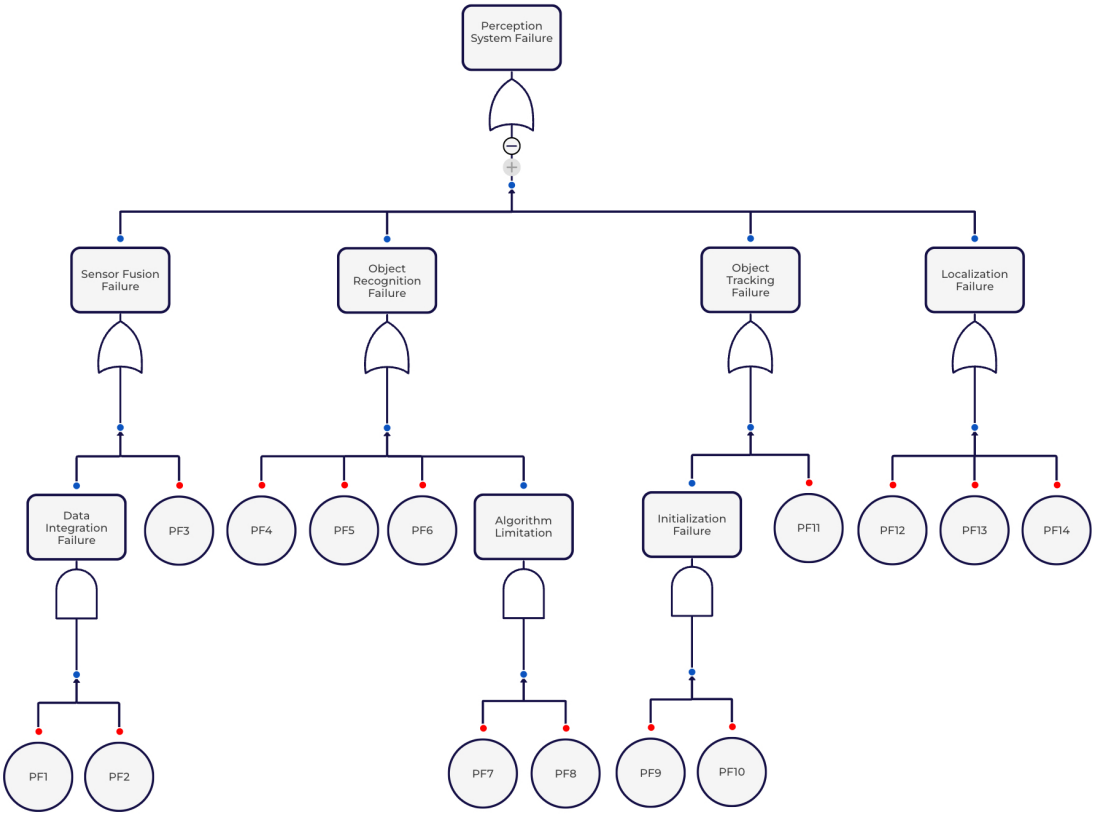

Figure 3.4: Fault Tree Diagram of Perception System in Collision as Top Event

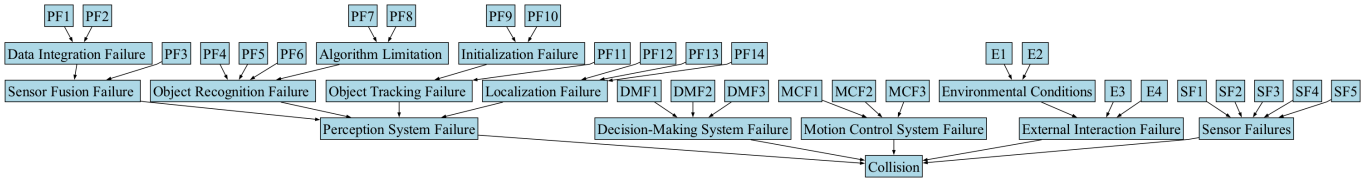

Figure 3.5: Integration of Fault Tree with Bayesian Network

collision failure probability of 0.001 and ensure the BN reproduces the same probability. The probability of collision calculated in both FTA and BN is identical at 0.001, confirming the consistency and accuracy of the two models. Table 3.1 and Fig. 3.6 present the failure rates of the basic events with 95% confidence levels. These provide a refined understanding of the system's vulnerabilities and highlight critical components contributing to the overall collision risk.

Among the analyzed basic events, several basic events are identified as high-risk contributors. PF5 (6.42 ± 0.843 FIT) and PF6 (6.56 ± 0.454 FIT) highlight object detection algorithm limitations in perception system, particularly under adverse conditions. The failure rate of DMF2 is 6.56 ± 1.08 FIT, indicating that latencies in decision-making processes substantially increase collision risk. Similarly, DMF3 demonstrated a failure rate of 6.31 ± 0.923 FIT, highlighting the need for enhanced decision-making systems to handle complex driving scenarios effectively. Beyond internal system failures, E4 emerges as a significant external failure event, highlighting the growing risk of cybersecurity threats to autonomous vehicles. Cyberattacks have the potential to interfere with perception and decision-making algorithms, compromise vehicle control systems, and cause incorrect actions, which raises the collision risk.

Some events show high variability in failure rates, such as PF8 (4.27 ± 1.05 FIT), indicating significant uncertainty in handling rare or extreme scenarios. This highlights the need for better AI training on diverse datasets to address edge-case performance issues. Moreover, the sensor subsystem shows particularly high variability in failure rates, which can be attributed to factors such as environmental influences, calibration issues, or gradual degradation of sensor components over time. Since the AND gate is used to model sensor dependencies, the overall probability of sensor subsystem failure remains relatively low. To mitigate these uncertainties and improve system robustness, it is essential to implement redundant sensor configurations, advanced sensor fusion methodologies, and real-time diagnostic mechanisms. These measures will enhance fault tolerance and ensure more reliable perception performance under diverse operating conditions.

The perception system (40.06 FIT) is the most significant contributor to collision risk, followed by decision-making (18.85 FIT), motion control (16.16 FIT), and external interaction subsystems (18.93 FIT). By enhancing object detection accuracy and expanding the training datasets with more diverse and representative scenarios, the system can better handle edge cases and adverse conditions, such as extreme weather or poor visibility. These enhancements will help mitigate failures such as PF5, and PF6, which are critical contributors to the overall collision risk. A well-trained perception system will reduce the likelihood of misidentification and failure to detect objects, especially in challenging environments.

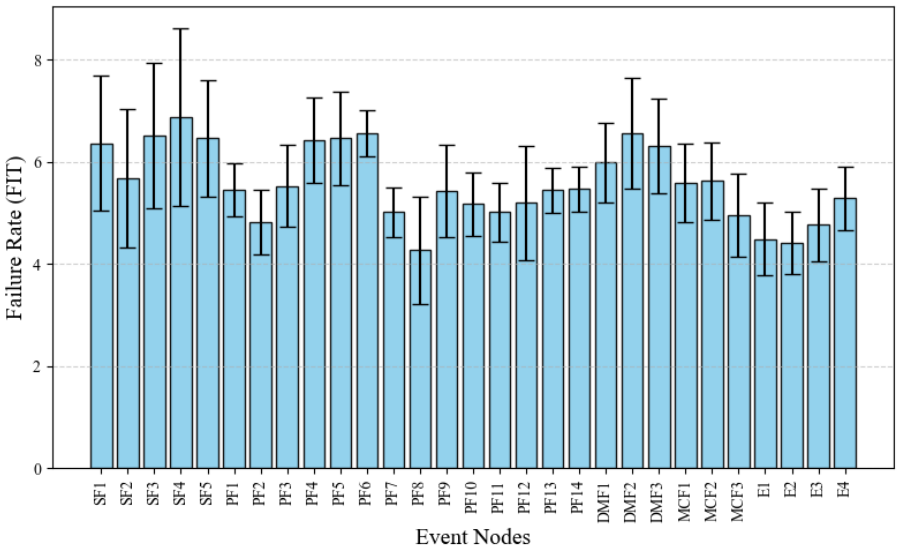

Figure 3.6: Failure Rates of Basic Events

The implementation of real-time validation mechanisms for cross-checking data from multiple sensors can greatly reduce sensor fusion failures. This approach is particularly effective in preventing failures PF1 by ensuring that inconsistencies in data alignment are detected and corrected. Moreover, introducing redundancy in perception systems helps mitigate single-sensor malfunctions, preventing cascading failures that could disrupt object tracking and motion planning. This approach is particularly valuable for mitigating failures such as PF9 and PF11, which can disrupt the tracking and continuity. In the decision-making subsystem, failures such as DMF3 can be mitigated by using predictive decision algorithms and fail-safe override mechanisms. These enhancements allow the system to anticipate dynamic obstacles and respond immediately to avoid collisions. For the motion control subsystem, which includes components like brakes and steering, redundancy is critical. Introducing backup control systems and deploying predictive maintenance can reduce risks associated with MCF2 and MCF3. The external interaction subsystem, which includes environmental factors like E2, can benefit from adaptive driving algorithms that dynamically adjust to varying road conditions, minimizing their impact on vehicle safety. By focusing on specific failure events using these mitigation techniques, the subsystems can operate more effectively, improving the safety and robustness of autonomous vehicles.

3.5 Conclusion

In this chapter, we have presented a risk-based safety assessment methodology for autonomous vehicles collision risk analysis by integrating FTA and BN. Traditional FTA provides a structured breakdown of failure events but assumes static failure probabilities and lacks the ability to model interdependencies. To address this limitation, we have applied Bayesian inference to refine failure rates and establish ASIL B target failure rates for critical autonomous vehicles components. We construct a fault tree based on autonomous vehicles subsystem architecture, defining collision as the top event. Failure rates are assigned while ensuring the collision remained within 100 FIT, which is the predefined safety threshold. By converting to BN, Bayesian inference is used to update posterior probabilities, repeated over 10 times and analyzed at a 95% confidence level. Results indicate that perception system failures (46.06 FIT) contribute most to collision risk, followed by decision-making (18.85 FIT), motion control (16.16 FIT), and external interactions (18.93 FIT). Delayed response to obstacles (DMF2), misclassification (PF6), and failure to detect existing objects (PF5) are identified as critical failure points. Mitigation strategies are proposed for each subsystem, focusing on improving sensor redundancy, enhancing perception algorithms, and optimizing decision-making processes. The integration of FTA and BN enables dynamic risk quantification, providing system designers with refined failure rate targets. The proposed methodology can scale from subsystem-level failures to system-level hazards. However, the inclusion of Bayesian inference introduces additional nodes and conditional dependencies, increasing computational complexity and requiring more detailed parameterization. While effective for design-time analysis, scalability becomes a concern as interdependencies grows. Future work will explore real-time Bayesian inference implementation and enhanced environmental modeling to improve autonomous vehicles safety assessment.

Chapter 4

A SOTIF-Oriented Framework for Safety Assessment in Autonomous Vehicles Using Integrated Fault Tree and Bayesian Network Analysis

This chapter is the outcome of the following publication: Lansu Dai, and Burak Kantarci, "A SOTIF-Oriented Framework for Safety Assessment in Autonomous Vehicles Using Integrated Fault Tree and Bayesian Network Analysis". (Under Submission)

Ensuring the safety of autonomous vehicles requires addressing not only hardware and software failures, but also functional insufficiencies arising from limitations in perception, decision-making, and sensor performance. While traditional FTA provides a systematic framework for identifying failure pathways, it does not account for the dynamic dependencies and uncertainties inherent in real-world driving scenarios. To overcome these limitations, this chapter proposes an integrated methodology that combines FTA with BN from the perspective of the SOTIF. The proposed framework enables both qualitative and quantitative safety analysis, capturing logical structural as well as probabilistic dependencies among triggering conditions, functional insufficiencies, and failure modes. A case study on object detection failure in autonomous vehicles is conducted to demonstrate the applicability of the methodology. The results indicate that environmental factors weather and occlusion are major contributors to object detection failures, which contributes to 45.76% and 58.72%, respectively. The results indicate that environmental factors such as weather and occlusion are major contributors to object detection failures. Additionally, posterior

probability analysis highlights the influence of specific variables and provides actionable insights for prioritizing mitigation strategies. This study contributes a novel framework that enhances safety assessment capabilities in SOTIF contexts and provides system designers with practical guidance to improve the reliability and robustness of autonomous vehicle systems.

4.1 Introduction

As autonomous vehicle technology continues to advance, ensuring safety in complex and dynamic environments has become a critical challenge. Traditional safety frameworks, such as ISO 26262 [43], primarily address hazards resulting from hardware and software malfunctions. However, they do not adequately capture risks that arise when the system functions correctly but fails due to performance limitations or unforeseen interactions with the environment. To address these gaps, the ISO 21448 standard, commonly known as the SOTIF [44], extends the safety scope to include hazards caused by functional insufficiencies and reasonably foreseeable misuse. It aims to ensure system safety by addressing non-fault-based hazards, particularly in perception and decision-making functions that are fundamental to autonomous vehicle performance.

FTA is a widely used deductive technique for modeling hazardous events and identifying their root causes through logical structure. However, FTA assumes static and independent failure events, limited in modeling dynamic, probabilistic, and interdependent relationships among system components [63]. BN provide a probabilistic graphical framework that captures causal dependencies and supports reasoning under uncertainty, enabling dynamic updates and inference over complex failure scenarios. While integrating FTA with BN has been successfully applied in varies safety-critical domains, such as unmanned aerial vehicles [138] and flare systems [64], its application within the SOTIF framework for autonomous vehicles remains largely unexplored.

