Autonomous Vehicle Risk Management Profile for Traffic Sign Recognition

by
Christine Carlton

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
(Information Systems and Technology)
in the University of Michigan-Dearborn
2023

Master's Thesis Committee:
Assistant Professor Birhanu Eshete, Chair
Professor Hafiz Malik
Associate Professor Anys Bacha

Dedication

I humbly dedicate this thesis with deep gratitude. First and foremost, my profound appreciation goes to God, for the countless blessings that have been bestowed upon me. I extend my heartfelt thanks to my cherished family, whose support has been the cornerstone of my journey. To my beloved husband, Jason, and my children, Amelie, Peyton, Hayden, Zoe, Massimo, Mia, and Kai, your love, patience, and unwavering understanding have been the keys to my success. Jason, your presence in my life is a constant source of inspiration. I must also express my gratitude to my Mother and Father, whose unceasing encouragement and invaluable feedback have guided me throughout life, and throughout my academic journey.

Acknowledgements

First and foremost, I would like to extend deep gratitude to my thesis advisor, Dr. Birhanu Eshete for his guidance, invaluable insight, and significant contributions toward this work. Furthermore, I would like to express my appreciation to the thesis reviewers who have played a pivotal role in shaping the traffic sign recognition user profile. They include: Stephanos Loizou, Justin Teems, Animesh Sarkar, Amanda Fischer, and Amin Fadel. Finally, I would like to offer special thanks to my leadership, Anthony Cataldo and Don Choma, for their consistent support, and push for continued growth.

Table of Contents

List of Tables

List of Figures

Glossary of Key Terms

AI Actors: individuals or organizations who are identified as contributors or users of an AI system, and thus engaged in the risk management processes.

Artificial Intelligence (AI) (a term coined by emeritus Stanford Professor John McCarthy in 1955) was defined by him as "the science and engineering of making intelligent machines" [1]. Today, AI is a focused domain of computer science, studying, designing, and developing machines to learn, and reason as humans do.

Autonomous (Automated) Vehicle a vehicle equipped with automation technology utilized for driving and operating a motor vehicle (as defined in SAE J3016). This applies to vehicles fitted with any level of automation as defined by SAE J3016 [2].

Incident Report is a tool used to document incident details for an organization pertaining to a specific product. It should include how incidents were identified, associated risk (if applicable), incident response strategy, assigned service level agreement (SLA), impact assessment, and communication plans.

Levels of Autonomy are defined in SAE J3016 and range from Level 0 (no driving automation) to Level 5 (full driving automation) in the context of motor vehicles and their operation on roadways [3]. (Referenced in figure 4-8.)

Risk can have a positive or negative impact. Requires identification from stakeholders and further definition on the impact, likelihood, and potential mitigation strategies. NIST's AI RMF states, "negative impact or harm can be experienced by individuals, groups, communities, organizations, society, the environment, and the planet" [4].

Risk Management refers to coordinated activities to direct and control an organization regarding risk [5].

Risk Register is a tool used to collect risk details for an organization, product (or project). It contains valuable details including risk name, risk description, risk owner, likelihood of occurrence, impact, items impacted, probability, anticipated timing, response strategy, and resolution. The risk register is updated throughout the product (or project) lifecycle.

Risk Tolerance "refers to the organization's or AI actor's readiness to bear the risk to achieve its objectives. Risk tolerance can be influenced by legal or regulatory requirements" [4].

Traffic Sign Recognition (TSR) is functionality that provides vehicles or traffic management systems with the ability to recognize traffic signs and respond to them. TSR will vary depending on the vehicle's level of automation, from the ability to display the sign on the dashboard in lower levels of automation, to the ability to identify and respond to traffic signs in higher levels of automation.

Trustworthy AI systems are seen as “valid and reliable, safe, secure, and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed” [4].

Abstract

For decades Americans have been dreaming of driverless automobiles, and today, those dreams are becoming reality; with autonomous taxis operating in cities, and many of today's vehicles equipped with semi-autonomous functionality, i.e., lane keeping, traffic sign recognition (TSR) and object detection. In the next few years, these implementations are expected to expand, with more and more vehicles adapting levels 3+ automation.

Continued development of AI technologies is allowing for advancement in autonomous vehicles. Such advancement provides added safety benefits to consumers. Features such as TSR will help propel the industry into level 5+ automation. Managing the risks for this technology, including the cybersecurity implications is crucial.

To provide guidance on risk management, and in response to the National Artificial Intelligence Initiative Act of 2020, the National Institute of Standards and Technology (NIST) developed the AI Risk Management Framework (AI RMF 1.0). The Framework, which was released in January of 2023, was written as a guideline for organizations in all industries, to utilize in managing risks throughout the AI system lifecycle.

This work leverages the NIST AI RMF framework's functions, categories, and subcategories as a comprehensive guideline. It combines this framework with extensive research on the associated automated technologies to develop a temporal AI risk management use case profile for TSR. This profile is designed to offer valuable insight into effectively managing risk throughout the AI lifecycle, specifically focusing on the associated technologies of the use case.

To ensure accuracy and relevance, we have sought feedback from industry professionals involved in the development of AI technologies for autonomous vehicles.

Additionally, we provide insight gathered from a survey we conducted on the public's perception of the trustworthiness of autonomous vehicles, the helpfulness of TSR, and what organizations can do to enhance the public's comfort level with the adoption of fully autonomous vehicles utilizing TSR functionality.

Keywords: Risk Assessment, Automotive, Autonomous Vehicle, Risk Management Framework, NIST AI RMF 1.0, AI AV Profile, Traffic Sign Recognition, Incident Management, GOVERN, MAP, MEASURE, MANAGE

Chapter 1: Introduction

Managing technological risk is essential. With the rapid expansion of artificial intelligence (AI) products being designed, developed, deployed, and operationalized there are inherited risks that have unique impacts. To provide guidance on risk management, and in response to the National Artificial Intelligence Initiative Act of 2020, the National Institute of Standards and Technology (NIST) developed the AI Risk Management Framework (AI RMF 1.0). The Framework, which was released in January of 2023, is "intended to be voluntary, rights preserving, non-sector specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors and throughout society to implement the approaches in the framework" [4].

Continued development of AI technologies is allowing for advancement in autonomous vehicles (AVs). Such advancement provides added safety benefits to consumers. Features such as TSR will help propel the industry into levels 4+ of automation required for full autonomy. Managing the risks for this technology, including the cybersecurity implications is crucial.

1.1 Approach Overview

This work leverages the NIST AI RMF framework's functions, categories, and subcategories as a comprehensive guideline. It combines this framework with extensive research on the associated automated technologies to develop a temporal AI risk management use case profile for TSR. This profile is designed to offer valuable insight into effectively managing risk throughout the AI lifecycle, specifically focusing on the associated technologies of the use case. It is important to note that the identified risks are expected to evolve over time as technology progresses. To understand the public's perception of AVs, and their trustworthiness, we surveyed

159 individuals. Additionally, to ensure accuracy and relevance, we have sought feedback from industry professionals involved in the development of AI autonomous vehicle technologies.

1.2 Contributions

This work provides overviews of the NIST AI RMF framework and autonomous vehicle architecture and technical stack. We contribute a use case profile on TSR, which has been peer reviewed, and developed utilizing the AI RMF as a foundation. This profile will be provided to the NIST to consider for adoption and publication. In addition, we conducted a survey on the public's perception of the trustworthiness of AVs, their thoughts on whether TSR is a helpful functionality and what organizations can do to enhance their comfort level with to adopt a fully autonomous vehicle that utilizes TSR. Finally, we solicited input from industry professionals on the current risk management processes they employ, suggestions for improving the content provided in the profile, and feedback on if additional AI RMF profiles would assist their organizations in improving current risk management practices.

1.3 Thesis Organization

The rest of the thesis is organized in the following way: Chapter 2 provides an overview of the NIST RMF framework, with special emphasis on the AI RMF Core. Chapter 3 provides an overview of the autonomous vehicle's architecture and technical stack. Chapter 4 introduces a use case profile on TSR. Chapter 5 shares feedback obtained from industry professionals, and Chapter 6 concludes the paper with suggestions for future research opportunities.

Chapter 2: NIST AI Risk Management Framework

2.1 Framework Overview

AI provides a substantial opportunity to transform society, through technological advances in commerce, health, transportation, and cybersecurity, to name a few. Like other technologies, opportunities in AI can also introduce risks. The NIST AI Risk Management Framework (RMF) is developed as a sector and technology agnostic framework to help address this risk. The Framework is designed to assist organizations and individuals "with approaches that increase the trustworthiness of AI systems, and to help foster the responsible design, development, and use of AI systems over time" [4].

The Framework was created as a collaborative effort, over an 18-month period with more than 240 contributing organizations from all sectors (private, academia, government) [6]. The Framework is intended to be a living document, with a regular revision cycle. To avoid duplication with other resources, NIST crosswalks between previous works, referencing developed material from ISO, OECD and NIST.

Before the RMF can be applied, NIST sets a foundation for the actors using the Framework, by outlining foundational information. In section 1, explanations are provided regarding how to frame risk, how to address the risks, impacts and harms, the challenges for risk management, risk tolerance, risk prioritization, organizational integration, and management of risks. The explanations are broad, providing an opportunity to be widely utilized, with the goal of developing more trustworthy AI systems. Key definitions from this section, and remaining sections of this work can be found in the glossary.

2.1.1 AI Risks and Trustworthiness

Risk is inherent to AI, making trustworthiness a critical factor in development and deployment of AI systems. The AI RMF 1.0 provides approaches that will enhance AI trustworthiness, while also seeking to reduce negative risks. Firstly, the AI RMF defines characteristics of trustworthy AI systems, which include: "valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful biases managed" [4]. These characteristics (which are each further defined in the Framework) require human judgement when defining the metrics associated, combined with the organization's defined threshold for risk. As part of the GOVERN function that is further explained in section 2.2, (GOVERN subcategory 1.2), these characteristics are integrated into the organizational polices, processes, procedures, and practices.

2.2 Applying the AI Risk Management Framework

To apply the framework, we look to the AI RMF Core, which highlights activities that should be continuously conducted throughout the entire AI system's lifecycle. The Core, which is illustrated in figure 2-1, consists of four functions: GOVERN, MAP, MEASURE and MANAGE. Each of these functions is broken down further into categories and subcategories, with 19 categories and 72 subcategories in total. The functions can be implemented iteratively in any order, but typically begin with the GOVERN function, then continue with MAP, MEASURE and MANAGE. The following sections will provide a high-level overview of each. (Additional detailed information, which is included in our TSR use case profile, can be found in the AI RMF Playbook.)

Figure 2-1 shows the GOVERN function at the center of the AI RMF Core. This is intentional as governance is designed to be part of each function and a function of its own. It is

infused within the other functions as it is critical for cultivating a culture of risk management within an organization. An organization's governing authority will provide helpful oversight in the "organization's mission, goals, values, culture and risk tolerance" [4]. All of these are critical for AI actors participating in the development, deployment, and operations of AI systems to have a clear understanding of the expectations, policies, and procedures for risk management strategies.

Figure 2-1: AI RMF Core [4]

With governance in place, AI actors can begin to conduct the MAP function. This function is key to identifying and defining the risks associated with the AI system. All AI actors should be involved in the MAP function as AI actors may have varying levels of visibility to risks introduced into the system, throughout the system's lifecycle. The risks identified in this function can be negative or positive. The key here is incorporating diverse perspectives, from the designers and the developers to the End Users and members of the community. Once properly conducted, this function will trigger a "Go-No-Go" decision about whether the system should be further developed or deployed. If the decision is to continue, then users of the framework would typically move into the MEASURE and MANAGE functions, while continuing to apply the GOVERN function.

Using the MAP function as an input, “the MEASURE function employes quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts” [4]. The key to the MEASURE function is tracking metrics that impact trustworthiness, society, and human-AI configurations. The outputs of this function (which will be utilized in the MANAGEMENT function) include “objective, repeatable, or scalable test, evaluation, verification and validation (TEVV) processes including metrics, methods, and methodologies” [4].

The MANAGE function takes all the work completed in the MAP and MEASURE function, and allocates the risk as defined by the GOVERN function. Through the MANAGE function, risk monitoring, and treatment (if required) is conducted; this includes facilitating the plans and strategies for overseeing the risks, such as mitigation measures, and communication plans. As with all aspects of risk management, and the AI RMF core, these actions need to be continuous throughout the lifecycle. The AI RMF states, “It is incumbent on Framework users to continue to apply the MANAGE function to deployed AI systems as methods, contexts, risks, and needs or expectations from relevant AI actors evolve over time” [4].

2.3 AI RMF Profiles

To further develop the AI RMF, NIST is encouraging organizations and individuals to produce AI RMF Profiles. These profiles are implementations of the AI RMF Core functions (GOVERN, MAP, MEASURE and MANAGE) along with their associated categories and subcategories for specific applications. “Profiles may illustrate and offer insights into how risks can be managed at various stages of the AI lifecycle or in specific sector, technology or end-use applications” [4]. NIST identifies 3 types of profiles:

- Current – How AI is currently being managed and the related risks in terms of current outcomes.
- Target – Indicates the outcomes needed to achieve the desired AI risk management goals.
- Temporal – Descriptions of the current state, or the desired target state of AI risk management activities within a given sector, industry, organization, or application.

To progress profile development, NIST formed a public working group with the intent of creating a profile for Generative AI. To continue to grow in knowledge and understanding of the framework, and approaches for profile development, I joined this group, serving as a contributor for the Governance and Incident sections. Based on experiences from that group, and learnings from how the GenAI profile was developed, the TSR profile was adjusted to better reflect the NIST profile, and format provided in the AI RMF Playbook.

The following two chapters of our work focus on the development of a temporal profile on TSR utilized in AVs, with automation level 3+. Chapter 3 explains the technical stack and architecture of AVs. Chapter 4 includes our contribution of a use case profile, with tailored guidance for risk management as it pertains to TSR. To complete this use case, we employ NIST's AI RMF playbook, an accompaniment to the AI Risk Management Framework to implement the AI RMF Core, GOVERN, MAP, MEASURE and MANAGE.

Chapter 3: Autonomous Vehicle Architecture

3.1 Autonomous Vehicles Overview

The world has been dreaming driverless, since 1925, when Francis P. Houdina sent his driverless car, "American Wonder" through the streets of New York City. In 1939, General Motors (GM) presented a scale model of a futuristic city where the cars drove themselves. GM predicted at that time that the technology would be available before 1960 [7]. Sixty years later, we are now seeing autonomous taxis operating in cities, with many of today's vehicles equipped with semi-autonomous functionality, i.e., lane keeping, traffic sign recognition and object detection. In the next few years, these implementations are expected to expand, with more and more vehicles adapting levels 3+ automation.

Advances in AI have provided tools for further development of AVs with the intent to enhance road safety and provide traffic efficiency [8]. The system architecture of the AV is comprised of four sections that include perception, decision and planning, control, and chassis. The following section provides a high-level overview of the system architecture with specific emphasis on the perception layer, before explaining the technical stack that serves as the inputs (the eyes and ears) for perception and sensing.

3.2 System Architecture

Considerable progress has been made on the AV system architecture, encompassing both software and hardware components. Representative architectures can be found in [9], [8] and [10]. In the context of this work, we reference work conducted in [11] as indicated in figure 3-2. The

emphasis of this design is on the system architecture, while the vehicle hardware plays a supporting role. The system architecture is divided into four sections, each dedicated to specific tasks, namely perception, decision and planning, control, and chassis.

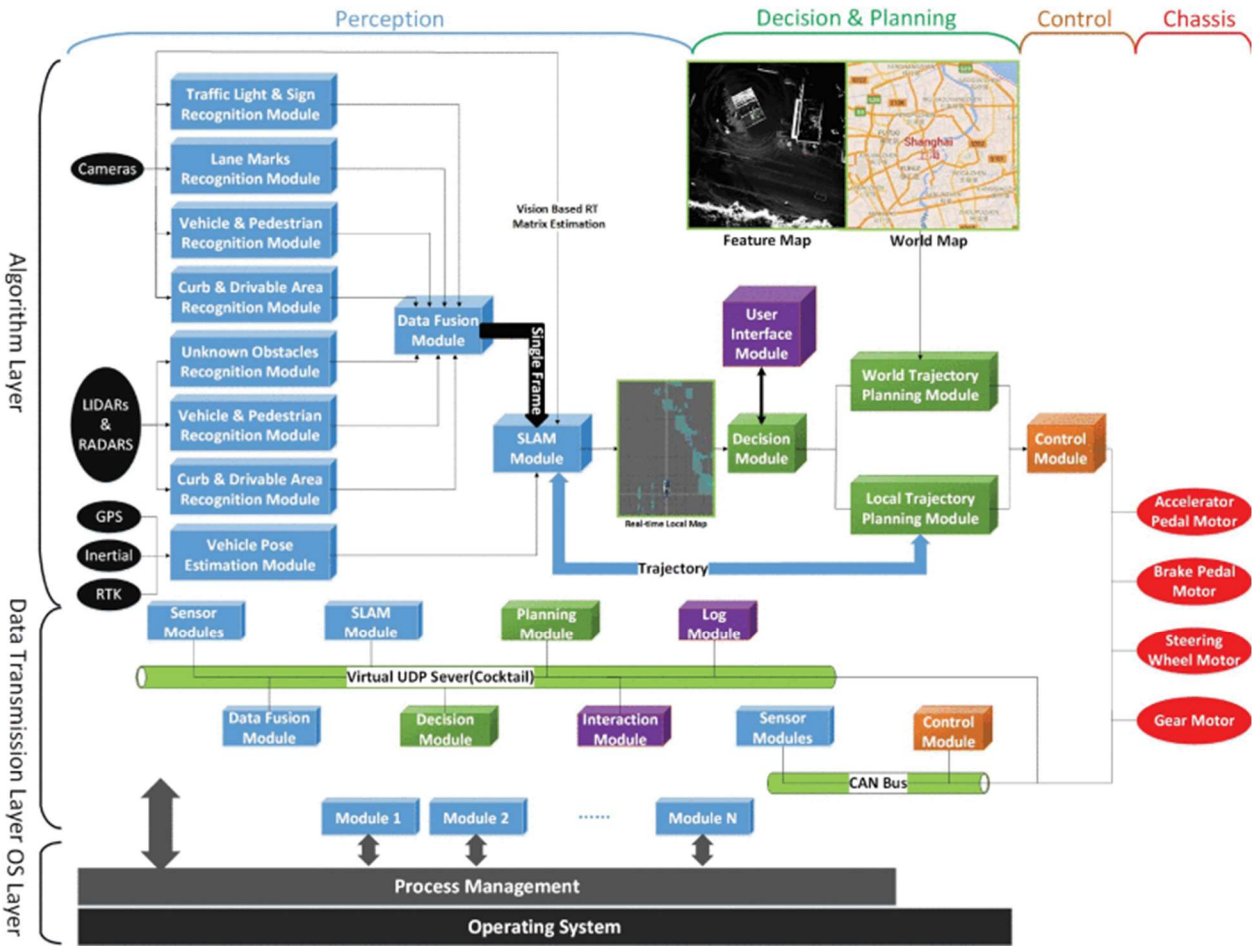

Figure 3-2: Autonomous Vehicle System Architecture [11]

3.2.1 Perception

Through the perception task, the autonomous vehicle senses its surroundings. The AV requires in-depth analysis of the environment to make decisions for safe navigation [8]. Perception uses the vehicles hardware (cameras, LIDAR, Radar, etc.) to see and perceive vehicles, pedestrians, lane markers and traffic signs [11]. Information obtained from the sensors is processed in the data fusion module. Figure 3-3 illustrates how the perception model works for object

detection. The authors of [8] divide the processes in the perception framework into three main steps which include object detection from the sensor (camera and LiDAR), calibration between the two, and fusion mapping, stating that the core idea of perception is to determine the location and size of an object in a dynamic environment. This is key in identifying and recognizing objects, including traffic signs, pedestrians, cyclists, or other vehicles.

