A Safety Case Framework for Automated Vehicle Safety Evaluation

by

Sunder Swaminathan

A Thesis Presented in Partial Fulfillment of the Requirements for the Degree
Master of Science

Approved November 2024
by the Graduate Supervisory Committee:

Junfeng Zhao, Co-Chair

Jeffrey Wishart, Co-Chair

David Wichner

Qiaoning Carol Zhang

Aviral Shrivastava

ARIZONA STATE UNIVERSITY

December 2024

ABSTRACT

As numerous new developers and operators are entering the field of Automated Vehicles (AVs), promising safer, faster, and more efficient modes of transportation, it is crucial to ensure these vehicles are being developed and deployed keeping safety as a priority. One such method is by establishing a Safety Case Framework (SCF), which provides a structure that helps AV organizations furnish claims and evidence to support the argument that an AV is safe and reliable to operate on public roads. This work aims to develop a dynamic SCF that consists of the three pillars:

i. Safety Management System (SMS) pillar addresses organizational safety by managing process risks identified throughout the lifecycle of the system. It establishes protocols for hazard identification, risk assessment, and mitigation strategies to ensure comprehensive safety management, and promotes and educates employees and stakeholders on the importance of safety.

ii. Design Methods pillar emphasizing rigorous engineering practices, ensuring that the design of AVs and their subsystems meets stringent safety requirements. It includes methodologies such as Hazard Analysis and Risk Assessment (HARA), Failure Modes and Effects Analysis (FMEA), and adherence to standards like ISO 26262 for functional safety.

iii. Testing pillar focuses on validating and verifying the behavioral competency and safety of AV through various testing methods. These methods range from simulation-based testing to closed course and on-road testing, ensuring that the AV performs reliably under various conditions and scenarios.

This work additionally focuses on Technology Readiness Levels (TRLs) to ensure safety at each phase of development and includes change risk management for post-deployment risk. This framework is aligned with the TRL scale, which ranges from early conceptualization (TRL 1) to full operational deployment (TRL 9), requiring specific safety objectives and documentation for each TRL stage.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to everyone who has supported and guided me throughout my journey in developing this thesis, titled "A Safety Case Framework for Automated Vehicle Functional Safety Evaluation."

First and foremost, I am deeply grateful to my co-chairs, Dr. Jeffrey Wishart, and Dr. Junfeng Zhao for their invaluable guidance, encouragement, and insightful feedback. Their expertise and mentorship have been instrumental in shaping my research and in helping me navigate the complexities of this work. I am also thankful to the members of my thesis committee, David Wichner, Dr. Qiaoning Carol Zhang, and Dr. Aviral Shrivastava for their constructive criticism and support.

Finally, I am incredibly thankful to my family and friends for their unwavering support, patience, and encouragement. Their belief in me has been a source of motivation and strength throughout my Master's journey.

To all those who have contributed to my work, directly or indirectly, I extend my deepest appreciation. Thank you for helping make this thesis a reality.

TABLE OF CONTENTS

Page

Page

LIST OF ABBREVIATIONS

Abbreviations	**Definitions**
ADS	Automated Driving Systems
ADS-DV	Automated Driving System-Dedicated Vehicle
AI	Artificial Intelligence
ANSI	American National Standards Institute
AR	Augmented Reality
ASIL	Automotive Safety Integrity Level
ASPICE	Automotive Software Process Improvement and Capability dEtermination
AUR	Absence of Unreasonable Risk
AV	Automated Vehicle
AVSC	Automated Vehicle Safety Consortium
AV-TEP	Automated Vehicle Test and Evaluation Process
BAS	Brake Assist System
CARLA	CAR Learning to Act
CRM	Change Risk Management
DA	Driving Assessment
ESC	Electronic Stability Control
FAA	Federal Aviation Administration
FMEA	Failure Modes and Effects Analysis
FTA	Fault Tree Analysis
FTD	Fallback Test Driver
HARA	Hazard Analysis and Risk Assessment

HAZOP	Hazard and Operability Study
HMI	Human-Machine Interface
ICAO	International Civil Aviation Organization
IFTD	In-vehicle Fallback Test Driver
IRL	Integration Readiness Level
ISO	International Organization for Standardization
ML	Machine Learning
NHTSA	National Highway Transport Safety Administration
ODD	Operational Design Domain
OEM	Original Equipment Manufacturer
ORAD	On-Road Automated Driving committee
PHA	Preliminary Hazard Analysis
QMS	Quality Management Systems
SA	Safety Assurance
SAE	Society of Automotive Engineers
SCF	Safety Case Framework
SMM	Safety Management Manual
SMS	Safety Management System
SOTIF	Safety of the Intended Functionality
SP	Safety Promotion
SPI	Safety Performance Indicators
SPO	Safety Policy and Objectives
SRM	Safety Risk Management
TRA	Technology Readiness Assessment

TRL	Technology Readiness Level
USDOT	United States Department of Transportation
VSSA	Voluntary Safety Self-Assessment

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

INTRODUCTION

The automotive industry has seen substantial advancements in technology over the years since its invention in the 1860s. Automated Driving System (ADS) leads in technology aiming to revolutionize transportation by enhancing safety, efficiency, and accessibility, with significant environmental and societal benefits. To gain public trust and widespread adoption of this technology, robust safety protocols at the company level are necessary along with stringent industry wide guidelines prescribed by regulatory agencies. Trust in the system can be gained when the system is designed and operated to be safe and reliable, and assurance that the organization has addressed potential hazards and failures that would compromise the safety of the vehicle. Addressing operational deployment risk is a key factor in this thesis work and it provides methods for real-time risk management, incident detection, weather and environmental hazards. Here, a Safety Case Framework helps the AV developer systematically identify and assess hazards and develop counter measure to control risks while maintaining assurance of effective mitigation and provide justifiable evidence to the argument stating the AV is safe for deployment on public roads. Utilizing a safety case-based approach to validate the safety of the AV, helps in tailoring risk management strategies, promoting traceability and accountability, and provide comprehensive coverage over the lifecycle of the AV facilitating continuous improvement and monitoring.

Brief History of Safety

The advent of the Industrial Revolution in 1750 along with the invention of the steam engine gave rise to the first age of safety referred to as 'Technology age'. This age helped in identifying the component of failure and prevent single component failures. The technology age transitioned to the second age ('the human'), when investigations conducted after the meltdown of the Three Mile Island (TMI) reactor pointed out the cause for incident to be people-related problems rather than equipment failure. During this age, systems were designed to be tolerant to human errors which collectively helped in minimizing the failures caused by human error and single component failures. Until the incidents of Chernobyl and Challenger space shuttle accident where investigations identified organizational factors to have contributed to the calamity. The 'organization age' viewed human errors and component failure as a consequence rather than a cause. The organization age aimed to remove weaknesses higher up in the organization through new safety management models and assessment methods. The fourth age of SMS, referred to as the 'system age', is a holistic approach that aims to establish an understanding of the interconnection and interdependencies between the 3 previously mentioned 'ages', providing a clear picture of the context, facilitating control measures that are more effective and efficient in avoiding accidents. These industrial safety practices transitioned to more comprehensive regulatory frameworks in the early 20th century, which were later adopted by industries post-World War II [1].

Safety and technology in the automotive industry have seen remarkable advancements since Carl Benz patented his "vehicle powered by a gas engine" in 1886, with most safety

features used in modern cars beginning to be introduced as early as the 1950s. Laws and regulatory guidelines were implemented in the 1960s with the intent of improving the safety of the occupants. The 1970s and 1980s saw the establishment of the National Highway Traffic Safety Administration (NHTSA), and few manufacturers introduced airbags in their vehicles. Basic electronic safety features like Brake Assist Systems (BAS), Electronic Stability Control (ESC) were introduced in the 1990s, and crash testing for all new vehicles became mandatory. In the 2000s, automotive technology took a giant leap forward and introduced various driver assistance features that help the driver in mitigating and avoiding crashes with the help of cameras and motion sensors that monitor blind spots, provide lane departure warnings, and emergency braking systems [2].

Safety has always been a critical component of many industries. Organizations benefitted from safety regulations, as they provided guidelines to periodically monitor and mitigate risks and continue the reduction in the rate of accidents. Following a significant regulatory model established by Occupational Safety and Health Administration (OSHA) in enforcing workplace safety regulations, the International Civil Aviation Organization (ICAO) introduced the concept of Safety Management System (SMS) in the aviation industry in its Safety Management Manual (SMM) [3] in 2006. Witnessing a tremendous expansion across various industries in the 2000s, today, SMS is essential for enhancing safety, risk management, assurance, and promotion of safety culture.

Automated Vehicles (AVs)

Continuing with the same pace of growth, Google introduced its first fully automated ride on public roads in 2015 [4], with the goal to provide a safer mode of transportation for the public. Since then, AVs have been a major focus of the automotive industry, and every major automobile OEM and technology company is racing and competing to deploy their AV on public roads. To ensure they deploy safe systems, AV companies go through rigorous testing ranging from simulation-based testing, closed-course testing, to on-road testing on public roads. The AV companies evaluate their Automated Driving System (ADS) in numerous conditions and various iterations of scenarios the AV would encounter during its operation. The magnitude of testing depends on the level of automation that the organization is developing.

The industry standard SAE J3016 [5] document titled "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles" defines 6 levels of autonomy starting from SAE Level 0: No Driving Automation, SAE Level 1: Driver Assistance, SAE Level 2: Partial Driving Automation, SAE Level 3: Conditional Driving Automation, SAE Level 4: High Driving Automation, SAE Level 5: Full Driving Automation (shown in Figure 1).

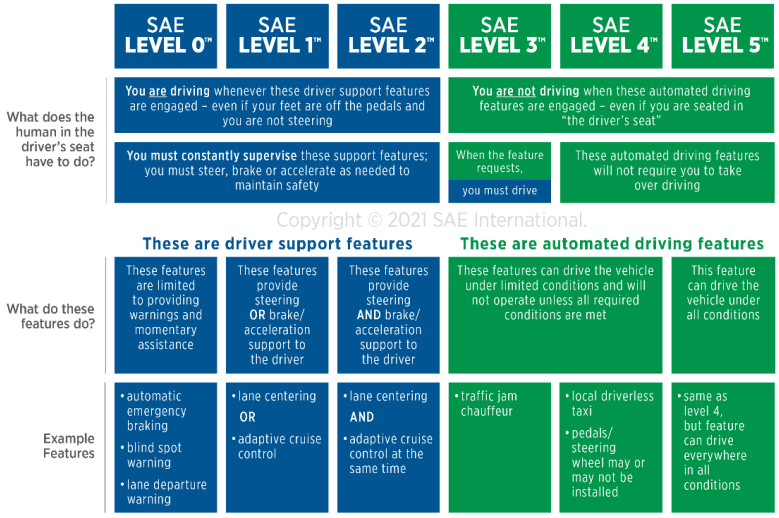

Figure 1: SAE levels of automation
Source: [5]

Automated Vehicle Test and Evaluation Process (AV-TEP) Mission

The AV-TEP mission is an effort conceived and initiated by one of my thesis co-chairs, Dr. Jeffrey Wishart at the Science Foundation Arizona and in collaboration with Arizona State University to develop a safety case-based approach to help AV developers in the safe deployment of ADS-equipped vehicles. For the AV-TEP mission, the SCF is structured in three pillars that address the AV developer's organization, the AV design, and the performance of the AV. An overall solicitation of evidence documentation from the AV developer is developed for each pillar in the SCF, and these documents are called "Specification Forms." These forms are developed leveraging works from relevant organizations such as International Organization for Standardization (ISO 26262, ISO 21448, ISO 4804/5083), Underwriters Laboratory (UL4600), Automated Vehicle Safety Consortium (AVSC SMS, AVSC CRM, AVSC Best Practices on ODD and Behavioral

Competency), SAE documents (SAE J3016, SAE J3018, SAE J3237, SAE J3131, etc.), and many other documents and standards from safety-critical industries.

This work for the AV-TEP mission focuses on the dynamic driving task (DDT) safety of the AV with attention to safety engineering. The AV-TEP mission safety case does not cover all the aspects of operational safety of an AV such as cybersecurity, vehicle maintenance, and passenger-initiated emergency stops. These operational safety requirements are left to future work.

Statement of Purpose

As the AV industry is in its nascent stages of development, the need for an exhaustive SCF that factors in strong regulatory guidelines along with a healthy safety culture is crucial to help AV developers deploy safe modes of transportation onto the public roads. Establishing trust in AV technology requires a proactive approach to managing risks, validating system safety, and adhering to high standards of accountability.

This thesis aims to cater to the evolving field of ADS technology to provide an SCF in the form of specification forms that has been developed by harmonizing standards and best practices from various industries such as aviation, nuclear, automotive, etc. This SCF is structured around three fundamental pillars that collectively form the backbone of safety assurance: Safety Management System, Design Methods, and Testing. This SCF includes Technology Readiness Level (TRL), a system that communicates the maturity level of a technology that is used to also stage-gate the AV development process.

Through this holistic approach, the proposed SCF enables developers to demonstrate the safety and reliability of their ADS technology, offering a clear, evidence-based pathway to gaining regulatory approval and public trust. The framework's adaptability ensures that it can evolve alongside emerging advancements in AV technology while maintaining a steadfast focus on safety.

Safety Case

A Safety Case is defined as a structured way of demonstrating how the organization determines its ADS is acceptably safe for deployment on public roads. A safety case is structured in the form of formal arguments put forward by the organization and the stakeholders, followed by evidence to support these arguments. A safety case must demonstrate that reasonably foreseeable risks associated with the vehicle have been identified and sufficient evidence is provided that each one has been adequately addressed. It should also include essential elements like design and operational principles such as absence of unreasonable risk (AUR), continuous monitoring and improvement, transparency and accountability, operational risk mitigation, and compliance with relevant regulations such as European Commission (EU) regulation 2022/1426 [6], industry standards and best practices. A few companies and regulatory agencies involved in the development and regulation of ADS are employing a safety case-based approach to validate the safety of their ADS.

Three Pillar Framework

The three pillars (shown in Figure 2) of the AV-TEP SCF are Safety Management System (SMS), Design Methods, Testing pillars. SMS is an approach derived from the aviation industry to enhance organizational decision-making using a combination of safety principles, processes, and practices systematically and comprehensively. This process helps in the routine identification of potential risks and hazards and develop mitigation strategies to address these risks.

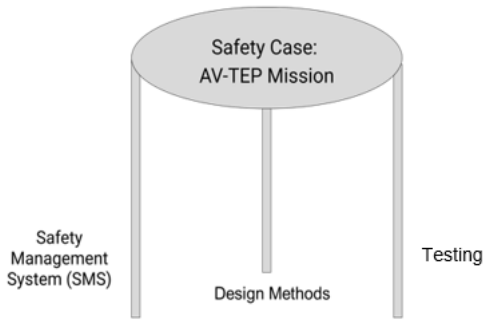

Figure 2: Three Pillar Safety Case Framework
Source: [7]

The SMS pillar comprises four fundamental components that have been implemented in various safety-critical industries, these include Safety Policy and Objectives (SPO), Safety Risk Management (SRM), Safety Assurance (SA), and Safety Promotion (SP), adapted from the AVSC SMS model [8].

The second pillar, Design Methods helps the developers in ensuring the safety, reliability, and performance of their AV, as well as providing requirements for compliance with regulatory frameworks and industry standards. The AV-TEP Design Methods specification form is developed by leveraging ISO 26262, ISO 21448, SAE J3131, and other relevant

documents to define system requirements, design constraints, identify stakeholders, and develop a concept of operation. This approach includes considerations for Hazard Analysis and Risk Assessment (HARA), Design Verification, Traceability, Documentation of safety requirements and Change Management.

The Testing pillar deals with multiple methods of testing the subject vehicle in real-world scenarios and understanding how it behaves in different conditions. The test methods have different uses, as some test specific ADS sub-systems, and different fidelities, as shown by the continuum (shown in Figure 3). The testing method changes as the AV progresses through different stages of design, production and safety levels. In general, simulation, closed course, and on-road testing are the most common test methods. Simulation-based testing is the first and primary testing methodology and is universally used. Once competency in simulation is achieved, a prototype can be built and tested in a closed course with the option of AR-enhanced testing, which allows the developer to test the AV in AR-generated scenarios that are not possible with the equipment and facilities at the closed course. As a next step, the AV can be driven on public roads in Shadow Mode [9], where a human driver operates the vehicle and the system records data generated by the perception module. Following this, the AV is optionally tested on public roads with AR enhancements to generate complex traffic scenarios, and later transitions to on-road testing without AR enhancements. Later, the AV is tested on public roads with the assistance of a trained, In-vehicle Fallback Test Driver (IFTD) who intervenes when necessary. When the AV shows satisfactory competency with prior testing methods, it is testing on public roads without the IFTD. It should be noted that other test methods, such as vehicle-in-the-loop (ViL),

model-in-the-loop (MiL), software-in-the-loop (SiL), and hardware-in-the-loop (HiL), collectively known as XiL are also possible.