In this chapter, we propose an integrated methodology that combines FTA with BN to support both qualitative and quantitative safety analysis from a SOTIF perspective. This integration preserves the logical clarity and deductive structure of FTA while leveraging the probabilistic reasoning capabilities of BN. The resulting framework enables the modeling of multi-state variables, interdependent relationships, and dynamic updates among three critical safety dimensions emphasized by SOTIF: triggering conditions, functional insufficiencies, and failure modes.

In this paper, we propose an integrated methodology that combines FTA with BN to support both qualitative and quantitative safety analysis from a SOTIF perspective. This

integration preserves the logical clarity and deductive structure of FTA while leveraging the probabilistic reasoning capabilities of BN. The resulting framework enables the modeling of multi-state variables, interdependent relationships, and dynamic updates among three critical safety dimensions emphasized by SOTIF: triggering conditions, functional insufficiencies, and failure modes. The main contributions of this paper are as follows:

- We present a novel methodology that integrates FTA and BN within the SOTIF framework to enable both qualitative structure-based analysis and quantitative probabilistic reasoning. The approach effectively models complex and interdependent safety scenarios with multi-state variables and uncertain causal relationships.
- We apply the proposed methodology to a case study on object detection failure in autonomous vehicles, identifying significant risk factors and critical failure pathways.
- We provide mitigation strategies based on the analysis outcomes to assist system designers in enhancing the safety and robustness of autonomous vehicle systems.

To validate the proposed methodology, we conduct a case study focusing on object detection failure, which is one of the most important safety functions in autonomous vehicles' perception systems. The case study demonstrates how the integrated FTA-BN model can identify key contributors to detection failures and map the corresponding failure pathways. The results show that adverse weather and occlusion are the important contributors to the detection failure. These insights help to prioritize safety improvements and provide actionable guidance for system designers and safety engineers.

The remainder of the paper is organized as follows: Section 4.2 reviews related work on Fault Tree Analysis, Bayesian Networks, and their applications in autonomous vehicle safety. Section 4.3 introduces the proposed methodology for integrating FTA and BN within the SOTIF framework. Section 4.4 presents the proposed approach through a case study on object detection failure in autonomous vehicles, and Section 4.5 present the qualitative and quantitative analysis results. Finally, section 4.6 concludes the paper and highlights the direction for future research.

4.2 Related Work

This section reviews related work on the safety assessment of autonomous vehicle systems, with a focus on SOTIF, Fault Tree Analysis, and Bayesian Networks. We discuss the differences between ISO 26262 and ISO 21448, the strengths and limitations of FTA, and

how BN and their integration with FTA improve probabilistic reasoning and risk assessment under uncertainty. We also identify specific research gaps in each area to motivate the proposed framework.

4.2.1 Safety of the Intended Functionality

The ISO 21448 standard defines SOTIF [44], which addresses hazards that arise not from system malfunctions, but from the system operating as intended under certain insufficiently considered conditions. This framework is particularly significant in automated and autonomous vehicle systems, where safety-critical decisions are increasingly dependent on perception systems and AI-based decision-making algorithms. Unlike traditional safety standards like ISO 26262 FuSa [43], which focus on hazards caused by hardware or software malfunctions, SOTIF extends the scope of safety analysis to include scenarios where the system operates as intended but still results in unsafe outcomes [98]. It focuses on hazards resulting from functional insufficiencies and reasonably foreseeable misuse.

Functional insufficiencies refer to situations in which the system behaves as designed but produces unsafe outcomes due to performance limitations [41]. SOTIF categorizes into two primary types of functional insufficiencies. The first involves hazardous behavior that emerges when functional limitations are exposed by specific triggering conditions, such as poor lighting, adverse weather, occluded objects, or unusual traffic scenarios. The second is the functional insufficiency leading to the inability to prevent the reasonably foreseeable indirect misuse, which includes drivers' behavior deviating from the intended use but remains predictable. For example, over-reliance on driver assistance features [119].

SOTIF serves as a complementary framework to ISO 26262, rather than a replacement [44, 104]. While ISO 26262 focuses on preventing unintended behavior due to faults, ISO 21448 emphasizes safety assurance in the absence of faults. A comparative overview is provided in Table 4.1. For instance, in an electronic braking system, ISO 26262 would ensure that the system transitions into a safe state (e.g., fallback braking) when a wheel speed sensor fails. ISO 26262 ensures fail-safe operation due to hardware or software malfunctions through diagnostics, redundancy, and systematic design verification. ISO 21448 encompasses several critical aspects, including the performance limitation of sensors, inaccuracies in AI-based perception and decision-making systems, the influence of unpredictable environmental conditions, and unintended drivers' behaviors. For example, even if the camera is working properly, it may fail to detect lane markings on a snow-covered road [54]. This can result in unexpected lateral movement or failure to provide steering assistance. ISO 21448 ensures that functional insufficiencies caused by environmental conditions are identified, tested, and mitigated during design and validation processes.

While SOTIF provides a conceptual framework for identifying and categorizing non-fault-based hazards, existing studies focuses on specific scenarios and scenario generation methods rather than comprehensive and generalizable modeling approaches [16]. There is a lack of comprehensive modeling that captures the relationships between triggering conditions, functional insufficiencies, and hazards. This gap highlights the need for methodologies that can systematically represent these relationships and quantify their combined impact on safety.

Table 4.1: Comparison of ISO 26262 Functional Safety and ISO 21448 Safety of Intended Functionality

Standard	**Focus on**	**Scope**	**Applicability**
ISO 26262	Fault-based hazards (hardware or software malfunctions)	Address system failures, including hardware failures and systematic design errors	Safety-critical system
ISO 21448	Non-fault-based hazards (functional insufficiencies, reasonably foreseeable misuse)	Address performance limitations such as sensor degradation, perception and decision-making limitation, actuator imprecision, and human misuse	ADAS and AI-based functions

4.2.2 Fault Tree Analysis

Fault Tree Analysis is a deductive, top-down safety analysis method for systematically identifying the causal factors leading to a specific undesired event, known as the top event, and systematically tracing its potential causes. This is achieved by modeling the causal relationships between events using Boolean logic gates, such as AND and OR. The fault tree consists of three types of events: basic events, intermediate events, and the top event [110]. Basic events represent individual component-level failures, while intermediate events correspond to subsystem or functional-level failures that result from combinations of basic events. The top event occurs when specific combinations of basic and intermediate events satisfy the logical conditions defined by the fault tree. The hierarchical representation allows for a clear and structured visualization of how combinations of lower-level failures interact and propagate through the system to produce a hazardous outcome.

FTA consists of both qualitative and quantitative analysis [9]. The qualitative FTA aims to understand the structural vulnerability of a system without involving the probabilistic information. The key objective of qualitative analysis is to identify minimal cut

sets. An minimal cut set is the smallest set of basic events that can cause the top event [42]. A cut set is defined as any set of basic events whose simultaneous occurrence leads to the occurrence of the top event. Formally, let T denote the top event and $\mathcal{B} = \{B_1, B_2, \ldots, B_n\}$ denote the set of basic events. A cut set $C \subseteq \mathcal{B}$ satisfies:

$$\bigwedge_{B_i \in C} B_i \implies T$$

where $\bigwedge$ denotes the logical AND operator. A cut set is minimal if no proper subset of it is also a cut set. Formally, C is a minimal cut set if:

$$\forall C' \subset C, \quad \bigwedge_{B_i \in C'} B_i \implies T$$

This minimality condition ensures that every event in the set is necessary for the top event to occur. This analysis provides insight into the logical interdependencies within the system and supports the development of risk mitigation strategies during the design and verification phases. Quantitative FTA extends the qualitative analysis by incorporating numerical failure data to estimate the probability of occurrence of the top event [35]. Probabilities are assigned to basic events based on historical failure data and expert judgment [110]. The overall system failure probability can be evaluated using Boolean algebra and probabilistic computation. This enables a numerical risk assessment, which supports compliance with safety standards and guides resource allocation for safety-critical functions.

FTA is widely used across safety-critical domains, including automotive [23, 51], chemical [58, 92], and nuclear power [53]. In the context of automotive systems, FTA is explicitly recommended by the ISO 26262 standard as part of the functional safety assessment process [43]. As a result, numerous applications of FTA are explored in the domain of autonomous vehicle systems. For example, Chen et al. [23] apply FTA to analyze failure during takeover scenarios in Level 2 and 3 automated driving. Their study introduces a taxonomy based on the ODD boundaries and human information processing stages, helping identify failure components during control transitions between the automated system and the driver. Sreeraj et al. [125] conduct a functional safety assessment of the battery management system in autonomous electric vehicles, combining Hazard Analysis and Risk Assessment (HARA) with quantitative FTA to ensure the system's compliance with ISO 26262 and to mitigate risks associated with battery thermal and voltage failures. Moreover, Bhavsar et al. [14] apply FTA to assess the risks of autonomous vehicles in mixed traffic environments by modeling failures in vehicular and infrastructure components. This analysis highlights the most critical combination of events that could lead to autonomous vehicle failures.

FTA can be further extended to align with ISO 21448, also known as the SOTIF. Since SOTIF focuses on hazards that arise from functional insufficiencies and reasonably foreseeable misuse, rather than system malfunctions, its integration requires a broader scope of analysis. Valerij et al. [118] apply FTA to analyze failure scenarios in automated valet parking systems. Their work emphasizes hardware failures and hazards resulting from sensor limitations and environmental uncertainties, which are relevant to SOTIF considerations. The direction is further supported by Kaiser [65], who explores the practical integration of ISO 26262 and ISO 21448 using model-based tools such as medini analyze. His approach suggests that traditional FTA can be extended with environmental conditions and performance limitations to better capture SOTIF-relevant hazards.

Traditional FTA remains an effective tool for systematically analyzing system-level failures. However, it has several significant limitations when applied to the safety assessment of autonomous vehicles [32]. One notable limitation is its static representation [46]. FTA assumes a fixed architecture, making it difficult to model the dynamic behaviors and environmental interactions that are inherent in autonomous driving scenarios. Furthermore, FTA typically assumes statistical independence among failure events, limiting its ability to capture complex interdependencies between components [63]. In practice, autonomous vehicle systems consist of tightly coupled interacting subsystems, including sensors, perception, decision-making, and motion control. These interdependencies are critical to the emergence and propagation of failures. To overcome these limitations, FTA is frequently combined with complementary analytical techniques, such as probabilistic graphical models [48, 138] or simulation-based validation methods [37, 128], to enhance its applicability in complex and adaptive systems.

4.2.3 Bayesian Network

Bayesian networks are probabilistic graphical models that use a directed acyclic graph to represent the conditional dependencies between variables [126]. There are three types of nodes in BN: parent nodes, intermediate nodes, and child nodes. Each node in the network corresponds to a random variable [100]. Edges represent the relationship between nodes. Nodes with no incoming edges are known as parent nodes, and they are associated with prior probability distributions since their states are not dependent on any other variables. Nodes with one or more incoming edges are referred to as child nodes, and they are associated with CPTs, which specify the likelihood of their states given the states of their parent nodes. The intermediate nodes serve as both child and parent nodes within the network also require CPTs.

BN provides a structured approach for modeling uncertainty, interdependencies, and causal relationships, making them highly suitable for safety analysis in autonomous vehicle systems [34, 67]. Autonomous vehicles rely on the coordinated operation of multiple interconnected subsystems, such as perception, decision-making, and motion control [101, 103]. BN effectively captures the probabilistic dependencies among these components, offering insight into how failures or degraded performance propagate and impact overall system behavior. Furthermore, BN supports both forward and backward analysis [20]. Forward analysis, also known as predictive inference, estimates the likelihood of downstream events based on known conditions. Backward analysis focuses on computing the posterior probabilities of variables given observed evidence. It enables the inference of the most probable underlying causes of an observed system failure, which is especially valuable for fault diagnosis and system monitoring in safety-critical applications.

The integration of FTA and BN provides a more comprehensive risk assessment framework by systematically decomposing failure events and modeling complex dependencies. This hybrid approach combines the strengths of FTA's hierarchical and deductive structure with the probabilistic reasoning and inference capabilities of BN. Xiao et al. [138] apply the integration of FTA and BN to evaluate public safety risks associated with UAV. Their analysis identifies the primary contributing risk factors and offers insights on developing effective regulations to improve public safety in UAV operations. Sakar et al. [113] apply a similar methodology to the analysis of grounding accidents in maritime systems. By mapping a fault tree into a Bayesian Network, they enable both predictive and diagnostic inference under uncertain conditions, allowing for the identification of critical contributing factors and the dynamic updating of risk levels based on observed evidence. Their work demonstrates the practical benefits of the integration of FTA and BN in enhancing the flexibility and real-time applicability of risk assessment in safety-critical domains. Despite its success in these fields, FTA-BN integration is still underdeveloped in autonomous vehicles safety research. Our previous work [34] presents an integrated FTA-BN framework for autonomous vehicles safety assessment in alignment with the ISO 26262 standard. The proposed approach enables dynamic, subsystem-level risk quantification, supporting more informed safety engineering decisions. Since the integration approach demonstrates the effectiveness under the ISO 26262 framework, it can be further extended to align with ISO 21448, which addresses safety concerns arising from functional insufficiencies and reasonably foreseeable misuse in the absence of system faults.