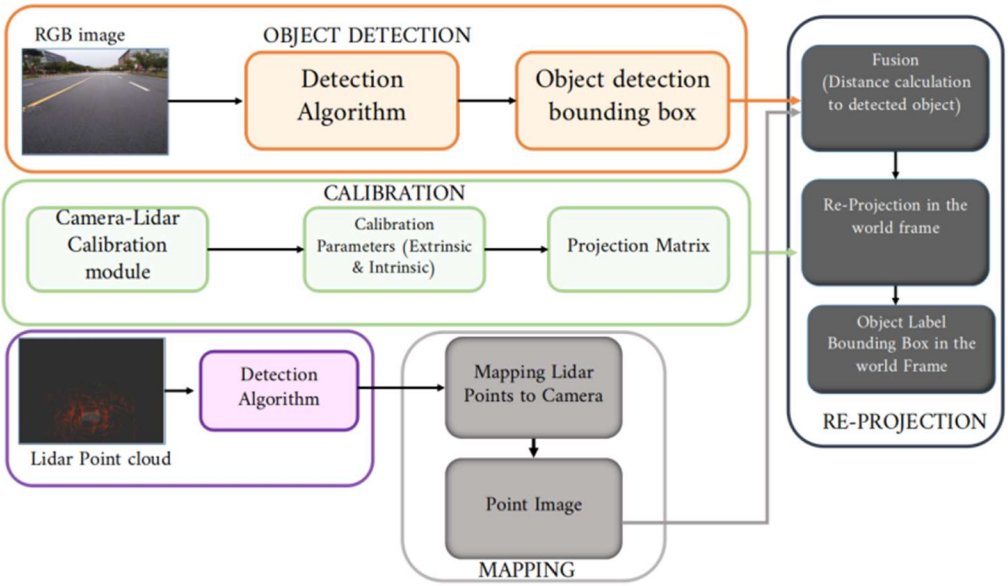

Figure 3-3: Example Perception Framework [8]

3.3 Technical Stack

The perception layer leverages the vehicles infrastructure as inputs which are essential for the recognition and estimation modules. To delineate the risks associated with the infrastructure of AVs the authors of [12] identified three key elements of the AV technical stack and their corresponding components, as illustrated in figure 3-4. These elements encompass the automotive control system, autonomous driving system components and vehicle to everything (V2X) communication. The context of use for each of these systems is defined below.

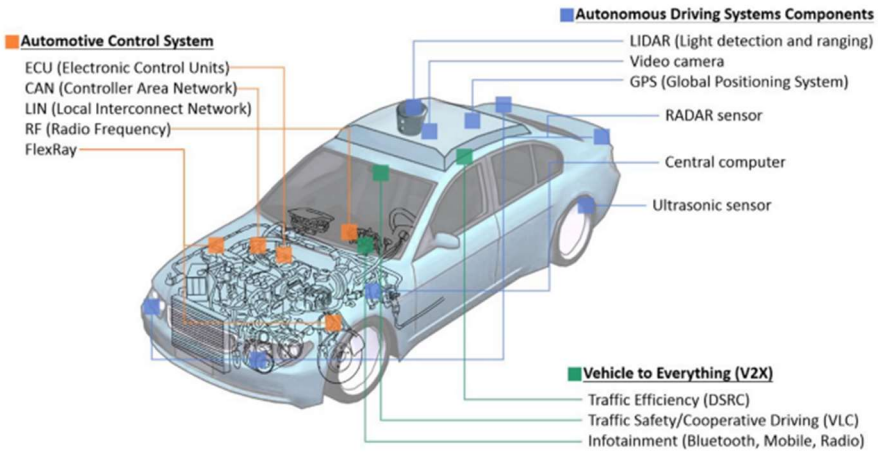

Figure 3-4: Key Elements of Autonomous Vehicles [12]

3.3.1 Automotive Control System

The automotive control system contains key components that are seen in all vehicles. This includes the electronic control unit (ECU). "An ECU is an embedded computer in vehicles to control mechanical or electronic systems or subsystems. It is an electronic component, which takes an input from its sensors or other ECUs and uses actuators to control functionalities of the vehicle. For example, the Anti-Lock Brake System (ABS) can take inputs from the Powertrain Control Module (PCM) to check whether the traction control is working" [13]. Today's vehicles contain millions of lines of code, which execute on 70 to 100 ECUs throughout the vehicle [14].

To enable input transmission and reception, an ECU utilizes various network protocols such as the Controller Area Network (CAN), Local Interconnect Network (LIN) and FlexRay. These protocols offer distinct functionalities and performance levels tailored to meet the specific needs of the systems they serve. For instance, FlexRay is better suited to manage the vehicle's more crucial functionalities, as elaborated below.

CAN (originally introduced in 1986) remains the prominent communication protocol between ECUs. It is a broadcast communication bus system, which allows ECUs to communicate with one another [15]. The CAN system includes a communication network system, gateway unit, vehicle power distribution and failure diagnosis system. Breaking these down further, they also contain and support the engine electronic control system, automatic transmission control system, automatic transmission control system, etc. [16].

The Local Interconnect Network (LIN), was standardized in 2000, is used to support subsystems, i.e., all the functionality of a door (window control, door locks, etc.). LINs are interconnected with the CAN through a LIN/CAN gateway. They are a lower cost, simpler solution to complement the existing portfolio of automotive networks, providing communication needed for non-safety related subsystems [17] [18].

FlexRay is the newest of the vehicle network communication protocols. It was designed to support future automotive application needs. FlexRay supports higher data rates, up to 10 Mbps, compared to the 1 Mbps supported by CAN [19]. It is the most expensive of the bus systems, with the ability to support real-time, critical applications, including steer-by-wire, drive-by-wire, brake-by-wire, adaptive cruise control, etc. [20].

3.3.2 Autonomous Driving System Components

The Autonomous Driving System Components include LIDAR, camera, GPS (maps), radar, ultra sonic sensor, and central computer; with the LIDAR, camera, GPS, radar, and ultra sonic sensors serving as data sources that are input into the central computer, processed through the perception layer. Each component and their basic functionality are expanded upon below.

LIDAR sensors (Light Detection and Ranging) use lasers to map the environment [21]. In [22], the authors explain that there are varying approaches to deploying LIDAR, and propose a neutral overview of the technology, that we will reference here. The authors reference the three principles that LIDAR uses for measurement, which include pulsed, amplitude modulation of a continuous wave (AMCW) and frequency-modulated continuous wave (FMCW). The first principle uses time-of-flight (TOF) to measure depth by counting time delays in events in light emitted from a source. To do this, an optical signal is projected through pulses onto a target and the reflected signal is detected and processed to determine the object's distance. The second principle AMCW, uses the intensity modulation of a continuous light wave. After reflection is received from the target, the detector collects the signal, and a phase meter measures the distance. The final principle FMCW emits instantaneous optical frequency that is periodically shifted by varying the power applied to the source. The reflected signal combines with the emitted source, resulting in a beat frequency that serves as an indicator of the probe distance. The differences and use cases for the three principles are highlighted in table 3-1.

Table 3-1: The Three LIDAR Principles [22]

	Pulsed	**AMCW**	**FMCW**
Parameter measured	Intensity of emitted and received pulse	Phase of modulated amplitude	Relative beat of modulated frequency, and Doppler shift
Measurement	Direct	Indirect	Indirect
Detection	Incoherent	Incoherent	Coherent
Use	Indoor/Outdoor	Only indoor	Indoor/Outdoor
Main advantage	Simplicity of setup; long ambiguity range	Established commercially	Simultaneous speed and range measures
Main limitation	Low SNR of returned pulse	Short ambiguity distance	Coherence length/Stability in operating conditions (e.g., thermal)
Depth resolution (typ)	1 cm	1 cm	0.1 cm

Note: maximum attainable range has been avoided as it needs the definition of several other parameters (instantaneous FOV, reflectivity of target, eye safety level, etc.)

Vehicle video cameras are utilized to read traffic lights, road signs, and detect objects, i.e., pedestrians and obstacles [12]. The authors of [23] provide an overview of various aspects of in-vehicle cameras. Highlighting the five types of cameras, which are deployed based on their application and include a front view camera, side-view camera, rear-view camera, surround-view camera and built-in camera. The type of camera selected varies by application. The risk assessment provided in this work will focus on camera types used for the TSR use case, and thus exclude the risks associated for use cases utilizing the built-in camera. Figure 3-5 provides an overall architecture for cameras and how they collect and feed data through the perception algorithm to support the applications.

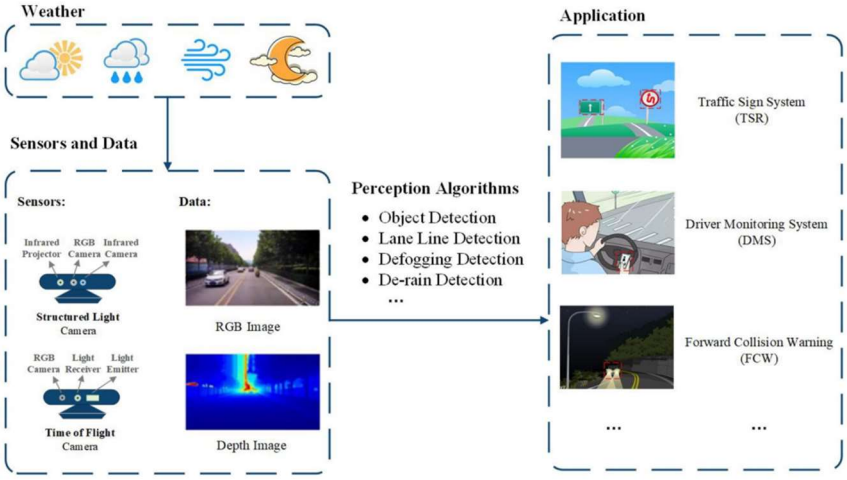

Figure 3-5: Camera Architecture [23]

Global positioning systems (GPS) is a developed technology used by consumers to provide mapping services, asset tracking, and more. AVs rely on GPS to determine their precise location by receiving signals transmitted from three or more GPS satellites. By measuring the distance to each satellite and knowing their respective positions, the vehicle utilizes trilateration to calculate its own position [12]. GPS satellites provide a critical role in automated functionality, providing vehicles with information to determine directions to destinations, speed limits, and vehicle speed.

In [12], the authors elaborate on the utilization of radio detection and range (RADAR) by AVs to instantly perceive their surroundings. RADAR sensors function as a detection system, utilizing radio waves to gauge the distance, direction, angle, and velocity of a target, although it offers slightly lower accuracy than LIDAR. Despite being known for accuracy in inclement weather, RADAR encounters certain drawbacks, such as limited pedestrian and object detection capabilities, along with the possibility of false alarms triggered by interference from other vehicles [24].

Ultra-sonic sensors (also known as sonar), measure the position of objects through echolocation to determine if they are in the range of the sensor. This is accomplished by using the time taken by the signal to come back to the sensor after emitting it, as illustrated in figure 3-6 [25]. Ultra-sonic sensors are seen as being a cheaper alternative to LIDAR [21]. The authors of [21] explain that the sensors send ultrasonic impulses, which are reflected from the nearby obstacle, back to the vehicle where it is processed by the vehicle to determine the data associated with the obstacle. The range for ultra sonic sensors is 5.5 meters and are typically used for functions such as park assist.

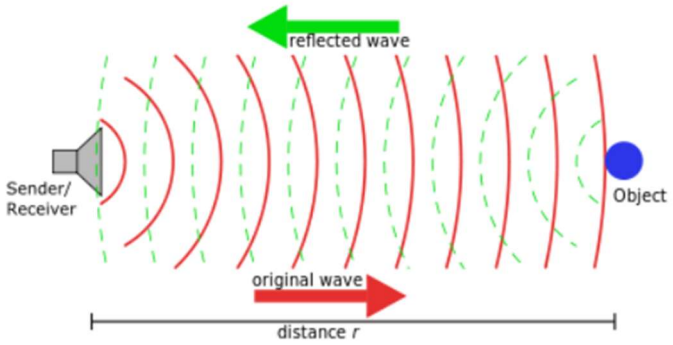

Figure 3-6: Echolocation used by ultrasonic sensor [25]

The central computer brings all the infrastructure together by serving as the core intelligence. In [12], the authors explain that it acquires comprehensive data from the sensors and

leverages this information to regulate crucial aspects such as steering, acceleration and brakes. Within the central computer resides the vital software responsible for interpreting sensor signals and discerning road conditions. Hence, the computer necessitates robust computing power to accommodate the vast influx of sensor data and vehicle to everything (V2X) communications. Centralized computers must exhibit reliability and responsiveness surpassing that of human drivers, as they process colossal amounts of data with minimal latency. Notably, table 3-2 showcases the typical data size generated by various autonomous vehicle sensors per second [26].

Table 3-2: Typical data size generated by AV sensors per second [26].

GPS	50 kB
Ultrasonic radar	10–100 kB
Camera	20–40 MB
Radar	10–100 kB
LiDAR	10–70 MB

3.3.3 Vehicle to Everything

Leveraging vehicle-to-everything (V2X) communication offers significant potential for enhancing perception capabilities by enabling the exchange of information among vehicles and roadside units. The real time data, when coupled with dynamic mapping enhances the vehicle's awareness of its surroundings. It proves particularly valuable in situations where there is no direct line of sight, leading to improved traffic efficiency, and reduction accident rates [26].

In their work [27], the authors explain the two main types of communication technologies used for V2X: Dedicated Short-Range Communication (DSRC) and Long-Term Evolution for V2X (LTE-V2X). The DSRC encompasses IEEE and SAE standards and utilizes the 802.11p protocol at the physical and MAC layers. The simplifies authentication and enables vehicles to broadcast security information directly to nearby vehicles, pedestrians, and objects. On the other

hand, LTE-V2X is a wireless communication within the LTE mobile network defined by 3GPP. It encompasses two types of working modes of cellular communication, Uu (cellular to vehicle) and PC5 (direct vehicle to vehicle communication). Both communication technologies contribute significantly to improving traffic safety, facilitating cooperative driving, and improved traffic efficiency.

Chapter 4: Use Case Profile for Traffic Sign Recognition

TSR is a required functionality of AVs. TSR works by using the visual information, provided from vehicle cameras, including color segmentation (or adjustment), selection of region of interests (ROI) where traffic signs are present, and identification of traffic signs through data classification conducted via deep learning (DL) [28]. Figure 4-7 highlights a sample of a possible TSR architecture, and associated processes.

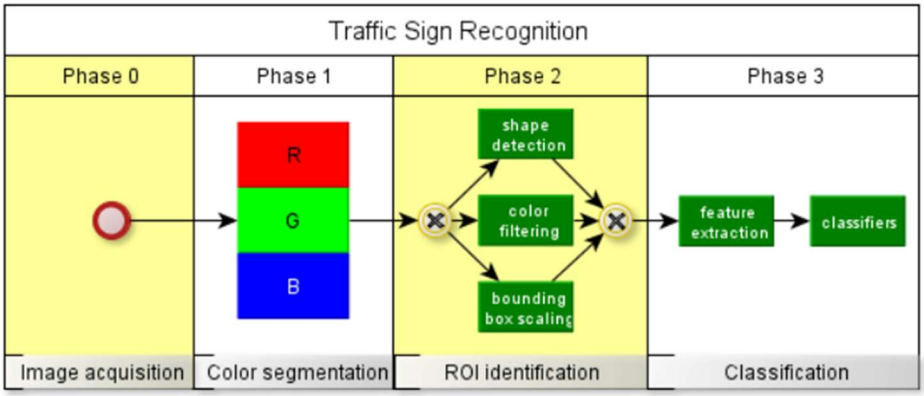

Figure 4-7: Traffic Sign Recognition Architecture [28]

The following section applies NIST's AI RMF functions, categories, and subcategories into the context of TSR. The profile specifically addresses the risks associated with the integration of AI systems within AVs' TSR function in automation levels 3+. To accommodate ongoing technological advancements, our profile adopts DL model agnostic approach. The temporal nature of our profile provides desired state of AI risk management activities related to TSR utilization.

The profile is organized into several sections, each addressing a crucial aspect of the topic:

- **Section 1**: This section highlights the GOVERN function and outlines how Original Equipment Manufacturers (OEMs) can effectively implement governance processes to mitigate AI risks within the TSR system.
- **Section 2**: Here, a comprehensive review of product specifications for TSR is presented. This includes a comparison of current functionalities with anticipated future capabilities, along with end user expectations.
- **Section 3**: Focuses on TEVV (Testing, Evaluation, Verification, and Validation) for the TSR system; this section encompasses all integral components such as software, hardware, data, models, and cybersecurity considerations.
- **Section 4**: Delving into TSR risk management, this section provides an in-depth risk table accompanied by corresponding risk management strategies. This detailed analysis offers a comprehensive understanding of risk mitigation.
- **Section 5**: Concluding the profile, this section provides valuable insights into TSR incident response protocols. These details serve to guide effective actions in the event of TSR-related incidents.

SECTION 1: TSR Governance

The following section provides comprehensive guidance for organizations to reference when governing risk associated with TSR throughout the entire lifecycle encompassing design, development, deployment, operation, monitoring, and test, evaluation, verification, and validation. The section shares the known legal and regulatory requirements, along with key recommendations for organizations implementing TSR systems to incorporate risk management into their processes and procedures, including roles and responsibilities of common AI actors. The governance section

should be revisited regularly throughout the entire TSR system lifecycle, and through the AI RMF core processes, including MAP, MEASURE and MANAGE.

GOVERN 1.1

ABOUT

TSR systems deployed in AVs need to ensure legal and regulatory compliance and follow industry best practices pertaining to the design, development, and deployment of the systems. Automotive OEMs should consult with their Compliance Experts and Auditors teams throughout the entire TSR lifecycle process. As AI actors, these legal professionals are responsible for providing valuable insights and conducting regular reviews to ensure adherence to applicable laws, regulations, and best practices, thus reducing the associated AI risk. Table 4-3 provides an overview of prominent US national requirements, and voluntary standards that should be considered. Additional legal requirements vary by state and would require a much further detailed analysis by an OEM's legal team, Compliance Experts and Auditors.

Table 4-3: Applicable Legal and Regulatory Requirements

Issued By	Legal/Regulatory Requirement Name	Details	References
National Highway Traffic Safety Administration	49 CFR Part 571 Docket No. NHTSA-2021-0003 RIN 2127-AM06 Occupant Protection for Vehicles with Automated Driving Systems	Advanced Driver Systems must provide the same levels of occupant protection that current vehicles provide.	[29]
The International Organization for Standardization (ISO)	ISO: 23150 Road vehicles — Data communication between sensors and data fusion unit for automated driving functions — Logical interface	Specifies the logical interface between vehicle perception sensors, i.e., a camera used for TSR, and the fusion unit which interprets the data.	[30]
The International Organization for Standardization (ISO)	Intelligent transport systems – Geographic Data Files ISO: 20524-1: 2022 Part 1: Application independent map data shared between multiple sources. ISO: 20524-2: 2022	Specifies conceptual and logical data model and physical encoding formats for geographical databases for intelligent transport	[31]

	Part 2: Map data used in automated driving systems, Cooperative ITS, and multi-modal transport.	systems applications and services. _____ Defines the map database exchange format for ITS utilized in vehicle navigation systems. (This includes the capabilities required to use map databases, which could assist TSR systems.)	[32]

SUGGESTED ACTIONS

- Ensure all applicable AI and autonomous vehicle regulations are documented and reviewed by Compliance Experts.
- Share regulation information with AI actors involved in design, development, deployment, and testing of the system.
- Develop detailed training of legal and regulatory requirements for AI actors.
- Ensure all AI actors participate in training and confirm knowledge of legal and regulatory requirements.
- Audit compliance with training completion, and compliance with legal and regulatory requirements.
- Organizational Management should provide support to ensure adherence.
- Define who the directly responsible individual (DRI) is for communicating changes and updates to regulations.
- Define who the DRI is if the TSR system is found to be in non-compliance with a legal requirement.