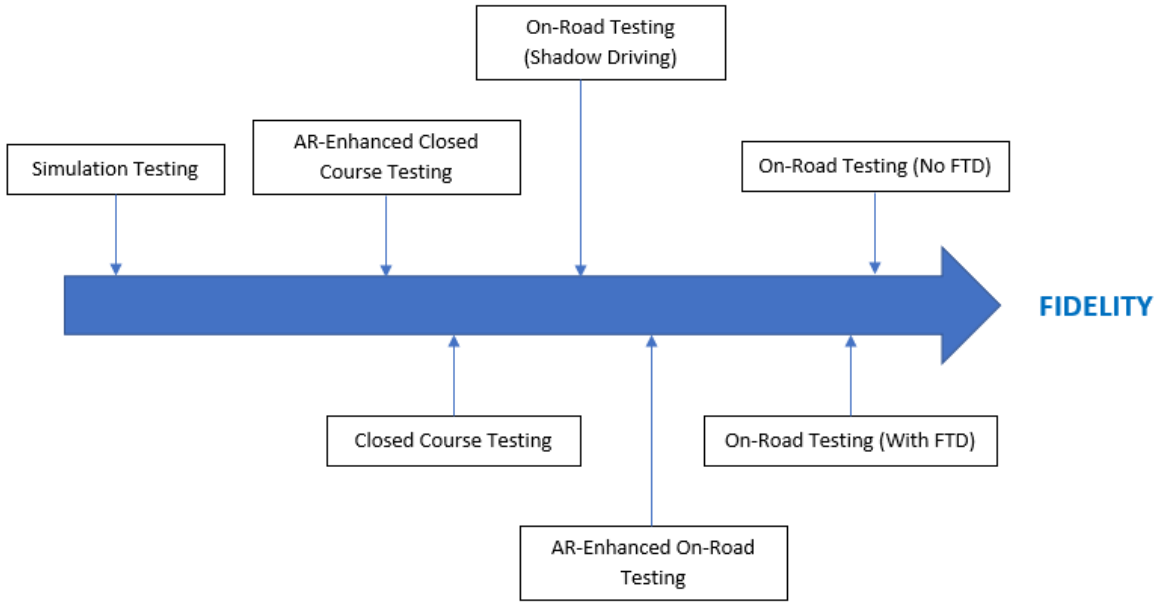

Figure 3: Progression of testing methodologies
Source: [10]

As part of the Testing pillar, an exhaustive set of scenarios can be created based, in part, on the list of pre-crash scenarios prescribed by the National Highway Transport Safety Administration (NHTSA) [11] that have varying conditions and parameters in terms of road networks, weather conditions, pedestrian interactions. CARLA is used to simulate scenarios, as it is an open-source software with great graphics and flexible customization options for vehicle physics and weather conditions and has an exhaustive suite of detailed sensors. Running these scenarios on the simulator generates a metrics file with information like speed and positions of the vehicles, distance between the ego and adversary vehicle, that is fed through a Python script which has the formulations for calculating the score for each metric within the black box, grey box, and white box. Once the individual score for

each metric (0 for “good”, and 1 for "bad” denoting a violation of a threshold) is calculated, the severity of each violation is determined, and the Unweighted Driving Assessment (DA) Score is calculated; the Weighted DA Score by providing weight for the complexity, relevancy of the scenario, and the fidelity of the testing method. The two DA Scores out of 100 are then assigned to the scenario navigation.

LITERATURE REVIEW

Safety Case efforts

Various participants in the automotive industry are contributing to the effort of developing a safety case framework with the interest of ensuring the safety and reliability of AVs. These safety cases published by various organizations give an understanding of how the organization addresses the concept of safety and what mitigation strategies have been implemented to address safety concerns. Publishing such documents encourages the public to trust and adopt the technology while enjoying the benefits it brings. These efforts focus on adhering to industry standards and best practices, to demonstrate that the vehicle has been designed, developed, tested, and validated to operate safely in various conditions.

Waymo's Safety Case Approach [12] from 2023, focuses on the concept of the Absence of Unreasonable Risk (AUR) which emphases the operation and deployment of AVs without posing excess risk to the passengers, and other road users. The document is structured in three sections, the first section, Layered Approach to Safety decomposes AUR into three sub-components of risk due to architectural, behavioral, and in-service operational hazards with each sub-component requiring a set of risk assessment criteria to be specified to assess if residual risk is achieved or further mitigation is required. The Dynamic Approach to Safety section talks about the dynamic and iterative nature of the safety determination life cycle. In the absence of a means of establishing the credibility of a safety case, the Credibility Approach to Safety section proposes the notion of a case credibility assessment that provides a top-down justification for the structure of argument, and a bottom-up approach for level of credibility afforded by the existing mitigation strategies. In [13],

Waymo highlights the strong engineering foundation and the thorough ODD-specific risk analysis process that help Waymo determine the readiness of its AVs. In this document, Waymo elaborates on the criteria for performance evaluation of its AVs which include crash avoidance, completion of trips in driverless mode, and ability to adhere to applicable driving rules. While Waymo addresses the need for a hazard analysis process that is tailored to applicable to SAE Level 4 vehicles, Waymo follows ISO 26262 guidelines for its HARA. This document addresses the entire lifecycle of the ADS, and the approach to safety focuses on the ADS, safety of base vehicle, and safety of AV operation. The document is divided into three sections, namely Hardware Layer, Behavioral Layer, and Operations Layer. The Hardware Layer details the system engineering process that is incorporated to define safety requirements and includes the verification and validation procedures for the components, sub-system and system elements. The Behavioral Layer is a multi-faceted approach covering hazard analysis, scenario-based verification, and large-scale simulated deployment to assess the capability of the ADS to safely handle the driverless movements of the AV. The Operations Layer discusses the fault management protocols and field safety processes that are necessary for fleet operation post-deployment. Prior to the transition of one level of development to another, a thorough readiness assessment is to be performed that is focused on the ODD, use case, and vehicle platform, and Waymo applies all prior methodologies before implementing the transition. Waymo establishes a performance criterion for its vehicle that helps in assessing the readiness for deployment.

Aurora's safety report [14] elaborates on its approach to prove the safety of its ADS. The Trustworthy section in this report talks about the safety culture and the safety management system implemented that helps the organization to develop AVs in a responsible and dependable manner. The SMS approach is influenced by proven best practices in various safety-critical industries. Along with the SMS, Aurora has implemented human factors, fatigue management, and incident reporting to promote safe on-road testing and identify opportunities for growth. A dedicated safety culture program is in place that encourages employees to participate in safety programs. When comparing the works of Waymo and Aurora in terms of technical rigor, Waymo focuses on developing original methodologies and statistical assessments, while Aurora adapts existing approaches and provides more detail on their technical components.

Edge Case Research (ECR) published their version of a SCF titled "Open Autonomy Safety Case Framework (OASCF)" [15]. This SCF follows a similar three pillar structure as the AV-TEP SCF, with the pillars being Live It Right, Engineer It Right, and Operate It Right. The goal of the OASCF is to assist safety engineers in building a SCF that follows the structure prescribed UL4600 and establish arguments that the system development process and result is compliant with relevant industry standards. This paper talks about the SCF being an iterative document that is constantly evolving that takes into consideration new hazards and safety concerns that arise with every change to the system or post-deployment of the vehicle. The OASCF uses templated Safety Performance Indicators (SPIs) that help in assessing the credibility of the claims in being true or false in different situations. OASCF categorizes SPIs into behavioral and operational; with behavioral SPIs covering

autonomy functions such as perception, prediction, planning, and control, and operational SPIs covering SMS process and safety culture. Analyzing these SPIs helps in improving the safety performance of the AV, evaluate effectiveness of mitigation measures, increase confidence in the system, and identify unknown issues before they occur.

OASCF proposes a seven-step strategy that helps the AV developer in designing a safe system and provide evidence to claims in a comprehensive safety case, addressing the issue of disparate and incomplete view of risk among internal teams in an organization and regulators responsible for public safety. The Live It Right pillar can be compared to the SMS pillar of the AV-TEP SCF as it argues that the organization building and operating the AV is safe, and the Engineer It Right pillar to the Design Methods pillar of the AV-TEP SCF, where the autonomy technology is argued to be safe by design, and the development and safety assurance of the system is founded on rigorous engineering practices. The Operate It Right pillar argues that the AV is operated safely based on the contention that all operational controls are in place, and that the Quantitative Fault Tree Analysis (performed in the second pillar) describes rates of causal factors that occur during operation, forming the argumentation that the safety assurance process of SMS is effective.

Toyota published its Voluntary Safety Self-Assessment (VSSA) [16] for testing of SAE L4 and L5 vehicles in 2020. The elements covered in this VSSA follow the 12 safety elements described by NHTSA in its "Automated Driving Systems 2.0: A Vision for Safety" document in 2017. Within these 12 elements, Toyota covers its system safety design practices including identifying and mitigation of risks and safety performance of the system, determining ODD and associated sensor suite, behavioral competencies of the

ADS in handling complex situations and interactions, training and education of consumers and personnel, and compliance with regulations. This approach helps ensure Toyota's vehicles meet rigorous safety standards and incorporate robust safety measures, underscoring Toyota's commitment to advancing automotive safety across its lineups.

Torc Robotics, a company specializing in automated driving systems, is aiming to develop and deploy SAE Level 4 autonomy in trucks with the goal of transforming safety and efficiency in the trucking industry. Torc published its safety report [17] in the form of a voluntary safety self-assessment titled "Innovating Safety and Efficiency" in 2021. This report covers Torc's approach to developing and deploying safe ADSs, following the 12 safety elements defined by NHTSA. As of date of the report, Torc tests its ADS vehicles with the presence of an IFTD and elaborates on the rigorous training and selection protocol that is followed to hire and upskill the IFTD in taking over the controls of the vehicle in the case of ADS malfunction. The safety concept is categorized into 3 pillars which are vehicle, ADS, and IFTD, having individual requirements that need to be satisfied as part of the system safety design. Torc utilizes Automotive SPICE Process Assessment model to develop the hardware and software components and incorporate safety in production. The validation methodology follows a 5-step process starting with unit testing, and progressing through component, software and system integration testing, and system validation as the final step. Torc's mission to save lives drives its commitment to safety, with robust systems in place to identify and mitigate risks early on, ensuring safer deployment of their self-driving trucks.

Uber published its “Safety Guidelines for Autonomous Mobility and Delivery Providers on the Uber Platform” [18] outlining key safety features and considerations incorporated into the organization to ensure safe deployment and operation of AVs and delivery robots. The safety plan assessment framework is made up of three pillars which detail the safety considerations within the organization, vehicle platform, and operation of AV. The Safety of organization pillar follows a similar structure to the AVSC SMS in incorporating safety policy, SRM, safety assurance, and safety promotion. The Safety of the Autonomous Vehicle covers system design practices, behavioral competency of the AV, cybersecurity, and verification and validation procedures. The Safety of operation pillar discusses the possibility of failures and the necessity for continuous monitoring and improvements of the AV post-deployment. The key concepts highlighted in this section include remote assistance, vehicle misuse, incident response and data collection, fleet operation, and regulatory approvals. Additionally, in this document, Uber recommends relevant standards and industry best practices that could be used by an AV developer to achieve the safety requirements within each pillar. Incorporating AV safety principles such as real-time monitoring, robust testing protocols, and stringent partner reviews, Uber’s guidelines prioritize transparency and public trust, ensuring that every deployment meets strict safety requirements.

UL 4600 [19] published as a joint effort by the American National Standards Institute (ANSI) and Underwriters Laboratory (UL), helps an AV developer in ensuring the safety of fully automated vehicles, through a safety case-based approach to address the safety-critical concerns, and is structured in mandatory, required, highly recommended, and pitfall

prompt elements. The overall objective of the document is to demonstrate by documentation that an AV is acceptably safe to operate. UL 4600 is explicitly non-prescriptive regarding any overall design process or safety methodology. UL 4600 requires a three-step process to validate the safety of AV through making measurable safety claims, argue that the claim is true, and provide evidence to establish that the system performs as expected. This document is designed to be technology agnostic and can be used along with any other automotive related standards and best practices.

This literature review of safety case frameworks developed by various AV developers shows the importance of a documented approach to show the system performance in varying real-world conditions. It is a common practice for these safety case frameworks to adhere to and establish compliance with relevant industry standards and best practices. The following sections of the literature review cover such standards and best practices and categorizes them into the 3 pillars of the SCF based on the guidelines that talk about the organizational approaches to ensure systematic approach to safety for the SMS pillar, technical design and architectural choices for the Design Methods pillar, and verification and validation through simulation-based testing methods for the Testing pillar. The literature review additionally includes TRL section which explores existing frameworks and their importance in an organization.

Safety Management Systems Pillar

The aviation industry has been the pinnacle of safety in the transportation industry, bringing in revolutionary methodologies and practices that have been adopted by industries in various disciplines. Numerous passenger aircraft accidents led to the implementation of the Safety Management System (SMS) by the International Civil Aviation Organization (ICAO) in its publication entitled "Safety Management Manual (SMM)" in the 1990s. ICAO issued a comprehensive document titled "Safety Management Manual (SMM) [3]" to guide states and aviation service providers in developing a safety management system. It promotes a systematic, proactive, and data-driven approach to managing risks and improving safety performance across the global aviation system. The SMM introduces the fundamentals of safety management covering concepts such as *accident causation* elaborating on the famous 'Swiss cheese' model [20] which illustrates that accidents involve successive breaches of multiple system defenses, *practical drift*, a theory used in aviation to understand how the baseline performance of a system "drifts away" from original specification due to unforeseen circumstances, *SHELL model*, a conceptual tool used to analyze the interaction of multiple system components. Following the fundamentals, the SMM dives into the elements of a healthy SMS which would include a robust safety reporting and investigation program, management of change, collecting safety data and analyzing it to identify hazards and develop control measures, identifying SPIs. The SMM categorizes the safety culture into organizational, professional, national, reporting cultures characterized by the beliefs, values, biases and their resultant behavior that are shared by members of the organization. ICAO defines an SMS as a systematic approach to managing safety, including the necessary organizational structures,

accountabilities, policies, and procedures. The FAA SMS framework has four key components: Safety Policy and Objectives, Safety Risk Management, Safety Assurance, and Safety Promotion. The SMM provides a pathway to implement an SMS which begins with a system description, outlining the existing operational processes and structures. Next, integration of management systems ensures alignment with other frameworks like quality and environmental management. A gap analysis then identifies deficiencies in current practices. The SMM is a dynamic framework, recognizing that safety management must evolve with technological advancements, operational changes, and safety risks. Although the introduction of SMS in the aviation industry was late, as years went by, the SMS gained traction among different industries and evolved into a robust top-down risk-based approach, integrating technology, data analysis, and safety measures to mitigate identified hazards.

Aviation industry has seen dramatic advancements in safety performance since the inception of the Federal Aviation Administration (FAA) in 1958. Since then, the responsibilities for the FAA have grown and now include establishing flight rules and regulations, licensing pilots, and developing manufacturing standards. Following the Aviation Deregulation Act, the aviation industry has seen key transformations in business and safety requirements. Following the introduction of the concept of SMS in the aviation industry by ICAO, the FAA published its version of a Safety Management System (SMS) [21] providing guidelines and recommendations encompassed within the four components namely Safety Policy, Safety Risk Management (SRM), Safety Assurance, and Safety Promotion (shown in Figure 4). Safety Policy establishes a documented approach covering

the organization's commitment to safety, which defines safety objectives, accountability, and responsibility of the organization and employees. The SRM component ensures that appropriate levels of hazard assessment are performed and associated risks are identified, steps to develop mitigation strategies to address the identified hazards, and track identified hazards and monitor implemented safety risk controls. In the Safety Assurance, the organization is expected to assure that the mitigation strategies are functioning as expected, and it achieves the intended safety performance targets. To perform this, the FAA SMS recommends collecting operational data and analyzing the safety performance of the system based on observed Safety Performance Indicators (SPIs). Following the analysis, FAA suggests implementing corrective measures to address new hazards. The Safety Promotion component is a combination of training and communication of safety information to support the implementation and operation of an SMS.