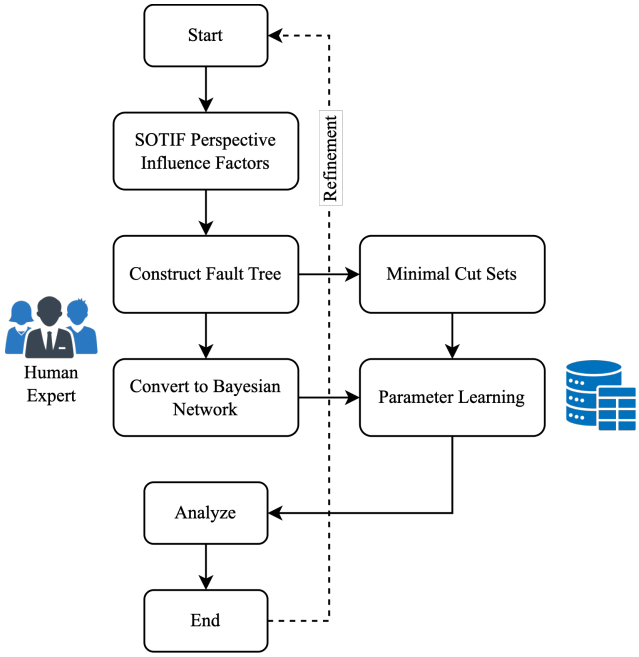

Figure 4.1: The Flowchart of the Proposed Methodology

4.3 Methodology

Fig. 4.1 presents the flowchart of the proposed methodology, which integrates FTA with BN to support safety assessment from the SOTIF perspective. The process begins by identifying SOTIF-relevant influence factors and constructing the fault tree with guidance from domain experts. Minimal cut sets are used to identify critical failure combinations of basic events that may lead to system-level hazards. The fault tree is then systematically transformed into a Bayesian Network for probabilistic modeling. Parameter learning incorporates prior data or expert judgment, supporting both forward and backward inference. The framework also supports iterative refinement based on the results of the analysis, allowing for continuous improvement in the safety assessment process.

4.3.1 SOTIF Perspective Influence Factors

ISO 21448 [44] provides a non-exhaustive list of scenario factors that represent various contextual and operational conditions influencing the intended functionality of a system. This list serves as the starting point of the safety analysis in the absence of fault. SOTIF

categorizes these factors into two types: functional insufficiencies and triggering conditions.

Functional insufficiencies refer to inherent performance limitations of the system. Examples include poor generalization in AI-based perception algorithms and insufficient response to edge-case scenarios. Triggering conditions include external environmental or operational variables, such as adverse weather, occlusion, lighting variations, and complex road layouts, that can expose these limitations.

The identification of relevant influence factors is selected with guidance from domain experts, ensuring alignment with the system's ODD and its intended functionality. These factors are then incorporated into the fault tree structure to represent contributing causes of hazardous behavior. For instance, if an autonomous vehicle's camera system has limited detection capabilities in snowy weather, and the road markings are obscured, the combination of sensor limitation (functional insufficiency) and snow (triggering condition) represents a SOTIF-relevant scenario factor. This process enables a structured qualitative assessment of the causal relationships leading to unsafe behavior.

4.3.2 Fault Tree Construction

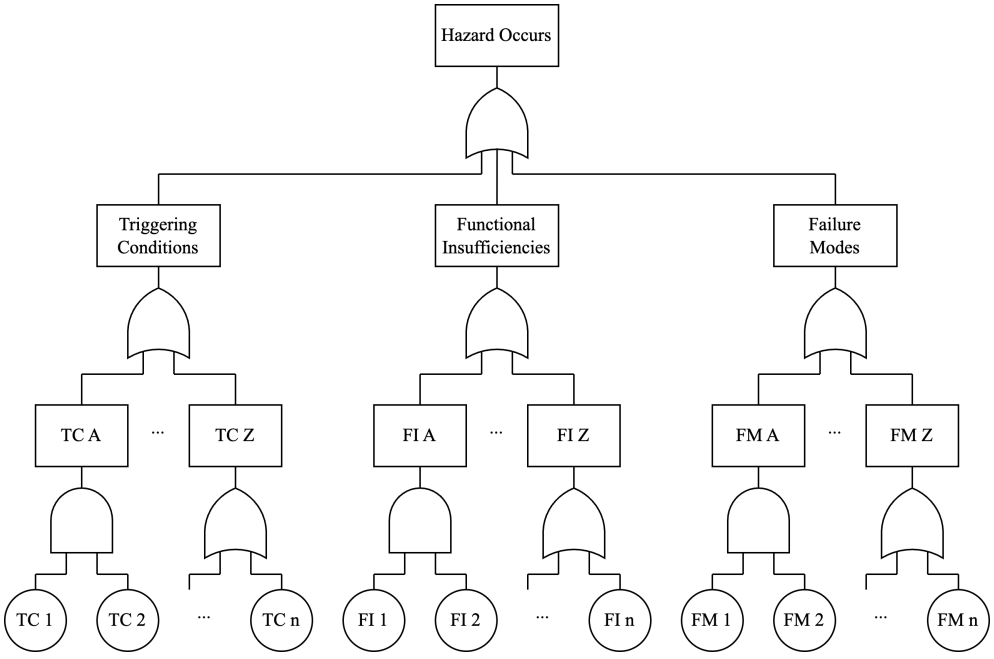

Figure 4.2: The Conceptual Structure of the Fault Tree for SOTIF-Related Hazard Analysis

Once the functional insufficiencies and triggering conditions relevant to the SOTIF perspective are identified, the next step is to systematically determine how the factors contribute to hazardous behavior. In the proposed methodology, these SOTIF-related influence factors are incorporated into a fault tree structure alongside traditional failure modes, enabling a comprehensive representation that addresses both ISO 26262 and ISO 21448 concerns. Fig. 4.2 shows the conceptual structure of the fault tree, where the hazard occurs at the top level and three major branches: triggering conditions, functional insufficiencies, and failure modes. Each branch decomposes further into intermediate and basic events, capturing the interactions among them that may lead to a hazard.

The construction of the fault tree is performed with active involvement from domain experts to ensure its relevance and correctness. Their insights help define the logical structure and relationships between contributing factors, enhancing the fault tree's applicability to the specific ODD and intended functionality of the autonomous vehicle systems.

A key outcome of the qualitative FTA is the identification of minimal cut sets, which represent the smallest combinations of basic events that can independently lead to the top-level hazards [110]. Let T denote the top event, and let $\{E_1, E_2, \ldots, E_n\}$ represent the set of basic events in the system. Each minimal cut set $C_i \subseteq \{E_1, E_2, \ldots, E_n\}$ consists of basic events whose simultaneous occurrence is sufficient to cause the top event. The top event can be expressed as the logical disjunction of all such minimal cut sets, where each cut set is represented by the conjunction of its constituent basic events. Mathematically, this relationship can be formulated as:

$$T = \bigvee_{i=1}^{m} \left(\bigwedge_{j \in C_i} E_j \right) \tag{4.1}$$

where m is the total number of minimal cut sets, $\bigwedge$ and $\bigvee$ represent logical AND and OR gate, respectively. Minimal cut sets analysis provides insight into the structural vulnerabilities of the system and supports risk mitigation efforts by highlighting the most critical contributors to the hazardous event.

Quantitative FTA extends the qualitative FTA approach by using probabilistic information to estimate the likelihood of system-level hazards [78]. It assigns failure probabilities to basic events based on expert judgment or historical data. These probabilities are propagated through the fault tree using Boolean logic gates, such as AND and OR, to calculate the overall probability of the top event. This analysis provides a numerical estimate of system-level risk. In our methodology, the computed failure probability of the top event is used to validate the accuracy and consistency of the corresponding BN model.

4.3.3 Bayesian Network Conversion

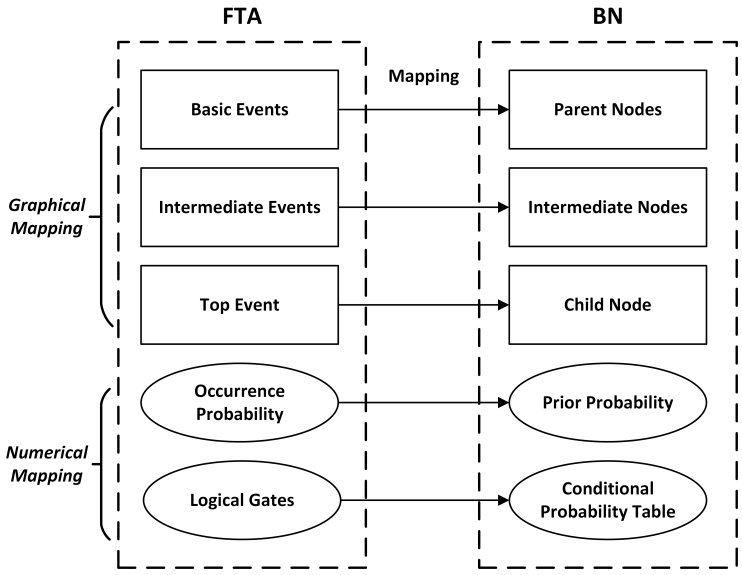

Figure 4.3: The Graphical and Numerical Mapping from FTA to BN

Any fault tree can be systematically converted into a Bayesian Network by applying the transformation rules [113]. Fig. 4.3 shows the general conversion rule from fault trees to Bayesian Networks. Fig. 4.4 demonstrates how FTA logic gates (AND and OR gates) are converted into their equivalent CPTs in BN. This transformation preserves the causal structure of the original fault tree while enabling probabilistic reasoning.

Following the structural conversion, the topology of the BN is established based on the fault tree but introduces several modeling advantages. Traditional FTA is limited to binary logic representations and assumes a static structure with independent basic events. In contrast, BN supports both discrete and continuous variables, allowing nodes to capture multi-state behavior, which is especially relevant in safety-critical systems such as autonomous vehicles. Moreover, BN explicitly models probabilistic dependencies between variables, allowing the representation of both direct and indirect causal relationships that are not fully expressed by traditional fault trees.

One significant limitation of traditional FTA is its assumption of statistical independence among basic events [22]. This restricts its ability to capture common cause failures that the scenarios where multiple components fail due to a shared root factor. Such failures are prevalent in real-world systems and are critical to address in comprehensive safety

assessments. BN offers a flexible framework that explicitly models dependencies among events [80]. By introducing a shared parent node to represent the common cause, BN can capture how a single factor simultaneously influences multiple variables. This allows for a more realistic representation of interdependencies, supporting both forward and backward inference. This allows for updating beliefs based on observed evidence, making the model responsive to real-time data. The joint probability distribution over all variables in the BN can be computed. Assuming a set of parameters $X_1, X_2, \ldots, X_n$ in the network, the joint probability distribution can be calculated as follows:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i)) \tag{4.2}$$

Once the BN structure is defined, expert knowledge is employed to initialize the CPTs associated with each node. These initial probabilities may incorporate historical reliability data or domain-specific assumptions. To enhance the accuracy and reliability of the model, these prior values can be refined through parameter learning techniques, such as Maximum Likelihood Estimation (MLE) or Bayesian Estimation, depending on the availability and quality of the observational data [61].

In MLE, the goal is to determine the parameter set θ that maximizes the likelihood of the observed dataset $\mathcal{D}$. This process assumes no prior information about the parameters and is expressed mathematically as:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} P(\mathcal{D} \mid \theta) \tag{4.3}$$

where $\hat{\theta}_{\text{MLE}}$ represents the optimal parameter values that best explain the observed data.

In contrast, Bayesian Estimation incorporates prior beliefs about the parameters and updates them using observed data. To incorporate evidence and perform probabilistic updates, Bayes' Theorem is applied. For a specific variable X_i, the posterior probability given observed evidence E is calculated as:

$$P(X_i \mid \text{E}) = \frac{P(\text{E} \mid X_i) \cdot P(X_i)}{P(\text{E})} \tag{4.4}$$

where $P(X_i \mid \text{E})$ denotes the posterior probability of variable X_i after considering the evidence E, $P(\text{E} \mid X_i)$ is the likelihood of observing the evidence assuming X_i is true, and $P(X_i)$ is the prior probability of X_i. The marginal probability of evidence $P(\text{E})$, used for normalization, is given by:

$$P(\text{E}) = \sum_{j} P(\text{E} \mid X_j) \cdot P(X_j) \tag{4.5}$$

where X_j iterates over all possible values of the variable X_i in the network. These parameter learning approaches allow the BN to incorporate both expert judgment and empirical data, resulting in a model that is both structurally interpretable and probabilistically robust.