TRANSPARENCY AND DOCUMENTATION

- Are all legal and regulatory requirements documented?

- Have roles and responsibilities been clearly defined?
- How will new legal and regulatory requirements be communicated to AI actors?
- Who is responsible for communicating new risks to the assigned AI Risk Management Specialist?
- What is the documented process to amend violations?
- At what frequency will violation status be communicated until resolution?

GOVERN 1.4

ABOUT

Organizations developing and deploying TSR systems must review established risk management processes, procedures, and other required controls, and if needed make additions, or adjustments. Effective documentation management plays a crucial role in risk and incident management and will be more effective if it is standardized across the organization. This section provides guidance on documentation management policies, including the risk register, and other TSR project/product related documents, last day of support (LDOS) artifacts, and public disclosures.

SUGGESTED ACTIONS

- Ensure all documents are stored in a centralized repository accessible to all AI actors involved in the TSR system's product lifecycle.
- Confirm that as products reach their retirement phase, all risk-related documents (including the risk register) are relocated to the organization's electronic data management system (EDMS) as per the timeline set forth by the organization's Compliance Experts.

- Uphold regulatory standards by having Auditors review the filed documents.
- Assign an AI Risk Management Specialist to maintain the risk register, which includes all the identified risks, their likelihood, impact, risk score, responsible actors, systems impacted and strategy. The risk register is a dynamic document requiring regular updates throughout the TSR system's lifecycle.
- Make sure as part of TSR development, project documentation, including project charter (with business justification, assumptions, and limitations), communication plan, change management plan, project status reports, RACI matrix, stakeholder (AI actor) contact list, lessons learned documents and project closure reports are be maintained in a central repository available to all relevant parties. Projects utilizing the Waterfall methodology should also maintain project plans, a work breakdown structure, and a Gantt chart. Projects utilizing the Agile methodology should maintain backlogs, user stories, burndown charts, release plans, roadmaps, worklogs and retrospective action items.
- Communicate clear expectations for documentation to AI actors involved in the design, development, testing, evaluation, and maintenance of the TSR system. For example: Software Developers must properly comment on all methods, classes, and functions. Model Engineers must describe and characterize the training data, and reasoning behind feature selection. TEVV Experts record all tests plans, expected results and actual results, including explanatory visualizations.
- Establish a retention period (as required for legal and regulatory requirements) for all TSR product documentation, including those that have become last day of support (LDOS).

- Disclose relevant information on how the TSR system has been designed, developed, maintained, and evaluated to the public to solicit trust. In our public survey, 57.2% of users reported that product specifications would make them feel more comfortable with adoption, with 56.6% of users stating they would feel more comfortable viewing test data, and 64.2% of users wanting to have an understanding all AI risks associated with TSR. (additional details on the survey can also be found in the appendix of this thesis). Public disclosure should happen at regular intervals, beginning with the product launch. In addition, automotive OEMs should include detailed product specs, including instructions for use in the automobile's manual.

TRANSPARENCY AND DOCUMENTATION

- Are roles and responsibilities for document management clearly understood?
- How will documentation management be monitored?
- Have retention periods been established by or reviewed with Compliance Experts?
- Are document repositories pre-existing, or will new ones need to be established?
- Has the organization determined what information will be disclosed publicly?
- Are intervals for public disclosure clearly defined?
- Does the information have the appropriate security classification (public, proprietary, confidential, or secret)?

GOVERN 1.5

ABOUT

Risk management requires ongoing monitoring, with the risk register remaining a dynamic document. Continuous monitoring of risks throughout the TSR lifecycle will ensure that risks are closed when they are no longer likely, and added when new risks are identified. All AI actors are responsible for reporting risks that they discover to the assigned AI Risk Management Specialist. As the impact of a TSR malfunctioning in an AV could result in loss of human life, risks are expected to be communicated immediately.

SUGGESTED ACTIONS

- Ensure risks that could impact End Users, which have not been properly mitigated, are communicated to the public, until a solid mitigation plan has been developed and deployed.
- Establish communication channels between End Users and the organization. End Users and the public should be able to immediately report incidents, for further investigation. Incidents that are defined to have a high impact, such as loss of human life, should be communicated with all AI actors, End Users, and the public. It is recommended that this communication include immediate remediation strategies (when possible), until a full mitigation plan has been developed and deployed.

TRANSPARENCY AND DOCUMENTATION

- Are roles and responsibilities for risk monitoring clearly documented and communicated?
- Have communication channels between End Users and the organization been established?

GOVERN 2.1

ABOUT

TSR risk management requires organized accountability structures, to ensure that AI actors are engaged throughout the TSR system's lifecycle and have a clear understanding of their responsibilities as it pertains to mapping, measuring, and managing risks associated with their expertise. As a developing AI technology, the list of AI actors involved in the design, development and deployment of TSR is evolving. Table 4-4 outlines standard expected AI actors, their associated roles and responsibilities, and the stages of the AI lifecycle where they should be engaged in the risk mapping and measuring process. It is assumed that AI actors will vary by organization, and the table below provides suggestions only, and not meant to be prescriptive.

Table 4-4: Traffic Sign Recognition AI Actors

AI Actor	Roles and Responsibilities	AI Lifecycle Stage
Organizational Management	● Responsible for decisions about risks associated with TSR's AI system lifecycle. ● Define organizations risk tolerance. ● Give input into risk register. ● Provide feedback on AI risk measurements. ● Contribute to risk mitigation strategies.	● Plan and Design ● Deploy and Use ● Operate and Monitor ● Use or Impacted By
AI Risk Management Specialist (Impact Assessors)	● Collaborate with AI Actors to define and document risks. ● Record and communicate defined risk tolerance. ● Maintain risk register. ● Document risk management strategies and measurements. ● Maintain a document of all AI actor's contact information. ● Communicate with all stakeholders regularly. ● Oversee AI risk management training.	● Plan and Design ● Deploy and Use ● Operate and Monitor ● Use or Impacted By
Product Manager Product Owner	● Gather, understand, manage, and communicate end user requirements. ● Formulate and maintain product vision. ● Provide leadership to the cross functional team through the entire TSR AI system lifecycle. ● Develop and manage communications plans. ● Give input into risk register. ● Provide feedback on AI risk measurements. ● Provide ongoing risk monitoring. ● Contribute to risk mitigation strategies.	● Plan and Design ● Deploy and Use ● Operate and Monitor ● Use or Impacted By
UI/UX Designers	● Gather, understand, manage, and communicate end user requirements. ● Evaluate user experience. ● Give input into risk register. ● Provide feedback on AI risk measurements.	● Plan and Design ● Deploy and Use ● Operate and Monitor ● Use or Impacted By

	• Provide ongoing risk monitoring. • Contribute to risk mitigation strategies.	
Domain Experts	• Provide subject matter expertise. • Identify opportunities for TSR improvement. • Give input into the risk register. • Provide feedback on TSR risk measurements.	• Plan and Design • Collect and Process Data • Build and Use Model • Verify and Validate • Deploy and Use • Operate and Monitor
AI Designers	• Document system requirements. • Design AI systems. • Provide operation and monitoring support. • Give input into the risk register.	• Plan and Design • Operate and Monitor
TEVV Experts	• Examine TSR system for defects and potential problems. • Validate AI system model. • Validate AI system components. • Provide ongoing risk monitoring. • Oversee TSR system testing. • Give input into the risk register.	• Plan and Design • Collect and Process Data • Build and Use Model • Verify and Validate • Deploy and Use • Operate and Monitor
Product Developers	• Mitigate risk through detailed design considerations. • Provide detailed design documentation. • Give input into the risk register. • Provide feedback on AI risk measurements. • Contribute to TEVV.	• Plan and Design • Collect and Process Data • Build and Use Model • Verify and Validate • Deploy and Use
Data Scientist	• Gather, validate, and clean data for model development. • Give input into the risk register.	• Collect and Process Data • Build and Use Model • Verify and Validate
Data Engineers	• Build and maintain required databases and associated structures. • Give input into the risk register.	• Collect and Process Data
Human Factor Experts	• Ensure TSR design is user centric by conducting detailed analysis on user behaviors and potential sources of error. • Conduct a detailed task analysis to identify potential risks. • Document the workflow design. • Contribute to end user training and documentation. • Evaluate user experience. • Give input into the risk register. • Provide feedback on AI risk measurements. • Provide ongoing risk monitoring. • Contribute to risk mitigation strategies.	• Collect and Process Data • Deploy and Use
Model Engineers	• Select algorithms. • Train models. • Validate model output. • Give input into the risk register. • Provide feedback on AI risk measurements. • Provide ongoing risk monitoring.	• Build and Use Model • Verify and Validate
System Integrators	• Contribute to integration documentation and testing, compliance testing and validation. • Give input into the risk register. • Provide feedback on AI risk measurements. • Provide ongoing risk monitoring.	• Deploy and Use
Systems Engineers	• Incorporate risk mitigation into system design.	• Deploy and Use

	• Give input into the risk register. • Provide feedback on AI risk measurements. • Contribute to risk mitigation strategies. • Participate in TEVV. • Provide ongoing risk monitoring.	
Software Engineers	• Incorporate risk mitigation into software design. • Incorporate error handling into the AI system design. • Give input into the risk register. • Provide feedback on AI risk measurements. • Contribute to risk mitigation strategies. • Participate in TEVV. • Provide ongoing risk monitoring.	• Deploy and Use
Procurement Experts	• Conduct risk assessment on third party suppliers. • Give input into the risk register. • Contribute to risk mitigation strategies. • Provide ongoing risk monitoring.	• Plan and Design • Deploy and Use
Evaluators	• Review AI models to ensure results are as expected through ongoing data analysis. • Give input into the risk register. • Contribute to risk mitigation strategies. • Communicate with AI Risk Management specialist and other key stakeholders. • Provide ongoing risk monitoring.	• Operate and Monitor
Compliance Experts	• Maintain database of legal and regulatory requirements. • Give input into the risk register. • Provide feedback on AI risk measurements. • Communicate regularly with all AI actors. • Contribute to training of organizational staff, including those working on design, development, and deployment activities.	• Plan and Design • Operate and Monitor
Auditors	• Audit design, development, deployment, and operations. • Ensure compliance with legal and regulatory requirements. • Ensure training is conducted at regular intervals.	• Plan and Design • Operate and Monitor
System Operators	• Participate in the system concept design and development. • Oversee the operations. • Continuously assess issues, and impacts. • Communicate when mitigation is required.	• Plan and Design • Operate and Monitor • Use or Impacted By
Cyber Security	• Give input into the risk register. • Provide feedback on AI risk measurements. • Contribute to risk mitigation strategies. • Participate in TEVV. • Provide ongoing risk monitoring.	• Plan and Design • Operate and Monitor • Use or Impacted By
Human Resources	• Participate in AI Actor hiring. • Ensure a diverse talent pool, with varying expertise. • Give input into the risk register. • Provide feedback on AI risk measurements. • Contribute to risk mitigation strategies. • Provide ongoing risk monitoring.	• Plan and Design • Deploy and Use • Operate and Monitor
DEI Experts	• Participate in AI Actor hiring.	• Plan and Design

	● Ensure a diverse talent pool, with varying expertise. ● Develop DEI policies. ● Assess the TSR systems inclusivity. ● Ensure bias and accessibility issues are addressed. ● Give input into the risk register. ● Provide feedback on AI risk measurements. ● Contribute to risk mitigation strategies. ● Provide ongoing risk monitoring.	● Deploy and Use ● Operate and Monitor
Third Party Suppliers	● Conduct internal risk assessment on provided product. ● Share results of risk assessment and known documented risks. ● Share documented proof of compliance with legal and regulatory requirements. ● Provide ongoing risk monitoring. ● Communicate regularly with AI actors, including AI Risk Management Specialist.	● Deploy and Use ● Operate and Monitor
End Users	● Utilize TSR as a driver in a vehicle. ● Provide feedback, including reporting incidents. ● Participate in focus groups.	● Plan and Design ● Operate and Monitor ● Use or Impacted By
The General Public	● Directly experience positive or negative impacts of TSR. ● Provide feedback, including reporting incidents. ● Participate in focus groups.	● Operate and Monitor ● Use or Impacted By

Not all AI actors have good intentions. Cybercriminals are adversaries who seek to exploit AI technology. They achieve this by utilizing various methods, such as data poisoning or model attacks. Table 4-5 showcases well-known adversary types who may target TSR systems, their goals, and capabilities.

Table 4-5: TSR Adversaries

Adversary Type	**Goals**	**Capabilities**
Adversarial Design Attacker	● Intentionally manipulate the TSR system. ● Exploit vulnerabilities.	● Conduct reverse engineering of system to learn vulnerabilities. ● Model extraction. ● Manipulation of TSR's recognition capabilities. ● Exploit vulnerabilities to gain control of the system.
Machine Learning (ML) Attacker	● Tamper or modify ML models.	● Model extraction. ● Data poisoning. ● Input manipulation.
Poisoning Adversary	● Inject malicious data into the training dataset, with the intent of causing the system to incorrectly identify traffic signs.	● Model manipulation. ● Data poisoning. ● Modify inputs.

Spoofing Attacker	● Manipulate or interfere with the signals sent or received by the AV sensors used to identify traffic signs.	● Manipulate signals received or sent by the sensors. ● Cause the vehicle to incorrectly identify (or miss) traffic signs.
Denial of Service Attacker	● Interrupt the vehicle's system, to cause an interference between the sensors and central compute, creating a denial of service.	● Overload the sensors. ● Create communication interference. ● Overload the central compute.
Malware Author	● Design malicious software with the intent of compromising the TSR system, stealing data, or causing harm to the operator and public.	● Data theft. ● Data manipulation. ● Propagation of other vehicle systems. ● Interference with TSR functionality. ● Ransomware. ● Remote control of the vehicle.
Sign Vandals	● Vandalize traffic signs with the intent of confusing the TSR system from recognizing and correctly identifying the sign, and associated action.	● Access to traffic signs and materials for vandalism. ● Knowledge of specific techniques to damage a sign, modifying the system inputs, and potentially poisoning training data.
Traffic Sign Thieves	● Remove/steal signs from their fixed location causing the TSR system to fail.	● Ability to physically remove a sign from the roadside.

SUGGESTED ACTIONS

- Reference table 4-4 as a basis for developing an AI actors table for your organization.
- Verify that all roles and responsibilities, as shown in the second column of table 4-4 are accounted for.
- Establish policies that promote regular communication with the AI actor assigned as the AI Risk Management Specialist.
- Ensure all AI actors understand their roles and responsibilities.
- Establish policies which enable AI actors to report any conflicts of interest. The documented conflicts should become dynamic artifacts, actively maintained by

the organization, and shared with the AI Risk Management Specialist and Organizational Management.

- Reference table 4-5 as a basis for developing a bad actors table for your organization.
- Develop and document a plan for counteracting bad actors mentioned in table 4-5.
- Ensure Cyber Security Experts regularly update the bad actors table based on newly identified threats.
- Communicate any cyber threats to the assigned AI Risk Management Specialist.

TRANSPARENCY AND DOCUMENTATION

- Have the roles and responsibilities been clarified with all stakeholders?
- Has the organization defined desired frequency of communication for each AI actor? (i.e., using a RACI)?
- Are roles and responsibility for the TSR system's governance and reporting clearly defined and understood?

GOVERN 2.2

ABOUT

In addition to the legal and regulatory training requirements outlined in GOVERN 1.1, all AI actors should be trained in the organization's TSR risk management processes (or broader AI risk management processes), including their roles and responsibilities for identifying and reporting risks. The overarching goal is for risk management to become part of the organizational culture, and a forethought during all aspects of the TSR system lifecycle.

Concurrently, organizations developing and deploying TSR systems should set high proficiency standards for AI actors contributing to design, development, deployment, and testing of said systems. Tools to provide actors with the ability to enhance their technical aptitude of ever-changing technology will help maintain this standard. Examples of this include tuition assistance programs, our online training sources, such as Pluralsight, Coursera, or LinkedIn Learning.

SUGGESTED ACTIONS

- Develop comprehensive risk management training that can be shared with AI actors when they join the TSR system's team.
- Ensure legal and regulatory requirements are also covered in the TSR system's risk management training.
- Define expectations for the organization to update training to ensure changes to legal and regulatory requirements are captured and communicated.
- Ensure a process is in place for AI actors to acknowledge completion of training and commitment to following the TSR system's AI risk management processes.
- Develop proficiency standards for AI actors based on assigned roles.
- Ensure additional training tools for AI actor skill development are in place, and readily available for AI actors to continue skill development.
- Establish a required frequency for AI actors to retake AI risk management training to ensure relevant content and changes are communicated and understood.

TRANSPARENCY AND DOCUMENTATION

- Are all AI actors aware of required risk management training?
- Do all AI actors understand their role in the TSR risk management process?

- How will AI actor proficiency be measured?

GOVERN 2.3

ABOUT

Organizational Management (OM) plays a vital role in the lifecycle of the TSR system. They serve as the executive sponsors for the TSR product, overseeing the assignment of an AI Risk Management Specialist, and defining the organization's risk tolerance level. The final decisions about risk management processes and mitigation strategies lie within the responsibilities of the OM. To ensure effectiveness of risk management efforts, OM must provide full support, fostering a culture that encourages employees to feel comfortable in vocalizing risks, and actively engaging in regular AI risk management training.

SUGGESTED ACTIONS

- Define the criteria for an executive sponsor.
- Assign an executive sponsor to the TSR product.
- Ensure the executive sponsor clearly understands their roles and responsibilities.
- Clearly communicate the organization's risk tolerance level.
- Review all roles and responsibilities of the AI actors.
- Define what level of risks need to be communicated to the OM in governance meetings.

TRANSPARENCY AND DOCUMENTATION

- Does the OM need to sign off on all mitigation strategies?
- At what level of risk can the AI actors decide their own mitigation strategy?

GOVERN 3.1

ABOUT

A key to effective risk management is utilizing diverse opinions. As stated previously, and identified in table 4-4, all the AI actors are expected to participate in TSR risk identification, and management throughout the TSR system's lifecycle. The best way to ensure that all risks are identified is to obtain feedback from a diverse team, who vary in demographics and backgrounds, combined with experts who have a broad range of experience and expertise. Obtaining such a varied talent pool must be engrained in Automotive OEMs and third-party suppliers hiring practices. Human Resources are responsible for ensuring that diversity goals are met, and monitoring risks of AI actors' attrition.

SUGGESTED ACTIONS

- Commit to providing Diversity, Equity, and Inclusion (DEI) throughout the organization and within the TSR AI product, will help to facilitate inclusivity and integration of a variety of insights into the risk management process.
- Review hiring processes and team development to ensure diversity goals are met.
- Validate that Human Factor Experts account for a diverse user group.
- Establish policies that promote inclusive design.

TRANSPARENCY AND DOCUMENTATION

- Does the organization have an established DEI office (or assigned individual) that can be utilized for establishing policies that encourage AI actors to provide feedback?
- Is a diverse user group being accounted for in TSR system development?
- Has the TSR system design been reviewed for bias?

GOVERN 4.1

ABOUT

As mentioned in GOVERN 2.3, Organizational Management is responsible for creating a culture where AI Actors working on designing, developing, and evaluating the TSR system are committed to consider and communicate AI risk, and promptly report incidents. The following section dives further into steps management can take to further develop this culture, including through incentivized risk management.