To summarize, the FAA SMS is intended to help participants in the aviation industry in systematically identifying and addressing hazards and risks, ensure product is developed in compliance with industry best practices and standards, saving running cost by proactively managing risks and preventing accidents, fostering a culture of safety and continuous improvement. The FAA SMS is mandatory for various sectors of the aviation industry, including commercial airlines, and airports, ensuring that safety is prioritized across the entire industry.

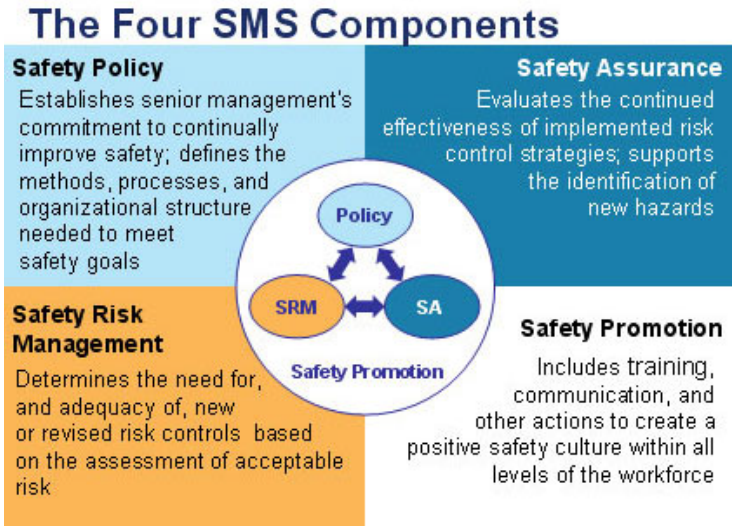

Figure 4: Elements of FAA SMS
Source: [21]

Adapting the learnings and findings from the aviation industry, AVSC published an SMS Information Report (AVSC, 2021) [8] to provide the AV industry with an initial SMS framework which the authors believe can serve as a valid starting point for any AV company. It was intended to provide basic guidance to organizations on elements they might use to assess and manage safety risks, evaluate effectiveness of risk control strategies, promote healthy safety culture, and establish safety policies and objectives in the organization during the testing and evaluation phase of ADS development of SAE Level 4 and 5 vehicles. The AVSC SMS framework is based on FAA SMS framework and comprises four components: Safety Policy and Objectives (SPO), Safety Risk Management (SRM), Safety Assurance (SA), and Safety Promotion (SP). A similar structure is adopted for the SMS pillar of this thesis work.

Identifying the need for a comprehensive SMS framework that harmonizes standards, best practices, and its associated learnings and suggestions from various industries, and tailor it

to the AV industry, the SAE On-Road Automated Driving (ORAD) Committee established a Task Force to develop a Recommended Practice [22] which builds upon the AVSC SMS Information Report. As the AV industry is at its nascent stages without the backing of a reliable regulatory framework, utilizing just the aviation SMS is not sufficient to handle the relatively novel fusion of technologies including automotive technology, AI, and operations. This recommended practice emphasizes safety culture management, a topic which is often overlooked in other widely used industry best practices, that has an influence over safety structures and expectations of the overall SMS. It later delves into subsequent elements of the framework which include Safety Planning, Safety Governance, Safety Risk Management, and Safety Lifecycle Management, covering all phases in the AV value chain that include SAE Level 3-5 AVs.

Outside of the aviation and automotive industry, academic research efforts have been actively exploring methods of implementing SMS and promote the importance of incorporating a healthy safety culture within the organization. Safety Management Systems and their Origins [23] is a book by Corinne Bieder, published in 2023, that takes a deep dive into the concept of SMS introduced in the aviation industry and presents three different perspectives of the SMS currently prevalent in the industry. The first chapter gives a detailed introduction to SMS and the associated history. The book uses a multidisciplinary approach combining document analysis, literature review, historical research, and interview. The first perspective describes SMS as a new approach to managing safety in a proactive manner and considering the role of management in achieving safe performance. It introduces the four main components of the SMS as defined

in ICAO which includes Safety Policy, Safety Risk Management, Safety Assurance, and Safety Promotion, which covers the processes of developing safety policies, appointing key personnel, conducting hazard analysis and risk assessment, setting safety performance indicators, and promoting a healthy safety culture. The second perspective of SMS is described to be a more critical version of SMS as a safety approach highlighting the conceptual, methodological, and practical limitations of implementing an SMS. The book identifies key issues to be oversimplification of complex systems and practices, and overreliance on quantitative indicators. The third perspective of SMS presents a socio-historic viewpoint of SMS by considering overall context, highlighting the emergence of SMS as a response to technological evolution, fragmented enterprises, and economic pressures. The author describes that the concept of SMS was influenced by different schools of thought in safety risk management, and the concept spread through various industries and countries through exchanges between communities and regulatory initiatives. As a conclusion, the author recommends possible directions the SMS could head in the future which includes suggestions such as incorporating a broader view of risk management, acknowledging uncertainties and its impact, consider safety in the broader organizational context, foster synergies between safety and organizational goals, and the inclusion of concepts of social science and organizational psychology that were absent in the basic aviation model.

Several students from various parts of the world have taken up the effort of analyzing the pros and cons of existing SMS frameworks and assessing relevant audit tools and published their work as part of their thesis. The thesis work [24] by Kathleen Fox, examines how the

introduction of SMS is affecting the way safety risks are managed in the transportation industry. The thesis is structured as a transcript of an interview/conversation with people from the industry and includes case studies. It primarily focuses on management decision-making, mentioning that it is difficult to identify accidents proactively, rather it is easier in retrospect to identify contributing factors. This work notices that the major contributing factor for organizations to settle for acceptable performance is the pressure towards cost-effectiveness. The methodology employed in this work includes identifying shortcomings in risk management practices that may have contributed to an accident, as well as changes in how managers position themselves relative to their decisions, risk, and management. This thesis talks about how the motivation for adopting an SMS was prompted by high profile, tragic accidents and realization that organizations need to find a better way to prevent them, and that traditional approaches to safety management was insufficient. When comparing the transportation industry with Highly Reliable Organizations (HRO) [25], the author states that the leaders in the latter prioritized safety, reliability, incorporating significant levels of redundancy, strong organizational culture, and continuous training and improvement to cater to evolving safety concerns, while providing sources that discuss the key factors that would encourage a good safety culture within the organization.

[26] is another thesis work that focuses on evaluating the audit tools in assessing the effectiveness of the implementation of SMS and measures the reliability of these tools in mitigating risks. The results observed by the author show that safety audit tools are helpful, but do not ensure reliable and valid audit results. The author puts forward three hypotheses statements which are 1) A reasonably high reliability can be achieved when auditors are

familiar with audit tools, national legislations, company's culture, 2) A properly constructed audit tool can improve the reliability of audit results, and 3) Safety activities are at a higher level in the USA compared to Finnish companies using D&S assessment tools. The author discusses the importance of evaluating the SMS through procedures such as measurement of safety performance, safety audits, and management reviews. This thesis work highlights the roles/goals for everyone in an organization to achieve safety objectives. These goals are given based on relevant industry standards. It describes a few safety audit tools and explains the pros and cons of each. Prospects for future work as mentioned by the author include developing a new SMS by integrating health and safety, getting a better understanding of safety culture with the overall aim of achieving zero accidents.

The author of [27] develops an SMS audit tool by triangulating information from industry best practices and interviews with fleet managers and drivers regarding fleet safety practices. The audit tool covers 5 topics which are 1) management, systems and processes, 2) monitoring and assessment, 3) employee recruitment, training, and education, 4) vehicle technology, selection and maintenance, and 5) vehicle journeys. Each of these categories has one to three sub-categories which are individually scored from a scale of 1 to 4, with 1 being the highest level of safety performance and 4 being the least. These individual scores are aggregated to provide a total score out of 36, which can tell how well the organization is performing. To assess the effectiveness and validate the properties of the audit tool, the author emphasizes the application of the tool by fleet safety researchers, ensuring that the tool stays relevant and effective in improving fleet safety management practices. Overall, the audit tool represents a significant step towards enhancing safety management in light

vehicle fleets by providing a systematic approach to evaluate and improve practices within organizations.

While it is important to have an SMS within the organization to identify and mitigate operational risks, it is equally important to have a Quality Management System (QMS) that ensures the consistency and reliability of products and services. This standard [28] from the International Organization for Standardization (ISO) focuses on the quality management and control of products manufactured, and states that the impact of quality can extend beyond customer satisfaction, and it can have a direct impact on the organization's reputation, showing the importance of having a reliable quality management system (QMS) within the organization. This standard when implemented in conjunction with the International Automotive Task Force (IATF) 16949 [29] standard helps in enhancing the quality management practices for the automotive industry. The QMS process consists of activities that help the organization identify the goals and develop methods of achieving these goals, with the focus of meeting customer requirements and exceeding expectations. The key principles covered in this document include customer focus, leadership goals, importance of engaging people within the organization, process approach to deliver consistent and reliable results, improvement strategies, evidence-based decision making, and relationship management visually represented in Figure 5. ISO 9001 is applicable to organizations of any size, in any industry, and serves as a framework for building a culture of quality, fostering continuous improvement, and delivering long-term value to customers and stakeholders. The QMS model proposed by this standard suggests that the model be flexible and dynamic considering changes and improvements over a

period. The standard also promotes risk management and strong stakeholder relationships, helping organizations improve efficiency, meet regulatory requirements, and enhance overall performance. Together, the SMS and QMS form the backbone of resilient, high-performing organizations where safety and quality go hand in hand together, they form the backbone of resilient, high-performing organizations where safety and quality go hand in hand. The holistic integration of SMS and QMS has various beneficial functional overlaps such as risk assessment and management, continuous improvement, documentation of procedures and processes, employee training, competence management, and leadership commitment to safety.

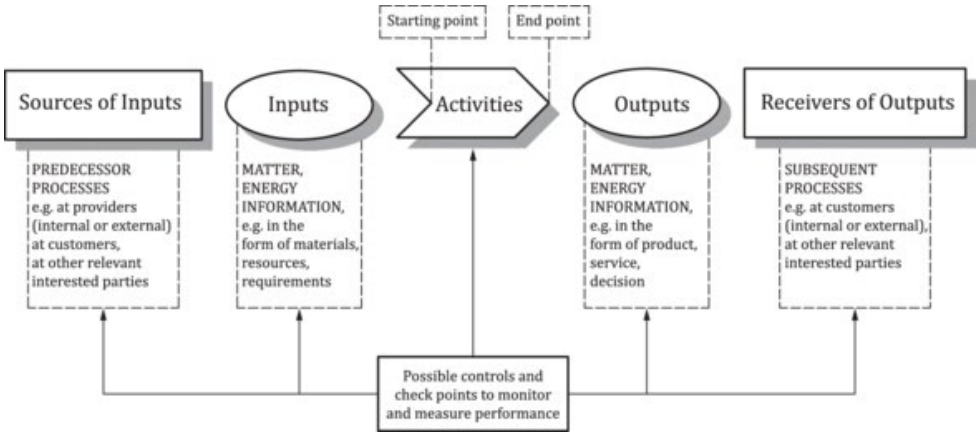

Figure 5: Visual representation of QMS elements

Source: [28]

Design Methods Pillar

Good design and engineering practices ensure the organization maintains safety standards at every stage of the product development process. The concept of safety by design emphasizes early identification and mitigation of hazards, instead of addressing them post-deployment. This proactive approach consists of rigorous engineering practices such as fault-tolerant design, redundancies, and hazard analysis methodologies like Failure Modes and Effects Analysis (FMEA), Fault Tree Analysis (FTA), and System-Theoretic Process Analysis (STPA) as a control-centric risk assessment model to help the organization in significantly reducing system failures that could be catastrophic.

SAE J3016 [5], first published in 2014, titled "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles" provides overarching standardized definitions of terminologies and levels of automation used in the world of ADS-equipped vehicle. This taxonomy can be seen used widely in this thesis work as well. This document defines 6 levels of automation starting from no automation and progressing to full automation. SAE J3016 is critical for creating a common understanding of automated driving capabilities and limitations, aiding in the development of regulatory frameworks, vehicle design, and public awareness. The document takes a technology-neutral stance and does not dictate specific technology choices for achieving automation. The focus is on defining what the system can do, not how it should be implemented, allowing for flexibility in how ADS technologies evolve, with the aim of providing a robust framework for categorizing and understanding the scope, capabilities, and limitations of driving automation systems.

Multiple industry standards and best practices exist that guide the engineers in adhering to safety by design principles. One such standard which is widely used in the automotive industry is the ISO 26262: Road Vehicles- Functional Safety [30] standard that helps in achieving Functional Safety by addressing hazards and risks at the component level. This standard defines Functional Safety as the "Absence of Unreasonable Risk (AUR) due to hazards caused by malfunctioning behavior of electric/electronic systems". AUR is defined as the absence of "risk judged to be unacceptable in a certain context according to valid societal moral concepts". This standard consists of twelve parts that provide guidelines for identifying, categorizing, and addressing hazards caused by malfunction in safety-related electrical or electronic systems in passenger vehicles over the lifecycle of the system. The goal of this standard is to avoid/mitigate the effects of system failures to ensure 'Functional Safety' through achieving the AUR categorize hazards by risk levels by incorporating Automotive Safety Integrity Levels (ASILs) classification and provide a reliable framework to ensure safety through development and validation.

Out of the twelve parts available, ten are applicable to the automotive industry and they are:

i. ISO 26262-1: Road vehicles – Functional safety – *Part 1: Vocabulary:* covers terms and definitions [30]
ii. ISO 26262-2: Road vehicles – Functional safety – *Part 2: Management of functional safety*: provides overarching guide focusing on management of safety requirements [31]

iii. ISO 26262-3: Road vehicles – Functional safety – *Part 3: Concept phase*: focuses on the concept phase of product development [32]

iv. ISO 26262-4: Road vehicles – Functional safety – *Part 4: Product development at the system level*: covers system level development [33]

v. ISO 26262-5: Road vehicles – Functional safety – *Part 5: Product development at the hardware level*: focuses on hardware aspects of system design and implementation [34]

vi. ISO 26262-6: Road vehicles – Functional safety – *Part 6: Product development at the software level*: focuses on software aspects of system design and implementation [35]

vii. ISO 26262-7: Road vehicles – Functional safety – *Part 7: Production, operation, service, and decommissioning*: Details requirements for system production, operation, and installation [36]

viii. ISO 26262-8: Road vehicles – Functional safety – *Part 8: Supporting processes*: defines requirements for processes that support development effort, including change management, V&V, and tool qualification [37]

ix. ISO 26262-9: Road vehicles – Functional safety – *Part 9: Automotive Safety Integrity Level (ASIL)-oriented and safety-oriented analysis*: requirements and guidance regarding safety analyses, and ASIL [38]

x. ISO 26262-10: Road vehicles – Functional safety – *Part 10: Guideline on ISO 26262*: guidance on applying ISO 26262 [39]

The ISO 26262 standard uses V-model (shown in Figure 6), a lifecycle process model that has been established for the development of complex safety-critical systems. The V-model is designed in such a way that ensures all aspects of the project are considered, and that each stage of development builds upon the previous stage and emphasizes verification and validation.