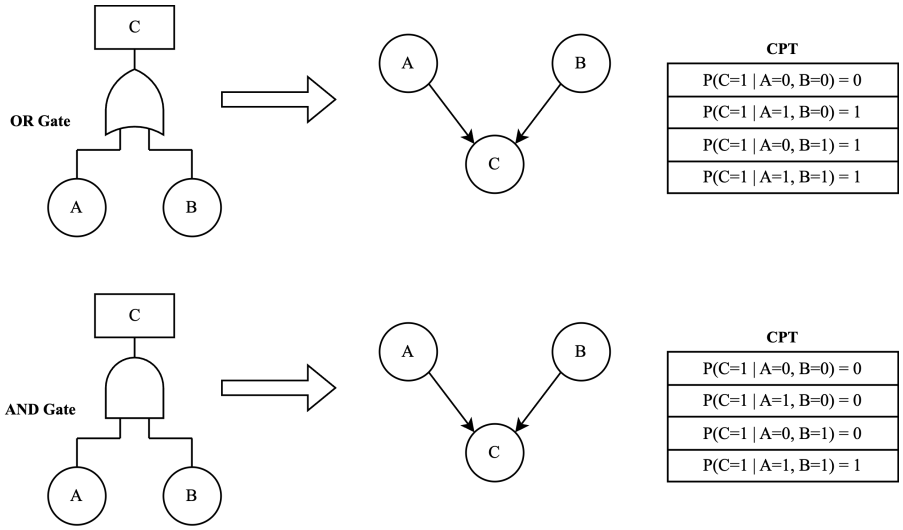

Figure 4.4: Transformation of AND and OR gates from FTA to CPTs in BN [34]

4.3.4 Refinement

The refinement step supports the iterative modification of the model structure and input factors, especially in response to newly available data or updated knowledge. In the context of SOTIF, new operational insights or test outcomes may reveal additional influence factors, system behaviors, or previously unconsidered hazards. These findings may require updates to the list of SOTIF-related influence factors and failure modes. As a result, the initial fault tree and the corresponding BN must be updated to reflect the new causal relationships. This continuous refinement process ensures that the model remains consistent with real-world observations and adapts to the evolving system characteristics.

4.4 Case Study on Object Detection in Autonomous Vehicles

In this section, we apply the proposed methodology to the analysis of object detection failures in autonomous vehicles, highlighting its practical implementation and effectiveness.

4.4.1 Experimental Setup

To demonstrate the applicability of the proposed methodology, a case study is conducted, focusing on the analysis of object detection failures in autonomous vehicles. Following the structured methodology, a fault tree is constructed to represent the relevant failure pathways, including SOTIF-related influence factors and failure modes. To support the development and refinement of the Bayesian Network, we use the PointPillars model [77] for 3D object detection, executed on an NVIDIA RTX 4080 GPU. This configuration provides efficient processing for large-scale inference tasks and makes it easier to extract probabilistic dependencies from the model's behavior.

The experimental evaluation is conducted using the KITTI dataset [49], a widely used benchmark in autonomous driving research. KITTI dataset provides synchronized data from multiple sensors, including one Velodyne HDL-64E LiDAR scanner and two Point Grey Flea 2 FL2-14S3C-C cameras, each with a resolution of 1.4 megapixels. These sensors enable the modeling of realistic perception scenarios and analyze the dependencies between failure events in object detection. For variables that cannot be obtained directly from the dataset, relevant values and dependencies are estimated based on existing literature, historical data, or expert judgment.

4.4.2 SOTIF Perspective Influence Factors

According to ISO 21448 (SOTIF), the identification of potential functional insufficiencies and triggering conditions must be based on a systematic analysis of the system's ODD boundaries. The ODD defines the set of environmental and operational conditions under which the system is expected to function safely.

In this study, the 3D object detection model is designed to operate in daytime, clear weather in urban and suburban driving scenarios with moderate traffic conditions. The system utilizes input from both LiDAR and cameras, and is capable of detecting objects such as cars, pedestrians, and cyclists within a range of 70 meters [143].

Based on the defined ODD, two categories of SOTIF-relevant influence factors are identified:

- **Functional Insufficiencies** refer to inherent performance limitations of the perception model. These include the model's inability to generalize to edge-case scenarios, such as uncommon object types and rare driving situations, and sensor fusion limitations arising from time synchronization error and sensor fusion algorithm limitations.
- **Triggering Conditions** represent external factors that may activate or expose these insufficiencies. In this study, the key triggering conditions include adverse weather, such as fog, rainy, or sunny conditions, and varying levels of occlusion, ranging from partial obstruction by nearby objects to complete visual blockage. These conditions can significantly degrade detection performance, even when the system is operating as intended.

These SOTIF perspective influence factors serve as the foundation for constructing the fault tree that systematically captures the causal pathways to object detection failures, as discussed in the following subsection.

4.4.3 Fault Tree Construction

Building upon the previously identified SOTIF influence factors, the fault tree is constructed to represent the causal relationships leading to object detection failure in autonomous vehicles. The top event of the fault tree is defined as *Object Detection Failure*. Following the methodology, it is decomposed into intermediate events corresponding to the categories of triggering conditions, functional insufficiencies, and failure modes.

Triggering conditions are modeled as independent environmental factors, including *Weather* and *Occlusion*, which can degrade detection performance. Functional insufficiencies, such as *Edge Case Limitations* and *Sensor Fusion Limitations*, are represented as internal system vulnerabilities. To comprehensively capture the causal pathways to unsafe behavior, the constructed fault tree also incorporates specific failure modes, including *Detecting Non-existent Objects*, *Failure to Detect Objects*, and sensor failures such as *Camera Errors* and *LiDAR Errors*. In particular, the sensor-related failures are connected using an AND gate. Since the simultaneous failure of the camera and LiDAR will lead to complete sensor failure.

We use Pathfinder from reasonX Labs to model the fault tree. Fig. 4.5 shows the complete fault tree structure, capturing how variables propagate to the top-level hazard. Additionally, Table 4.2 summarizes the basic events used in the fault tree analysis.

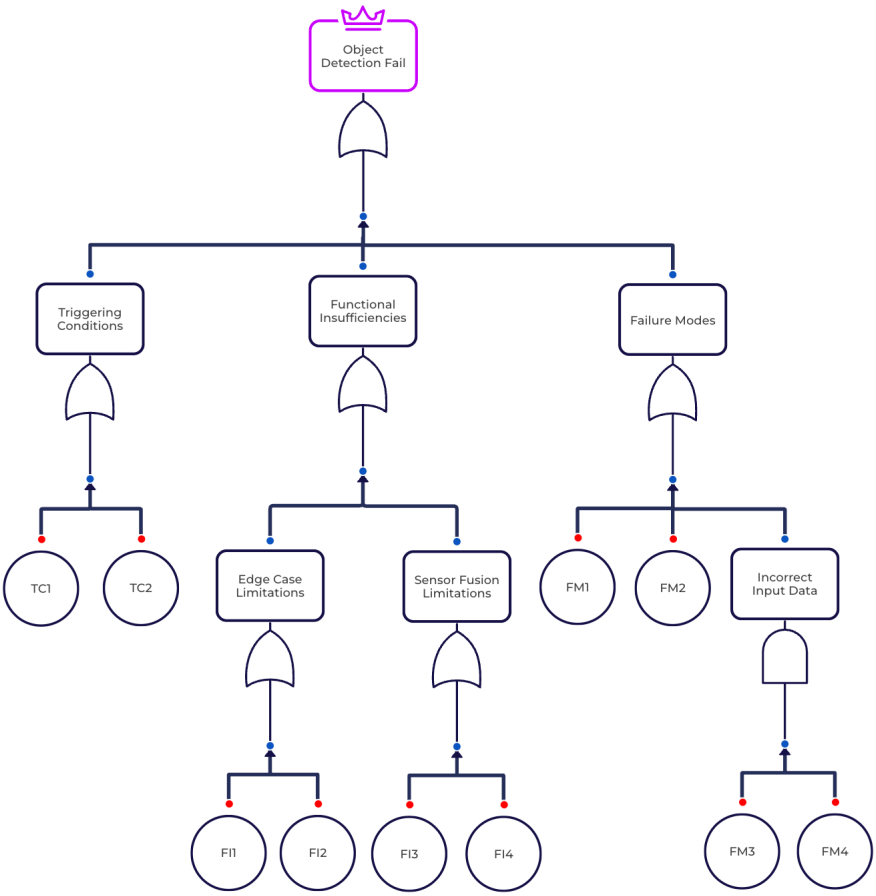

Figure 4.5: Fault Tree Diagram of Object Detection Failure as Top Event

Table 4.2: List of Event Nodes with Their Corresponding Posterior Probabilities

Event Node	**Name of Event**	**Posterior Probability (%)**	**Event Node**	**Name of Event**	**Posterior Probability (%)**
TC1	Weather	45.76 ± 1.62	FI4	Sensor Fusion Algorithm Limitations	23.89 ± 2.26
TC2	Occlusion	58.72 ± 1.90	FM1	Detect Non-Existent Objects	7.91 ± 1.39
FI1	Unseen Object Types	13.30 ± 1.81	FM2	Failure to Detect Objects	17.30 ± 0.89
FI2	Rare Driving Scenarios	18.04 ± 1.81	FM3	Camera Errors	5.60 ± 0.71
FI3	Time Synchronization Error	5.69 ± 0.99	FM4	LiDAR Errors	10.51 ± 1.28

4.4.4 Bayesian Network Conversion

Following the transformation rules outlined in Section 4.3.3, the fault tree of object detection failure is systematically converted into a BN. Each logic gate in the fault tree is mapped to a corresponding set of CPTs, preserving the structural logic of the original model while enabling probabilistic reasoning. The BN construction and analysis are implemented using the Python library pyAgrum, which supports efficient creation, handling, and computation of Bayesian Networks.

In this case study, the BN not only reflects the topology of the fault tree but also incorporates additional causal dependencies derived from expert knowledge. While traditional fault trees assume independent basic events, the BN is extended to capture shared influencing actors that affect multiple failure pathways. Specifically, domain experts identified that adverse weather conditions and occlusion significantly contribute to the failure to detect objects. These dependencies are explicitly modeled in the BN by introducing directed edges from the *Weather* and *Occlusion* nodes to the *Failure to Detect Objects* node. This enhancement allows the BN to more accurately reflect the influence of environmental factors on object detection failure, which are not fully captured by traditional fault trees.

Additionally, the fault tree is inherently limited to binary logic. In contrast, the BN

allows multi-state representations of key variables. In this study, the *Weather* node is modeled with four states: sunny (0), cloudy (1), rainy (2), and foggy (3). *Occlusion* is represented with four levels: fully visible (0), partly occluded (1), largely occluded (2), and unknown (3). This multi-state modeling capability allows for a more detailed and realistic safety analysis.

After the BN structure forms, parameter learning is performed using the corresponding data instance to estimate the CPTs. These learned probability distributions combined with expert assumptions enable the BN to serve as an effective tool for analysis and estimation. The refined BN is shown in Fig. 4.6, it provides a more expressive and flexible framework for safety analysis, supporting multi-state variable modeling, forward and backward inference, and real-time probabilistic updates.

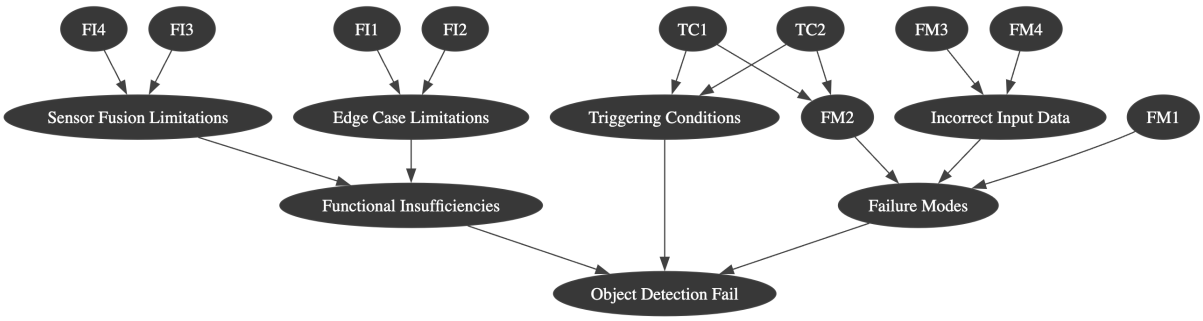

Figure 4.6: Refined Bayesian Network of Object Detection Failure

4.5 Results

This section presents the results obtained through the application of the proposed methodology. The analysis is divided into two parts: qualitative analysis and quantitative analysis. The qualitative analysis focuses on identifying critical failure pathways and understanding the causal relationships among SOTIF-relevant factors. The quantitative analysis involves computing failure probabilities using the constructed BN, enabling a probabilistic assessment of object detection failure in autonomous vehicles.

4.5.1 Qualitative Analysis

The qualitative analysis focuses on understanding the structure of the fault tree of the object detection failure case study. This analysis aims to identify minimal cut sets, which are the combinations of basic events that can independently or jointly lead to the top-level event, which in this case is object detection failure. The minimal cut sets provide insights into the critical failure paths and prioritize risk mitigation strategies.

In the constructed fault tree, a single order-2 minimal cut set is identified: FM3 (Camera Error) and FM4 (LiDAR Error). These events are connected by an AND gate, indicating that both sensor failures must occur simultaneously to cause object detection failure, while the failure of only one does not critically impair the system. This structure highlights that redundancy between camera and LiDAR inputs helps prevent single-sensor failures from escalating into system-level hazards. This also implies that the object detection remains resilient to the failure of one sensor under normal conditions. However, if both sensors fail due to hardware faults, calibration drift, or environmental interference, the system loses its perceptual capabilities. This emphasizes the importance of cross-sensor health monitoring, robust sensor fusion algorithms, and fail-safe mechanisms capable of detecting and responding to simultaneous sensor degradations.

In contrast, all remaining basic events in the fault tree are identified as single points of failure, meaning that the occurrence of any one of these events alone is sufficient to cause the top-level hazard. Environmental factors such as adverse weather and occlusion can directly impair sensor performance, resulting in degraded object detection. To mitigate these risks, object detection models should be trained with augmented datasets that reflect diverse environmental conditions. Moreover, incorporating sensor modalities less affected by visibility (e.g., radar) and employing context-aware filtering can enhance robustness in degraded environments.