SUGGESTED ACTIONS

- Utilize Auditors to ensure all systems are designed in a clear and easy to interpret manner, adhering to legal and regulatory requirements, and fortified with robust cyber security measures.
- Ensure cyber security teams are performing continuous monitoring and auditing of the TSR system, checking for existing or potential vulnerabilities.
- Record all identified vulnerabilities in the risk register and assign a risk score based on the likelihood and impact.
- Assign a DRI to own monitoring of the risk, development of a response strategy and action plan, should it become an active risk or incident.
- Establish or enhance policies that incentivize AI actors for participating in the risk management process. These incentives could take the form of safety awards, or special recognitions from Organizational Management. Such awards should be presented publicly to encourage others on the team to follow suit.

TRANSPARENCY AND DOCUMENTATION

- Do organizational policies encourage AI actors to effectively utilize critical thinking and a safety-first mindset for TSR systems?
- Are incentive programs well documented and shared throughout the organization?

GOVERN 6.1

ABOUT

When third-parties are contracted by an OEM to provide software, data, algorithms, and support of vehicle components risks associated with the obtained product or service must be properly documented. Proper risk management when third-party systems are integrated into a vehicle can be complicated. To ensure that risks associated with third-party TSR system components are properly documented, both organizations (the OEM and third-party supplier) must align on expectations.

SUGGESTED ACTIONS

- Ensure the third-party is engaged in the AI risk management process.
- Utilize contracts to ensure expectations, such as key performance indicators (KPI), are clearly communicated to the third-party regarding required internal risk assessments on their provided products. This includes providing honest and reliable feedback on impact and likelihood of the risks.
- Ensure third-party representatives remain engaged throughout the TSR system's lifecycle, communicating regularly with the assigned AI Risk Management Specialist, and providing ongoing risk monitoring and incident reporting.

- Document proof of third-party compliance with legal and regulatory requirements, including risks of infringement of the third-party's intellectual property is mandatory.
- Engage Auditors, Compliance Experts and Procurement Experts to review all provided risks, legal, ethical, and other issues pertaining to procurement and use of the third-party systems.

TRANSPARENCY AND DOCUMENTATION

- Are third-party policies clearly documented?
- Is there an established traceability process?
- Can the third-party's TSR system components be audited?
- Have any additional risks that are introduced from using a third-party supplier been documented?
- Are Procurement Experts, Auditors and Compliance Experts in agreement with the third-parties' terms and conditions?

SECTION 2: TSR Product Specifications

To effectively GOVERN, MAP, MEASURE, and MANAGE risks, a comprehensive grasp of product specifications is vital. In the realm of governance, this entails documentation of TSR system inventories and processes, encompassing the methodology for implementing updates and navigating the system beyond its support lifecycle. When it comes to mapping, a thorough review of the TSR system is essential, encompassing both present and future deployments. Furthermore, articulating the system requirements for automation levels 3+ and delving into details about the system's knowledge limitations becomes imperative. Facilitating effective measurement involves a comprehensive explanation and documentation of the TSR AI model. Lastly, proficient system management necessitates the establishment of regular monitoring protocols, particularly in relation to the deployed model. The following section will highlight all the aforementioned areas in further detail.

GOVERN 1.6

ABOUT

As part of the AI risk management process for the TSR system, ongoing monitoring and regular reviews should be planned. The TSR system and components remain a dynamic technology, constantly evolving. The development of the inventory is essential for cross-referencing with the TSR's AI risk register, ensuring that all systems' risks are identified and documented.

SUGGESTED ACTIONS

- Create and maintain system inventories as part of the TSR system design process. (Typically, this role is completed by the AI Designer.)

- Develop an inventory plan to review with AI actors which will help ensure all appropriate inventories are captured.
- Ensure all AI actors involved in the TSR system design and development contribute to the inventory. For instance, the Data Scientist and Data Engineers should provide links to the utilized data, while Model Engineers should provide the explanations of model selection and source code of the model design.
- Make updates to the inventory as they occur.
- Establish a regular monitoring cadence for the TSR system.
- Incorporate inventory reviews and audits as part of the organization's annual file review (AFR) process.

TRANSPARENCY AND DOCUMENTATION

- Have all the roles and responsibilities for system inventories been clearly assigned?
- Is there a clear traceability between components?
- Has a regularly established review cadence been established?

GOVERN 1.7

ABOUT

As part of the TSR system planning, processes and procedures should be in place for providing system upgrades, repairing defects, and communicating if a system becomes its last day of support (LDOS). To maintain trust with End Users and the public, notifications of product upgrades, issues, defects, or LDOS systems must be communicated according to legal requirements.

SUGGESTED ACTIONS

- Provide consumers with options to upgrade current compute systems (including software and ML algorithms), through over-the-air (OTA) updates, and hardware via dealership main shops.
- Upon issue identification communicate the estimated-time-to resolve identified issues to consumers directly, via email, and direct mail communications.
- Plan for the organization to cover costs that ensure the meeting of regulatory, or legal requirements, with End Users containing the costs of other enhanced feature upgrades.
- Communicate to End Users when a TSR system is going LDOS.
- Provide End Users with options to upgrade their system for continued support.
- Establish a process for handling when an end user refuses a regulatory update. (i.e., disabling other vehicle features.)

TRANSPARENCY AND DOCUMENTATION

- Are there well documented communication plans for product upgrades, issues, defects or LDOS systems?
- Have all the communication methods been identified?
- Are roles and responsibilities clearly understood?
- Does the communication plan meet legal and regulatory requirements?

MAP 1.1

ABOUT

TSR deployments can vary greatly depending on the level of automation. As mentioned previously, the focus of this profile is automation levels 3+, as highlighted in figure 4-8 below. The following section reviews a TSR system overview, including intended purpose, beneficial uses, norms and expectations, end user and operator expectations, current system deployments, future deployments, suggested actions and transparency and documentation.

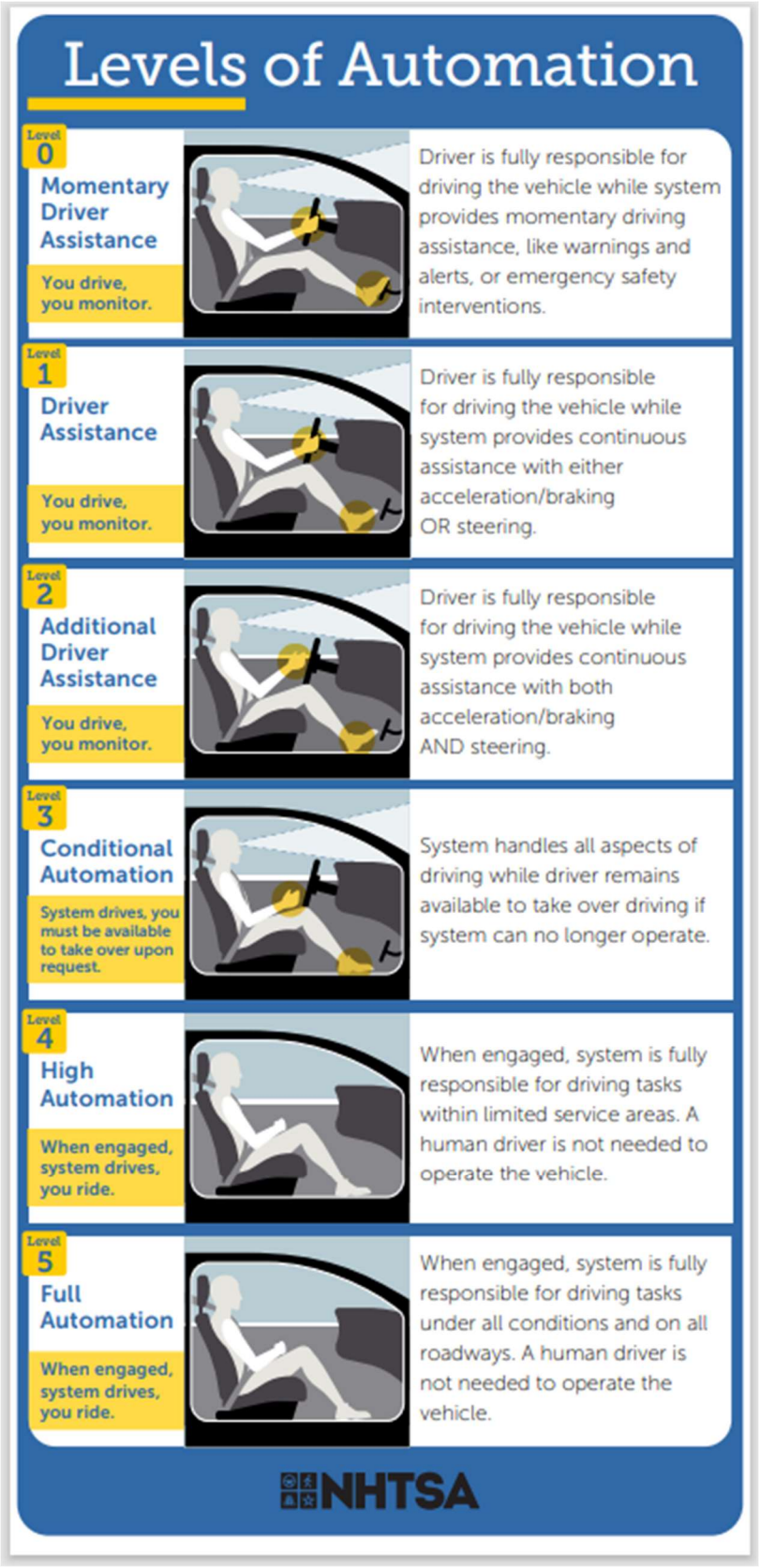

Figure 4-8: Levels of Automation [33]

Intended Purpose

TSR is a key functionality of AVs, playing a vital role within autonomous driver assistance systems (ADAS). Its primary purpose is to enhance safety by actively monitoring the environment, identifying relevant traffic signs, and communicating required response actions to the vehicle to execute. In automations levels 0-2, TSR acts as a supplementary pair of eyes for the driver, relaying essential information garnered from the system, such as real-time speed limits indicated by road signs. In levels 3+, the TSR system operates exclusively without the need for human intervention.

Current TSR Deployment

As of the writing of this profile, there are several examples of TSR, currently deployed in vehicles, from automakers such as Ford, Mercedes, Volvo, and BMW (to name a few.) Results from our public survey show that 17.6% of End Users have ridden in a vehicle that utilizes TSR, with a majority of those having utilized Ford (43%), followed by Tesla (21%) and GM (18%). The level of autonomy that comes with each system varies by automaker and are examples of level 1-2 automation requiring the driver to monitor and control the system.

For example, Ford utilizes TSR to enable their intelligent adaptive cruise control system, which manually adjusts the speed of the vehicle. This works, once this vehicle's speed is set (through cruise control), if the system identifies a speed limit sign below the set speed, it automatically adjusts the vehicle speed accordingly and will resume to the preset speed once a new sign has been detected. [34]

The system currently provided by Volvo performs slightly differently. Using road sign information (RSI), select Volvo vehicles automatically detect speed limit signs and display them in the instrument panel, providing drivers with speed limit awareness. Beyond this, there is also

an option for the vehicle to alert the driver with speed warning alerts in the form of a flashing light or sound, if the identified speed limit is exceeded. [35]

Finally, Tesla's Model Y is currently testing a traffic light and stop sign control feature, which is in BETA at the time of this writing. This feature is designed to recognize and respond to traffic lights and stop signs, by slowing the Model Y to a stop when the driver is utilizing the Traffic-Aware cruise control, or Autosteer. This functionality utilizes the vehicle's forward-facing cameras, in conjunction to GPS data, and slows the car for all detected traffic lights, including those which are green or blinking yellow. As the vehicle approaches the intersection, the touchscreen display notifies the driver of its intention to slow down. The driver then must confirm if they wish to continue, otherwise the vehicle will stop at the red line displayed on the screen. If the light is green, and the driver wishes to proceed, they must press the accelerator pedal to give the vehicle permission to do so [36]. An example of how this works is highlighted below in figure 4-9.

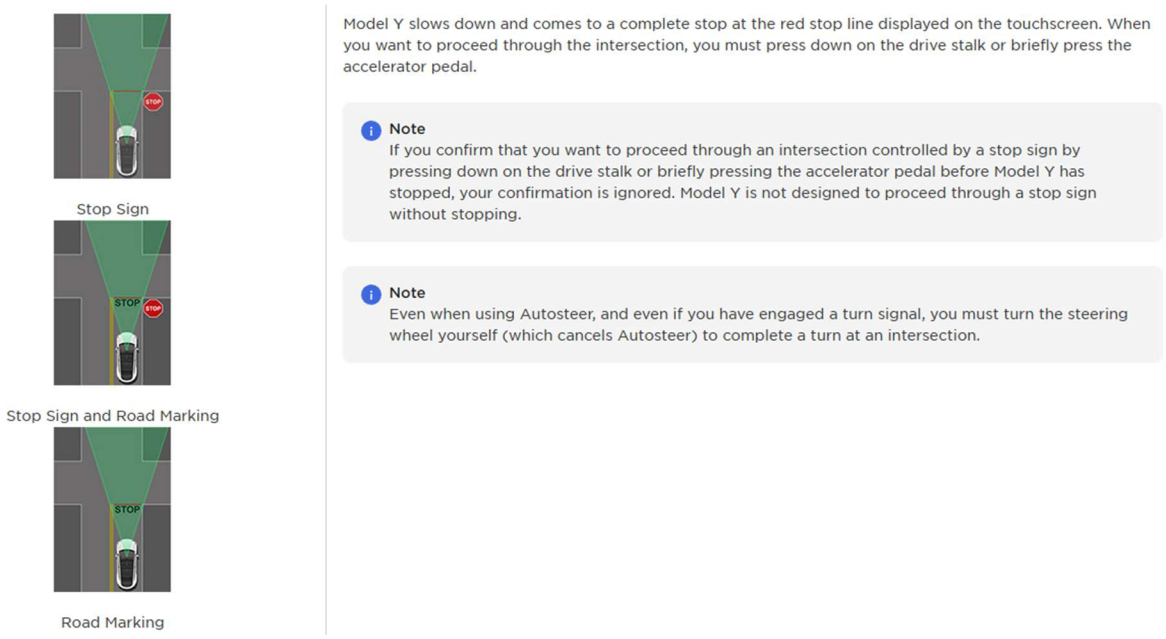

Figure 4-9: Tesla Model Y Traffic Light and Stop Sign Control [36]

Future TSR Deployment

Future TSR deployments will utilize levels 3-5 automation. For such deployments, the vehicle will provide full automation, with human interaction only occurring in level 3, in the event of a system failure. To work, the TSR system will utilize cameras, and information received from GPS satellites to recognize traffic signs and respond accordingly. Our public survey results show that 6.9% of people believe that this technology is available now, with 3% of survey respondents stating that they have ridden in a fully autonomous vehicle. For the remaining respondents, 27.7% believe fully AVs with built in TSR will be ready within the next 5 years, 35.8% believe it will be ready within the next 10 years, 19.5% believe it will be the next 10+ years, with the remaining 10.1% believing it will never be ready. Examples of level 4 automation can be seen currently in Waymo deployments, which do not require a human driver to operate.

Waymo, which is currently offered in select cities in the United States as a taxi service, uses a combination of information obtained from perception sensors (lidar, cameras, radar), combined with detailed territory maps (that are compiled prior to operating in the area) and computes the safest route. The vehicle is driven by the Waymo Driver, who "is the embodiment of fully autonomous technology that is always in control from pickup to destination" [37]. Waymo riders are not required to operate the vehicle. "They can sit in the back seat, relax, and enjoy the ride with the Waymo Driver getting them to their destination safely" [37]. Figure 4-10 highlights an example of a current Waymo vehicle that is deployed.

Figure 4-10: Waymo, Level 4 Automation [37]

Beneficial Uses

According to The National Highway Traffic Safety Administration (NHTSA) in 2022, there were 42,795 deaths related to automobile accidents [38]. To address this, the U.S. Department of Transportation (DOT) announced a National Roadway Safety Strategy, which includes a Safe System Approach, serving as the guiding paradigm to address roadway safety [39], [40]. This approach accepts that humans make mistakes or decisions that can lead to accidents, and transportation systems should be designed to accommodate these and correct when possible. This can be done through the development and deployment of safer vehicles that include ADAS functionality, such as TSR [40]. When asked, if AVs were able to perform 10 times better than a human driver, would you fully trust an autonomous vehicle, 57% of respondents stated, no, showing that the general public has much higher safety expectations for AVs, and organizations should design testing accordingly, with passing results showing greater than 10 times improvement compared to a human driver.

In addition to the safety implications, a benefit of advanced levels of ADAS systems (4+), is the increased accessibility for those who may not have the means, or ability to operate a vehicle. Systems that can recognize traffic signs, and respond accordingly, provide needed functionality to obtain full automation.

Norms and Expectations

TSR systems (in advanced levels of automation 3+) are expected to see, properly perceive traffic signs, and then act accordingly. This requires robust models to integrate seamlessly with systems in all environments. For the system to work consistently, the model must include proper specifications about traffic signs, and the required actions the vehicle must take for each sign. In addition, models must be able to quickly adapt to changes in local laws and regulations as they arise.

Below we highlight some examples of signs that are required to be incorporated in the training data for the model, and in turn direct the vehicle to behave correctly. The provided examples for United States deployments only, with automotive OEMs and supporting third-party suppliers required to obtain and program models with the fully extensive list, based on the deployment locations. Vehicles with deployments in multiple regions are expected to have added complexities, requiring additional data, testing, evaluation, and verification.

- **Regulatory Signs** – Regulatory signs are imperative and reflect local laws. They instruct the vehicle on mandatory actions such as adhering to speed limits, when to stop, yield, where to park (or not park) and which roads are identified as one way. Failure to comply with regulatory signs increases the risk of accidents and invites potential citations from law enforcement.

Figure 4-11: Regulatory Sign Examples [41]

- **Warning Signs** – Shown as yellow and black signs, warning signs help provide awareness to drivers and vehicles about potential hazards, such as school or pedestrian crossing, upcoming hills, intersection details, or even animal crossing.

Figure 4-12: Warning Sign Examples [41]

- **Guide Signs –** Identified by their green and white coloring (with occasional yellow), guide signs include information such as street names and exit numbers. These signs help drivers and vehicles with route information, including where to turn or depart the highway.

Figure 4-13: Guide Sign Examples [41]

- **Construction Signs and Markers –** Construction signs can vary; typically seen as orange and black, with orange and white cones, construction areas may also employ the use of boards with light for information, or as arrows to indicate the need to merge or proceed with caution. Varying from traditional signage, construction signs may also come in the form of a handheld sign, held by a worker in the road to indicate stopping or proceeding with caution. Drivers and vehicles need to be able to process all these signs, and their varying locations.

Figure 4-14: Construction Sign Examples [41]

- **Route Markers –** Shown in black and white, route markers indicate the way the vehicle is traveling, or directions to a specific route.

Figure 4-15: Route Marker Sign Examples [41]

- **Service Signs –** Indicated in blue and white, service signs provide information regarding what services are provided at the upcoming exit, i.e., restrooms, gas stations, hospital, or lodging.

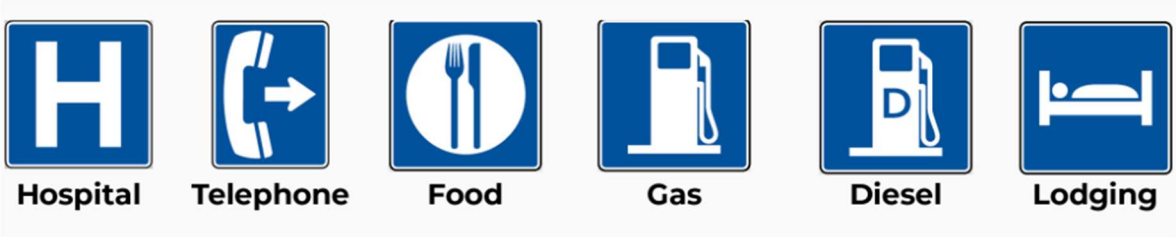

Figure 4-16: Service Signs Examples [41]

- **Vehicle Signage –** Vehicles must be able to identify signs on vehicles, including stop signs seen on school buses.