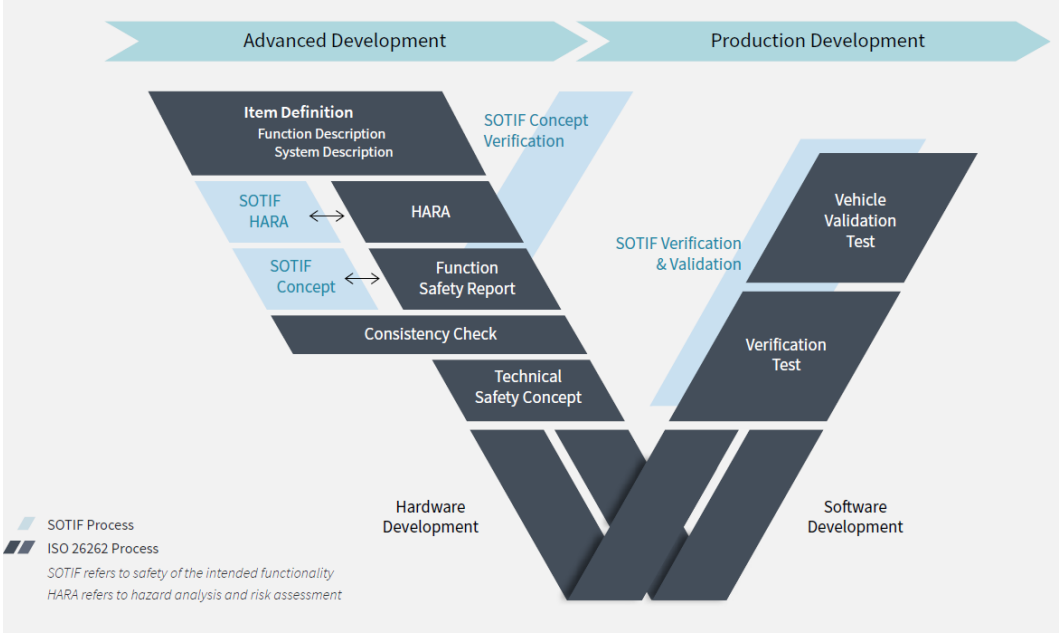

Figure 6: V-model product development

Source: [17]

While ISO 26262 helps the AV developer in addressing hardware and software failures with the aim of achieving functional safety, ISO 21448 [40] complements it by covering the Safety of the Intended Functionality (SOTIF) of the system. This standard brings up the question of whether the system's intended behavior is safe in the likely conditions of its usage. It offers guidelines for assessing hazards that may arise from insufficiencies in

the system's intended functionality or from foreseeable misuse of the system. SOTIF is applicable to SAE L1 and L2 systems and can be applied to higher levels with additional measures. To demonstrate compliance with this standard, the objectives can be listed and provide arguments that the objectives have been achieved. This standard categorizes scenarios that are within the use case of the system into four sections which are known safe, known unsafe, unknown unsafe, and unknown safe scenarios, as shown in Figure 7. An assumption made by ISO 21448 is that the electrical/electronic hardware failures and systematic faults of the systems are addressed using ISO 26262. ISO 21448 is crucial in the development of automated and assisted driving systems, ensuring that these technologies can safely handle complex, real-world driving situations. It plays a vital role in advancing the safety of systems that rely on perception, decision-making algorithms, and sensors, making it a cornerstone for manufacturers developing AVs.

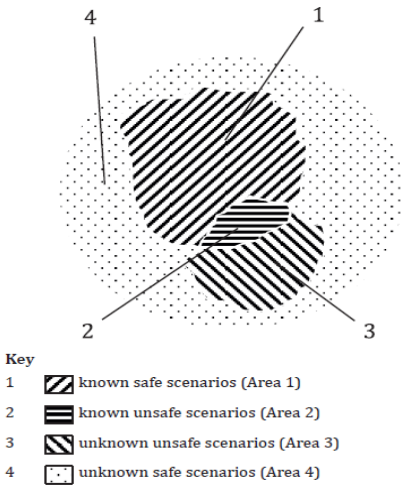

Figure 6: Scenario categorization in ISO 21448

Source: [40]

In the automotive industry, best practices and information reports developed by organizations such as SAE and AVSC, help engineers and developers in addressing safety-critical concerns in their system design, giving them guidance in the absence of regulation. This information report [41] titled “Taxonomy and Definition of Safety Principles for Automated Driving System (ADS)” classifies and defines a set of 18 safety principles that are to be considered during the development and deployment of ADS. It accounts for the unique challenges posed by large vehicles and commercial transportation while aligning with the broader SAE J3016 framework for automation in all vehicles. For each of the principles, this document provides suggestions and recommendations that should be considered to satisfy each respective principle. The key safety principles covered in this document include safety performance criteria, ODD selection, failure management and fallback driver, human-machine interaction (HMI), and verification and validation of safety.

SAE J3171 Identifying Automated Driving Systems-Dedicated Vehicles (ADS-DVs) Passenger Issues for Persons with Disabilities [42] addresses the issues with ADS-DV design for people who were not able to secure a driver’s license due to some form of impairment. This document carries out a review of the requirements for obtaining a driver’s license and conducts stakeholder interview to identify ways of overcoming this obstacle. The document highlights the importance of inclusive design to ensure that automated vehicles are accessible, safe, and user-friendly for all passengers, including those with disabilities. The document proposes design considerations to make the ADS-DVs more equitable, flexible, requiring low effort, simple, and intuitive to use by passengers with

disabilities. Based on the literature review and interviews, the document identifies issues that are unique to ADS-DV and passengers with visual impairments, wheeled mobility device users, and other forms of mobility/sensory related disabilities. The document recommends a communication channel to be established and to offer real-time support to passengers if necessary. It additionally suggests that the ADS-DVs offer customizable features to accommodate individual disabilities, ensuring a comfortable and independent travel experience.

AVSC Information Report for Change Risk Management [43] published in 2023 covers methods of identifying and handling new hazards and associated risks that may arise after deployment of the ADS due to planned or unplanned changes. The process of Change Risk Management (CRM) is iterative, meaning for every new change that has a potential to introduce new hazards, CRM needs to be performed, and this process is shown in Figure 8. The CRM process is applicable to both the SMS and Design Methods pillars of the SCF. In the SMS pillar, the CRM is used to address the process management aspect of post-deployment changes, and in the Design Methods pillar, it is used to address the risk that arises due to technical changes to the system. This document recommends widely used analysis methods such as model-based systems engineering, bowtie model, event sequence diagram, Subject Matter Expert (SME) brainstorming to perform impact analysis and identify CRM triggers. Following the impact analysis, a hazard analysis is performed, which later feeds into the risk analysis process to determine the severity and likelihood of the hazardous event. Based on the safety risk assessment, developers can identify, implement, and validate safety measures that is capable to mitigate the risks, and these

measures are prioritized by their influence in managing the risk. As a final step, the level of residual risk is measured, and if this value is unacceptable, the change is reverted and/or the safety measures applied are rejected.

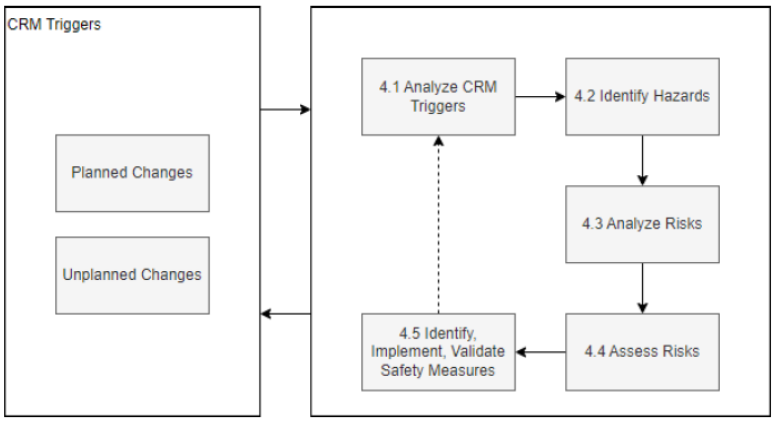

Figure 7: Process of Change Risk Management

Source: [43]

The AVSC Best Practice for Continuous Monitoring and Improvement after Deployment [44] focuses on the analysis of safety performance data collected during the operation of the ADS, to help in identifying new hazards and risks that arise due to changes in the environment. The continuous monitoring section in this document covers sample data sources that can be collected and methods of collected such data, and additionally includes examples of possible irregularities and changes in the environment that could lead to such irregularities. Following the monitoring stage, the document recommends continuous improvement in the system performance by minimizing or mitigating irregularities resulting in a confidence boost in the design assumptions. The continuous improvement sections highlight suggested responses that can be developed, tested, and implemented to mitigate the hazards identified while monitoring the safety performance data. This best

practice proposes a framework covering the flow of the working process for analyzing the performance data and developing improvement strategies. It suggests that the changes to the environment can be both known and unknown, and it is important as a developer to identify these changes. Steps to do so along with a comprehensive workflow framework are covered in the document. Compared to the AVSC CRM document, this document focuses on monitoring the real-world performance of the vehicle rather than evaluating risks associated with proposed changes to the system. This document deals with broad, ongoing safety assurance processes, and feedback loops based on operational data, ensuring long-term safety, reliability and operational efficiency of the ADS.

The NHTSA, the U.S. government agency responsible for ensuring road safety by setting and enforcing vehicle performance standards, and providing oversight on regulations, published a comprehensive framework in its document titled "Automated Driving Systems 2.0: A Vision for Safety" [45] that provides guidance to AV developers and states for the safe development, deployment, and operation of ADS vehicles. This document contains twelve safety design elements such as system safety, human-machine interface, crashworthiness, and post-crash behavior. The document additionally provides safety goals and approaches that could help the AV developer in satisfying these safety elements. It promotes best practices and a safety-first approach during the design and testing of ADS technologies. In this effort, NHTSA offers a framework for states to create consistent regulations and policies for testing and deploying automated vehicles, encouraging collaboration between local governments and industry stakeholders to maintain safety while advancing innovation.

Testing Pillar

Testing has always been a crucial step in the process of developing a new product, or when making modifications to an existing product. This step helps engineers validate the performance and efficiency of the product, the organization to have confidence that new features have AUR and compare the vehicle performance with the initial estimations defined in the concept phase and ensure adherence to relevant standards and stakeholder requirements. When it comes to AVs, different testing methodologies exist. A widely used method for the testing of AVs is by driving it with or without the help of an FTD[1] for thousands of miles around public roads or closed-course test tracks and collecting data. This data is then analyzed and compared with expected performance criteria set by the organization. When we compare the number of miles driven and the accident or failures encountered in those miles, the percentage is very less, which makes this method an unviable option to test out the vehicle prior to releasing it on public roads [46]. Based on statistical methods proposed in [46], it takes approximately 8.8 million miles involving a fleet of 100 AVs to test the AV in all possible test case. This is practically and logistically impossible to implement, as it takes about 400 years to complete the test with cars being driven around the clock. Contrastingly, on the practical side of AV deployment, Waymo has driven 40 million miles [47] showing significant reduction in air-bag deployment, injury-causing accidents, and police-reported crashes in the first 25 million miles compared to a human-driven vehicle for the same distance.

[1] AV is tested in the presence of an IFTD in closed-course circuit following satisfactory behavioral competency in simulation-based testing. The organization transitions to testing the AV without an IFTD on public roads when a satisfactory behavioral competency and performance is observed in closed-course testing.

With the logistical limitations involved in choosing public road testing as the first step in the validation process, simulation-based testing is the go-to option for AV developers as it helps them address major safety concerns and mitigate them before advancing to closed-course testing, with the benefits of being cost-effective and flexible in testing the vehicle in a multitude of edge cases. Running simulation scenarios generates data that can be collected and analyzed to understand how the vehicle behaves and maneuvers through the scenario. Wishart, et.al, proposed a set of metrics called Operational Safety Assessment (OSA) metrics [48] that uses the data collected from the simulation to score the vehicle's performance and operational safety based on individual score in a set of metrics. This work categorizes the leading and lagging metrics into black box, grey box, and white box metrics depending on the level of ADS data required to calculate the metric. With black box requiring no ADS data and white box requiring sufficient ADS data to present satisfactory results. This work was later expanded to include a wide variety of metrics with the aim of providing a more comprehensive scoring methodology, with the OSA metrics renamed to Driving Assessment (DA) metrics and to be published as a best practices document in SAE. Along with works like this, SAE has relevant recommended practices documents such as SAE J3018 [49] titled "Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles" first published in 2015 and revised in 2020, provides a framework for testing the performance and competency of the prototype ADS equipped vehicle in various scenarios, and provides guidance for training the In-vehicle Fallback Test Driver (IFTD). This recommended practice divides the testing of prototype ADS-equipped vehicle into early-stage prototype and late-stage prototype

depending on if the system prototype has been demonstrated in operational environment and if the actual system completed and qualified through test and demonstration. With respect to IFTD training, this document recommends progressing through different training steps such as classroom training, simulation training, test track training, and on-road training, additionally explaining the importance of workload management for IFTDs. SAE J3018 also recommends following a safety protocol before the on-road testing of ADS-equipped vehicle, and it includes selection of test routes, pre-trip checklist and test vehicle inspection, IFTD state monitoring, post-trip de-briefing, shadow mode testing, and incident response protocol. SAE J3018 is designed to foster the safe development of automated driving technologies by promoting best practices for real-world testing.

ISO 34502 [50] provides guidance for evaluating the performance of AV by establishing a scenario-based testing framework that includes a comprehensive list of hazardous scenarios the vehicle could encounter. This standard categorizes the scenarios into three phases: perception, decision, and control. For the perception phase, the standard considers sensor limitations and failures, environmental conditions that affect sensor perception, and unexpected obstacles in the path of the test vehicle. In the decision phase, the difficulty and the ability of the AV to maneuver through complex traffic situations, how the vehicle reacts to unexpected behavior of other road users, and challenging road conditions and structures are addressed. As the AV navigates in complex scenarios, it dynamically changes the driving path and decisions based on real-time information. Extensive research is undertaken in this topic to overcome the challenges that include latency, uncertainty, and complexity in the interactions by developing robust real-time motion planning algorithms

[51]. The control phase provides test scenarios that address vehicle dynamics and handling, actuator limitations and failures, and unexpected changes in vehicle performance. Once a sufficient level of behavioral competency is achieved with the simulation-based testing, the test vehicle progresses to other modes of testing such as closed-course testing and public road testing, and ultimately resulting in deployment.

Technology Readiness Level (TRL)

Technology Readiness Level (TRL) is a process developed by NASA [52] to assess the maturity of technology throughout the development lifecycle of the product. It consists of nine levels starting from TRL 1 which encompasses basic research principles to TRL 9 where the product is deployed in its usage environment. This process helps the developer identify, assess, and mitigate all hazards and satisfy all requirements before transitioning to the next technology level. Numerous publications on this topic discuss the benefits and drawbacks of implementing TRL, and some have also detailed steps for conducting an accurate TRL assessment and implementing TRL in the organization.

Technology readiness assessments: A retrospective [53] discusses the process of conducting a Technology Readiness Assessment (TRA) and talks about the history of TRAs and TRLs. The authors identify key data that can be beneficial in conducting an effective TRA which include performance objectives, TRLs, research and development degree of difficulty which essentially attempts to get an understanding of remaining hurdles and projected uncertainties. The paper additionally mentions challenges that could form obstacles when conducting the TRA such as achieving the same level of technology

maturity across all subsystems and components, establishing the right metrics, and the need for technology management figure of merit that measures the riskiness of new technology. Some advantages of incorporating a formal TRA over individual evaluation are consistency in the results, enhancement in communication and collaboration, improved risk management due to the implementation of performance metrics to gauge uncertainty. Technology Readiness Levels: A White Paper [54] provides a detailed description of each level of the TRL. Technology Readiness Levels at 40: A Study of State-of-the-Art Use, Challenges, and Opportunities [55] aims to expand on the 30 years retrospective paper [53] and provides more detailed observations based on evidence from diverse industries. This paper identifies that implementing TRL will have substantial benefits such as incorporating a shared understanding of technology maturity and relevant risks associated with each level of the product development, and possibility of TRLs providing a systematic approach and model for assessing the maturity level of the technology.

US Department of Transportation (USDOT) published a TRL guidebook [56] that provides a comprehensive guide to understanding and conducting TRL assessment. The guide explains TRL scale, its purpose, its use cases, and a step-by-step process for conducting an assessment. USDOT identifies key advantages of implementing TRLs which are ensuring a common language of technology maturity assessment across the organization and identify gaps that are addressed before proceeding to the next step. USDOT advises against the usage of TRLs for risk analysis, or as the only source of indicator to validate the safety performance. Following the successful assessment, the results are used to make strategic funding decisions, research planning, and identify potential collaborators. [57] explores

challenges and shortcomings with implementing TRLs. It conducts a 3-stage study, starting with interviews with practitioners, and collecting data. This data upon analysis, revealed 15 key challenges which were later categorized into system complexity, planning and review, and assessment validity. Some suggested improvements include expanding the TRL definition to include steps for integration, develop a standardized Integration Readiness Level (IRL) to assess the maturity of interfaces, and establish a clear guidance for assessing modified technologies.

Literature Review: Thoughts and Discussion

After an extensive literature review of existing standards and best practices from various industries, several important themes emerge, each contributing to a comprehensive and structured approach to ensure the safety of AVs

Various AV developers have adopted the concept of SCF as a method of showing a body of evidence demonstrating safety at every stage of product development and operation. These documents can be seen as living documents that adapt and evolve as the AV progresses through design, testing, and real-world deployment. By centering these SCFs on strong industry standards and best practices, developers can show that they have addressed foreseeable risks and ensure compliance with relevant standards.

Standards and best practices relevant to implementing an SMS framework brings an important layer of integration of the organizational practices and technical implementation, emphasizing routine risk assessment, change management, personnel training and competence management, and a structure approach for safety assurance.