Another critical limitation is the inability to recognize unseen object classes or respond effectively to rare driving scenarios. These gaps often stem from insufficient representation in training data. To address the limitations, the training dataset should be expanded using synthetic data generation and simulation environments to cover rare conditions. Furthermore, integrating online and continuous learning mechanisms allows the system to adapt to new scenarios over time.

Temporal misalignment between sensor data and limitations in fusion algorithms can compromise the accuracy of object localization and classification. These issues can be addressed by adopting precise time-stamping mechanisms, enforcing synchronization constraints, and developing fusion strategies that incorporate uncertainty modeling and temporal filtering. These failure modes need targeted countermeasures, including adaptive

perception algorithms that adjust to environmental variability and robust post-processing mechanisms to reduce false detections. Detecting non-existent objects and failing to detect existing objects represent false positives and false negatives, respectively. Both are critical failure types that can directly lead to unsafe behaviors such as sudden braking or failure to avoid obstacles.

Each single-point failure reflects a specific vulnerability that must be systematically addressed. The fault tree structure provides a clear and formal framework for identifying and analyzing these vulnerabilities. To enhance the safety and reliability of object detection in autonomous driving, designers must prioritize targeted mitigation strategies, including architectural redundancy, robust machine learning techniques, and comprehensive scenario validation. The identification of these critical paths through fault tree analysis provides a structured foundation for systematically addressing object detection hazards within the context of autonomous driving.

4.5.2 Quantitative Analysis

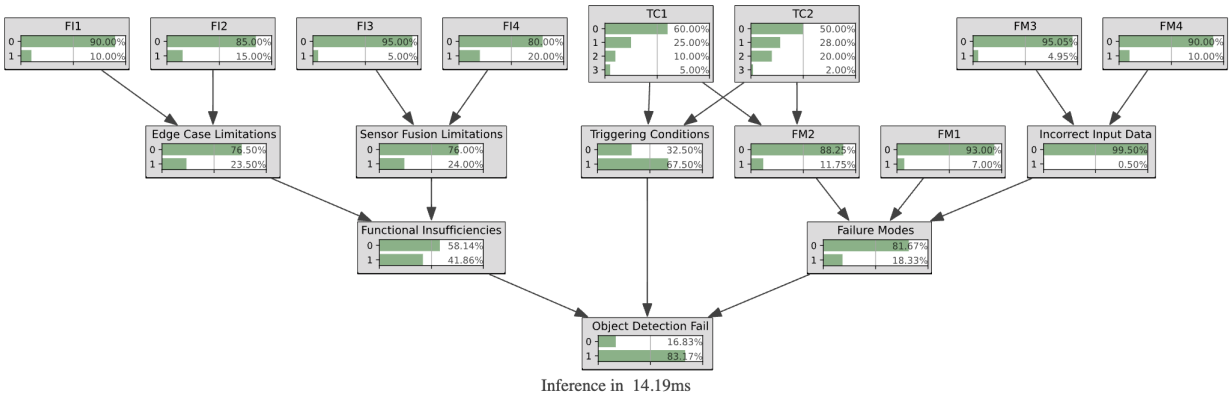

Figure 4.7: Bayesian Network for Object Detection Failure in Autonomous Vehicles

The BN structure with associated prior probabilities for each node produced by the pyAgrum library is shown in Fig. 4.7. These prior probabilities represent the initial beliefs about the likelihood of each contributing factor. They are derived from a combination of experimental results based on the PointPillars model in the KITTI dataset, relevant literature [29, 136], and expert judgment. The structure captures the causal relationships and conditional dependencies among environmental conditions, functional insufficiencies,

and system-level failure modes that may lead to object detection failure in autonomous vehicles.

To evaluate the influence of each contributing factor in the context of a failure scenario, evidence is introduced into the BN by conditioning the child node *Object Detection Fail* to true. The resulting posterior probabilities with 95% confidence level for relevant parent nodes are represented in Fig. 4.8 and Table 4.2. These updated probabilities are computed through probabilistic inference and parameter learning techniques, allowing us to quantify the impact of each factor on the observed system failure.

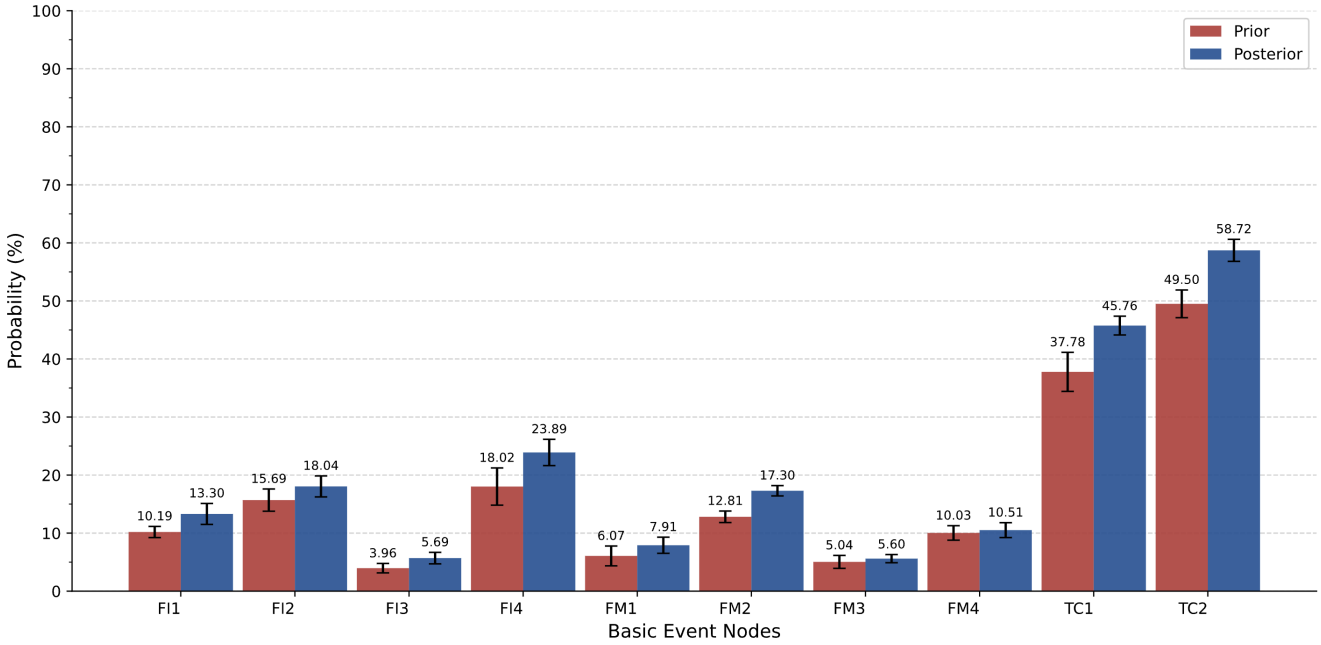

Figure 4.8: Prior and Posterior Probabilities of All Basic Event Nodes at 95% Confidence Level

Several nodes have significant changes in their posterior probabilities, indicating a strong contribution to object detection failure. In particular, TC1 (Weather) and TC2 (Occlusion) show significant increases in probability. TC2 rises from 49.50% (±2.39%) to 58.72% (±1.90%), a change of +9.22%. TC1 increases from 37.78% (±3.36%) to 45.76% (±1.62%), a change of +7.98%. These two nodes represent triggering conditions which, in line with SOTIF, do not indicate a malfunction but rather external environmental contexts that may impair perception performance. The significant increase in the posterior values suggests that environmental conditions play an important role in object detection failures.

This emphasizes the importance of scenario-based testing under varied weather and visibility conditions during system validation. As a mitigation strategy, it is recommended to enhance robustness through the inclusion of adverse weather scenarios in training datasets, deploy weather-adaptive models, and integrate redundancy across sensors with differing sensitivities to environmental variability.

FI4 (Sensor Fusion Algorithm Limitation) also shows a pronounced increase, rises from 18.02% (±2.26%) to 23.89% (±3.21%). This node represents a functional insufficiency. The insufficiency is a fundamental concept in SOTIF, where the perception system may fail to correctly interpret multi-sensor input despite functioning as intended. The increase in posterior probability implies that limitations in the fusion algorithm can significantly contribute to unsafe outcomes. Algorithmic improvements such as uncertainty-aware sensor fusion, dynamic confidence weighting, and fail-over mechanisms should be considered as mitigation strategies.

Furthermore, at the failure mode level, FM1 (Detect Non-existent Object) and FM2 (Failure to Detect Existing Object) both show increased posterior probabilities. FM2 increases from 12.81% (±0.99%) to 17.30% (±0.89%), which corresponds to false negatives, is particularly critical from the SOTIF perspective. It represents a situation where an object is present in the environment, but the system fails to detect it due to performance limitations. This failure can be particularly dangerous because it may lead to unsafe decisions by the planning or control modules without any warning or fault signals. To mitigate false negatives, supplementing training data with edge cases, applying ensemble detection models, and implementing temporal consistency checks across frames can be applied to enhance robustness and reduce missed detections. FM1 rises from 6.07% (±1.39%) to 7.91% (±1.71%), which is related to false positives. While false positives can lead to unnecessary actions or degraded driving comfort, it is generally considered less critical than false negatives in terms of safety impact. However, in complex scenarios, an accumulation of false positives may still degrade system reliability and trust. Techniques such as post-processing filter based on semantic context and scene understanding can help reduce false positive detections.

Overall, the findings emphasize the need to address both functional insufficiencies and triggering conditions in safety analysis. These aspects are foundational to SOTIF and are not fully covered by conventional fault-based approaches. The application of BN in this context enables a structured and probabilistically grounded method for quantifying the impact of influence factors. Targeted mitigation strategies include expanding ODD-aligned weather and occlusion testing, improving sensor fusion robustness and reducing false negatives through targeted data augmentation and temporal consistency checks. By identifying high-impact nodes and recommending targeted mitigation strategies, this anal-

ysis supports the development of more resilient and trustworthy perception systems in autonomous vehicles.

4.6 Conclusion

In this chapter, we have proposed an integrated methodology that combines Fault Tree Analysis with Bayesian Networks to assess safety from the perspective of the Safety of the Intended Functionality in autonomous vehicles. While traditional FTA provides a structured qualitative framework for hazard identification, it lacks the capability to model interdependencies and dynamic causal relationships among variables. To overcome these limitations, we incorporate BN to enable probabilistic reasoning, support multi-state variables, and capture conditional dependencies across different safety dimensions.

Our methodology extends FTA by aligning the analysis with the three key perspectives outlined in ISO 21448 and ISO 26262: triggering conditions, functional insufficiencies, and failure modes. FTA is used to construct the hierarchical structure of causal factors, while BN allows for quantitative evaluation through the computation of posterior probabilities. This framework enables both predictive analysis and diagnostic inference of failure pathways.

We validated the proposed methodology through a case study on object detection failure in autonomous vehicles. The results demonstrate that the environmental conditions, such as adverse weather and occlusion are the most significant contributors to detection failures, which contributes to 45.76% and 58.72%, respectively. These findings underscore the importance of accounting for performance limitations even in the absence of hardware or software faults. Moreover, we identified mitigation strategies, including the inclusion of adverse weather scenarios in training datasets, the use of ensemble detection models, and temporal consistency checks to support system designers in enhancing detection robustness.

One limitation of the current study lies in the initialization of prior beliefs in the BN. Some of the prior probability estimates used to construct the model are based on expert judgments, which may vary depending on the expert. The subjectivity introduces uncertainty and bias into the model. However, BN provides mechanisms to incorporate different expert opinions by introducing auxiliary nodes into the parent set of nodes of interest, allowing for sensitivity analysis and model refinement.

Another limitation relates to the scalability and complexity of the proposed approach. The introduction of multi-state variables and explicit interdependencies in the BN substantially increases the size of the CPTs and the computational cost of inference. While this

enhances the expressiveness and realism of the analysis, it also creates challenges in scaling the methodology to larger system models or broader operational domains. Addressing this trade-off between model accuracy and computational efficiency remains an important direction for future work.

In conclusion, this study provides a structured and extensible methodology for SOTIF-aligned safety analysis, offering both qualitative and quantitative analysis. It supports proactive risk assessment and the identification of critical failure contributors, resulting in more informed safety design and validation in the development of autonomous vehicle systems. Future work can be extended by incorporating empirical data obtained from real-world scenarios or in-vehicle testing to refine the conditional probability tables, thereby increasing the model's reliability. Furthermore, the use of Dynamic Bayesian Networks is a promising direction to capture temporal dependencies and evolving system behaviors, which are critical in dynamic driving environments.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Ensuring the safety of autonomous vehicles in complex and unpredictable environments remains a critical and ongoing challenge. Hazards can occur even when the system operates according to its intended design. While the ISO 26262 standard provides a robust framework for addressing hazards that originate from hardware and software malfunctions, it does not account for performance limitations or functional insufficiencies. These aspects are central to ISO 21448, the SOTIF standard, which can be viewed as a natural extension of ISO 26262. ISO 21448 suggests FTA as a potential tool for safety assessment, making it relevant to the SOTIF framework. FTA is effective for conducting hierarchical, fault-based safety analysis, but has limited ability to represent uncertainty, capture interdependencies between events, and address hazards that depend on specific operational contexts.