Figure 4-17: School Bus Stop Sign [42]

End User and Operator Expectations

As stated previously, the public expectation for TSR functionality in AVs is high, with demonstrated lack of trust. Results from our survey confirm this, with 66% of users stating that they do not trust AVs that remove the need for a human driver, with 29% of users trusting the AV somewhat, and 5% fully trusting an autonomous vehicle. Some examples of public expectations include, that the TSR system should:

- Be safe, having been through rigorous testing.
- Perform more than 10 times better than a human driver.

- Be protected from cyberattacks.
- Risks are fully documented, monitored and shared by automotive OEMs, and third-party suppliers.
- Protect users' privacy, not sharing data without explicit permissions.
- Provide clear steps for operator remediation in the event of a failure.
- Provide product training through videos and onsite dealership support.
- Have the ability to disable autonomous features and take control of the vehicle.

SUGGESTED ACTIONS

It is the role and responsibility of the UX designer to clearly document user experience expectations as part of the product development and share these expectations with the product team. As the TSR technology evolves with levels of automation, these expectations will continue to develop and should be revisited. Additional actions for the team to take:

- Maintain awareness of traffic laws, and legal standards for AVs and TSR functionality requirements.
- Ensure traffic sign data inputs include the most recent regional data to encompass changing regulations and signage variations.
- Maintain awareness of technological updates, and potential improvements.
- Benchmark TSR industry standards.
- Ensure Human Factor experts are engaged to review the design and provide feedback to the AI Risk Management Specialist regarding any potential risks.
- Ensure all TSR System product documents clearly define the system's tasks and intended purpose.

- Review TSR System benefits posted in the About section above and continue to update as the technology progresses.
- Review TSR system norms and expectations in the About section above and update as needed to include expectations of End Users and the general public when the TSR system is operating in production environments.
- Perform context analysis as it pertains to safety concerns, geographic area of deployment, physical environmental limitations, and intended setting of operation.
- Gain and maintain awareness about the roles the end user and public perform in TSR system operation.
- Identify and document where the TSR system will replace human decision making.

TRANSPARENCY AND DOCUMENTATION

- What is the expected behavior for each traffic sign, in a particular situation?
- Have the AI actors documented the intended uses of their associated pieces in the system?
- Who is the person responsible for reviewing and maintaining the list of ethical considerations throughout the TSR system lifecycle?

MAP 1.6

ABOUT

A critical functionality of a TSR system is that it respects the privacy of its users, and that the design includes all socio-technical implications needed to address AI risks. The Product

Manager is responsible for documenting system requirements with input from other AI actors, including the UI/UX Experts, Human Factor Experts, Compliance Experts, Software Engineers, Cyber Security Experts, etc. By ensuring that the design incorporates the End Users' needs, and other safety factors, i.e., cyber security breaches, the overall system will have enhanced trustworthiness.

SUGGESTED ACTIONS

The Product Manager should ensure:

- The requirements incorporate trustworthy characteristics, i.e. security measures, and fail safes.
- Detailed feedback documenting human factors and extensive testing that includes End Users, and the public.
- Risks documented during requirement development are fed to the AI Risk Management Specialist for inclusion in the risk register.
- Dependencies between the hardware, software, model, and other system components are captured, along with potential impacts for partial or full system failure.

TRANSPARENCY AND DOCUMENTATION

- Does the TSR system contain any end user information that could be accessible to others?
- What type of information on the TSR system design, operations and limitations are shared publicly?
- Is the company providing enough transparency regarding the limitations of the TSR system?

- Where will relevant information be provided to End Users, including how the system operates, how it was designed, and what the potential limitations are?
- How will cybersecurity experts address potential cyber risks that may disrupt the TSR system?

MAP 2.2

ABOUT

A TSR system's knowledge is limited to the training data. A vehicle's system's knowledge is expected to contain regional information to allow for intracontinental travel. If such data was not included in the training data, such information should be documented, and End Users must be informed.

SUGGESTED ACTIONS

- Document knowledge limits, any associated risks, and impacts. (Typically, this is the responsibility of the Data Scientist, Data Engineers, and Model Engineers.)
- Provide risks, likelihoods and impacts to the AI Risk Management Specialist for inclusion in the risk register.
- Ensure documentation provides sufficient information, which is available to all AI Actors to review, for informed decision making as it pertains to system design, and testing.
- Communicate known limitations to the End User.

TRANSPARENCY AND DOCUMENTATION

- Is the purpose of the model clearly documented and understood?
- Is the expected behavior clearly identified?

- Are the interdependencies between the model and the overall TSR system clear and easy to understand for all AI Actors?
- Has the deployment setting, and environment, i.e., region of use, been clearly documented?
- Where is the training data stored for others to review?
- Are the data and knowledge limits explainable?
- What is the plan for adjusting data if the knowledge limit is identified to impact system performance?
- Does the system account for all legal requirements?
- Who owns the go-no-go decision for the model pre- and post-testing?
- Has the training data been reviewed by DEI and Human Factor Experts been deemed inclusive?

MEASURE 2.9

ABOUT

The model used in the TSR system must be easy to explain and validate. Model Engineers are responsible to properly document details, (including a detailed description, explanations of model selection and expected outputs) of the model, providing access to the details to all other AI actors to review. This information will then be used in the testing, and evaluation of the system. Ensuring that the details provided are explained properly will assist testers, including Cybersecurity Experts, with information to properly diagnose risks, and their potential impacts.

SUGGESTED ACTIONS

- Verify that the model is clearly explained.

- Review the documentation, to ensure that any acronyms are removed, and that information provided can be easily understood by any AI actors who may need to review the model details.
- Document details around model training, and reasoning for model selection.
- Make model documentation available to all pertinent AI actors for review and feedback.
- Document and share relevant test data. (Details around training data used, i.e., region of traffic signs utilized, any metrics established on the effectiveness of the model, tools, i.e., model cards, and model development processes, including those used to prevent potential manipulation of the TSR system should be included.)
- Plan to test the TSR system's model over time.
- Develop a plan to handle future system updates (i.e., from regulatory changes).

TRANSPARENCY AND DOCUMENTATION

- Is the TSR system's model documentation clearly explained to allow for easy interpretability?
- Is model selection reasoning clearly rationalized?
- What are the checks and testing plans for the model's effectiveness?
- Is the learning data clearly labeled?
- Does the learning data have any constraints?

MANAGE 3.2

ABOUT

After the TSR system's model has been defined, and trained, it must be monitored, and undergo regular system maintenance. Conducting regular evaluations on a TSR system in production could be challenging. In [43], the authors highlight a strong need for organizations to move from a reactive to proactive approach, by following a comprehensive strategy. To do this, they suggest five different steps that organizations should take, which include: firstly, defining model performance metrics with labels for inference data. They state, "The availability of labels enables calculating and analyzing common model validation metrics, such as false positive/negative rates, error/loss functions, AUC/ROC, precision/recall and so on." Secondly, they recommend establishing granular behavior metrics of the model's outputs. This is because the output behavior observed can help to indicate problems that are barely detectable elsewhere. Thirdly, collect metadata to properly segment metric behavior, stating, "to truly realize the value of monitoring, behavioral metrics have to be looked at for subsegments of model runs." This will assist with tasks, i.e., compliance assessments, and root cause analysis. Finally, track data during training, test, and inference time, as "forward thinking teams expand the monitoring scope to include training and test data."

SUGGESTED ACTIONS

- Collect the source of training data used by AI actors involved in the model development, including sign classifications must be shared and properly documented to help with identifying risks and assist with incident management.

- Establish a process for TSR system monitoring, which includes risk and incident management, and TSR system decommission when risk tolerances are exceeded.
- Assign a team for monitoring.
- Ensure that all pre-trained models, and associated data are included in the system's inventory.

TRANSPARENCY AND DOCUMENTATION

- Have the roles and responsibilities been clearly assigned?
- How are the TSR system's data sources, labels, dependencies, and constraints documented?
- Who is responsible for ensuring that all applicable traffic signs have been included in the data set?
- Does the data sent include intraregional signs (for travel in the continent)?
- How does the model plan to handle new or obscure traffic signs?
- How does the model respond to signs that have been vandalized or damaged?

SECTION 3: Pre-Deployment Testing

Pre-deployment testing, evaluation, verification, and validation (TEVV) is critical for TSR systems to be safely deployed. The following section reviews organizational policies for eliciting feedback from AI actors, utilizing UX/UI design principles, considerations for ensuring scientific integrity for the TSR system's TEVV, defined performance metrics, and pre-deployment assessment.

GOVERN 5.1

ABOUT

For the most effective risk management, AI Actors are required to remain engaged throughout the TSR system lifecycle, including testing, evaluation, verification, and validation. As mentioned in GOVERN 2.1, organizational policies should require Actors to participate in distinct phases, including TEVV based on their assigned roles and responsibilities, and providing feedback. The following section highlights the use of UX/UI design principles, experimental design, and cyber security procedures which can be used to enhance the effectiveness of the TSR System's TEVV phase.

The UI/UX paradigm states that the usability of a system and the achievement of its goal of use is largely determined by the quality of design and the physical arrangement of interface elements [44]. For a TSR system to be successfully implemented, it is critical for designers to understand how users will interact with the system and validate through testing, evaluation, and verification. Henry Ford said it best when he stated, "If there is any one secret of success, it lies in the ability to get the other person's point of view and see things from that person's angle as well as from your own" [45].

Additionally, a clear and decisive standard for experimental design must be in place. In [46], the authors propose a testing framework capable of virtually evaluating the closed-loop properties of an autonomous vehicle system, including ML components. TSR TEVV Experts are responsible for designing such a method, with a comprehensive test plan to effectively conduct performance and adversarial tests.

SUGGESTED ACTIONS

- Prioritize human centered design principles. Documenting and testing the way End Users perceive and interact with the TSR system is critical for successful implementation. Neglecting user engagement, and acceptance, could potentially undermine the safety advantages that a TSR system aims to deliver [47].
- Utilize human factor experts to conduct user research and gain a deep understanding of user needs, behaviors, and preferences.
- Facilitate the creation of prototypes that align with End Users' expectations.
- Ensure Human Factor experts collaborate closely with the Product Owner, ensuring that valuable feedback is integrated into the TSR system's development and test plans.
- Analysis should be conducted by the UI/UX Designers to document how the TSR system will be utilized by users, to ensure a successful implementation.
- Engage UI/UX Designers early in the TSR system design, to assist Human Factor Experts with user research.
- Ensure that after the initial system is developed, the UI/UX Designers work with the Evaluators to engage End users and Developers building planning and control modules for system testing and validation, providing feedback to the design team,

and associated AI actors. Throughout this process, any risks that are revealed in the research must be provided to the assigned AI Risk Management Specialist.

TRANSPARENCY AND DOCUMENTATION

- Which stakeholders are engaged in UI/UX and cyber threat analysis?
- Is there a need to employ mitigation strategies, i.e., bug bounties?

MAP 2.3

ABOUT

Ensuring the scientific integrity of the TSR system's testing, evaluation, verification, and validation (TEVV) is vital. Performance criteria should be expressed in qualitative and quantitative measures with pre-deployment tests conducted through simulations and real-world physical tests, that mimic the actual deployment setting. Results should clearly confirm that the system is operating as designed, providing enhanced safety features to End Users. A commitment to ongoing testing of the TSR system post-implementation is essential to ensure continual assessment throughout the system's lifecycle.

SUGGESTED ACTIONS

- Document full test plans, and associated AI actors' roles and responsibilities.
- Ensure that testing occurs throughout the entire TSR system lifecycle, starting in designing, and continuing through post-implementation.
- Identify simulation tools for conducting model testing. (Examples include: Autoware, Carla and ns-3.)
- Define where and how physical system testing will occur.
- Document the required performance specs and acceptable levels of latency.

- Define system testing as it relates to individual systems and how the systems interact with one another during TEVV. (Refer to figure 4-18, for an overview of how to think of system interactions.)
- Leverage the expertise of Model Engineers, Data Engineers, and Data Scientist to develop scenarios that test model behaviors.
- Engage Software and Systems Engineers to provide performance standards, expected behaviors and documented test plans.
- Engage Cyber Security experts to document well-known cyber threats, develop vulnerability assessments, and plans for standardized tests, i.e., penetration testing to ensure security of the TSR system.
- Measure the TSR system's ability to sense traffic signs, recognize them correctly, and act accordingly.
- Evaluate system performance using the Automated Driving Systems Interaction Evaluation (ADSIE framework) as referenced in figure 4-19.
- Deploy testing protocols, i.e., Failure Mode and Effect Analysis (FMEA) to discover potential failures.
- Ensure that test results are written in a way that is easy to interpret and understand.
- Document assumptions used for testing, along with expected and actual results.
- Establish methods for regular communications regarding test results.
- Document and share information regarding model test data used for TSR system training. This includes any data excluded, and associated reasonings.
- Document known system limitations, and associated risk mitigation plans.

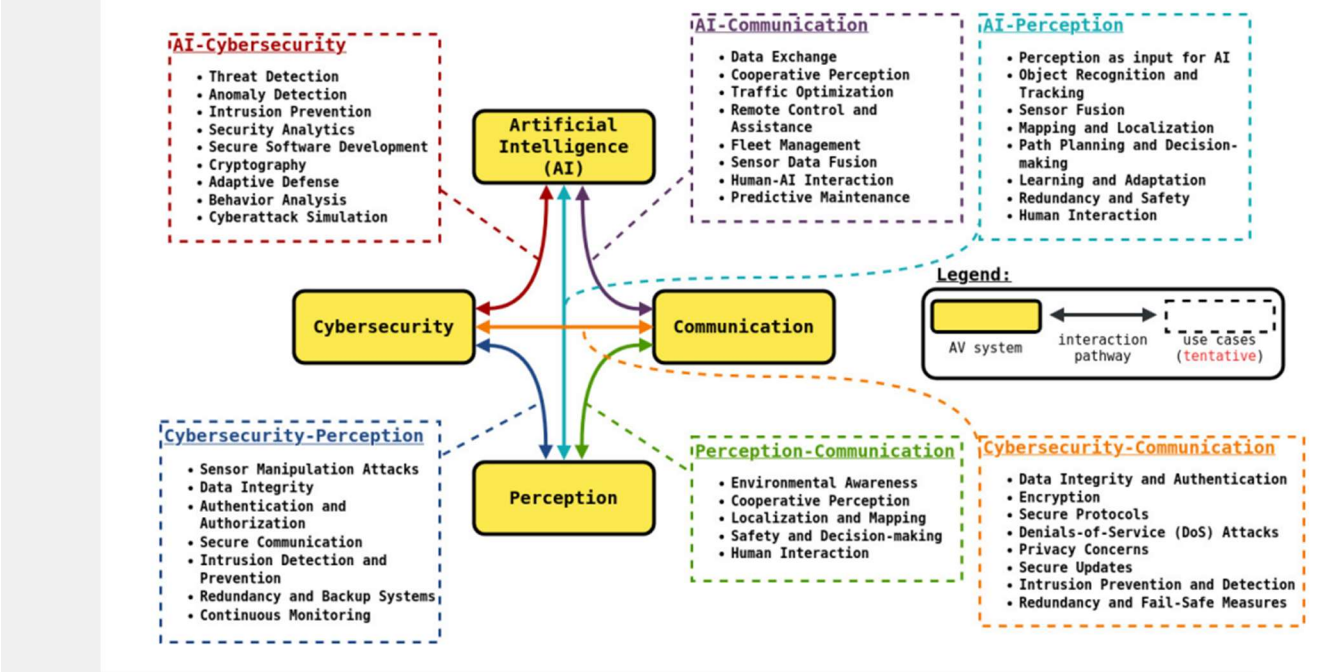

Figure 4-18: Systems Interactions [48]

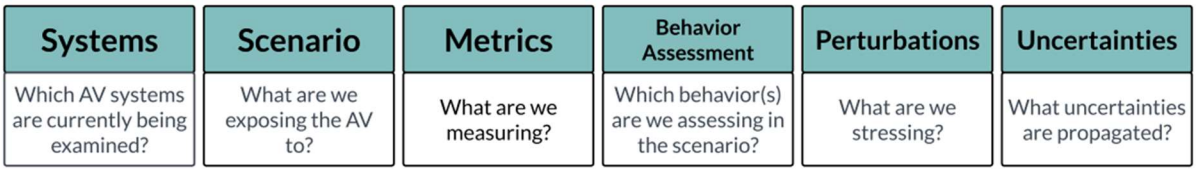

Figure 4-19: ADSIE Framework [48]

TRANSPARENCY AND DOCUMENTATION

- Do the test results comply with the organization's accepted level of risk tolerance?
- Has a mitigation plan been developed for cyber threats that may impact the vehicle's ability to perceive traffic signs correctly?
- Do the test results show enhanced safety functionality?
- Do the tests confirm system reliability?
- How was the model data (traffic sign information, and associated actions) collected? Does the scope of the data align with the vehicles planned deployment region?

- Are plans in place to review model data throughout the TSR system's lifecycle?
- Are plans in place for auditors (external to the TSR team) to review the TSR system documentation, and test results?

MEASURE 2.1

Before initiating the TEVV process, it is essential to record all TSR test data, including expected results, and comprehensive methods about the testing techniques and tools used. The expected results should be expressed in quantifiable terms. For example, a model may misclassify a sign with an affixed sticker, that a human could easily identify. Evaluating the system's ability to correctly identify signs, including those that have been altered, and assessing the time of identification compared to a human will be beneficial.

SUGGESTED ACTIONS

- Utilize well documented test plans that are easily interpreted and understood by Auditors.
- Provide clear standards for measurement, that can be replicated through multiple tests.
- Follow industry standard best practices for testing autonomous vehicle technology, including the use of simulation and physical tests that mimic real world scenarios.
- Utilize methods such as FMEA for testing the effects of failure, documenting results in a measurable manner.
- Regularly assess the methods used for testing to determine if they need to be updated.

TRANSPARENCY AND DOCUMENTATION

- Are key performance indicators of success defined, documented and communicated to all AI actors?
- Have testing intervals for the TSR model been defined?
- Did all AI actors have input into the test plan and metrics used?
- How do the metrics used to define success compare to the results of a human driver?

MEASURE 2.3

ABOUT

The TSR system performance needs to be measured quantitatively, to demonstrate that the autonomous vehicle can perform more than 10 times better than a human driver, providing a safer solution. Model and system tests conducted in both simulations and physical environments should all be designed based on the expected context of uses. Details regarding the context will come from the UI/UX and Human Factor experts.

Tests that produce lower quantitative performance should be documented as a risk, with associated likelihood and impact, and mitigated to improve performance. Details regarding the tests conducted and associated conditions the tests were conducted in must be included in the testing documentation.

SUGGESTED ACTIONS

- Assign TEVV Experts who will engage all AI actors in developing comprehensive test plans with clear results that can be quantified.
- Define KPIs to measure the test results.

- Ensure regular engagement between TEVV Experts, UI/UX Experts, and Human Factor Experts.
- For TSR systems that will be deployed in multiple regions, maintain a demographically diverse team, to ensure local and regional factors are considered.
- Ensure test plans account for documented risks, i.e., traffic sign is damaged or tampered with.
- Conduct testing with End Users to gather feedback on the TSR system's trustworthiness.
- Document any differences between the test settings (simulation and/or physical) and the actual deployment environments.
- Ensure the measurements are written in a way that they can be easily audited by a third-party.
- Conduct tests on the TSR model to validate how the training data changes over time.