A very important dynamic exists between the design and testing related standards, in a way that system design standards provide guidelines to build a system with fault tolerance, redundancy, and resilience at each stage of product development, and testing best practices incorporate rigorous testing methodologies to validate these components in real-world scenarios.

The 3 pillars of the SCF work together in way that when testing reveals new safety considerations or edge cases, SMS framework ensure these insights lead to modifications in both design and operational practices, forming a continuous improvement loop that ensures the safety of the AV throughout its lifecycle.

In essence, these observations highlight the fact that no single standard can help in achieving AV safety. An adaptive safety case, proactive SMS processes, robust design principles, and thorough testing must all be used in tandem for true safety, ensuring compliance with varies standards. As AV technologies become an active component of our transportation system, this integrated strategy not only prepares AVs for real-world deployment but also supports industry efforts to achieve and maintain a strong safety balance.

METHODOLOGY

Safety Management System (SMS) Pillar

The SMS pillar is designed to support organizational safety systematically and comprehensively by combining safety principles, processes, and practices to enhance organizational decisions based on safety risk. The SMS process involves defining the safety objectives of the organization, identifying potential safety hazards, evaluating, and managing safety risks, implementing controls and mitigations to address those risks, promoting and educating the users and stakeholders on the importance of safety and the different functionalities of the system, encompassed within the components of Safety Policy and Objectives, Safety Risk Management (SRM), Safety Assurance, and Safety Promotion.

SMS has been a reliable source of assurance that ensures the system is developed with safety as the priority. Following the implementation in the aviation industry, multiple different industries have adopted this methodology to cater to the safety-critical needs. When it comes to the AV industry, this is a relatively new industry when compared with aviation, nuclear, and Petro-chemical. Hence, this industry lacks the strong and robust regulatory frameworks and guidelines that are present in other industries.

The AV-TEP SMS pillar harmonizes learnings and guidelines from various industries and establishes an SMS approach that is tailored for the AV industry that caters to the development, operation, and personnel level safety considerations and recommendations, covering the entire lifecycle of the product.

The **Safety Policy** component within the SMS pillar addresses concepts such as defining the context of the organization, identifying and creating safety roles within the organization with specific responsibilities, and establishes a communication channel that encourages employees to communicate safety related concerns and report occurrences of hazardous events without the fear of reprisal. The safety policy component creates a broad and comprehensive understanding of what the organization is doing and what are the goals and objectives, the organization aims to achieve with the help of a talented group of personnel selected through rigorous competence assessment. This component encourages employees to work cooperatively as a team towards achieving the shared goal of safety and discourages unacceptable behavior within the organization. It addresses the need to have constant engagement with stakeholders to share progress and ensure that the system development is aligned with the stakeholder requirements. The organization is encouraged to document safety policies and objectives, planned safety activities, proposed system safety approach, and establish a statement of compliance showing adherence to relevant safety standards and stakeholder requirements.

The **SRM** component is designed to manage risks through a process consisting of describing the system, identifying hazards, and assessing, analyzing, and controlling risks based on safety risk assessment. Safety risks are conceptually assessed as acceptable, tolerable, or intolerable. As a first step in the SRM process, a detailed description of the system is provided which covers the concepts and technologies of the system and subsystems, characteristics of the Operational Design Domain (ODD) including the system boundaries and functional range, and additionally describes system limitations and

dependencies, to give the user and stakeholder a broader understanding of the system functionalities and limitations. A Hazard Analysis and Risk Assessment (HARA) process is performed to systematically identify hazards in the system, and evaluate the associated risks, and develop mitigation strategies that address the identified hazards, essentially minimizing the number of hazards to an acceptably low level. Popular hazard analysis techniques such as Failure Modes and Effects Analysis (FMEA), Fault Tree Analysis (FTA), Hazard and Operability Study (HAZOP), Preliminary Hazard Analysis (PHA). Following successful mitigation of identified hazards, a verification and validation procedure is performed to assess the effectiveness of the mitigation strategies and validate to ensure the risk is within the acceptable level which is defined in consultation with the stakeholders. All these processes and strategies are documented for future reference and approved by the dedicated safety personnel.

Safety Assurance component ensures the implementation and effectiveness of safety risk mitigation and checks if the organization meets or exceeds its safety objectives through the collection and analysis of safety performance data. This process involves collecting data from deployment/testing, and analyzing this data to identify safety performance discrepancies, and implement mitigation strategies to ensure the system operates within acceptable safety parameters. It also evaluates the efficacy of implemented safety risk controls and assesses their sustained effectiveness over time, ensuring that the strategies continue to mitigate risks as intended, allowing for timely updates to safety protocols. Regular audits and evaluations are conducted to ensure adherence to relevant standards, policies, and best practices, reinforcing a culture of continuous safety improvement.

Confirmation reviews are conducted for various critical work products to ensure thorough safety evaluations. These include impact analysis at the item level, hazard analysis and risk assessment, and reviews of key safety documentation such as the safety plan, functional safety concept, and technical safety concept. Additionally, integration and test strategies, safety validation specifications, safety analyses, and dependent failure analyses undergo confirmation reviews. Finally, the review of the overall safety case ensures that all aspects of functional safety have been adequately addressed and validated.

The **Safety Promotion** component helps the organization conduct education and training programs that promote safety awareness among employees. This approach is aimed at fostering a positive safety culture throughout all levels and roles of the organization. Having a good safety culture helps ensuring a good overall SMS that is reliable. Effective communication across all levels of the organization is essential for promoting a strong safety culture. This includes clearly conveying the organization's safety objectives and keeping employees informed about the status of safety-related activities and events. Promoting positive safety leadership throughout the organization helps reinforce this culture. Additionally, the organization’s safety promotion efforts must ensure that safety objectives are communicated not only to internal employees but also to vendor employees, fostering a unified approach to safety across all stakeholders.

This SMS pillar also incorporates a requirement to establish a safety culture. **Safety Culture** is defined as the way safety is perceived, valued, and prioritized in an organization. Safety Culture is characterized by the beliefs, values, biases, and resultant behavior that are shared by members of a society/organization. A healthy safety culture

relies on a high degree of trust and respect between personnel and management and must therefore be created and supported at the senior management level. The definition of safety culture should encompass the roles of both safety and non-safety staff, management, suppliers, and all lifecycle participants in ensuring overall safety. It emphasizes compliance with regulatory requirements while embracing safety as a core organizational value. A strong safety culture encourages employee engagement and ownership by promoting non-punitive reporting systems ensuring timely responses to safety concerns, and punitive actions for willful disregard caused by problems in the safety culture. Additionally, it fosters a systematic approach to risk management, ensuring that all employees clearly understand and contribute to the organization's safety goals.

Design Methods Pillar

The Design Methods pillar proposes a systematic approach to developing, designing, and verifying AV systems, which includes considerations for Human-Machine Interface (HMI), and functional safety aspects of the system. A comprehensive design methods process ensures that the system is designed not only to operate safely under normal circumstances but also to take into consideration foreseeable malfunctions. The AV-TEP Design Methods pillar focuses on system design with a concentration on driving safety aspects, including risk assessment and management, and verification and validation, and ensures compliance with industry standards and regulatory requirements. The structure of this specification follows the V-Model product development process which covers the lifecycle of the product through concept, development, production, deployment, and post-deployment phases.

Defining high-level functionalities of the system, outlining safety and regulatory requirements, and finally documenting all these processes, are covered in the concept phase of the design methods pillar. This gives the AV developer and the associated stakeholders an understanding of the various functionalities of the system and a plan for countermeasures that can be developed to address the hazards identified during the preliminary analysis phase. A comprehensive understanding of the system's intended context of use is established, ensuring that its functionality aligns with real-world applications. It is also important to describe and analyze the technical and organizational control over the system to ensure its proper governance. Access to enabling systems or services must be obtained to support the system's operation. Additionally, gathering relevant legal, social, and environmental information is crucial to assessing the system's overall feasibility.

The product development phase in the design methods pillar helps in translating the conceptual design devised in the previous phase into a functional prototype. The steps involved in this process include finalizing the system design, integrating the hardware and software components, and ensuring that the interaction between them is as expected, and develop mitigation strategies that arise due to this integration. Upon building the prototype of the AV, rigorous verification and validation process is followed to assess the effectiveness of the mitigation strategies and identify any residual risk and failure modes that can affect the performance of the AV. Once a successful testing of the prototype is completed, a production plan is developed that allocates necessary resources that enable scaling up the production process, and ultimately ensures compliance with relevant

industry standards. The production plan can include safety-related characteristics, configuration of the product, and competence of the personnel in handling the production process.

Continuously monitoring the performance of the AV post-deployment ensures it meets requirements, and performs as expected, and a communication channel is established to gather feedback from these deployments. Safety performance data is collected to assess functionality, while regular updates and maintenance procedures are implemented to ensure the system continues to function properly. When implementing software updates and maintenance changes, it is critical to identify new hazards and risks that may arise and develop appropriate mitigation strategies. A well-defined change management plan and process should be established, that supports proactive risk management by identifying, assessing, and mitigating potential risks associated with the changes. This structured approach ensures that changes do not introduce new hazards or negatively impact the system's performance, aiming to maintain safety and reliability standards. Authorized personnel must analyze and evaluate each change request through an impact analysis, making decisions based on its potential effects. Once approved, the changes and updates should be implemented, with thorough documentation for future reference. Field data must be collected and analyzed to detect functional safety issues and triggering events, followed by the development of strategies to address these concerns. It is equally important to ensure the safe and systematic deactivation of the ADS, removing it from service without causing disruption or safety risks. Real-time monitoring systems should be in place to detect failures as they occur and diagnose their root causes. Each ADS unit must undergo end-of-

line testing procedures and be validated against predefined benchmarks to ensure compliance with safety and performance standards.

Testing Pillar

The Testing pillar focuses on validating the requirements and recommendations suggested in the SMS and Design Methods pillars. As the first step in the testing process, a digital twin of the test vehicle is created using Blender and this model is imported into CARLA (CAR Learning to Act) for this research work. CARLA [58] is an open-source simulator developed for the testing and validation of automated vehicles in different test cases with varying scenario elements. CARLA includes a variety of sensors, such as cameras, LIDAR, and RADAR, and supports integration with various testing frameworks, allowing users to create complex driving scenarios for validation, including adverse weather conditions, dynamic traffic, and pedestrian behaviors. With the digital twin model developed in Blender, necessary sensors were integrated into the model to mimic the test vehicle.

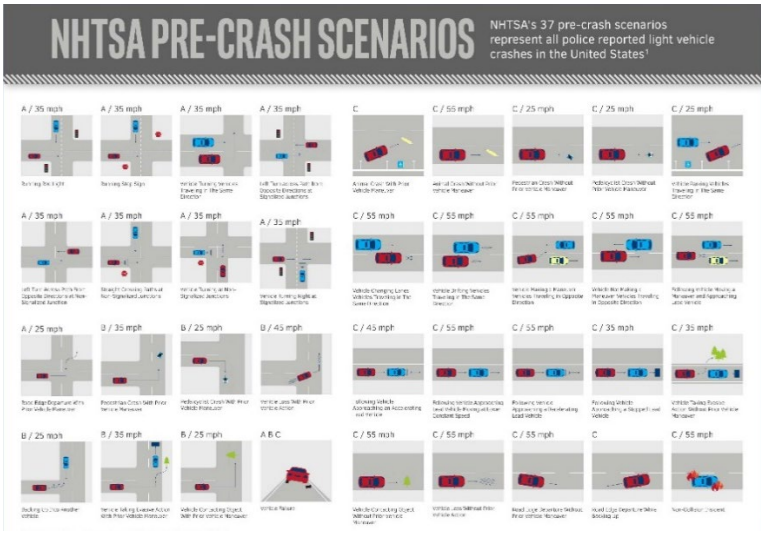

Figure 8: NHTSA pre-crash scenarios

Source: [11]

The initial test set consisted of the 37 pre-crash scenarios (shown in Figure 9) defined by NHTSA [11] involving scenarios with other road vehicles, vulnerable road users (pedestrians and bicyclists). Simulating these scenarios generates data that consists of various parameters of the vehicle and information regarding the scenario including the speed, position, heading direction and angle, distance between the ego and subject vehicle/pedestrian, etc. This information is fed through a script that evaluates the performance of the vehicle in that scenario based on various metrics [48] such as Time-to-collision, Traffic Law Violation, Lane Departure Violation, Safety Envelope Violation, Aggressive Acceleration Violation, with these metrics categorized into Black box, Grey box, and White box metrics based on the minimum data required from the ADS to compute the particular metric. Based on the score that is calculated per metric, an overall Driving Assessment (DA) score is calculated for the scenario. Additional information such as complexity, relevance, fidelity, and severity. *Complexity* indicates the difficulty level of a scenario that the vehicle is trying to navigate. *Relevance* describes how often the vehicle is expected to encounter the scenario throughout its operational lifespan. *Fidelity* reflects the level of confidence in the accuracy of the scenario's outcome measurement. *Severity* refers to the importance of a metric violation when it occurs. The DA Score equation is [59]:

$$DA\ Score = \left((Complexity * Relevance * Fidelity) * \left(1 - \sum (Severity_i * Metric_i)\right)\right) x100\%$$

As the AV shows satisfactory competency with the simulation-based testing, the next stages of the testing pillar are implemented. When the vehicle is designed and produced in adherence to relevant industry standards and best practices as suggested in the Design Methods pillar, and when the organization is ready to transition to the next stage in the testing, the vehicle is tested in closed course test track with the help of Augmented Reality (AR) enhancement that would assist with creating virtual scenarios and obstacle using AR overlays with traffic, and pedestrians. Once a satisfactory behavioral competency is observed, the test vehicle can transition to closed course testing without AR overlays. The testing method progresses to on-road testing with the help of a safety driver, trained and monitored in accordance with requirements suggested in the SMS pillar conforming with relevant best practices. Shadow mode testing is suggested as a first step while advancing to on-road testing, where the safety driver drives the test vehicle around the selected test route giving the sensor suite on the vehicle to capture and learn about its surroundings.

Technology Readiness Level

Technology Readiness Levels (TRLs) act as a point of reference to the AV developer before they can transition from one technological level to another. Prior to making the transition, this structured approach helps AV developer ensure that all requirements and objectives relevant to the current TRL are satisfied and prevents premature transition. The TRL framework is categorized by development phases as follows:

- Basic Research Phase: TRL 1 to TRL 3
- Applied Research Phase: TRL 4 to TRL 5
- Development Phase: TRL 6 to TRL 8
- Deployment Phase: TRL 9

In the AV-TEP SCF, the requirements for each TRL are organized into three foundational pillars: Safety Management Systems (SMS), Design Methods, and Testing. Additionally, specific requirements are provided for the safety case at the end of each technology phase. By categorizing requirements into these pillars, the framework improves clarity and readability, ensuring that AV developers can systematically address all requirements under one pillar before progressing to the next. However, to achieve a successful TRL transition, the AV developer must satisfy the requirements of all three pillars, ensuring a holistic evaluation of the system's readiness.

The inclusion of safety case requirements at each stage plays a pivotal role in ensuring that AV developers not only meet technical milestones but also establish a preliminary safety case early in the process. This safety case evolves continuously throughout the development lifecycle, reflecting system modifications and safety improvements. At the

culmination of TRL 9, a final safety case is published, providing comprehensive safety assurance for the system's operational deployment. This approach ensures that the safety case is a living document that guides safe development practices from the early phases of research through to commercial deployment.

The list of relevant standards and best practices provided in the previous section act as benchmarks to show conformance, in the absence of industrial regulations, bridging gaps in regulation ensuring clear safety and reliability assurances. Furthermore, to validate the integrity of the safety case, developers are encouraged to conduct an internal audit of the preliminary safety case. Upon reaching the final TRL, a third-party evaluation is recommended to ensure an unbiased, rigorous assessment of the system's safety.

Another key element of the SCF is the integration of Safety Performance Indicators (SPIs) within each pillar. These SPIs provide a measurable and data-driven approach to evaluating the effectiveness of risk mitigation strategies and the overall safety performance of the AV system. By monitoring SPIs, organizations can continuously assess whether the system's risk controls are functioning as intended and identify areas for further improvement.

One key challenge in the implementation of this framework is it being post hoc, meaning developers who have already advanced without TRL considerations, enforcing it retroactively might seem unnecessary or counterproductive. However, some approaches could foster broader adoption among existing and future developers. This TRL framework is customized to the AV industry factoring in current industry best practices and technology constraints, making the adoption of the framework easier. The AV-TEP TRL framework

helps in reducing developmental risk and improve project outcomes, which could facilitate increase in project funding, hence motivating developers to adopt this framework.