To address these limitations, this thesis introduces an integrated safety analysis framework that combines FTA with BN. This integration is designed to enhance the capability of traditional fault-based safety methods in context of autonomous vehicles, and in alignment with the SOTIF framework. With the increasing complexity of autonomous vehicle systems and emergence of hazards not caused by hardware or software malfunctions, there is pressing need to go beyond ISO 26262 and align with SOTIF framework outlined to ISO 21448.

This thesis provides a comprehensive literature review of and its relevance and limitations when applied to performance-related and context-dependent risks addresses by SOTIF. The review underscores the lack of existing methods that can simultaneously represent hierarchical system failures and capture probabilistic dependencies and uncertainty.

To address this gap, the thesis introduces a novel framework that integrates FTA with BN. This integration preserves the deductive structure of FTA while leveraging the probabilistic reasoning capabilities of BN to model causal relationships and uncertainties. The framework is designed to support both qualitative and quantitative analysis and is compatible with the principles outlined in ISO 21448.

The proposed methodology is validated using two representative case studies. In the first case study, which focuses on collision in autonomous vehicles, a fault tree is constructed based on the subsystems of autonomous vehicles. Failure rates are assigned to basic events while ensuring the overall system failure rate did not exceed 100 FIT. Bayesian inference is used to update posterior probabilities based on model structure. The results show that perception system failure, with total 46.06 FIT, is the most significant contributors to collision risk. The most critical events identifies are failure to detect existing object and object misclassification. Based on this analysis, mitigation strategies are proposed across multiple subsystems, including sensor design, perception algorithms, decision-making logic, and motion control. This case study demonstrates the feasibility and effectiveness of using the integrated FTA-BN framework to model and analyze fault-based failures in autonomous driving.

The second case study focuses on object detection failure to assess the framework's ability to support SOTIF-oriented safety analysis. From the first case study, we found that perception system is one of the most important perception tasks. We construct a fault tree that top event as object detection failures, including three branches: triggering conditions, functional insufficiencies, and failure modes. The results show that weather and occlusion are major contributors to object detection failures, which contributes to 45.76% and 58.72%, respectively, highlighting the importance of accounting for performance limitations even in the absence of system faults. The study also identifies practical mitigation strategies, including incorporating challenging weather scenarios in training datasets, using ensemble detection models, and applying temporal consistency checks to improve robustness.

The proposed methodology has the ability to incorporate rare but high-impact events. Traditional FTA typically modeled rare event as low-probability basic events, but their influence on overall system safety can be understated due to the static and binary nature of the method. By integrating BN, rare scenarios such as aggressive driving by other road users, or unusual environmental conditions, can be explicitly modeled as probabilistic variables. Their likelihoods can be dynamically updated when new evidence is observed, allowing the analysis to reflect their true impact on system-level hazards. This capability enhances the framework's relevance to SOTIF, where accounting for infrequent but critical scenarios is essential.

Overall, the findings from both case studies confirm the value of integrating FTA with BN to support both fault-based and performance-based risk assessment. This dual capability aligns with the complementary perspectives of ISO 26262 and ISO 21448, providing a structured and probabilistic foundation for comprehensive safety analysis in autonomous vehicles.

5.2 Future Work and Open Issues

While the proposed methodology demonstrates significant promise, several limitations and open research questions remains.

First, the CPTs in the BN were primarily initialized using expert assumptions or hypothetical failure rates due to the lack of publicly available real-world data. Incorporating operational data from field testing and perception logs would significantly enhance the accuracy and credibility of the model. Future work could explore the use of Bayesian parameter learning or data-driven estimation techniques to refine CPTs more robustly.

Another promising direction for future work is the integration of deep probabilistic modeling into the proposed framework. While the current BN relies on expert assumptions or simplifies estimates to initialize conditional probability tables, deep probabilistic models can provide data-driven uncertainty estimates directly from high-dimensional perception data. These models are capable for quantifying both epistemic and aleatoric uncertainty, particularly in rare or adverse scenarios that are central to SOTIF. Incorporating their outputs as evidence nodes in BN would allow failure probabilities to reflect real-world perception performance under varying operational conditions.

The scope of validation is limited to two case studies. Although these scenarios are representative, broader validation across additional autonomous vehicles functions, such as planning, control and human machine interaction, would strengthen the generalization of the framework. Application to larger system models could also test the scalability of the approach and its integration with model-based design tools.

Continuous improvement can be achieved by decentralizing the learning and obtaining an aggregated global model, thereby preserving privacy while scaling the collection of SOTIF-related evidence. Recent studies on federated learning in smart city sensing demonstrate how collaborative, decentralized training can address privacy and scalability challenges in large-scale, heterogeneous environments while maintaining model performance [62]. Applying federated learning in the context of autonomous vehicles would allow

BN to be refined continuously with fleet-wide operational evidence, improving calibration while preserving the explainability of the FTA structure.

In addition, future research could investigate the adoption of Machine Learning Operations (MLOps) and Artificial Intelligence for IT Operations (AIOps) practices to establish pipelines for continuous data collection, deployment, and monitoring. These approaches ensure that learning models remain up-to-date with changing operational environments, while also maintaining traceability, reproducibility, and compliance with safety standards [12,94]. Integrating such lifecycle management techniques with the FTA–BN framework would provide a systematic way to ensure that both probabilistic safety models and machine learning components evolve in a coordinated and controlled manner.

Future research can explore how this framework can support real-time safety monitoring and adaptive fault tolerance as part of a system's lifecycle. Its integration with runtime monitoring architectures has the potential to transform safety assessment from a one-time design activity to a continuous operational capability. This shift would enhance the long-term reliability and resilience of autonomous vehicle systems.

References

[1] Mohamed Abdelgawad and Aminah Robinson Fayek. Fuzzy reliability analyzer: Quantitative assessment of risk events in the construction industry using fuzzy fault-tree analysis. *Journal of Construction Engineering and Management*, 137(4):294–302, 2011.

[2] Asim Abdulkhaleq, Daniel Lammering, Stefan Wagner, Jürgen Röder, Norbert Balbierer, Ludwig Ramsauer, Thomas Raste, and Hagen Boehmert. A systematic approach based on stpa for developing a dependable architecture for fully automated driving vehicles. *Procedia Engineering*, 179:41–51, 2017. 4th European STAMP Workshop 2016, ESW 2016, 13-15 September 2016, Zurich, Switzerland.

[3] Stefan von der Decken Alexander Börger, René Hosse. Sotif - a new challenge for functional testing. *ATZ Electron Worldw*, 15:56–60, 2020.

[4] Matthias Althoff and Alexander Mergel. Comparison of markov chain abstraction and monte carlo simulation for the safety assessment of autonomous cars. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1237–1247, 2011.

[5] Matthias Althoff, Olaf Stursberg, and Martin Buss. Safety assessment of driving behavior in multi-lane traffic for autonomous vehicles. In *2009 IEEE Intelligent Vehicles Symposium*, pages 893–900, 2009.

[6] Toshiaki Aoki, Daisuke Kawakami, Nobuo Chida, and Takashi Tomita. Dataset fault tree analysis for systematic evaluation of machine learning systems. In *2020 IEEE 25th Pacific Rim International Symposium on Dependable Computing (PRDC)*, pages 100–109, 2020.

[7] Koorosh Aslansefat, Sohag Kabir, Youcef Gheraibia, and Yiannis Papadopoulos. Dynamic fault tree analysis: state-of-the-art in modeling, analysis, and tools. *Reliability management and engineering*, pages 73–112, 2020.

[8] Pavanaditya Badida, Yakesh Balasubramaniam, and Jayapriya Jayaprakash. Risk evaluation of oil and natural gas pipelines due to natural hazards using fuzzy fault tree analysis. *Journal of Natural Gas Science and Engineering*, 66:284–292, 2019.

[9] Ahmed Ali Baig, Risza Ruzli, and Azizul B Buang. Reliability analysis using fault tree analysis: a review. *International Journal of Chemical Engineering and Applications*, 4(3):169, 2013.

[10] Abhishek Balasubramaniam and Sudeep Pasricha. Object detection in autonomous vehicles: Status and open challenges, 2022.

[11] Gourav Bathla, Kishor Bhadane, Rahul Kumar Singh, Rajneesh Kumar, Rajanikanth Aluvalu, Rajalakshmi Krishnamurthi, Adarsh Kumar, R. N Thakur, and Shakila Basheer. Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. *Mobile Information Systems*, 2022(1):7632892, 2022.

[12] Denis Baylor, Eric Breck, Heng-Tze Cheng, Noah Fiedel, Chuan Yu Foo, Zakaria Haque, Salem Haykal, Mustafa Ispir, Vihan Jain, Levent Koc, Chiu Yuen Koo, Lukasz Lew, Clemens Mewald, Akshay Naresh Modi, Neoklis Polyzotis, Sukriti Ramesh, Sudip Roy, Steven Euijong Whang, Martin Wicke, Jarek Wilkiewicz, Xin Zhang, and Martin Zinkevich. Tfx: A tensorflow-based production-scale machine learning platform. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 1387–1395, New York, NY, USA, 2017. Association for Computing Machinery.

[13] Joseph R. Belland. Modeling common cause failures in diverse components with fault tree applications. In *2017 Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–6, 2017.

[14] Parth Bhavsar, Plaban Das, Matthew Paugh, Kakan Dey, and Mashrur Chowdhury. Risk analysis of autonomous vehicles in mixed traffic streams. *Transportation Research Record*, 2625(1):51–61, 2017.

[15] Igal Bilik, Oren Longman, Shahar Villeval, and Joseph Tabrikian. The rise of radar for autonomous vehicles: Signal processing solutions and future research directions. *IEEE Signal Processing Magazine*, 36(5):20–31, 2019.

[16] Lukas Birkemeyer, Christian King, and Ina Schaefer. Is scenario generation ready for sotif? a systematic literature review. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pages 472–479, 2023.

[17] Sungil Byun, Mayorkinos Papaelias, Fausto Pedro García Márquez, and Dongik Lee. Fault-tree-analysis-based health monitoring for autonomous underwater vehicle. *Journal of Marine Science and Engineering*, 10(12), 2022.

[18] Carmen Carlan, Noah Carlson, Chris Dwyer, Manoja Hirannaiah, and Michael Wagner. The SOTIF Meta-Algorithm: Quantitative Analyses of the Safety of Autonomous Behaviors . In *2024 IEEE 35th International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pages 191–198, Los Alamitos, CA, USA, Oct 2024. IEEE Computer Society.

[19] Laura Carnevali, Lorenzo Ciani, Alessandro Fantechi, Gloria Gori, and Marco Papini. An efficient library for reliability block diagram evaluation. *Applied Sciences*, 11(9), 2021.

[20] Enrique Castillo, Zacarías Grande, Elena Mora, Xiangdong Xu, and Hong K. Lo. Proactive, backward analysis and learning in road probabilistic bayesian network models. *Computer-Aided Civil and Infrastructure Engineering*, 32(10):820–835, 2017.

[21] Andrea Ceccarelli and Francesco Secci. RGB Cameras Failures and Their Effects in Autonomous Driving Applications . *IEEE Transactions on Dependable and Secure Computing*, 20(04):2731–2745, 2023.

[22] Qi Chang, Changcong Zhou, Haodong Zhao, Wenxuan Wang, and Zhufeng Yue. How dependent basic events with interval-valued probabilities affect fault tree analysis. *Quality and Reliability Engineering International*, 39(1):382–411, 2023.

[23] Kuan-Ting Chen, Winnie Chen, Ann Bisantz, Su Shen, and Ercan Sahin. Where failures may occur in automated driving: A fault tree analysis approach. *Journal of Cognitive Engineering and Decision Making*, 17, 08 2022.

[24] Lei Chen, Jian Jiao, and Tingdi Zhao. A novel hazard analysis and risk assessment approach for road vehicle functional safety through integrating stpa with fmea. *Applied Sciences*, 10(21), 2020.

[25] Liang Chen, Fu Zheng, Xiaopeng Gong, and Xinyuan Jiang. Gnss high-precision augmentation for autonomous vehicles: Requirements, solution, and technical challenges. *Remote Sensing*, 15(6), 2023.

[26] Serena H. Chen and Carmel A. Pollino. Good practice in bayesian network modelling. *Environmental Modelling Software*, 37:134–145, 2012.

[27] Shanzhi Chen, Xinghua Hu, Jiahao Zhao, Ran Wang, and Min Qiao. A review of decision-making and planning for autonomous vehicles in intersection environments. *World Electric Vehicle Journal*, 15(3), 2024.

[28] Wai-Ki Ching and Michael K Ng. Markov chains. *Models, algorithms and applications*, 650:111–139, 2006.

[29] Huazhen Chu, Lisha Mo, Rongquan Wang, Tianyu Hu, and Huimin Ma. Visibility of points: Mining occlusion cues for monocular 3d object detection. *Neurocomputing*, 502:48–56, 2022.

[30] Tabitha S. Combs, Laura S. Sandt, Michael P. Clamann, and Noreen C. McDonald. Automated vehicles and pedestrian safety: Exploring the promise and limits of pedestrian detection. *American Journal of Preventive Medicine*, 56(1):1–7, 2019.

[31] Mirko Conrad and Georg Schildbach. Analysis of functional insufficiencies and triggering conditions to improve the sotif of an mpc-based trajectory planner, 2024.

[32] G Cristea and DM Constantinescu. A comparative critical study between fmea and fta risk analysis methods. *IOP Conference Series: Materials Science and Engineering*, 252(1):012046, oct 2017.