TRANSPARENCY AND DOCUMENTATION

- Does your test plan clearly define success?
- Are the metrics easy to understand?
- Do the metrics demonstrate a reliable and trustworthy system?
- Have systems obtained by third parties gone through similar tests?

MANAGE 1.1

ABOUT

Prior to deploying the TSR system in an autonomous vehicle, a determination must be made as to whether the system achieves the intended purpose of being able to correctly identify and respond to traffic signs. System performance is a critical feedback point that Organizational Management will use to determine if the TSR system's performance is trustworthy for deployment, or if risks identified in testing need further mitigation.

SUGGESTED ACTIONS

- Define what characteristics make the TSR system trustworthy.
- Review and further mitigate risks that may hinder a successful deployment.
- Review Section 4 for further details on TSR risk management.
- Maintain the risk register for all risks throughout the TSR system's lifecycle.

TRANSPARENCY AND DOCUMENTATION

- How do the test systems validate that the system is achieving the desired output?
- Do the test results consistently demonstrate compliance with legal and regulatory requirements?

SECTION 4: Risk Management

Risk management is the crux of the TSR use case profile. As stated previously, to properly and safely deploy a TSR system, organizations must have a comprehensive risk management plan in place. The following section reviews risk management policies and process for deploying a TSR system. It includes details for risk scoring based on risk tolerance, well known identified TSR risks, likelihood and magnitude of impacts and associated metrics for risk measurement.

GOVERN 1.2

ABOUT

It is imperative that organizations establish robust AI risk management policies that align with the characteristics of a trustworthy AI system. As part of these processes, the organizations deploying a TSR system should review current organizational policies and procedures to ensure AI risks, and strategies for dealing with AI risks are included.

SUGGESTED ACTIONS

- Adapt the organization's regular governance process to include data governance. (Key AI actors for leading data governance are the Data Scientist, Data Engineer, Model Engineers, Software Engineers, Compliance Experts, Auditors, and Third-Party Suppliers (when applicable)).
- Develop a detailed risk management plan. This process must be developed prior to beginning TSR system development and maintained throughout the product life cycle.
- Review risks regularly as part of an organization's regular product governance. This meeting should include Organizational Management, and key AI actors.

- Assign an AI Risk Management Specialist to document risks in a dynamic risk register and present the status of the current risks in governance meetings, including their anticipated timing and mitigation strategy. (This process should continue throughout the entire TSR system's lifecycle.)
- Encourage all AI actors to contribute to risk mapping. Specific details regarding which process the AI actors should be engaged in can be found in table 4-4. An example process for TSR is included below in figure 4-20.
- Conduct change management, including the documentation of new risks, as part of the organization's already established change management process.
- Maintain a dynamic risk register, adding or closing risks through the TSR system's lifecycle.
- Note changes and communicate per a prescribe cadence outlined in the communication plan to all impacted AI actors and continue to review risk status in regular governance meetings.
- Conduct regular engagement with all stakeholders, including End Users. Details regarding risks associated with each AI actor, and when they should be engaged throughout TSR AI Lifecycle phases are outlined in table 4-4 and figure 4-20.
- Establish a formal communication plan as part of the TSR's product development plan. This plan should include standards for communicating risk and incident management throughout the product's lifecycle.
- Utilize tools, such as a RACI to designate polices around how often AI actors are communicated to, and what specific risks or incidents should be communicated to them.

- Review the facets of data management which include data privacy and security, data quality and integrity, data retention, and legal and regulatory requirements.
- Documented data processes and have Auditors conduct regular reviews.
- Share any identified risks that arise to the AI Risk Management Specialist for inclusion in the risk register.
- Obtain feedback regarding processes, and opportunities for improvement at regular intervals, with a minimum predefined timing of quarterly.
- Communicate changes to predefined risk management processes to all organizational AI actors and review in governance meetings.
- Evaluate the efficacy of the risk management processes, with Organizational Management providing directions to the AI Risk Management Specialist if adjustments are required.
- Establish a comprehensive protocol that allows AI actors to discreetly report serious system concerns, without facing repercussions.
- Ensure whistleblower policies are widely and consistently communicated.
- Conduct risk measurement, after the risk mapping process is conducted. Risk measurement is multifaceted and includes reviewing the risks that have been identified in the MAP function, their assigned impact and likelihood, and assigning appropriate metrics. Subsequently, these risks are assigned a score based on their metrics (Impact + Likelihood = Risk Score). The resulting scores, as depicted in table 4-6, are then utilized in the mitigation planning phase. For detailed information on known risks, impacts and likelihood associated with TSR, please refer to MAP 5.1.

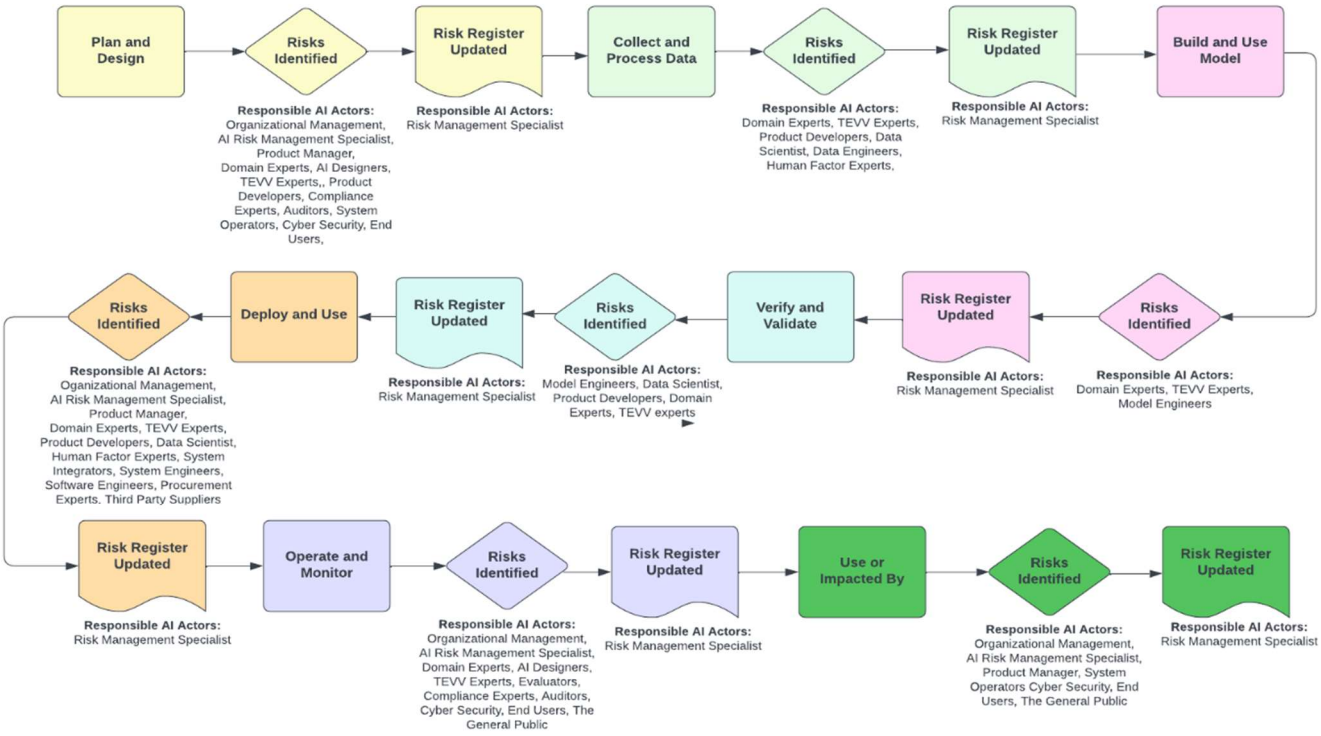

Figure 4-20: TSR Risk Mapping Process.

Table 4-6: TSR Risk Matrix

Risk Matrix		Impact		
		Minor (1)	Moderate (2)	Major (3)
Likelihood	Unlikely (1)	Low (2)	Low (3)	Moderate (4)
	Possible (2)	Low (3)	Moderate (4)	High (5)
	Likely (3)	Moderate (4)	High (5)	Critical (6)
	Certain (4)	High (5)	Critical (6)	Critical (7)

GOVERN 1.3

ABOUT

Risk tolerance, defined by NIST as “the organization’s or AI actor’s readiness to bear the risk in order to achieve its objectives” [4] is not prescribed by NIST’s AI RMF. Rather, it should be influenced based on the legal and regulatory requirements identified by the Compliance Experts, and risk deemed acceptable by the Organizational Management. Risk tolerance is expected to vary by use case and organization. As the malfunctioning of a TSR system poses great safety risk, the risk tolerance assigned to this use case is low, meaning that an enhanced level of risk management activities is required to ensure End User and public safety.

The safety of End Users and the public utilizing ADAS, relies heavily on the proper functioning of TSR systems. As a result, a thorough examination of all risks and their associated impacts is essential. As discussed in GOVERN 1.2 risk impact is determined by AI actors during the risk mapping process and added to the risk register.

To simplify the assessment, the recommended impacts are categorized as minor, moderate, and major. Each category is assigned score of 1 (minor), 2 (moderate), and 3 (major). Minor impact would have no-to minimal impact on End Users. Moderate impact would impact multiple users, but not all users. Major impact is system wide, with the risk of a vehicular accident, and potential risk of loss of life. These scores, in conjunction with the likelihood score, produce the overall risk score. Given the direct correlation between risk impacts and safety, continuous evaluation must be conducted throughout the TSR system’s lifecycle.

Accurate evaluation of risk likelihood is required to ensure proper risk scoring. AI actors are responsible for providing inputs into the risk likelihood as part of the risk mapping process. Additional data points collected from the testing, evaluation, and verification (TEVV) of models and real-world usage should be included as part of the likelihood metrics. With TSR the likelihood of risks may vary by country or region of deployment. The AI Risk Management Specialist should

consider these variabilities and may need to employ multiple scoring approaches based on the targeted areas of deployment.

SUGGESTED ACTIONS

- Establish and communicate policies for mapping and measuring the impacts of TSR system risks.
- Ensure all AI actors fully understand how risks are scored.
- Communicate the organization's risk tolerance level for the TSR system to all AI actors.
- Review GOVERN 5.1 to further understand some well-known and identified TSR risks, their likelihood and impact.
- Ensure compliance with regulatory and legal requirements by sharing the risk register, with identified risks, likelihoods, and impacts with Compliance Experts.

TRANSPARENCY AND DOCUMENTATION

- What organizational policies are in place to validate that the TSR system's impact assessments are consistent?
- Who is accountable for reviewing and assessing assigned impacts?
- What checks and balances are required to assure that risk impacts are documented and assessed properly?

MAP 5.1

ABOUT

As explained in GOVERN 1.2, risk mapping processes must be in place for the entire TSR system lifecycle. Risk mapping includes documenting the risk name, likelihood, impact, and associated AI actors. Table 4-7 highlights well-known identified TSR system risks, their associated

likelihood, impact, and actors. This table can be referenced when starting a risk register but is not intended to be an exhaustive list. Additional risks are expected to be identified as part of the TSR system design, development, deployment, and operation. As previously stated, all AI actors, including end users, are expected to contribute the contents of the risk register, which will remain a dynamic document throughout the system's entire lifecycle.

Table 4-7: TSR AI Risks

Identified Risk	Category	Description	Adversary	Likelihood	Impact	Actors Impacted	Strategy
Weakness in TSR System Design is Exploited	Cybersecurity attacks	If an adversary can find weakness in the TSR system, then they may be able to exploit the weakness, and in turn trigger any found vulnerabilities.	Adversarial Design Attacker	Unlikely	Major	AI Designers TEVV Experts Cybersecurity Evaluators Model Engineers Data Scientist Data Engineers System Operators End Users The General Public	Mitigate
Data Poisoning Cluster Attacks	Cybersecurity attacks	If an attacker can access the training data, then they may be able to inject malicious information into the dataset causing the system to incorrectly identify traffic signs.	Poisoning Adversary	Unlikely	Major	AI Designers TEVV Experts Cybersecurity Evaluators Model Engineers Data Scientist Data Engineers System Operators End Users The General Public	Mitigate
Sensor Spoofing	Cybersecurity attacks	If an attacker can spoof the sensors, then they can cause the vehicle to misinterpret traffic signs.	Spoofing Attacker	Unlikely	Major	Product Developers Systems Engineers AI Designers TEVV Experts Cybersecurity Evaluators System Operators End Users The General Public	Mitigate
Denial of Service (DoS)	Cybersecurity attacks	If an attacker can conduct a denial-of-service attack, then the	Denial of Service Attacker	Unlikely	Major	Software Engineers TEVV Experts Cybersecurity Evaluators	Mitigate

		vehicle will not be able to detect or identify traffic signs, potentially leading to a traffic accident.				Product Developers Systems Engineers System Operators End Users The General Public	
Malware	Cybersecurity attacks	If an attacker can deploy malware onto the system, then the TSR system will not be able to properly function, potentially leading the system to malfunction.	Malware Author	Unlikely	Major	Software Engineers TEVV Experts Cybersecurity Evaluators Product Developers Systems Engineers System Operators End Users The General Public	Mitigate
Public Privacy	Privacy	If a sensor feed inadvertently captures details of the surrounding environment, then the public's privacy may be breached.	Varies	Likely	Minor	Software Engineers TEVV Experts Cybersecurity Evaluators Product Developers Systems Engineers System Operators End Users The General Public	Mitigate
CAN Interruptions/Interference	Cybersecurity attacks	If a cybercriminal conducts an attack on a vehicle's CAN system, i.e., through signal jamming, then the vehicle may not be able to correctly identify traffic signs, potentially resulting in a traffic accident.	Varies	Unlikely	Major	Software Engineers TEVV Experts Cybersecurity Evaluators Product Developers Systems Engineers System Operators End Users The General Public	Mitigate
Distribution Shift Brittleness	Misclassification Model Training Error	If a TSR system misclassifies a traffic sign due to vandalism,	N/A	Unlikely	Major	AI Designers TEVV Experts Cybersecurity Evaluators Model Engineers	Watch

		environmental condition, aging signs, etc., then the system may not operate as expected, potentially resulting in a traffic accident.				Data Scientist Data Engineers System Operators End Users The General Public	
Users' dependence on the system may prevent them from noticing a system failure.	Over-reliance on system	If system operators or end users become too dependent on the system, and failure to identify system failures, then a traffic accident may result.	N/A	Possible	Major	System Operators End Users The General Public	Mitigate
TSR System Failure	System failure	The TSR system relies on cameras, on board computers, and deep learning models for processing data. If one system fails, then TSR will no longer work as expected.	N/A	Unlikely	Major	Product Developers Systems Engineers AI Designers TEVV Experts Evaluators System Operators End Users The General Public	Watch
Traffic sign is not visible due to being blocked by a tree or other object.	Environmental Factors	If a traffic sign is blocked by a tree or other object, then it will not be visible to the camera.	N/A	Likely	Major	System Operators End Users The General Public	Mitigate
Stop signs on school busses are not correctly identified and recognized by the system.	Misclassification Model Training Error	If a school bus stops and extends the stop sign and the vehicle does not recognize it, then the vehicle may try to pass the bus.	N/A	Possible	Major	Product Developers Model Engineers Systems Engineers AI Designers TEVV Experts Evaluators System Operators End Users The General Public	Mitigate

Traffic signs held by construction workers are not correctly recognized by the system.	Misclassification Model training error	If a worker is holding a sign in the street, and the vehicle does not recognize it, then the vehicle may not adhere to the sign requirements.	N/A	Possible	Major	Product Developers Systems Engineers AI Designers TEVV Experts Evaluators System Operators End Users The General Public	Mitigate
A traffic sign has fallen.	Environmental Factors	If a traffic sign has fallen, then the vehicle's on-board system will not be able to detect through visual means.	N/A	Possible	Major	System Operators End Users The General Public	Mitigate.
Traffic sign has been damaged or tampered with.	Cybersecurity attacks System failure Environmental Factors	If a traffic sign has been damaged or tampered with, then the vehicle's on-board system may not be able to correctly identify the sign, which could lead to a traffic incident. If traffic signs have been damaged or tampered with the intent of modifying system inputs, or ML data, then the system may malfunction, which could lead to a traffic incident.	Sign Vandals	Unlikely	Major	System Operators End Users The General Public	Mitigate
Traffic Sign is not registered in system.	False Reads	If a traffic sign is not registered in the system, then the vehicle will not recognize it and will not respond accordingly.	N/A	Unlikely	Major	AI Designers TEVV Experts Cybersecurity Evaluators Model Engineers Data Scientist Data Engineers System Operators End Users	Watch

						The General Public	
TSR system may register a false positive.	False Reads	If a TSR system identifies a sign that is not present, then the system may malfunction.	N/A	Unlikely	Major	AI Designers TEVV Experts Evaluators Model Engineers Data Scientist Data Engineers System Operators End Users The General Public	Mitigate
TSR systems are designed to recognize miles per hour, or kilometers per hour, and cannot easily transition between the two.	False Reads	If an end user travels between countries, where different speed measurements are used, then the system may not recognize the change, and will adopt speed incorrectly.	N/A	Unlikely	Moderate	Product Developers Systems Engineers AI Designers TEVV Experts System Operators End Users The General Public	Mitigate
TSR sign type is not correctly identified.	False Reads	If a vehicle fails to detect a sign, i.e., a stop sign, or wrong way sign incorrectly then there is an increased risk for an accident.	N/A	Unlikely	Major	AI Designers TEVV Experts Evaluators Model Engineers Data Scientist Data Engineers System Operators End Users The General Public	Watch
Regulatory and legal requirements are not fully documented or understood, introducing risk to the OEM.	Regulatory and Legal Requirements	If legal and regulatory requirements are not fully understood and defined, then the OEM is at risk of penalties and fines.	N/A	Unlikely	Moderate	Organizational Management Compliance Experts Auditors Third Party Suppliers	Mitigate
AI Actor Attrition	Human Resources	If an AI actor departs from the company during the AI system	N/A	Unlikely	Minor	Organizational Management Human Resources	Accept

		lifecycle, then expertise could be lost impacting product deliverables and causing cost implications to the organization.					
AI Actor Proficiency	Human Resources	If AI Actors are not proficient in their assigned responsibilities, the system may be designed with faults or not properly tested.	N/A	Possible	Moderate	Organizational Management Human Resources	Mitigate
Third Party Systems	Third Party	If third parties do not properly test provided components and effectively document risks, then unknown risks could be introduced into the TSR system.	N/A	Possible	Major	Organizational Management Procurement Experts Third Party Suppliers	Watch

SUGGESTED ACTIONS

- Establish a policy for assessing the likelihood and impact of risks by engaging experts and using a quantitative scale to garner the average response.
- Conduct regular assessments of risks to confirm if there has been a change in the risk, associated likelihood, or impact.
- Ensure that the risk register remains dynamic, through regular audits.

TRANSPARENCY AND DOCUMENTATION

- Have all the AI actors impacted by the risk been properly identified?
- Can independent assessment be conducted of the risk register to ensure validity?

MEASURE 4.1

ABOUT

TSR systems are intended to be deployed on public roads as part of an autonomous vehicle; A key to a successful deployment is for End Users to trust the system to operate as expected, detecting the traffic signs at a rate better than human drivers. UI/UX and Human Factor Experts can be used in conjunction with tools, i.e., surveys to collect feedback from End Users. Insights gained will help ensure metrics used in TEVV are accurate reflections of user expectations.