In summary, the structured approach of categorizing requirements into the three pillars, maintaining a dynamic safety case, adhering to relevant standards, and utilizing SPIs for ongoing evaluation ensures that AV developers meet all necessary criteria before advancing through TRLs. This methodology fosters not only technological progress but also rigorous safety assurance, culminating in a robust, validated, and deployable AV system.

The minimum set of standards and best practices that an AV must comply/conform with are provided below categorized by the pillars of the SCF. A detailed list of standards categorized by each TRL and pillar of the SCF is provided in Appendix C.

The AV developer is required to show compliance with the following list of standards and best practices to support the arguments of the SMS pillar of the TRL framework [60]:

- ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products [19]
- AVSC Information Report for Adapting a Safety Management System (SMS) for Automated Driving System (ADS) SAE Level 4 and 5 Testing and Evaluation [8]
- AVSC Best Practice for ADS Remote Assistance Use Case [61]
- AVSC Information Report for Change Risk Management [43]
- ISO 9001:2015 – Quality management systems [28]
- SAE J3018_202012 – Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles [9]

- SAE J3206_202107 – Taxonomy and Definition of Safety Principles for Automated Driving System (ADS) [41]
- SAE J3247_202403 – Automated Driving System Test Facility Safety Practices [62]
- European Commission (EU) regulation 2022/1426 [63]
- SAE J3320 – Safety Management System (SMS) application to SAE Level 3, 4, 5 ADS - Equipped Vehicles and Supporting Systems [64]

These additional set of standards and best practices can be used to bolster the safety case and evidence for the SMS pillar of the TRL framework [60]:

- Automotive Software Process Improvement and Capability determination [65]
- International Civil Aviation Organization (ICAO) Safety Management Manual (SMM) [3]
- ANSI/API Recommended Practice 1173: Pipeline Safety Management Systems [66]
- MIL-STD-882E, Department of Defense Standard Practice: System Safety [67]

The AV developer is required to show compliance with the following list of standards and best practices to support the arguments of the Design Methods pillar of the TRL framework [60]:

- Federal Motor Vehicle Safety Standards [68]
- ISO 26262:2018 – Road Vehicles – Functional safety [30]

- ISO 21448:2022 – Road Vehicles – Safety of the intended functionality [40]
- ISO 34503:2023 – Road Vehicles – Test scenarios for automated driving systems – Taxonomy for operational design domain [69]
- ANSI/UL 4600 [19]
- AVSC Change Risk Management [43]
- SAE J2980_202310 - Considerations for ISO 26262 ASIL Hazard Classification [70]
- SAE J3016_202104 – Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles [71]
- SAE J3131_202203 – Definitions for Terms Related to Automated Driving Systems Reference Architecture [72]
- SAE J3206_202107 – Taxonomy and Definition of Safety Principles for Automated Driving System (ADS) [41]
- ISO/TR 4804:2020 – Road vehicles – Safety and cybersecurity for automated driving systems – Design, verification and validation [73]
- ISO/CD TS 5083 – Road vehicles - Safety for automated driving systems – Design, verification and validation [74]
- SAE J3259 – Taxonomy & Definitions for Operational Design Domain (ODD) for Driving Automation Systems [75]

Additional list of standards and best practices that are helpful in showing compliance for the Design Methods pillar of the TRL framework [60] include the following:

- ISO/IEEE 7000 – Systems and software engineering – Life cycle management Part 7000: Standard model process for addressing ethical concerns during system design [76]
- SAE J3187_202202 – System Theoretic Process Analysis (STPA) Recommended Practices for Evaluations of Automotive Related Safety-Critical Systems [77]

To support the arguments defined by Dr. Wishart for the Testing pillar of the TRL framework [60], developers are required to comply/conform with the following list of standards and best practices:

- ANSI/UL 4600 Standard for Safety Evaluation of Autonomous Products [19]
- ISO 21448 Road vehicles - Safety of the intended functionality [40]
- ISO 26262 Road vehicles — Functional safety [30]
- ISO 34502:2022 – Road vehicles – Test scenarios for automated driving systems – Scenario based safety evaluation framework [69]
- ISO 34503 Road Vehicles – Test scenarios for automated driving systems – Taxonomy for operational design domain [78]
- ISO 34504:2024 - Road vehicles - Test scenarios for automated driving systems - Scenario categorization [79]

- SAE J3018- Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles [49]
- SAE J3237_202501 – Dynamic Driving Task Assessment (DA) Metrics for Automated Driving Systems [80]
- SAE J3247_202403 - Automated Driving System Test Facility Safety Practices [62]
- SAE J3279 - Best Practices for Developing and Validating Simulations for Automated Driving Systems [81]

PROOF OF CONCEPT

Specification Forms

For this thesis, the requirements for the SMS and Design Methods pillars of the SCF were structured into specifications forms that follows a questionnaire-based format where each sub-section of the specification form asks a question to the AV developer on the methods implemented within the organization or design of the product that would help them satisfy that requirement. This process ensures that the AV developer has considered all possible outcomes of the system and has designed and operated the vehicle keeping in mind the safety and stakeholder constraints.

These requirements and questions address the safety-critical objectives within each pillar, with the intent of giving the developer an opportunity to reflect on their compliance with safety protocols, regulatory standards, and risk mitigation techniques.

Within the SMS pillar, the specification form details on the setup of the organization, methods employed to identify, assess, and mitigate organizational hazards, promote and educate the concept of safety to the stakeholders and employees, and ensure that the mitigation strategies are working as expected. The design methods specification form talks about the requirements for engineering practices following the structure of the widely used V-model for product development. It focuses on defining the functionalities of the system, application of HARA techniques, considerations for operation of the vehicle, and change management for risks induced due to post-deployment modifications. The testing specification form assesses the behavioral competency of the ADS by discussing the

breadth and diversity of scenarios used, definition of the ODD of the test vehicle, and whether the ADS has been tested in edge cases.

Having the SCF structured in the form of specification forms established a transparent, traceable, and repeatable assessment process, regulators, auditors, and safety assessors can easily review the developer's responses, gaining insight into the specific measures taken to satisfy safety and regulatory constraints, fostering accountability.

The specification forms structured in the following way:

I. Item/Clause: Components or sub-components under different elements of the SMS pillar
    - Definition: Provides a brief description of the item and a top-level understanding of the applicable component or sub-component
    - Source: Provides reference information from which the description for the component and goals were derived.
    - Goal:
        - Specifies tasks AV developers should consider implementing in their safety culture and policy.
    - Argumentation/Question: Questions raised based on the goals mentioned above and suggestions to check for AV developer compliance and competence of the organization in handling the given component. It is essential to decompose each goal into tangible and measurable performance indicators.
    - Evidence/AV Developer response: Response to the questions provided above along with evidence showing competency to the component.

Sample element from the SMS Specification Form:

i. Documentation of System Safety Approach
    - Definition: It is the process of documenting the system safety approach for managing hazards as an integral part of the system engineering process. The program manager and contractor carry out this documentation process.

- Source: [67]
- Goal:
    - Describe the risk management effort and how the program is integrating risk management into the system engineering process.
    - Identify and document the prescribed and derived requirements applicable to the system.
    - Define how the appropriate risk acceptance authority formally accepts hazards and associated risks.
    - Document hazards with a closed-loop Hazard Tracking System (HTS). The HTS will include elements like identified hazards, associated mishaps, risk assessments, identified risk mitigation measures, selected mitigation measures, hazard status, verification of risk reductions, and risk acceptance.
    - Audit and update this documentation every year.
- Argumentation/Question: Based on the information above, how does your organization's system safety approach documentation integrate risk management with system engineering process to identify and mitigate any potential hazards?
- Evidence/AV Developer Response:

Sample element from the Design Methods Specification Form:

i. Item Definition

- Definition: The description of the system and its operating conditions is important in identifying and assessing all the risks and hazards the system would encounter during its operation.
- Source: [30] [40]
- Goal:
    - The following requirements can be made available:

        - Compliance with legal regulations and adherence to both national and global standards.
        - The operational characteristics at the vehicle level, encompassing modes or conditions of operation.
        - The necessary standards for functionality, encompassing quality, performance, and availability if relevant.
        - Restrictions related to the item, including functional interdependencies, reliance on other items, and the operating context.
        - Possible repercussions or outcomes.
    - Define boundaries for the item, and include assumptions, interfaces concerning its interactions with other items.
    - Describe operating ODD based on use cases, restrictions, and scenarios within that domain.
- Argumentation/Question: Based on the information above, provide a detailed description of the item
- Evidence/AV Developer Response:

Technology Readiness Level Table

The TRL table was created following the structure developed by NASA, with levels ranging from 1 to 9, representing the various stages of technology maturation. This effort to establish a TRL table led by Dr. Jeffrey Wishart provides definitions and requirements for each TRL that were adapted to suit the ADS industry. It was formatted into a user-friendly structure with columns for the technology phase, level, AV-TEP description for each level, relevant standards and best practices that the AV developer can show conformance/compliance to, and a final column for SPIs separately for each pillar. The technology phase column categorizes the entire table into 4 sections such as basic research, applied research, development, and deployment/implementation phase. This is further divided into the nine levels with levels 1-3 falling under basic research, levels 4 and 5 under applied research, levels 6-8 within the development phase, and level 9 corresponding to the deployment phase. The AV-TEP description highlights definitions, deliverables, and requirements for each TRL, providing an understanding of what elements constitute technology readiness at each stage, ensuring a focused progression toward deployment. The recommended standards column helps the AV developer ensure that their ADS meets or exceeds safety expectations at every stage of development. SPIs act as measurable metrics that evaluate the effectiveness of safety measures implemented throughout the development process. SPIs are of two types, Leading metrics that are predictive and proactive, helping the AV developer foresee and identify potential risks to the system, and Lagging metrics that outcome-based and reactive, reflecting the results of actions that have been already taken.

Table 1: AV-TEP TRL Description [60]

Technology Phase	TRL	AV-TEP Description
Basic Research Phase	1	**Basic principles and research**: The organizational objectives and plans are established, including operational design domain (ODD) and usage specification(s) for planned deployment(s), and the roadmap of regulations/standards/best practices to which conformance/compliance are planned is complete.
	2	**Technology concept formulated**: The SMS structure, design concept, and V&V plan are all established.
	3	**Proof of concept**: The proof of concept of the SMS and designed system have been validated.
Safety Case: **Preliminary version of the safety case, including SMS implementation, is complete.**		
Applied Research Phase	4	**Components validated in laboratory environment**: The assessment of sub-systems and components in a simulation and/or XiL environment is completed*
	5	**System demonstrated in laboratory environment**: The assessment of the system in a simulation environment is complete*, and the organization is ready for prototype build.
Safety Case: **Intermediate version of the safety case, including component and system V&V results from the laboratory environment, is complete and evaluated by an internal auditor.**		
Development Phase	6	**Prototype demonstrated in relevant environment**: The assessment of the prototype in a closed course environment is complete*, and the organization is ready to begin on-road testing with a fallback test driver.
	7	**Prototype demonstrated in operational environment**: The assessment of the prototype in an on-road environment (that is within the defined ODD) with a fallback test driver is complete*, and the organization is ready to finalize the system design and begin on-road testing without a fallback test driver.

	8	**Technology proven in operational environment**: The assessment of the finalized system in an on-road environment (that is within the defined ODD) without a fallback test driver is complete*, and the organization is ready to begin commercial build and deployment.
Safety Case: **Safety case is complete and evaluated by an independent auditor.**		
Deployment/ Implementation Phase	9	**Technology refined and adopted**: The organization is capable of commercially deploying the system in its defined ODD and Usage Specification(s), and there is demonstration of continual improvements in organization, design, and testing/deployment results.
Safety Case: **Safety case is continuously updated as changes to the organization, design, and test/deployment results occur/become available.**		

* Note that "complete" indicates that sufficient demonstration has occurred for the assessment at this TRL. However, it is eminently possible that additional testing using this test method will be included in later technological development phases to bolster the safety case.

The requirements for each TRL are structured in the following format:

Claims: The assertion that the AV and the organization is ready to transition from one TRL to the next.

Arguments: The case for why the claim is valid. The proposed approach is to use conformance with regulations and compliance with standards/best practices as the argument that the claim is valid, along with methods of managing risk specific to the claim.

Evidence: The proof to support the arguments. The proposed approach is to use the work products that demonstrate conformance with regulations and compliance with standards/best practices as the supporting content, including test results where applicable.

Sample TRL requirement from the framework:

TRL 6: Prototype Demonstrated in Relevant Environment

Claim: The assessment of the prototype in a closed course environment is complete[14], and the organization is ready to begin on-road testing with a fallback test driver.

SMS Pillar
Arguments

The arguments to support the SMS claims of TRL 6 include continued compliance with SAE J3018 [9], including training more FTDs if needed.

Design Methods Pillar
Arguments

The arguments to support the Design Methods claim for TRL 6 including demonstrating the capability of the vehicle prototype to satisfy all safety requirements, showing satisfactory performance in closed course environment, complying with Sections 4-6 of SAE J3247 [62]. Since the design is intended to begin on-road testing, the arguments also include compliance with FMVSS [68].

Testing Pillar
Arguments

The arguments to support the Testing claims of TRL 6 include the scenario-based testing components described in TRL 5 and applied to closed course testing, for which compliance with SAE J3247 [62] is required. It is expected that a percentage of scenarios tested in simulation will be validated using closed course testing. The AV-TEP ODD and behavioral competency specification form can be used to establish the closed course testing domain.

Simulation testing is not necessarily required at TRL 6. If simulation testing is conducted to support the Testing claims of TRL 6, the arguments include that the simulation tools

should comply with SAE J3279 [81], and that the AV should maintain or improve the DA scores for repeated scenarios and achieve the minimums required for new scenarios.

CONCLUSION

In conclusion, this thesis work helps in establishing a comprehensive and evolving safety case framework that caters to the unique challenges and requirements of the AV industry through the specification forms for the pillars of the SCF. The specification forms harmonize best practices and industry standards from various industries such as aviation, nuclear, and automotive ensuring that the AV developer and operator are considering safety at every stage of product development and post-deployment. The specification forms categorized into the SCF's three pillars- SMS, Design Methods, and Testing pillars- help in identifying, analyzing, and mitigating hazards and associated risks within the organization, and at the system level, and validating the behavioral competency of the AV in various critical scenarios through various modes of testing such as simulation-based, closed course, and on-road testing, ensuring the vehicle is safe to deploy on public roads. This process helps the AV developer in providing an assurance that the vehicle performs as expected within its ODD and helps in increasing public trust and scalability of the vehicle.

The specification forms help the developer in addressing safety concerns at every stage of the product development, encouraging healthy safety practices through documentation and ensuring compliance with relevant standards and best practices. The specification forms additionally help in improving traceability between the claims and the evidence provided by the developer that satisfies the safety arguments. The TRL table provides a step-by-step guidance for advancing AV technology from the concept phase to deployment and ensures that the AV developer satisfies all requirements relevant to one level before transitioning,

with the SPIs offering quantitative assessment to gauge the safety performance at every stage.

This work provides a structured approach to demonstrate the safety and readiness of AV technology, fostering public trust in AV systems by ensuring transparency and accountability. As the field of automated vehicles evolves over time, this SCF can adapt to the emerging technologies and regulations, ensuring relevance and ultimately supporting the safe and responsible deployment of AVs on public roads.

Contributions

This thesis work contributes to the AV-TEP mission by supporting the development of a comprehensive SCF through the creation of detailed specification forms for the SMS and Design Methods pillars. An example of these forms is provided in the appendix section. As mentioned in previous sections, the specification forms are developed after an extensive literature review of various standards and best practices from various industries, and harmonizing learnings and guidelines into key elements of the SMS and Design Methods pillars. Along with the specification forms, the major contribution of this thesis work to the AV-TEP mission is establishing the TRL framework under the guidance of Dr. Jeffrey Wishart. The arguments and the list of standards the developer can establish compliance for the SMS and Design Methods pillar were part of this thesis work. In the testing pillar, this thesis work contributed to the development of the digital twin of the subject vehicle, simulating pre-crash scenarios based on the list of pre-crash scenarios prescribed by

NHTSA in [11], analyzing performance metrics to assess the vehicle's behavior in navigating various scenarios.