[33] Yixin Cui, Shuo Yang, Chi Wan, Xincheng Li, Jiaming Xing, Yuanjian Zhang, Yanjun Huang, and Hong Chen. Continual adaptation for autonomous driving with the mixture of progressive experts network, 2025.

[34] Lansu Dai and Burak Kantarci. Advancing autonomous vehicle safety: A combined fault tree analysis and bayesian network approach, 2025.

[35] Erwin de Gelder, Hala Elrofai, Arash Khabbaz Saberi, Jan-Pieter Paardekooper, Olaf Op den Camp, and Bart de Schutter. Risk quantification for automated driving systems in real-world driving scenarios. *IEEE Access*, 9:168953–168970, 2021.

[36] Jordi Dunjó, Vasilis Fthenakis, Juan A. Vílchez, and Josep Arnaldos. Hazard and operability (hazop) analysis. a literature review. *Journal of Hazardous Materials*, 173(1):19–32, 2010.

[37] K. Durga Rao, V. Gopika, V.V.S. Sanyasi Rao, H.S. Kushwaha, A.K. Verma, and A. Srividya. Dynamic fault tree analysis using monte carlo simulation in probabilistic safety assessment. *Reliability Engineering & System Safety*, 94(4):872–883, 2009.

[38] Obieze E., Eze Nduka, and Chika Onuigbo. Reliability in engineering maintenance planning: Application of fmea and fta. *Explorematics Journal of Innovative Engineering and Technology*, 5(1):24–33, 04 2024.

[39] Frederick Ojiemhende Ehiagwina, Olufemi Oluseye Kehinde, Abubakar Sidiq Nafiu, Lateef Olashile Afolabi, and IkeolaSuhurat Olatinwo. Fault tree analysis and its modifications as tools for reliability and risk analysis of engineering systems–an overview. *International Journal of Research Publication and Reviews*, 2582:7421, 2022.

[40] Abyad Enan, Abdullah Ai Mamun, Jean Michel Tine, Judith Mwakalonge, Debbie Aisiana Indah, Gurcan Comert, and Mashrur Chowdhury. Basic safety message generation through a video-based analytics for potential safety applications. *ACM Journal on Autonomous Transportation Systems*, 1(4), August 2024.

[41] Víctor J. Expósito Jiménez, Georg Macher, Daniel Watzenig, and Eugen Brenner. Safety of the intended functionality validation for automated driving systems by using perception performance insufficiencies injection. *Vehicles*, 6(3):1164–1184, 2024.

[42] Hamed Fazlollahtabar and Seyed Taghi Akhavan Niaki. Fault tree analysis for reliability evaluation of an advanced complex manufacturing system. *Journal of Advanced Manufacturing Systems*, 17(01):107–118, 2018.

[43] International Organization for Standardization. Iso 26262 road vehicles – functional safety. Technical report, 2018.

[44] International Organization for Standardization. Iso 21448 road vehicles — safety of the intended functionality. Technical report, 2022.

[45] Marc Förster and Bernhard Kaiser. Increased efficiency in the quantitative evaluation of state/event fault trees. *IFAC Proceedings Volumes*, 39(3):255–260, 2006. 12th IFAC Symposium on Information Control Problems in Manufacturing.

[46] Luiz G. Galvão and M. Nazmul Huda. Pedestrian and vehicle behaviour prediction in autonomous vehicle system — a review. *Expert Systems with Applications*, 238:121983, 2024.

[47] Junru Gao, Bin Yao, and Tianfei Shen. Application of bayesian network based on fault tree in fault diagnosis of radio altimeter system. *Journal of Physics: Conference Series*, 2551(1):012014, jul 2023.

[48] Fausto Pedro García Márquez, Isaac Segovia Ramírez, Behnam Mohammadi-Ivatloo, and Alberto Pliego Marugán. Reliability dynamic analysis by fault trees and binary decision diagrams. *Information*, 11(6), 2020.

[49] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[50] Iago Pacheco Gomes and Denis Fernando Wolf. Health monitoring system for autonomous vehicles using dynamic bayesian networks for diagnosis and prognosis. *Journal of Intelligent & Robotic Systems*, 101(1):19, 2021.

[51] Jakfat Haekal. Quality control with failure mode and effect analysis (fmea) and fault tree analysis (fta) methods: Case study japanese multinational automotive corporation. *International Journal of Scientific Advances (IJSCIA), 3 (2), 227*, 234, 2022.

[52] Robert L Harrison. Introduction to monte carlo simulation. In *AIP conference proceedings*, volume 1204, page 17, 2010.

[53] Ahmad Hassan, Zahira Mokhtar, and Mazleha Maskin. Time dependent reliability analysis for a critical reactor safety system based on fault tree approach. *IOP Conference Series: Materials Science and Engineering*, 1231(1):012015, feb 2022.

[54] Haohui He, Cheng Wang, Yuxin Zhang, and Miao Zhang. A benchmark for sotif of lane marking detection algorithms of autonomous vehicles. In *2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, pages 1–6, 2023.

[55] Jehad Hedel, Nga Nguyen, and Ahmad Abuelrub. Reliability evaluation of autonomous electric vehicles using fault tree method. In *2023 North American Power Symposium (NAPS)*, pages 1–6, 2023.

[56] Chen Sun Kai Yang Dongpu Cao Jun Li Hong Wang, Wenbo Shao. A survey on an emerging safety challenge for autonomous vehicles: Safety of the intended functionality. *Engineering*, 33:17–34, 2024.

[57] Chunxi Huang, Ange Wang, Song Yan, and Dengbo He. Investigating the interrelationships among factors associated with automated vehicle crashes using additive bayesian network. *Transportation Research Record*, 0(0):03611981241274152, 0.

[58] Favour Ikwan, David Sanders, and Mohamed Hassan. Safety evaluation of leak in a storage tank using fault tree analysis and risk matrix analysis. *Journal of Loss Prevention in the Process Industries*, 73:104597, 2021.

[59] J Jaise, NB Ajay Kumar, N Siva Shanmugam, K Sankaranarayanasamy, and T Ramesh. Power system: a reliability assessment using fta. *International Journal of System Assurance Engineering and Management*, 4:78–85, 2013.

[60] Balaraju Jakkula, Govinda Raj Mandela, and Murthy Ch SN. Reliability block diagram (rbd) and fault tree analysis (fta) approaches for estimation of system reliability and availability–a case study. *International Journal of Quality & Reliability Management*, 38(3):682–703, 2021.

[61] Zhiwei Ji, Qibiao Xia, and Guanmin Meng. A review of parameter learning methods in bayesian network. In *International Conference on Intelligent Computing*, pages 3–12. Springer, 2015.

[62] Ji Chu Jiang, Burak Kantarci, Sema Oktug, and Tolga Soyata. Federated learning in smart city sensing: Challenges and opportunities. *Sensors*, 20(21), 2020.

[63] Sohag Kabir. An overview of fault tree analysis and its application in model based dependability analysis. *Expert Systems with Applications*, 77:114–135, 2017.

[64] Sohag Kabir, Mohammed Taleb-Berrouane, and Yiannis Papadopoulos. Dynamic reliability assessment of flare systems by combining fault tree analysis and bayesian networks. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 45(2):4305–4322, 2023.

[65] Bernhard Kaiser. An integrative solution towards sotif and av safety. In *IQPC SOTIF Conference*, pages 1–2, 2019.

[66] Samitha Khaiyum, Bishwajit Pal, and Y. S. Kumaraswamy. An approach to utilize fmea for autonomous vehicles to forecast decision outcome. In *3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, pages 701–709. Springer International Publishing, 2015.

[67] Nima Khakzad, Faisal Khan, and Paul Amyotte. Safety analysis in process facilities: Comparison of fault tree and bayesian network approaches. *Reliability Eng. & System Safety*, 96/8:925–932, 2011.

[68] Sakib Mahmud Khan, M Sabbir Salek, Vareva Harris, Gurcan Comert, Eric A. Morris, and Mashrur Chowdhury. Autonomous vehicles for all? *ACM Journal on Autonomous Transportation Systems*, 1(1), March 2024.

[69] Hoseon Kim, Jieun Ko, Cheol Oh, and Seoungbum Kim. Evaluation of autonomous driving safety by operational design domains (odd) in mixed traffic. *Sustainability*, 16(22), 2024.

[70] Jiyoon Kim, Bum-jin Park, and Jisoo Kim. Empirical analysis of autonomous vehicle's lidar detection performance degradation for actual road driving in rain and fog. *Sensors*, 23(6), 2023.

[71] O M Kirovskii and V A Gorelov. 534(1):012019, may 2019.

[72] Philip Koopman and Michael Wagner. Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety*, 4(1):15–24, 2016.

[73] Philip Koopman and Michael Wagner. Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine*, 9(1):90–96, 2017.

[74] Birte Kramer, Christian Neurohr, Matthias Büker, Eckard Böde, Martin Fränzle, and Werner Damm. Identification and quantification of hazardous scenarios for automated driving. In Marc Zeller and Kai Höfig, editors, *Model-Based Safety and Assessment*, pages 163–178, Cham, 2020. Springer International Publishing.

[75] Andreas Kuhn, Elvira Thonhofer, Simon Sigl, Jacqueline Erhart, Manfred Harrer, Dennis Boehmlaender, André Neubohn, and Steve Simon. System and odd descriptions and their consequences for vehicle safety and automated mobility, 01 2024.

[76] Johann Laconte, Abderrahim Kasmi, Romuald Aufrère, Maxime Vaidis, and Roland Chapuis. A survey of localization methods for autonomous vehicles in highway scenarios. *Sensors*, 22(1), 2022.

[77] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds, 2019.

[78] W. S. Lee, D. L. Grosh, F. A. Tillman, and C. H. Lie. Fault tree analysis, methods, and applications: A review. *IEEE Transactions on Reliability*, R-34(3):194–203, 1985.

[79] Shiqing Li, Michael Frey, and Frank Gauterin. An innovative technique for fault analysis of electric automated vehicles. *Vehicles*, 6(4):1995–2010, 2024.

[80] Yan-Feng Li, Hong-Zhong Huang, Jinhua Mi, Weiwen Peng, and Xiaomeng Han. Reliability analysis of multi-state systems with common cause failures based on bayesian network and fuzzy probability. *Annals of Operations Research*, 311(1):195–209, 2022.

[81] SATYAJIT LINGRAS, ARUNI BASU, ATHARV M KOLHAR, and STALEN RUMAO. Enhancing software dfmea processes through iso 26262 (automotive functional safety) and iso 21434 (automotive cybersecurity): Addressing rpn limitations with risk priority matrix and hazop integration. *JRE Jounals*, (7), 2025.

[82] Hu-Chen Liu, Long Liu, and Nan Liu. Risk evaluation approaches in failure mode and effects analysis: A literature review. *Expert Systems with Applications*, 40(2):828–838, 2013.

[83] Xingliang Liu, Zhichao Xing, Rui Fang, Lina Zhang, and Xuan Wu. Research on phase analysis method of fused security concept based on collaborative tree. In *2023 2nd International Conference on Automation, Robotics and Computer Engineering (ICARCE)*, pages 1–5, 2023.

[84] Kaushik Madala, Carlos Avalos-Gonzalez, and Gokul Krithivasan. Workflow between iso 26262 and iso 21448 standards for autonomous vehicles. *Journal of System Safety*, 57(1):34–42, Oct. 2021.

[85] Haneet Singh Mahajan, Thomas Bradley, and Sudeep Pasricha. Application of systems theoretic process analysis to a lane keeping assist system. *Reliability Engineering System Safety*, 167:177–183, 2017. Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis.

[86] Sumbal Malik, Manzoor Ahmed Khan, Hesham El-Sayed, Jalal Khan, and Obaid Ullah. How do autonomous vehicles decide? *Sensors*, 23(1), 2023.

[87] Bruce G Marcot and Trent D Penman. Advances in bayesian network modelling: Integration of modelling technologies. *Environmental modelling & software*, 111:386–393, 2019.

[88] Riccardo Mariani. An overview of autonomous vehicles safety. In *2018 IEEE International Reliability Physics Symposium (IRPS)*, pages 6A.1–1–6A.1–6, 2018.

[89] Helmut Martin, Kurt Tschabuschnig, Olof Bridal, and Daniel Watzenig. Functional safety of automated driving systems: Does iso 26262 meet the challenges? In *Automated Driving: Safer and More Efficient Future Driving*, pages 387–416. Springer, 2016.

[90] Farshad Mirzarazi, Sebelan Danishvar, and Alireza Mousavi. The safety risks of ai-driven solutions in autonomous road vehicles. *World Electric Vehicle Journal*, 15(10), 2024.

[91] Shuojie Mo, Xiaofei Pei, and Chaoxian Wu. Safe reinforcement learning for autonomous vehicle using monte carlo tree search. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6766–6773, 2022.

[92] Esmaeil Zarei Mohammad Yazdi. Uncertainty handling in the safety risk analysis: An integrated approach based on fuzzy fault tree analysis. *Journal of Failure Analysis and Prevention*, 18:392–404, 2018.

[93] John Molloy, Sepeedeh Shahbeigi, and John A. McDermid. Hazard and safety analysis of machine-learning-based perception capabilities in autonomous vehicles. *Computer*, 57(11):60–70, 2024.

[94] Paolo Notaro, Jorge Cardoso, and Michael Gerndt. A survey of aiops methods for failure management. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(6):1–45, 2021.

[95] Society of Automotive Engineers. J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Technical report, 2018.