SUGGESTED ACTIONS

- Define which AI actors are responsible for interacting with End Users to gain insights.
- Identify when End Users will be contacted, and what formats will be solicited for gathering information.
- Evaluate feedback obtained with all AI actors and review how it impacts identified risks, associated likelihood and impacts.
- Document how End User feedback will be integrated into the TSR systems product plans, and risk management activities.

TRANSPARENCY AND DOCUMENTATION

- Will the End User's feedback require adjustments to be made to the design?
- How does the feedback obtained align with TEVV plans?

MANAGE 1.2

ABOUT

To define strategies for dealing with risks, a risk score must be developed based on the assessed likelihood and impact of the risk. As mentioned in MAP 5.1, organizations should

establish a policy for assessing the likelihood and impact of risks by engaging experts, and then garner the average response. Table 4-8 builds upon risks identified in table 4-7 and assigns a risk score. As with all aspects of the risks, the scores are dynamic and may change throughout the TSR system's lifecycle. MANAGE 1.3 then utilizes the risk score to define treatment required.

Table 4-8: TSR Risk Management Score

Identified Risk	Likelihood	Impact	Risk Score
Weakness in TSR System Design is Exploited	Unlikely (1)	Major (3)	Moderate (4)
Data Poisoning Cluster Attacks	Unlikely (1)	Major (3)	Moderate (4)
Sensor Spoofing	Unlikely (1)	Major (3)	Moderate (4)
Denial of Service (DoS)	Unlikely (1)	Major (3)	Moderate (4)
Malware	Unlikely (1)	Major (3)	Moderate (4)
Public Privacy	Possible (2)	Minor (1)	Low (3)
CAN Interruptions/Interference	Unlikely (1)	Major (3)	Moderate (4)
Distribution Shift Brittleness	Unlikely (1)	Major (3)	Moderate (4)
Users' dependence on the system may prevent them from noticing a system failure.	Possible (2)	Major (3)	High (5)
TSR System Failure	Unlikely (1)	Major (3)	Moderate (4)
Traffic sign is not visible due to being blocked by a tree or other object.	Likely (3)	Major (3)	Critical (6)
Stop signs on school buses are not correctly identified and recognized by the system.	Possible (2)	Major (3)	High (5)
Traffic signs held by construction workers are not correctly recognized by the system.	Possible (2)	Major (3)	High (5)
A traffic sign has fallen.	Possible (2)	Major (3)	High (5)
Traffic sign has been damaged or tampered with.	Unlikely (1)	Major (3)	Moderate (4)
Traffic Sign is not registered in system.	Unlikely (1)	Major (3)	Moderate (4)
TSR system may register a false positive.	Unlikely (1)	Major (3)	Moderate (4)
TSR systems are designed to recognize miles per hour, or kilometers per hour, and cannot easily transition between the two.	Unlikely (1)	Moderate (2)	Low (3)
TSR sign type is not correctly identified.	Unlikely (1)	Major (3)	Moderate (4)
Regulatory and legal requirements are not fully documented or understood, introducing risk to the OEM.	Unlikely (1)	Moderate (2)	Low (3)
AI Actor Attrition	Unlikely (1)	Minor (1)	Low (2)
AI Actor Proficiency	Possible (2)	Moderate (2)	Moderate (4)
Third Party Systems	Possible (2)	Major (3)	High (5)

SUGGESTED ACTIONS

- Ensure the organization's risk tolerance level for the TSR system is communicated to all AI actors as stated in GOVERN 1.3.

- Plan for regular review of risks, including the organization risk tolerance, likelihood, and impact.
- Establish a plan for regular reviews and audits of the risk register and assigned score.

TRANSPARENCY AND DOCUMENTATION

- Is there a clear understanding of regarding what risk score should be reported to governance?
- Is Organizational Management aligned with the assigned risk scores?

MANAGE 1.3

After the risk identification is completed in MAP 5.1, and corresponding risk scoring completed in MANAGE 1.3, a risk response plan should be developed for all the TSR system's AI risks deemed high or critical. Risk response options include mitigate, transfer, avoid, watch, or accept. Table 4-9 reviews example strategies for each risk with a risk score of high or critical. It is the responsibility of the assigned AI Risk Management Specialist to work with the AI actors on developing risk management strategies for each risk, based on the risk score.

Table 4-9: TSR Risk Management Strategies

Identified Risk	Risk Score	Risk Response
Users' dependence on the system may prevent them from noticing a system failure.	High (5)	**Mitigate.** As part of the design, designers should plan to incorporate data from redundant systems, i.e., GPS or built in navigation systems to be concurrently with information obtained from cameras and sensors. By utilizing a combination of live sensor data, and pre-loaded traffic sign information, the TSR system will be able to make better decisions and decrease the likelihood of risk impact. As part of the design, if the TSR system detects an error, a visual and auditory alert will be triggered to notify end users. Data gathered from system failures must be sent to the TSR system's administrators through the vehicles modem for incident tracking.
Traffic sign is not visible due to being blocked by a tree or other object.	Critical (6)	**Mitigate.**

		As part of the design, designers should plan to incorporate data from redundant systems, i.e., GPS or built in navigation systems to be concurrently with information obtained from cameras and sensors. By utilizing a combination of live sensor data, and pre-loaded traffic sign information, the system will be able to make better decisions, and decrease the likelihood of risk impact.
Stop signs on school buses are not correctly identified and recognized by the system.	High (5)	**Mitigate.** AI Model Engineers must plan for this situation as part of the model development. TEVV experts are responsible for ensuring the model properly identifies stop signs on school busses and responds appropriately in both simulation and physical tests.
Traffic signs held by construction workers are not correctly recognized by the system.	High (5)	**Mitigate.** AI Model Engineers must plan for this situation as part of the model development. TEVV experts are responsible for ensuring the model properly identifies signs held by construction workers and responds appropriately in both simulation and physical tests.
A traffic sign has fallen.	High (5)	**Mitigate.** As part of the design, designers should plan to incorporate data from redundant systems, i.e., GPS or built in navigation systems to be concurrently with information obtained from cameras and sensors. By utilizing a combination of live sensor data, and pre-loaded traffic sign information, the system will be able to make better decisions, and decrease the likelihood of risk impact.
Third Party Systems	High (5)	**Watch** As part of the procurement process, Procurement Experts must ensure that risk management processes are in place for handling all Third-Party Systems risks. The ownership of the

SUGGESTED ACTIONS

- Ensure that a risk response is developed for all risks identified as High or Critical.
- Develop additional plans for risks assigned to be mitigated. (For example, public privacy can be protected through mitigating the pre-processing environment to ensure that there is not an unintended disclosure of private information.)
- Validate that risk responses aligns with goals of trustworthiness.
- Review risks, assign and prioritize response strategies.

TRANSPARENCY AND DOCUMENTATION

- Do the risk response strategies comply with legal and regulatory requirements?
- Do any of the risks require multiple response strategies?
- Are there any budget implications or limitations that may impact the ability to respond to the risks?
- If multiple risk response strategies are required, who is responsible for defining which order to execute the strategies?

SECTION 5: Incident Management

Incident management is a critical component of a successful TSR system deployment. Organizations should have a comprehensive, well documented plan for incident management, including established communication and incident response plans. The following section covers incident governance, incident priorities, third-party incidents, practices for regular feedback and engagement of stakeholders, mechanisms for TSR system decommissioning and suggested best practices for incident communication.

GOVERN 4.3

ABOUT

As TSR systems are key for autonomous vehicle functionality, well-documented practices need to be in place for incident management, including established communication plans. While many organizations may have established incident response plans, deploying an AI system, i.e., TSR may have additional complexities that should be considered. For example, incidents may occur within the model, or through the system's ability to learn. The following section provides considerations for adjusting incident management plans to help account for the new complexities associated with AI deployments.

SUGGESTED ACTIONS

- Review established incident response plans. If your organization does not have an incident response plan, one must be developed and deployed prior to deploying the TSR system.
- Ensure AI actors understand how to correctly identify incident priority, as demonstrated in table 4-10. Additionally, ISO/IEC 27035-2:2016 is a good

reference document for organizations to utilize when developing or reviewing Information Security Management and Incident response plans, providing detailed guidelines for teams to plan and prepare for incident response.

- Ensure expectations for incident communication are explicitly explained to all AI actors.
- Document all incident reports in an incident log and communicate with stakeholders.
- Communicate strategy for incident management if the incident came from a previously documented risk. If an incident has a direct impact on the public, Organizational Management are responsible for overseeing public disclosure and information sharing.

Table 4-10: TSR Incident Priority

Traffic Sign Recognition Incident Priority	**IMPACT**			
		HIGH	**MEDIUM**	**LOW**
URGENCY (Priority)	**HIGH**	Critical	High	Medium
	MEDIUM	High	Medium	Low
	LOW	Medium	Low	Low

TRANSPARENCY AND DOCUMENTATION

- Does Organizational Management have a clear policy for communicating identified incidents with the public?
- Are there open channels established for the End User to report incidents?

- Will vehicles be able to send incident information via connected systems to the organization?

GOVERN 6.2

ABOUT

When utilizing components from third-parties appropriate incident management plans must be in place for all risks deemed high to critical. Incident response plans and expected service level agreements (SLAs) should be included in the third-party's purchase agreements. Third-party suppliers are expected to share identified incidents with the automotive OEM promptly, to ensure proper response and communication plans are implemented.

SUGGESTED ACTIONS

- The Procurement Expert is responsible for ensuring that all incident response plans and SLAs are properly documented in the third-party's contract.
- Any incidents caused by the third-party system will be recorded by the AI Risk Management Specialist in the incident log.
- Resolutions when identified must be recorded.
- Clearly document which components have been obtained by third-parties in the system design.

TRANSPARENCY AND DOCUMENTATION

- Does the Procurement Expert have all the necessary information for creating the purchase agreement and third-party contract?
- Will Legal Experts be required to review the contract?
- Are third-party policies clearly communicated to AI actors?

MAP 5.2

ABOUT

To ensure a TSR system is operating as expected, organizations need to have an established feedback process to capture positive and negative feedback. This will entail utilizing tools including, but not limited to surveys, customer support hotlines, emails, social media, and focus groups. When feedback includes incident details, a member of the Product Management team will follow up with the AI actor who reported the incident to gather additional context that will be required for risk, incident, and impact analysis.

SUGGESTED ACTIONS

- As stated in GOVERN 1.5, organizations must establish communication channels between End Users and the organization. End Users and the public should be able to immediately report incidents, for further investigation. Incidents that are defined to have a high impact, such as loss of human life, should be communicated with all AI actors, and End Users. This communication should include immediate remediation strategies, until a full mitigation plan has been developed and deployed.
- Ensure that there are a mix of quantitative and qualitative methods to assess the TSR system's performance that may be defined as or become incidents. This can be done through monitoring the system, and having regular updates transmitted to the automotive OEM via connected vehicle transmissions.
- Solicit regular feedback from End Users to determine if the system performs in an untrustworthy manner.

- Compile reported incidents into an incident report, which includes details on impact, documented metrics and assigned AI actor responsible for addressing.

TRANSPARENCY AND DOCUMENTATION

- Do all AI actors have a clear way to share feedback on TSR system trustworthiness and incident details?
- Are third-party auditors established for reviewing deployed system performance?

MEASURE 3.3

ABOUT

Feedback documented in MAP 5.2 is measured by the established qualitative and quantitative measures. The risk register will have reference to the defined impact levels as shown in table 4-9. As mentioned previously, the risk register is a dynamic document, where risks and assigned likelihoods and impacts should be reassessed and updated as needed. Defined metrics from this subcategory will be used to assess which communication protocols will be used as defined in MANAGE 4.3.

SUGGESTED ACTIONS

- Measure incidents documented in MAP 5.2, assigning established qualitative and quantitative measure.
- Utilize UX Experts to monitor relevant metrics and share results with relevant AI actors.
- Ensure the risk register is reviewed and that impact levels are adjusted as needed.
- Include any metrics in the incident report.

- Establish incident response plan based on metrics. For example, if a certain number of End Users report the same incident and it is determined that the impact is critical where human life could be impacted, an immediate response, such as a software upgrade with immediate user communication will be required.
- Define measures of efficacy of incident reporting to ensure practices are effective.

TRANSPARENCY AND DOCUMENTATION

- Do End Users clearly understand how to provide feedback on incidents?
- Can all AI actors easily review the incident report?
- What information is available to stakeholders, including the public?
- Do incident response plans establish a feeling of trustworthiness for the public?
- Are incident measurements aligned with legal and regulatory requirements for AVs?

MANAGE 2.4

ABOUT

To effectively manage TSR related incidents organizations need to ensure appropriate processes and procedures are in place with roles and responsibilities assigned and understood. Incident response plans will go above and beyond the risk response in table 4-9. This may include the remediation, disengaging, or decommissioning of TSR systems.

As TSR functionality is critical for AVs with levels 3+ automation, the most likely response to incidents will be remediation. Based on the incident type, the responsible AI actors will need to define the response solution, and associated service level agreement (SLA) for solution

deployment with Organizational Management overseeing communication to End Users and the public. Communication must follow legal and regulatory requirements.

If the TSR AI system is identified as posing an unacceptable risk to End Users or the public which cannot be effectively mitigated by the OEM, or third-party suppliers, the decommissioning of TSR system should be initiated. Prior to decommissioning, details including root cause, impact, opportunities for mitigation and redeployment should be documented in the incident report. Furthermore, incidents warranting potential premature TSR system decommissioning should be brought forth to the organizational governance meeting for evaluation and will require sign-off of Organizational Management.

SUGGESTED ACTIONS

- Regularly review the incident log to ensure proper mitigations are being followed.
- Ensure all incidents are reported in regular governance, with organizational alignment on response strategies.
- Publish processes for remediation, disengaging, or decommissioning of TSR systems that are available for all AI actors to review to and comment.
- Ensure the team responsible for incident management understands how to properly conduct and document root case analysis.
- Retain documentation as required to meet legal and regulatory requirements.
- Document the upstream and downstream consequences if the TSR system requires decommissioning.
- Ensure test plans are developed for full system testing in the event of a required remediation.

TRANSPARENCY AND DOCUMENTATION

- Are all roles and responsibilities clearly defined for incident management?
- Does the team need to participate in incident management training, i.e., ITIL incident management?
- What additional methods are required to protect against cyber incidents, i.e., bug bounties, or red teaming?

MANAGE 4.3

ABOUT

The incident report is a dynamic document that requires regular updates, and detailed information, including how incidents were identified, associated risk (if applicable), incident response strategy, assigned SLA, impact assessment, and communication plans. Using a traceability matrix can assist with tracking risks and incidents amongst the product/project documentation, including the risk register, incident report and TEVV test plans to ensure AI actors have a full view of the system information documented to date.

SUGGESTED ACTIONS

- Ensure an AI resource is assigned to maintain traceability between all product documents, including the risk register, incident report, and any associated test plans.
- Review communication plans established in the GOVERN section are well understood by those in the organization.
- Maintain the incident details in a database, or other incident tracking system that is consistent with legal and regulatory requirements.

- Ensure document management includes processes, i.e., version history, name of updater, updates made, and retention period.

TRANSPARENCY AND DOCUMENTATION

- Are additional actions required to ensure all information is easily accessible by the TSR system's AI actors?
- Are plans in place to review previous impacts and lessons learned as part of planning for TSR system upgrades?
- Who is responsible for capturing any lessons learned and following up on incorporating them into the TSR system product plans?
- What type of incident information can be made publicly available?

Chapter 5: Industry Expert Feedback

To ensure the validity of the use case profile for traffic sign recognition, we sought to obtain feedback on our profile from industry experts, who varied in roles and responsibilities, with three presently employed by an automotive OEM, and one employed as a third-party supplier. Our reviewers included an ADAS Feature Owner, Senior Software Engineer (perception), ADAS Product Manager, and an ADAS Manager. Each reviewer provided unique insights into the profile, which was collected through a survey, emails, and marked copies of the profile. Much of the feedback obtained was included in the profile, with remaining insights shared below.

5.1 Survey Results

To understand the state of risk management at various organizations, we surveyed three industry professionals regarding current AI risk management processes. None of the three participating were familiar with NIST's AI Risk Management Framework. The learnings showed us that one of the professional's organizations had well established risk management processes, while the other two did not, or were unsure. In response to two questions, regarding whether the risk management processes were consistent throughout the organization with everyone having a clear understanding of their roles and responsibilities, two of the respondents said they were not sure, while the other replied, no. Additionally, we learned that risk management processes vary when products contain artificial intelligence, with the responsibility of capturing the risks, falling under the team who owns the specific tool being developed. Overall, the feedback obtained on today's practices solidified our stance on a strong need for established AI risk management

processes for autonomous vehicles, and the need for clear communication between all AI actors. One respondent's reply furthered this stance by stating, "…I am not aware of a unifying process that involves the communication specific to AI risks with all the various stakeholders."

5.2 Future Profile Opportunities

To understand future profile opportunities, we asked the three participants, "What is the single most important risk that you would like to see addressed in autonomous vehicles?" Two participants replied with, "Passenger and pedestrian safety," and "harm to people", while the third provided a more detailed response, "Before we see fully autonomous vehicles i.e., level 4,5 we will have a long period with a high consumer base of semi-autonomous i.e., level 2+ ADAS features on the road. The biggest challenge for such systems is a well-defined operational design domain, making sure the feature is activated in the defined operational domain where the developers have rigorously tested and mitigated maximum risks related to the feature. Human driver reliance on ADAS systems comes with a cost where too much reliance on the feature leads to more distracted driving and if the ADAS feature continues to operate in a non-ODD zone the risks are very high. Another challenge industry wide is transparency and communication to consumers about system limitations and performance. If consumers are aware and educated more about the system and its limitations it will lead to a safer as well as enjoyable cooperation between human and machine."

Chapter 6: Conclusion

Managing technological risk is essential. The increasing development and deployment of technology that utilizes AI capabilities is only increasing this need. NIST's AI Risk Management Framework (AI RMF 1.0) sets a strong foundation for organizations to use to establish processes and procedures that will assist them in implementing their own AI risk management processes.

Specifically, as it pertains to autonomous vehicles, our public survey results show that while 77.4% of people find TSR a helpful functionality, 65.4% do not trust an autonomous vehicle that removes the need for a human driver. Additionally, 64.1% of people want to have a better understanding of the risks associated with TSR systems.

Our work leverages the NIST AI RMF framework's functions, categories, and subcategories as a comprehensive guideline. It combines this framework with extensive research on the associated automated technologies to develop a temporal AI risk management use case profile for TSR, that has been peer reviewed, and helps to solve the issues of AV trustworthiness depicted in our survey results through elaborate AI actor and end user communication. This profile offers valuable insight into effectively managing risk throughout the AI lifecycle, specifically focusing on the associated technologies of the TSR use case.

References

[1] P. C. Manning, "Artificial Intelligence Definitions," Stanford University, September 2020. [Online]. Available: https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf. [Accessed 23 June 2023].

[2] U.S. Department of Transportation, "Preparing for the Future of Transportation: Automated Vehicles 3.0," 4 October 2018. [Online]. Available: https://www.transportation.gov/av/3. [Accessed 28 June 2023].

[3] SAE, "SAE Levels of Driving Automation™ Refined for Clarity and International Audience," 3 May 2021. [Online]. Available: https://www.sae.org/blog/sae-j3016-update. [Accessed 18 November 2023].

[4] National Institute of Standards and Technology (NIST), "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," January 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[5] ISO, "ISO 31000; Risk management - Guidelines.," ISO, Geneva, Switzerland, 2018.

[6] National Institute of Standards and Technology, "AI Risk Management Framework," U.S. Department of Commerce, 2023. [Online]. Available: https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF. [Accessed 8 June 2023].

[7] J. Whiting, Autonomous Vehicles, Mankato: Creative Education and Creative Paperbacks, 2020.