Additional contributions include publication of research papers in various conferences for AVs:

- Adapting Technology Readiness Level (TRL) for Automated Vehicles, SAE WCX 2025 (In Review) [60]
- Developing a Safety Management System for the Automated Vehicle Industry, SAE WCX 2025 (In Review) [82]
- Comprehensive Evaluation of Behavioral Competence of an Automated Vehicle Using the Driving Assessment (DA) Methodology, SAE WCX 2024 (Published) [83]
- A Proposed Safety Case Framework for Automated Vehicles with Pillars of Safety Management System, Design Methods, and Scenario-Based Testing, IEEE IAVVC 2023 (Published) [7]

FUTURE WORK

This thesis work can be expanded in the future to include considerations for human-machine interface and its associated hazards and risks. In the SMS pillar, future work could include improving the robustness of the Safety Culture component, and additionally include HRO principles.

Given the extensive reliance on AI systems, the future work for the Design Methods pillar could include safety assurance considerations for AI systems adhering to relevant industry standards such as ISO/PAS 8000: Road vehicles — Safety and artificial intelligence [84], ISO/TS 23792-1:2023: Intelligent transport systems — Motorway chauffeur systems (MCS) [85], ISO 23374-1:2023: Intelligent transport systems — Automated valet parking systems (AVPS) [86], SAE J3316 – Cooperative driving automation (CDA) Features: Common reference architecture, nomenclature and template [87], PAS 1883 - Operational Design Domain (ODD) Taxonomy for ADS Specification [88], IEEE 2846-2022: Standard for Assumptions in Safety-Related Models for Automated Driving Systems [89], ISO/DTS 5083: Road vehicles — Safety for automated driving systems — Design, verification and validation [74] to demonstrate that AI systems are safe, reliable, and are capable of handling real-world scenario posing minimal risk. Cummings proposes a Taxonomy for AI Hazard Analysis (TAIHA) [90] approach to investigate accidents caused by AI systems. Burton [91] discusses the challenges of safety assurance for machine learning (ML) in safety-critical systems. It proposes a framework for analyzing residual uncertainty in ML assurance arguments and identifies measures to address these uncertainties. These works by researchers and organizations could be beneficial for augmenting this thesis work in the future.

Once the protype vehicle shows sufficient behavioral competency in the simulation-based testing, the vehicle can be tested in more diverse scenarios in closed course and on-road testing facilities. As an AV heavily relies on software, connectivity, and real-time data processing, future research can work on including cybersecurity into the SCF.

This work can additionally be used as a point of reference for regulatory agencies and stakeholders in developing standardized approaches to handling safety-critical systems and processes in the AV industry.

REFERENCES

[1] Australian Radiation Protection and Nuclear Safety Agency, "History of Safety," [Online]. Available: https://www.arpansa.gov.au/regulation-and-licensing/safety-security-transport/holistic-safety/history#:~:text=The%20first%20age%20of%20safety,technology%20was%20safe%20to%20use.. [Accessed 15 06 2024].

[2] Clark Fountain, "The Evolution of Vehicle Safety," 11 Jan 2022.

[3] International Civil Aviation Organization, Safety Management Manual (SMM), 2013.

[4] Waymo, "Waymo Story," [Online]. Available: https://waymo.com/about/#story. [Accessed 5 11 2024].

[5] SAE International, "J3016 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," Recommended Practice, 2021.

[6] European Commission, COMMISSION IMPLEMENTING REGULATION (EU) 2022/1426, 2022.

[7] J. Wishart, J. Zhao, B. Woodard, G. O'Malley, H. Guo, S. Rahimi and S. Swaminathan, "A Proposed Safety Case Framework for Automated Vehicle Safety Evaluation," in *IEEE International Automated Vehicle Verification Conference*, Austin, TX, 2023.

[8] Automated Vehicle Safety Consortium, "AVSC Information Report for Adapting a Safety Management System (SMS) for Automated Driving System (ADS) SAE

Level 4 and Level 5 Testing and Evaluation," SAE Industry Technologies Consortia, 2021.

[9] SAE International, "SAE J3018_202012 - Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles," Recommended Practice, 2020.

[10] J. Wishart, *Personal communication,* 2024.

[11] National Highway Traffic Safety Administration, "Pre-Crash Scenario Typology for Crash Avoidance Research," 2007.

[12] Waymo, "Building a Credible Case for Safety: Waymo's Approach for the Determination of Absence of Unreasonable Risk," 2023.

[13] Waymo, "Waymo's Safety Methodologies and Safety Readiness Determinations," Waymo LLC, 2020.

[14] Aurora Innovation, Inc., "Safety Report," 2022.

[15] M. Wagner and C. Carlan, "The Open Autonomy Safety Case Framework," Edge Case Research Inc., 2024.

[16] Toyota, "Toyota Voluntary Safety Self-Assessment For SAE Level 4 And Level 5 Automated Vehicle Technology Testing On Public Roads," 2020.

[17] Torc, "Innovating Safety and Efficiency," 2021.

[18] Uber, "Safety Guidelines for autonomous mobility and delivery on the Uber Platform," 2022.

[19] Underwriters Laboratory, "UL 4600: Standard for Safety Evaluation of Autonomous Products," 2023.

[20] J. T. Reason, Managing the Risks of Organizational Accidents, Aldershot: Ashgate, 1997.

[21] Federal Aviation Administration, "Safety Management System Components Explained," Federal Aviation Administration, 17 Nov 2021. [Online]. Available: https://www.faa.gov/about/initiatives/sms/explained/components. [Accessed 08 Sep 2024].

[22] SAE International, "SAE J3320: Safety Management System (SMS) application to SAE Level 3, 4, 5 ADS - Equipped Vehicles and Supporting Systems," In preparation.

[23] C. Bieder, Safety Management Systems and their Origins: Insights from the Aviation Industry, CRC Press, 2023.

[24] K. Fox, "How has the implementation of Safety Management Systems (SMS) in the transportation industry impacted on risk management and decision-making?," Lund University, 2009.

[25] High Reliability Organizing, "Models of HRO: Weick and Sutcliffe," High Reliability Organizing, 2007. [Online]. Available: https://www.high-reliability.org/Weick-Sutcliffe#weick2007. [Accessed 7 November 2024].

[26] A. Kuusisto, "Safety management systems Audit tools and reliability of auditing," VTT PUBLICATIONS, 2000.

[27] R. Mitchell, R. Friswell and L. Mooren, "Initial development of a practical safety audit tool to assess fleet safety management practices," *Accident Analysis & Prevention,* vol. 47, no. doi.org/10.1016/j.aap.2012.01.021, pp. 102-118, 2012.

[28] ISO; International Organization for Standardization, ISO 9001: Quality Management Systems, 2015.

[29] International Automotive Task Force (IATF), IATF 16949:2016, 2016.

[30] ISO; International Organization for Standardization, ISO 26262-1:2018(en) - Road vehicles — Functional safety — Part 1: Vocabulary, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[31] ISO; International Organization for Standardization, ISO 26262-2:2018(en) - Road vehicles — Functional safety — Part 2: Management of functional safety, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[32] ISO; International Organization for Standardization, ISO 26262-3:2018(en) - Road vehicles — Functional safety — Part 3: Concept phase, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[33] ISO; International Organization for Standardization, ISO 26262-4:2018(en) - Road Vehicles - Functional Safety - Part 4: Product Development at the System level, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[34] ISO; International Organization for Standardization, ISO 26262-5:2018(en) - Road Vehicles - Functional Safety - Part 5: Product Development at the Hardware level,

Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[35] ISO; International Organization for Standardization, ISO 26262-6:2018(en) - Road Vehicles - Functional Safety - Part 6: Product Development at the Software level, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspectsw, 2018.

[36] ISO; International Organization for Standardization, ISO 26262-7:2018(en) - Road Vehicles - Functional Safety - Part 7: ISO 26262-7:2018(en) - Road Vehicles - Functional Safety - Part 7: Production, operation, service and decommissioning, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[37] ISO; International Organization for Standardization, ISO 26262-8:2018(en) - Road Vehicles - Functional Safety - Part 8: Supporting processes, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[38] ISO; International Organization for Standardization, ISO 26262-9:2018(en) - Road Vehicles - Functional Safety - Part 9: Automotive safety integrity level (ASIL)-oriented and safety-oriented analyses, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[39] ISO, International Organization for Standardization, ISO 26262-10:2018(en) - Road vehicles — Functional safety — Part 10: Guidelines on ISO 26262ty, Technical Committee ISO/TC 22,Road vehicles Subcommittee,Electrical and electronic components and general system aspects, 2018.

[40] ISO; International Organization for Standardization, ISO 21448:2022 - Road vehicles - Safety of the intended functionality, 2022.

[41] SAE International Information Report, "Taxonomy and Definition of Safety Principles for Automated Driving System (ADS), SAE Standard J3206_202107," 2021.

[42] SAE International Information Report, "Identifying Automated Driving Systems-Dedicated Vehicles (ADS-DVs) Passenger Issues for Persons with Disabilities, SAE Standard J3171_201911," 2019.

[43] Automated Vehicle Safety Consortium, "AVSC Information Report for Change Risk Management," SAE Industry Technologies Consortia, 2023.

[44] Automated Vehicle Safety Consortium, "AVSC Best Practice for Continuous Monitoring and Improvement after Deployment," 2023.

[45] National Highway Traffic Safety Administration (NHTSA), Automated Driving Systems 2.0: A Vision for Safety, United States Department of Transportation, 2017.

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

[47] Waymo, "Safety," [Online]. Available: https://waymo.com/safety/. [Accessed 7 November 2024].

[48] S. G. Como, J. Wishart, M. Elli and N. Kidambi, "Evaluating the Severity of Safety Envelope Violations in the Proposed Operational Safety Assessment (OSA) Methodology for Automated Vehicles," *SAE Technical Paper 2022-01-0819,* 2022.

[49] SAE International, "J3018 - Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles," *Ground Vehicle Standard,* 2020.

[50] ISO, International Organization for Standardization, "ISO 34502:2022- Road Vehicles - Test Scenarios for Automated Driving Systems - Scenario Based Safety Evaluation Framework," 2022.

[51] M. D. Tezerjani, D. Carrillo, D. Qu, S. Dhaka, A. Mirzaeinia and Q. Yang, "Real-time Motion Planning for autonomous vehicles in dynamic environments," 2024.

[52] NASA, "Technology Readiness Level | NASA," 1 April 2021. [Online]. Available: https://www.nasa.gov/directorates/heo/scan/engineering/technology/technology_readiness_level.

[53] J. C. Mankins, "Technology readiness assessments:A retrospective," *Acta Astronautica,* vol. 65, pp. 1216-1223, 2009.

[54] J. C. Mankins, "Technology Readiness Levels: A White Paper," Advanced Concepts Office, Office of Space Access and Technology, NASA, 1995.

[55] A. Olechowski, S. D. Eppinger and N. Joglekar, "Technology Readiness Levels at 40: A Study of State-of-the-Art Use, Challenges, and Opportunities," *Proceedings of PICMET '15: Management of the Technology Age,* 2015.

[56] US Department of Transportation, Technology Readiness Level Guidebook, 2017.

[57] A. L. Olechowski, S. D. Eppinger, N. Joglekar and K. Tomaschek, "Technology readiness levels: Shortcomings and improvement opportunities," *Systems Engineering,* vol. 23(4), pp. 395-408, 2020.

[58] CARLA, "CARLA: Open-source simulator for autonomous driving research," [Online]. Available: https://carla.org/.

[59] S. Como and J. Wishart, "Evaluating Automated Vehicle Scenario Navigation Using the Operational Safety Assessment (OSA) Methodology," *SAE Technical Paper 2023-01-0797,* 2023.

[60] S. Swaminathan, J. Wishart, J. Zhao and B. Russo, "Adapting the Technology Readiness Level (TRL) Framework to AV Development," *SAE Technical Paper,* 2025.

[61] Automated Vehicle Safety Consortium, "AVSC Best Practice for ADS Remote Assistance Use Case," SAE Industry Technologies Consortia, 2023.

[62] SAE International, "SAE 3247_202403 - Automated Driving System Test Facility Safety Practices," Recommended Practice, 2024.

[63] European Commission, Commission Implementing Regulation (EU) 2022/1426, 2022.

[64] SAE International, "SAE J3320 - Safety Management System (SMS) application to SAE Level 3, 4, 5 ADS - Equipped Vehicles and Supporting Systems," Recommended Practice, In preparation.

[65] VDA QMC Working Group 13 / Automotive SIG, "Automotive SPICE Process Assessment / Reference Model," 2016.

[66] American Petroleum Institute, ANSI/API Recommended Practice 1173 Pipeline Safety Management Systems, American Petroleum Institute, 2015.

[67] DEPARTMENT OF DEFENSE, MIL-STD-882E, STANDARD PRACTICE: SYSTEM SAFETY, 2012.

[68] National Highway Traffic Safety Administration (NHTSA), Federal Motor Vehicle Safety Standards.

[69] International Organization for Standardization (ISO), ISO 34503:2023 - Road Vehicles – Test scenarios for automated driving systems – Taxonomy for operational design domain, 2023.

[70] SAE International, "SAE J2980_202310 - Considerations for ISO 26262 ASIL Hazard Classification," Recommended Practice, 2023.

[71] SAE International, "SAE J3016_202104 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," Recommended Practice, 2021.

[72] SAE International, "SAE J3131_202203 - Definitions for Terms Related to Automated Driving Systems Reference Architecture," Recommended Practice, 2022.

[73] International Organization for Standardization (ISO), "ISO/TR 4804:2020 - Road vehicles - Safety and cybersecurity for automated driving systems - Design, verification and validation," 2020.

[74] ISO, International Organization for Standardization, ISO/DTS 5083 Road vehicles — Safety for automated driving sysems — Design, verification and validation, In preparation.

[75] SAE International, "SAE J3259 - Taxonomy & Definitions for Operational Design Domain (ODD) for Driving Automation Systems," In preparation.

[76] International Organization for Standardization (ISO), ISO/IEC/IEEE 24748-7000:2022 Systems and software engineering — Life cycle management Part 7000:

Standard model process for addressing ethical concerns during system design, 2022.

[77] SAE International, "SAE J3187_202202 – System Theoretic Process Analysis (STPA) Recommended Practices for Evaluations of Automotive Related Safety-Critical Systems," Recommended Practice, 2022.

[78] International Organization for Standardization (ISO), *ISO 34503 Road Vehicles – Test scenarios for automated driving systems – Taxonomy for operational design domain,* 2023.

[79] International Organization for Standardization (ISO), "ISO 34504:2024 - Road vehicles - Test scenarios for automated driving systems - Scenario categorization," 2024.

[80] SAE International, "SAE J3237 - Driving Safety Assessment (DSA) Metrics for Automated Driving Systems (ADS)," Recommended Practice, In preparation.

[81] SAE International, "SAE J3279 - Best Practices for Developing and Validating Simulations for Automated Driving Systems," Recommended Practice, In Progress.

[82] D. Wichner, J. Wishart, S. Swaminathan and J. Sergent, "Developing a Safety Management System for the Automated Vehicle Industry," *SAE Technical Paper,* 2025.

[83] T. Wang, S. Rahimi, S. Swaminathan, M. Zaidi, J. Wishart and H. Liu, "Comprehensive Evaluation of Behavioral Competence of an Automated Vehicle Using the Driving Assessment (DA) Methodology," *SAE Technical Paper,* 2024.

[84] ISO, International Organization for Standardization, ISO/PAS 8000: Road vehicles — Safety and artificial intelligence, Work in progress.

[85] ISO, International Organization for Standardization, ISO/TS 23792-1:2023 Intelligent transport systems — Motorway chauffeur systems (MCS), 2023.

[86] ISO; International Organization for Standarization, ISO 23374-1:2023 Intelligent transport systems — Automated valet parking systems (AVPS), 2023.

[87] SAE International, "J3316 – Cooperative driving automation (CDA) Features: Common reference architecture, nomenclature and template," In preparation.

[88] BSI, PAS 1883 - Operational Design Domain (ODD) Taxonomy for ADS Specification, 2020.

[89] IEEE, IEEE 2846-2022 Standard for Assumptions in Safety-Related Models for Automated Driving Systems, 2022.

[90] M. Cummings, "Identifying AI Hazards and Responsibility Gaps," 2023.

[91] B. Simon and H. Benjamin, "Addressing uncertainty in the safety assurance of machine-learning," *Frontiers in Computer Science,* vol. 5, no. 10.3389/fcomp.2023.1132580, 2023.

[92] KirkpatrickPrice, "The 5 Components of Risk Management," KirkpatrickPrice, 15 March 2021. [Online]. Available: https://kirkpatrickprice.com/blog/5-components-risk-management/.