[96] Sankar K Pal, Anima Pramanik, Jhareswar Maiti, and Pabitra Mitra. Deep learning in multi-object detection and tracking: state of the art. *Applied Intelligence*, 51:6400–6429, 2021.

[97] Mayur Anand Pandya, P C Siddalingaswamy, and Sanjay Singh. Fmea-based safety analysis of monocular depth estimation for autonomous vehicles. In *2025 International Conference on Artificial Intelligence and Data Engineering (AIDE)*, pages 907–912, 2025.

[98] Milin Patel, Rolf Jung, and Marzana Khatun. A systematic literature review on safety of the intended functionality for automated driving systems. In *SAE Technical Paper Series*, volume 1. SAE International, April 2025.

[99] Riccardo Patriarca, Mikela Chatzimichailidou, Nektarios Karanikas, and Giulio Di Gravio. The past and present of system-theoretic accident model and processes (stamp) and its associated techniques: A scoping review. *Safety Science*, 146:105566, 2022.

[100] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[101] Liang Peng, Boqi Li, Wenhao Yu, Kai Yang, Wenbo Shao, and Hong Wang. Sotif entropy: Online sotif risk quantification and mitigation for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 25(2):1530–1546, 2024.

[102] James Pickford, Rasadhi Attale, Siraj Shaikh, Hoang Nga Nguyen, and Lee Harrison. Systematic risk characterisation of hardware threats to automotive systems. *ACM Journal on Autonomous Transportation Systems*, 1(4), August 2024.

[103] Michele Pipicelli, Alfredo Gimelli, Bernardo Sessa, Francesco De Nola, Gianluca Toscano, and Gabriele Di Blasio. Architecture and potential of connected and autonomous vehicles. *Vehicles*, 6(1):275–304, 2024.

[104] Krystian Radlak, Michal Szczepankiewicz, Tim Jones, and Piotr Serwa. Organization of machine learning based product development as per iso 26262 and iso/pas 21448. In *2020 IEEE 25th Pacific Rim International Symposium on Dependable Computing (PRDC)*, page 110–119. IEEE, December 2020.

[105] Abu Mohammed Raisuddin, Tiago Cortinhal, Jesper Holmblad, and Eren Erdal Aksoy. 3d-outdet: A fast and memory efficient outlier detector for 3d lidar point clouds in adverse weather. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 2862–2868, 2024.

[106] Samik Raychaudhuri. Introduction to monte carlo simulation. In *2008 Winter Simulation Conference*, pages 91–100, 2008.

[107] Rhea C. Rinaldo and Timo F. Horeis. A hybrid model for safety and security assessment of autonomous vehicles. In *Proceedings of the 4th ACM Computer Science in Cars Symposium*. Association for Computing Machinery, 2020.

[108] Francisca Rosique, Pedro J. Navarro, Carlos Fernández, and Antonio Padilla. A systematic review of perception system and simulators for autonomous vehicles research. *Sensors*, 19(3), 2019.

[109] Michael Roth and Peter Liggesmeyer. Qualitative analysis of state/event fault trees for supporting the certification process of software-intensive systems. In *2013 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pages 353–358, 2013.

[110] Enno Ruijters and Mariëlle Stoelinga. Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Computer Science Review*, 15-16:29–62, 2015.

[111] Giedre Sabaliauskaite, Lin Shen Liew, and Jin Cui. Integrating autonomous vehicle safety and security analysis using stpa method and the six-step model. *International Journal on Advances in Security*, 11(1&2):160–169, 2018.

[112] Javier Saez-Perez, Qi Wang, Jose M. Alcaraz-Calero, and Jose Garcia-Rodriguez. Design, implementation, and empirical validation of a framework for remote car driving using a commercial mobile network. *Sensors*, 23(3), 2023.

[113] Cenk Sakar, Ali C. Toz, Muge Buber, and Burak Koseoglu. Risk analysis of grounding accidents by mapping a fault tree into a bayesian network. *Applied Ocean Research*, 113:102764, 2021.

[114] Ashkan Samadi, Marwan Ammar, and Otmane Ait Mohamed. Fault tree analysis and risk mitigation strategies for autonomous systems via statistical model checking. In *2021 IEEE International Conference on Autonomous Systems (ICAS)*, pages 1–5, 2021.

[115] Tobias Schmid, Stefanie Schraufstetter, Stefan Wagner, and Dominik Hellhake. A safety argumentation for fail-operational automotive systems in compliance with iso 26262. In *2019 4th International Conference on System Reliability and Safety (ICSRS)*, pages 484–493, 2019.

[116] Adam Schnellbach and Gerhard Griessnig. Development of the iso 21448. In Alastair Walker, Rory V. O'Connor, and Richard Messnarz, editors, *Systems, Software and Services Process Improvement*, pages 585–593, Cham, 2019. Springer International Publishing.

[117] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, 1:187–210, 2018.

[118] Valerij Schönemann, Hermann Winner, Thomas Glock, Eric Sax, Bert Böddeker, Geert Verhaeg, Fabrizio Tronci, and Gustavo García Padilla. Fault tree-based derivation of safety requirements for automated driving on the example of cooperative valet parking. In *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV) 2019*, 09 2019.

[119] Barbaros Serter, Christian Beul, Manuela Lang, and Wiebke Schmidt. Foreseeable misuse in automated driving vehicles-the human factor in fatal accidents of complex automation. Technical report, SAE Technical Paper, 2017.

[120] Omveer Sharma, N. C. Sahoo, and Niladri B. Puhan. Dynamic planning of optimally safe lane-change trajectory for autonomous driving on multi-lane highways using a fuzzy logic–based collision estimator. *ACM Journal on Autonomous Transportation Systems*, 1(1), March 2024.

[121] Barry Sheehan, Finbarr Murphy, Cian Ryan, Martin Mullins, and Hai Yue Liu. Semi-autonomous vehicle motor insurance: A bayesian network risk transfer approach. *Transportation Research Part C: Emerging Technologies*, 82:124–137, 2017.

[122] R. M. Sinnamon and J. D. Andrews. Improved accuracy in quantitative fault tree analysis. *Quality and Reliability Engineering International*, 13(5):285–292, 1997.

[123] R. M. Sinnamon and J. D. Andrews. Improved efficiency in qualitative fault tree analysis. *Quality and Reliability Engineering International*, 13(5):293–298, 1997.

[124] Yu Song, Madhav V. Chitturi, and David A. Noyce. Intersection two-vehicle crash scenario specification for automated vehicle safety evaluation using sequence analysis and bayesian networks. *Accident Analysis Prevention*, 176:106814, 2022.

[125] S.R. Sreeraj and T.J. Sarvoththama Jothi. Functional safety assessment of battery management system of autonomous electric vehicle. In *2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)*, pages 1–6, 2023.

[126] Luis Enrique Sucar. *Probabilistic graphical models*. Springer, 2021.

[127] Liangliang Sun, Yan-Fu Li, and Enrico Zio. Comparison of the hazop, fmea, fram, and stpa methods for the hazard analysis of automatic emergency brake systems. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 8(3):031104, 10 2021.

[128] Masoud Taheriyoun and Saber Moradinejad. Reliability analysis of a wastewater treatment plant using fault tree analysis and monte carlo simulation. *Environmental monitoring and assessment*, 187:1–13, 2015.

[129] Lei Tang, Ruijie Wang, Zhanwen Liu, Yunji Liang, Yuanyuan Niu, Wei Zhu, and Zongtao Duan. Scenario-based accelerated testing for sotif in autonomous driving: A review. *IEEE Internet of Things Journal*, 12(2):1453–1470, 2025.

[130] Jessica Van Brummelen, Marie O'Brien, Dominique Gruyer, and Homayoun Najjaran. Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies*, 89:384–406, 2018.

[131] Jessica Van Brummelen, Marie O'Brien, Dominique Gruyer, and Homayoun Najjaran. Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies*, 89:384–406, 2018.

[132] Jorge Vargas, Suleiman Alsweiss, Onur Toker, Rahul Razdan, and Joshua Santos. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors*, 21(16), 2021.

[133] Jun Wang, Li Zhang, Yanjun Huang, and Jian Zhao. Safety of autonomous vehicles. *Journal of Advanced Transportation*, 2020(1):8867757, 2020.

[134] Li Wang, Chun Yuan, Yuchen Lu, Yuxiang Xiao, Fanfeng Hong, and Zijian Zhang. A survey on perception, localization, planning, and control of autonomous vehicles: challenges and solutions. In Lijia Pan and Zaifa Zhou, editors, *Ninth International Symposium on Sensors, Mechatronics, and Automation System (ISSMAS 2023)*, volume 12981, page 129816C. International Society for Optics and Photonics, SPIE, 2024.

[135] Yijing Wang, Zhengxuan Liu, Zhiqiang Zuo, Zheng Li, Li Wang, and Xiaoyuan Luo. Trajectory planning and safety assessment of autonomous vehicles based on motion prediction and model predictive control. *IEEE Transactions on Vehicular Technology*, 68(9):8546–8556, 2019.

[136] Zhibo Wang, Xiaoci Huang, and Zhihao Hu. Attention-based lidar–camera fusion for 3d object detection in autonomous driving. *World Electric Vehicle Journal*, 16(6), 2025.

[137] Zi-Xiang Xia, Sudeep Fadadu, Yi Shi, and Louis Foucard. Robust long-range perception against sensor misalignment in autonomous vehicles, 08 2024.

[138] Qin Xiao, Yapeng Li, Fan Luo, and Hui Liu. Analysis and assessment of risks to public safety from unmanned aerial vehicles using fault tree analysis and bayesian network. *Technology in Society*, 73:102229, 2023.

[139] Xingyu Xing, Tangrui Zhou, Junyi Chen, Lu Xiong, and Zhuoping Yu. A hazard analysis approach based on stpa and finite state machine for autonomous vehicles. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 150–156, 2021.

[140] Xiaoxia Xiong, Long Chen, and Jun Liang. Vehicle driving risk prediction based on markov chain model. *Discrete Dynamics in Nature and Society*, 2018(1):4954621, 2018.

[141] Sichuan Xu, Hongjun Ding, Aimin Du, Chuanchuan Chu, Yeyang Han, Hongyun Li, and Zhongpan Zhu. A review of sotif research for human-machine driving mode switch of intelligent vehicles. In *2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, pages 1–6, 2022.

[142] Shinichi Yamaguchi and Yoshinobu Sato. Quantitative risk evaluation based on modeling of process safety and function accomplishment time for the assessment of sotif. *Next Research*, 2(1):100144, 2025.

[143] Zetong Yang, Zhiding Yu, Chris Choy, Renhao Wang, Anima Anandkumar, and Jose M. Alvarez. Improving distant 3d object detection using 2d box supervision, 2024.

[144] Botao Yao, Shuohan Huang, Peiyi Han, Jie Lin, Shaoming Duan, and Chuanyi Liu. Sotif-oriented risk assessment: A multi-dimensional model for autonomous driving. *IEEE Robotics and Automation Letters*, 10(2):1792–1799, 2025.

[145] Mohammad Yazdi, Sohag Kabir, and Martin Walker. Uncertainty handling in fault tree based risk assessment: State of the art and future perspectives. *Process Safety and Environmental Protection*, 131:89–104, 2019.

[146] Mohammad Yazdi, Javad Mohammadpour, He Li, Hong-Zhong Huang, Esmaeil Zarei, Reza Ghasemi Pirbalouti, and Sidum Adumene. Fault tree analysis improvements: A bibliometric analysis and literature review. *Quality and Reliability Engineering International*, 39(5):1639–1659, 2023.

[147] De Jong Yeong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6), 2021.

[148] Wenhao Yu, Jun Li, Li-Ming Peng, Xiong Xiong, Kai Yang, and Hong Wang. Sotif risk mitigation based on unified odd monitoring for autonomous vehicles. *Journal of Intelligent and Connected Vehicles*, 5(3):157–166, 2022.

[149] He Zhang, Jian Sun, and Ye Tian. Accelerated risk assessment for highly automated vehicles: Surrogate-based monte carlo method. *IEEE Transactions on Intelligent Transportation Systems*, 25(6):5488–5497, 2024.

[150] Jin Zhang, John Robert Taylor, Igor Kozine, and Jingyue Li. Analyzing influence of robustness of neural networks on the safety of autonomous vehicles. In *31st European Safety and Reliability Conference*, 09 2021.

[151] Xizhe Zhang, Siddartha Khastgir, Justin-Kiyoshi Tiele, Kazuhito Takenaka, Tasuku Hayakawa, and Paul Jennings. Odd and behavior based scenario generation for automated driving systems. *IEEE Access*, 12:10652–10663, 2024.

[152] Yuxiao Zhang, Alexander Carballo, Hanting Yang, and Kazuya Takeda. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:146–177, 2023.

[153] Yuxiao Zhang, Alexander Carballo, Hanting Yang, and Kazuya Takeda. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:146–177, 2023.

[154] Congcong Zhao, Tsz Leung Yip, Bing Wu, and Jieyin Lyu. Use of fuzzy fault tree analysis and bayesian network for occurrence likelihood estimation of navigational accidents in the qinzhou port. *Ocean Engineering*, 263:112381, 2022.

[155] Fei Zhao, Chengcui Zhang, and Baocheng Geng. Deep multimodal data fusion. *ACM Comput. Survey*, 56(9), 2024.

[156] Wenhao Zong, Changzhu Zhang, Zhuping Wang, Jin Zhu, and Qijun Chen. Architecture design and implementation of an autonomous vehicle. *IEEE Access*, 6:21956–21970, 2018.