APPENDIX A
SMS SPECIFICATION FORM

Appendix A provides a sample of the SMS specification form.

Contents of SMS Specification Form:

i. Safety Policy and Objectives (SPO)
ii. Safety Culture
iii. Safety Risk Management (SRM)
iv. Safety Assurance (SA)
v. Safety Promotion (SP)

Safety Policy and Objectives (SPO):

A safety policy in an SMS guides employees and can serve as the catalyst for acceptance and participation by employees in the SMS and its processes. Safety Policy outlines the principles, processes, and methods of the organization's SMS to achieve the desired safety outcome.

Goal:

- Reflect organization commitment toward safety.
- Successfully identify accountable executives for implementation and maintenance of SMS.
- Include safety reporting procedure.
- Document and communicate safety responsibilities.
- Periodically review and ensure relevancy of overall SMS.
- Commitment to achieving the highest safety standards and complying with applicable regulatory requirements.

Components of Safety Policy and Objectives:

ii. Context of the Organization

- Definition: It is important to introduce the structure of the organization which would include employee information, internal and external factors affecting the organization, scope of the SMS, basic outline of the business plan, information of stakeholders, compliance requirements.
- Source: [28]
- Goal:
    - Describe number of employees.
    - Define working conditions.
    - Provide information on the internal and external factors affecting the organization.

    - Describe cultural and social factors.
    - Briefly describe the financial conditions of the organization.
    - Address environmental responsibilities.
- Argumentation/Question: Based on the information above, how do internal and external factors affect the strategic planning in your organization?
- Evidence/AV Developer Response:

iii. Identifying or creating safety roles/teams with specific responsibilities

- Definition: This involves assigning individuals or teams to assess safety-related issues and enforce countermeasures to mitigate risks to achieve an Absence of Unreasonable Risk (AUR).
- Source: [8], [3]
- Goal:
    - Individuals with safety accountabilities
    - Appointment of key safety personnel
    - Involve multiple levels of management in investigating, assess safety risks and support the overall SMS.
    - Decision makers have direct and unbiased links to those who can provide safety information.
    - Personnel must enroll in mandatory safety training.
    - Ensure employees adhere to SOPs and safety directives.
- Argumentation/Question: Based on the information above, how does your organization define and assess responsibilities to individuals or teams to assess safety-related issues?
- Evidence/AV Developer Response:

Safety Risk Management (SRM):

The objective of SRM is to manage risks through a process consisting of describing the system, identifying hazards, and assessing, analyzing, and controlling risks based on safety risk assessment. Safety risks are conceptually assessed as acceptable, tolerable, or intolerable.

Goal:

- Elaborately describe the system and its operating environment.
- Identify all the hazards the system would encounter.
- Assess the identified hazards and categorize them as acceptable, tolerable, or intolerable.
- Implement mitigation measures for intolerable hazards and reduce tolerable hazards to acceptable levels.

Components of Safety Risk Management:

i. **System Description**

- Definition: The description of the system and its operating conditions is important in identifying and assessing all the risks and hazards the system would encounter during its operation.
- Source: [3] , [40]
- Goal:
    - Define all concepts and technologies of the system and subsystems.
    - Describe operating ODD based on use cases, restrictions, and scenarios within that domain.
    - Define system boundaries and functional range.
    - Provide description of the intended functionality of the ADS.
    - Describe dependencies on car driver, passengers, VRUs, environmental conditions, and interfaces with the road infrastructure.
    - Describe system limitations and its countermeasures.
- Argumentation/Question: Based on the information above, how does your organization utilize information about system's operating conditions and functional limitations to effectively mitigate hazards?
- Evidence/AV Developer Response:

ii. **Hazard Analysis and Risk Assessment (HARA)**

- Definition: It is the process of identifying and documenting potential risks and categorizing actual risks the organization faces. It is imperative to consider the risks that would arise in the organization's lifecycle.
- Source: [92]
- Goal:
    - Identify and implement a commonly used hazard identification technique such as:
        - Failure Mode and Effects Analysis (FMEA)
        - Failure mode, Effects and Criticality Analysis (FMECA)
        - Fault Tree Analysis (FTA)
        - Systems Theoretic Process Analysis (STPA)
    - Implement strategies to identify and assess known and unknown hazardous situations and minimize these scenarios to an acceptably low level through mitigation measures.
    - Measure the effectiveness of risk mitigation and evaluation of mitigation strategies.

- Argumentation/Question: Based on the information above, what are the processes that are in place to categorize and prioritize potential risks?
- Evidence/AV Developer Response:

APPENDIX B
DESIGN METHODS SPECIFICATION FORM

Appendix B provides a sample of the Design Methods specification form.

Contents of the Design Methods Specification Form:

i. Concept Phase
ii. Product Development Phase
iii. Post-Deployment phase and Change Management

Concept Phase

This is the initial stage of the development of an ADS, where fundamental ideas and foundational requirements are defined. This step typically consists of defining the system characteristics, and use cases, performing preliminary hazard analysis and risk assessment, defining high-level functionalities of the system, outlining safety and regulatory requirements, and finally documenting all these processes.

Goal:

- Define system objective and scope.
- Establish high-level requirements.
- Identify regulatory compliance requirements.
- Develop primary test and validation strategies.

Components of the Concept Phase

ii. Item Definition

- Definition: The description of the system and its operating conditions is important in identifying and assessing all the risks and hazards the system would encounter during its operation.
- Source: [40], [32]
- Goal:
    - The following requirements can be made available:
        - Compliance with legal regulations and adherence to both national and global standards.
        - The operational characteristics at the vehicle level, encompassing modes or conditions of operation.
        - The necessary standards for functionality, encompassing quality, performance, and availability if relevant.
        - Restrictions related to the item, including functional interdependencies, reliance on other items, and the operating context.

        - Possible repercussions or outcomes.
    - Define boundaries for the item, and include assumptions, interfaces concerning its interactions with other items.
    - Describe operating ODD based on use cases, restrictions, and scenarios within that domain.
- Argumentation/Question: Based on the information above, provide a detailed description of the item
- Evidence/AV Developer Response:

iii. Hazard analysis and Risk assessment

- Definition: The HARA process provides an automotive-specific risk-based approach for determining risk classes. The overall aim of the analysis is to identify and classify the potential hazards associated with the procedure and formulate safety goals related to the mitigation of these hazards.
- Source: [32]
- Goal:
    - Identification of hazards and formulating safety goals with respect to ASILs related to mitigation of risks to achieve absence of unreasonable risk.
    - Classify the identified hazards based on severity, exposure, and controllability.
    - Identification of Fault Models.
    - Risk Framework and evaluation and estimate the severity of potential harm based on defined rationale for each hazardous event.
    - Risk mitigation and evaluation of mitigation effectiveness.
- Argumentation/Question: How is the organization handling the HARA process in the concept phase of product development?
- Evidence/AV Developer Response:

Product Development Phase

This step in the development process of an ADS involves the translation of the conceptual design into a functional prototype. This phase includes detailed system design, hardware and software integration, testing to ensure system performance meets user and regulatory requirements.

Goal:

- Finalize system design, components, hardware and software specification.
- Integrate hardware and software elements.
- Build and test the prototype of the ADS.

- Perform a robust test and validation process for the safety of the prototype.
- Identify residual risks and failure modes at the system and sub-system level.
- Develop a plan for production and allocate necessary resources.
- Scale up manufacturing process.
- Ensure all produced AV components comply with regulatory standards.

Components of the Product Development Phase

i. Technical safety concept

- Definition: It is the aggregation of technical safety requirements and the system architectural design to provide a better understanding as to why the system architectural design can achieve the safety requirements.
- Source: [33]
- Goal:
    - Design safety considerations should include:
        - Design architecture.
        - Sensors and actuators
        - Communication failures
        - Software errors and inadequate control
        - Collision with environmental objects and VRUs
        - Loss of traction/stability
    - Define technical safety requirements in accordance with functional safety concept and system architectural design.
    - Specify safety mechanisms that can identify/mitigate faults at the output of the system that violate functional safety requirements., and this requirement applies to all ASILs.
    - The design of the component is based on item definition, functional safety concept and prior system architectural design, and this design should implement technical safety requirements.
- Argumentation/Question: based on the information provided above, provide a description of the technical safety concept.
- Evidence/AV Developer Response:

APPENDIX C

TRL TABLE WITH REQUIRED LIST OF STANDARDS

Table 2: TRL Table with Required List of Standards [59]

Technology Development Phase	TRL	Safety Case Claim	Developers are required to show compliance with the following standards and best practices
Basic Research Phase	1	Basic principles and research: The organizational structure and objectives are established, including usage specification(s), ODD, and behavioral competency portfolio for planned deployment(s), and the roadmap of regulations/standards/best practices to which conformance/compliance are planned is complete.	SMS pillar: • ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products [18] • AVSC Information Report for Adapting a Safety Management System (SMS) for Automated Driving System (ADS) SAE Level 4 and 5 Testing and Evaluation [7] • AVSC Best Practice for ADS Remote Assistance Use Case [60] • AVSC Information Report for Change Risk Management [42] • ISO 9001:2015 – Quality management systems [27] • SAE J3018_202012 – Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles [48] • SAE J3206_202107 – Taxonomy and Definition of Safety Principles for Automated Driving System (ADS) [40] • SAE J3247_202403 – Automated Driving System Test Facility Safety Practices [61] • European Commission (EU) regulation 2022/1426 [62] • SAE J3320 – Safety Management System (SMS) application to SAE Level 3, 4, 5 ADS - Equipped Vehicles and Supporting Systems [63] Design Methods pillar:

			• Federal Motor Vehicle Safety Standards [67] • ISO 26262:2018 – Road Vehicles – Functional safety [29] • ISO 21448:2022 – Road Vehicles – Safety of the intended functionality [39] • ISO 34503:2023 – Road Vehicles – Test scenarios for automated driving systems – Taxonomy for operational design domain [68] • ANSI/UL 4600 [18] • AVSC Change Risk Management [42] • SAE J2980_202310 - Considerations for ISO 26262 ASIL Hazard Classification [69] • SAE J3016_202104 – Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles [70] • SAE J3131_202203 – Definitions for Terms Related to Automated Driving Systems Reference Architecture [71] • SAE J3206_202107 – Taxonomy and Definition of Safety Principles for Automated Driving System (ADS) [40] • ISO/TR 4804:2020 – Road vehicles – Safety and cybersecurity for automated driving systems – Design, verification and validation [72] • ISO/CD TS 5083 – Road vehicles - Safety for automated driving systems – Design, verification and validation [44] • SAE J3259 – Taxonomy & Definitions for Operational Design Domain (ODD) for Driving Automation Systems [74] Testing pillar: • ANSI/UL 4600 [18] • ISO 21448 [39] • ISO 26262 [29]

			• ISO 34502:2022 – Road vehicles – Test scenarios for automated driving systems – Scenario based safety evaluation framework [49] • ISO 34503 *[68]* • ISO 34504:2024 - Road vehicles - Test scenarios for automated driving systems - Scenario categorization [78] • SAE J3018 [8] • SAE J3237_202501 – Dynamic Driving Task Assessment (DA) Metrics for Automated Driving Systems [79] • SAE J3247 [61] • SAE J3279 - Best Practices for Developing and Validating Simulations for Automated Driving Systems [80]
	2	Technology concept formulated: The SMS, system design, and V&V plan are all established.	SMS pillar: • AVSC SMS [7] • Section 5 of ISO 26262-2 [30] Design Methods pillar: • SAE J3131 [71] • AVSC Change Risk Management [42] Testing pillar: • SAE J3279 [80]
	3	Proofs of concept: The proof of concept of the SMS and designed system have been validated.	SMS pillar: • AVSC SMS [7] • Section 6 of UL 4600 [18] Design Methods pillar: • Sections 6 and 7 of ISO 26262-5 [33] • Sections 6 and 7 of ISO 26262-6 [34] • ISO 26262-9 [37] • Section 7 of ISO 21448 [39] • Section 14.3 of ANSI/UL 4600 [18]

			Testing pillar: • No standards are required for the testing pillar for this TRL
Safety Case: Preliminary version of the safety case, including SMS implementation, is complete.			To support the claims of the safety case, further arguments include complying with the following set of standards and best practices: • ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products [18] • AVSC Information Report for Adapting a Safety Management System (SMS) for Automated Driving System (ADS) SAE Level 4 and 5 Testing and Evaluation [7] • International Civil Aviation Organization (ICAO) Safety Management Manual (SMM) [3] • AVSC Best Practice for Continuous Monitoring and Improvement after Deployment [43] • ISO 26262:2018 – Road Vehicles – Functional safety [29] • SAE J3206_202107 – Taxonomy and Definition of Safety Principles for Automated Driving System (ADS) [40] • ISO 21448:2022 – Road Vehicles – Safety of the intended functionality [39]
Applied Research Phase	4	Components demonstrated in laboratory environment: The assessment of sub-systems and components in a simulation and/or XiL environment is completed*.	SMS pillar: • No standards are required for the SMS pillar for this TRL Design Methods pillar: • Section 10 of ISO 26262-5 [33] • Section 10 of ISO 26262-6 [34] Testing pillar: • EU Regulation 2022/1426 [62] • SAE J3279 [80]

			• ISO26262-5 [33] • ISO26262-6 [34] • ISO 34503 [68]
	5	System demonstrated in laboratory environment: The assessment of the system in a simulation environment is complete*, and the organization is ready to build a prototype and begin closed course testing.	SMS pillar: • Sections 4 and 5 of SAE J3018 [48] • SAE J3247 [61] Design Methods pillar: • Sections 9, 10, and 11 of ISO 21448 [39] • Section 12 of ANSI/UL 4600 [18] • Section 8 of ISO 26262-4 [32] Testing pillar: • SAE J3237 [79] • Section 16 of ANSI/UL 4600 [18] • ISO 21448 [39] • ISO 34502 [49] • ISO 34503 [68] • ISO 34504 [78] • Section 4 of ISO 21448 [39]
Safety Case: Intermediate version of the safety case, including component and system V&V results from the laboratory environment, is complete and evaluated by an internal auditor.			Organization is required to comply with the following documents: • Section 12 of ANSI/UL 4600 [18] • Section 9 of ISO 26262-8 [36]
Development Phase	6	Prototype demonstrated in relevant environment: The assessment of the prototype in a closed course environment is complete*, and the organization is ready to begin on-road testing with a fallback test driver.	SMS pillar: • SAE J3018 [48] Design Methods pillar: • Sections 4-6 of SAE J3247 [61] • FMVSS [67] Testing pillar:

			• SAE J3247 [61] • SAE J3279 [80]
	7	Prototype demonstrated in operational environment: The assessment of the prototype in an on-road environment (that is within the defined ODD) with a fallback test driver is complete*, and the organization is ready to begin on-road testing without a fallback test driver.	SMS pillar: • Section 4 and 5 of AVSC Best Practice for ADS Remote Assistance Use Case [60] Design Methods pillar: • Section 8 of ISO 21448 [39] Testing pillar: • SAE J3018 [8] • SAE J3279 [80] • SAE J3247 [61]
	8	Technology proven in operational environment: The assessment of the finalized system in an on-road environment (that is within the defined ODD) without a fallback test driver is complete*, and the organization is ready to begin commercial build and deployment.	SMS pillar: • Section 5.3, Component 3 of ICAO SMM [3] • Principles 11 to 15 of SAE J3206 [40] Design Methods pillar: • Section 12 of UL 4600 [18] • Section 9 of ISO 26262-8 [36] • Principles 1 to 10 of SAE J3206 [40] Testing pillar: • SAE J3018 [8] • SAE J3279 [80] • SAE J3247 [61]
Safety Case: Safety case is complete and evaluated by an independent auditor			• SAE J3206 [40] • ANSI/UL 4600 [18] • EU Regulation 2022/1426 [62] • ISO 26262-7 [35] • ISO 21448 [39]

Deployment Phase	9	Technology refined and adopted: The organization is capable of commercially deploying the system in its defined ODD and Usage Specification(s), and there is demonstration of continual improvements in organization, design, and testing/deployment results.	SMS pillar: • Section 5.3 of AVSC SMS [7] • Section 10.3 of ISO 9001 [27] • AVSC CRM [42] • ICAO SMM [3] Design Methods pillar: • Section 8 of ISO 26262-8 [36] • AVSC Continuous Monitoring and Improvement after Deployment [43] Testing pillar: • SAE J3279 [80] • SAE J3247 [61]
Safety Case: Safety case is continuously updated as changes to the organization, design, and test/deployment results occur/become available.			AVSC Continuous Monitoring and Improvement after Deployment [43]