[8] S. Azam, F. Munir, A. M. Sheri, J. Kim and M. Jeon, "System, Design and Experimental Validation of Autonomous Vehicle in an Unconstrained Environment," 22 October 2020. [Online]. Available: www.mdpi.com/journal/sensors. [Accessed 16 June 2023].

[9] M. N. Ahangar, Q. Z. Ahmed, F. A. Khan and M. Hafeez, "A Survey of Autonomous Vehicles: Enabling Communication," 21 January 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/3/706. [Accessed 16 June 2023].

[10] O. Sharma, N. Sahoo and N. Punhan, "Engineering Applications of Artificial Intelligence," 4 March 2021. [Online]. Available: https://pdf.sciencedirectassets.com/271095/1-s2.0-S0952197621X00036/1-s2.0-S0952197621000580/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEPD%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJGMEQCIAz4xUs%2F3AWyMrFo6Jzli4FquvSCD2782rsFWN95xfOVAiAvvkxb2a1up9. [Accessed 16 June 2023].

[11] W. Zong, C. Zhang, Z. Wang, J. Zhu and Q. Chen, "Architecture Design and Implementation of an Autonomous Vehicle," *IEEE Access,* vol. 6, no. doi: 10.1109/ACCESS.2018.2828260, pp. 21956-21970, 9 May 2018.

[12] K. Kim, J. S. Kim, S. Jeong, J.-H. Park and H. K. Kim, "Cybersecurity for autonomous vehicles: Review of attacks and defense," 5 January 2021. [Online]. Available: https://pdf.sciencedirectassets.com/271887/1-s2.0-S0167404821X00028/1-s2.0-S0167404820304235/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEPT%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQD5kR3btHG43lAIGAw8sG7qva4rskB9hxpVFp%2FsXgIZlwIgTB1vZisuX5. [Accessed 16 June 2023].

[13] M. S. U. Alam, S. Iqbal, M. Zulkernine and C. Liem, "Securing Vehicle ECU Communications and Stored," IEEE, 2019. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8762043. [Accessed 23 June 2023].

[14] R. N. Charette, "This Car Runs on Code," IEEE Spectrum, 01 February 2009. [Online]. Available: https://spectrum.ieee.org/this-car-runs-on-code. [Accessed 25 June 2023].

[15] S. Woo, H. J. Jo and D. H. Lee, "A Practical Wireless Attack on the Connected Car," IEEE, April 2015. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6894181. [Accessed 23 June 2023].

[16] S. L. X. H. Fang Zhou, "Development method of simulation and test system for vehicle body CAN bus based on CANoe," in *2008 7th World Congress on Intelligent Control and Automation*, Chongqing, China, 25-27 June 2008.

[17] T. Nolte, H. Hansson and L. L. Bello, "Automotive Communications - Past, Current and Future," IEEE, 2005. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1612631. [Accessed 23 June 2023].

[18] © LIN Consortium, 2010., "LIN Specification Package," 31 December 2010. [Online]. Available: https://www.lin-cia.org/fileadmin/microsites/lin-cia.org/resources/documents/LIN_2.2A.pdf. [Accessed 25 June 2023].

[19] Pal-Stefan and B. Groza, "Practical Security Exploits of the FlexRay In-Vehicle Communication Protocol," in *Lecture Notes in Computer Science*, Springer, Cham, Springer Nature Switzerland AG, 2019, pp. 172-187.

[20] J. Shepard, "What is the FlexRay communications network?," 15 June 2022. [Online]. Available: https://www.microcontrollertips.com/what-is-the-flexray-network-faq/. [Accessed 25 June 2023].

[21] R. Ayala and T. K. Mohd, "Sensors in Autonomous Vehicles: A Survey," *AMSE. J. Auton. Veh. Sys.,* Vols. JAVS-21-1008, no. https://doi.org/10.1115/1.4052991, p. 1(3): 031003 (12 pages), July 2021.

[22] S. R. a. M. Ballesta-Garcia, "An Overview of Lidar Imaging Systems for Autonomous Vehicles," *Applied Sciences,* vol. 9, no. 19, p. 4093, September 2019.

[23] C. Wang, X. Wang, H. Hu, Y. Liang and G. Shen, "On the Application of Cameras Used in Autonomous Vehicles," *Arch Computat Methods Eng ,* vol. 29, no. https://doi.org/10.1007/s11831-022-09741-8, pp. 4319-4339, 2022.

[24] J. V. Brummelen, M. O'Brien, D. Gruyer and Homayoun Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transportation Research Part C: Emerging Technologies,* vol. 89, no. https://doi.org/10.1016/j.trc.2018.02.012., pp. 384-406, 2018.

[25] B. S. Lim, S. L. Keoh and V. L. L. Thing, "Autonomous vehicle ultrasonic sensor vulnerability and impact assessment," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, Singapore, 2018.

[26] Y. Li and H. Shi, Advanced Driver Assistance Systems and Autonomous Vehicles, Singapore: Springer, 2022.

[27] J. Wang, Y. Shao and Y. G. a. R. Yu, "A Survey of Vehicle to Everything Testing," *MDPI,* Vols. 19, no. 2, no. https://doi.org/10.3390/s19020334, p. 334, January 2019.

[28] Z. T. Kardovacs, Z. Paroczi, E. Varga, A. Siegler and P. Lucz, "Real-time traffic sign recognition," in *2011 2nd International Conference on Cognitive Inforcommunications (CogInfoCom)*, Budapest, 2011.

[29] N. H. Traffic, "Occupant Protection for Vehicles With Automated Driving Systems," 30 March 2022. [Online]. Available: https://www.govinfo.gov/content/pkg/FR-2022-03-30/pdf/2022-05426.pdf. [Accessed 06 07 2023].

[30] I. 23150, *Road vehicles — Data communication between sensors and data fusion unit for automated driving functions - Logical interface,* BSI Standards Publication, 2023.

[31] ISO, *BS EN ISO 20524-1:2022,* BSI Standards Publication, 2022.

[32] ISO, *BS EN ISO 20524-2:2022,* BSI Standards Publication, 2022.

[33] NHTSA, "Levels of Automation," [Online]. Available: https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-05/Level-of-Automation-052522-tag.pdf. [Accessed 25 August 2023].

[34] Ford Motor Company, "What is Ford Speed Sign Recognition," Ford Motor Company, 2023. [Online]. Available: https://www.ford.com/support/how-tos/ford-technology/driver-assist-features/what-is-speed-sign-recognition/. [Accessed 11 August 2023].

[35] Volvo USA Car Support, "Road Sign Information," Volvo , 2019. [Online]. Available: https://volvo.custhelp.com/app/answers/detail/a_id/9572/~/road-sign-information-%28rsi%29. [Accessed 11 August 2023].

[36] Tesla, "Traffic Light and Stop Sign Control," Tesla, 2023. [Online]. Available: https://www.tesla.com/ownersmanual/modely/en_eu/GUID-A701F7DC-875C-4491-BC84-605A77EA152C.html#:~:text=Traffic%20Light%20and%20Stop%20Sign%20Control%20is%20designed%20to%20recognize,Aware%20cruise%20control%20or%20Autosteer. [Accessed 11 August 2023].

[37] https://waymo.com/waymo-driver/, "Waymo Driver," Waymo, [Online]. Available: https://waymo.com/waymo-driver/. [Accessed 8 September 2023].

[38] NHTSA, "NHTSA Estimates for 2022 Show Roadway Fatalities Remain Flat After Two Years of Dramatic Increases," 20 April 2023. [Online]. Available: https://www.nhtsa.gov/press-releases/traffic-crash-death-estimates-2022. [Accessed 4 September 2023].

[39] US Department of Transportation, "National Roadway Safety Strategy," 27 January 2022. [Online]. Available: https://www.transportation.gov/briefing-room/us-transportation-secretary-pete-buttigieg-announces-comprehensive-national-roadway. [Accessed 4 September 2023].

[40] US Department of Transportation, "What Is a Safe System Approach," 13 October 2022. [Online]. Available: https://www.transportation.gov/NRSS/SafeSystem. [Accessed 4 September 2023].

[41] PennDOT Driver & Vehicle Services, "Signs," 2023. [Online]. Available: https://www.dmv.pa.gov/Driver-Services/Driver-Licensing/Driver-Manual/Chapter-2/Pages/Signs.aspx. [Accessed 5 September 2023].

[42] A. RAO, "Passing school buses remains an issue," 27 February 2017. [Online]. Available: https://www.observertoday.com/news/page-one/2017/02/passing-school-buses-remains-an-issue/. [Accessed 5 September 2023].

[43] Y. Oren and I. B. Sinai, "The definitive guide to AI monitoring," Medium (Originally Published in Towards Data Science), 3 November 2020. [Online]. Available: https://towardsdatascience.com/the-definitive-guide-to-ai-monitoring-2427812cc1b. [Accessed 24 September 2023].

[44] L. Sirojetdinova, E. Kislitsyn and V. Gorodnicev, "UI / UX methodologies: Analysis and efficiency assessment," in *Melville: American Institute of Physics*, https://doi.org/10.1063/5.0162767, 2023.

[45] D. Carnegie, How to Win Friends & Influence People, New York: Pocket Books, a division of Simon & Schuster, 1981.

[46] C. E. TUncali, G. Fainekos, H. Ito and J. Kapinski, "Simulation-based Adversarial Test Generation for Autonomous Vehicles with Machine Learning Components," in *IEEE Inteligent Vehicles Symposium (IV)*, Changshu, Suzhou, China, 2018.

[47] E. R. B. R. R. E. P. C. W. A. S. D. F. N. M. Foroogh Hajiseyedjavadi, "Effect of environmental factors and individual differences on subjective evaluation of human-like and conventional automated vehicle controllers,," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 90, no. ISSN 1369-8478, https://0-doi-org.wizard.umd.umich.edu/10.1016/j.trf.2022.07.018., pp. 1-14, 2022.

[48] Z. (. I. o. S. a. T. Kootbally, "Progress on Systems Interaction in Automated Vehicles," 5-8 September 2023. [Online]. Available: https://www.nist.gov/system/files/documents/2023/09/25/Systems-Interaction-NIST-KOOTBALLY.pdf. [Accessed 1 October 2023].

[49] B. Sreeja, S. Bokka, G. Shravya and K. S. V. Vardini, "Traffic Sign Detection using Transfer learning and a Comparison Between Different Techniques," in *2nd Asian Conference on Innovation and Technology (ASIACON)*, Pune, India, 2022.

[50] J. Mishra and S. Goyal, "An Effective Automatic Traffic Sign Classification and Recognition Deep Convolutional Networks.," *Multimedia Tools and Applications 81.13 (2022),* vol. 81(13), no. doi:https://doi.org/10.1007/s11042-022-12531-w, pp. 18915-18934, 2022.

[51] "Traffic Signs Identification by Deep Learning for Autonomous Driving," in *Stevenage: The Institution of Engineering & Technology*, doi:https://doi.org/10.1049/cp.2018.1382, 2018.

[52] A. Magazinus, N. Mellegard and L. Olsson, "Bug Bounty Programs - a Mapping Study," in *45th Euromicro Conference on Software Engineering and Advanced Applications*, Kallithea, Greece, 2019.

Appendix

Public Survey Questions

To understand the public's perception of the trustworthiness of autonomous vehicles, and their perceptions on the technology, including when it would be ready, and what would make them feel more comfortable to adapt the technology, we conducted a survey of 159 users. The survey was conducted via the internet, with responses solicited via, social media platforms (LinkedIn and Facebook), and through direct conversations where the survey link was provided. The demographics of users participating were diverse, with a wide range of age ranges.

Question #1

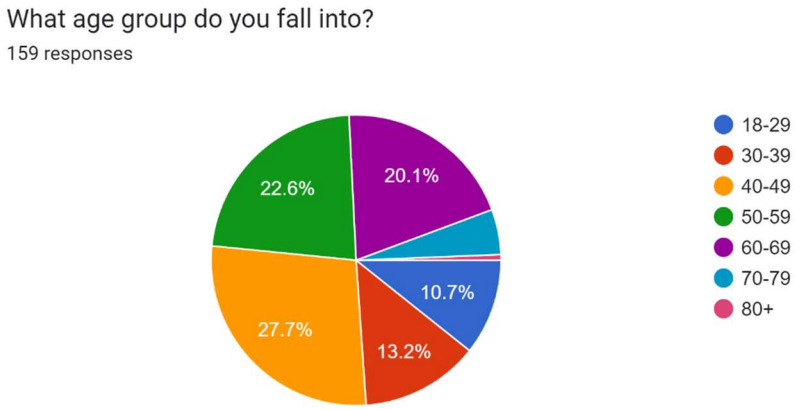

Figure A-21: Question 1 Results

Question #2

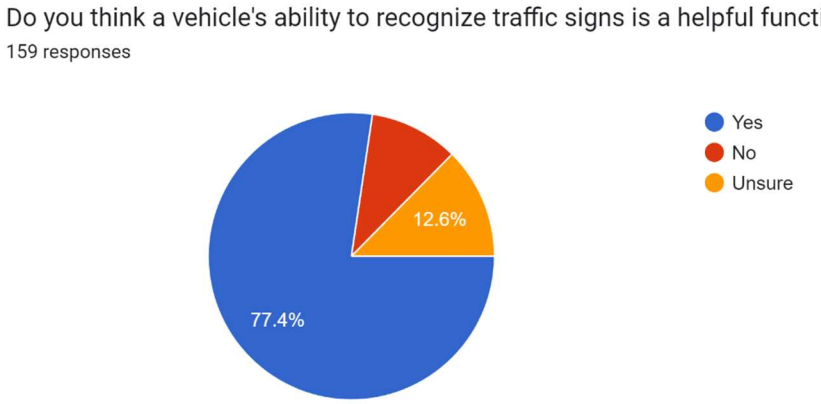

Figure A-22: Question 2 Results

Question #3

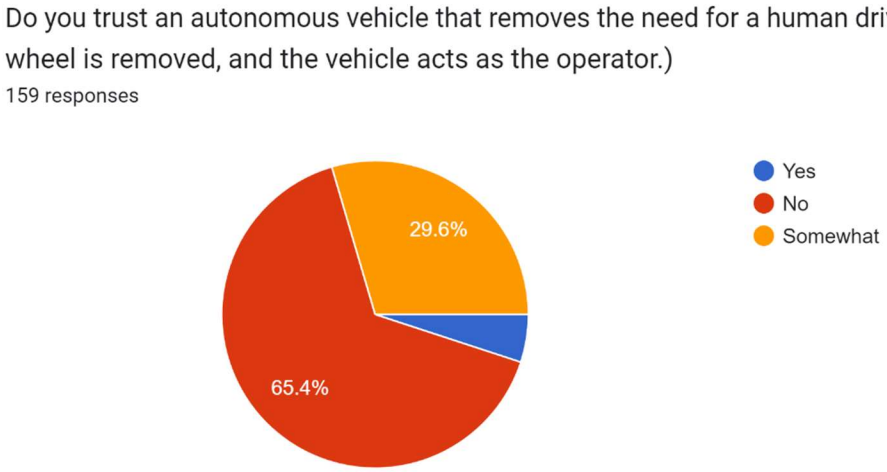

Figure A-23: Question 3 Results

Question #4

For the fourth question, we asked, "How can we enhance your comfort level with the adoption of fully autonomous traffic sign recognition functionality?" We provided 5 options:

- Reviewing the product specifications. (Understanding how the system works.)
- Reviewing the test data.
- Understanding all risks associated with the TSR functionality.
- Product training through videos, onsite dealership support, etc.

- Other

We received 38 other responses, which we reviewed and categorized. The most prevalent of these was strong cyber security at 5%, followed by 3.7% of respondents stating they will never trust autonomous vehicles. The following chart highlights the top answer categories received.

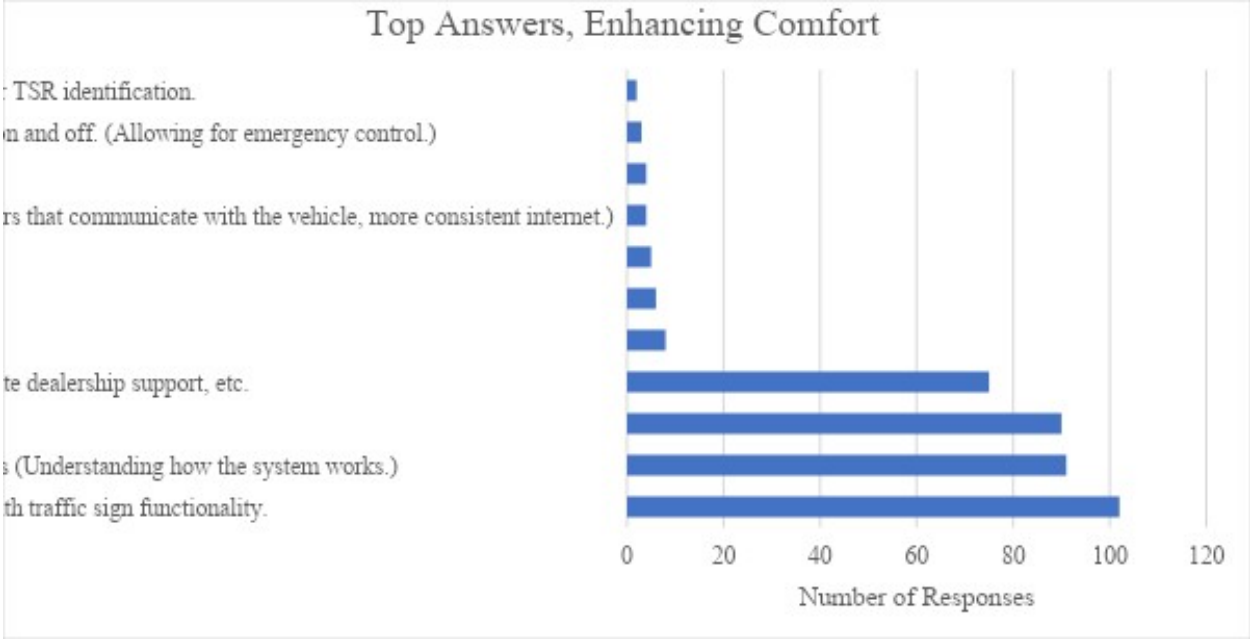

Figure A-24: Question 4 Results

Question #5

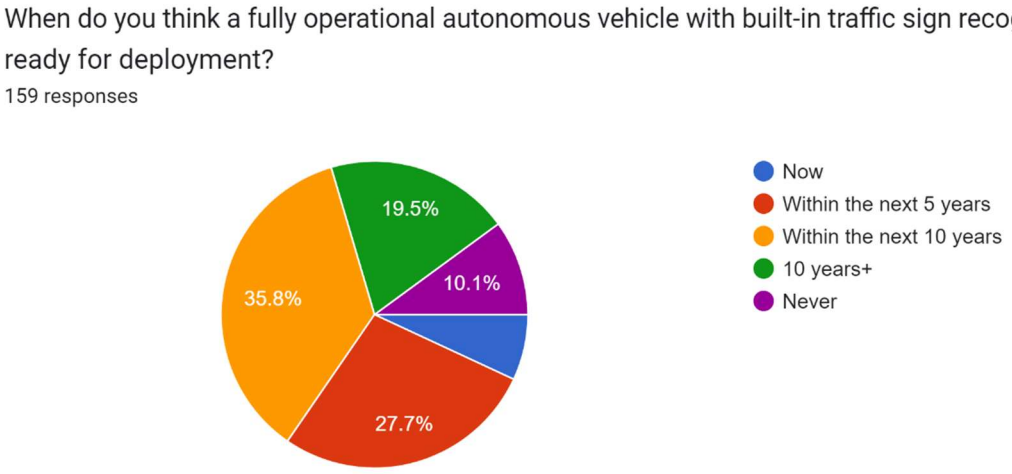

Figure A-25: Question 5 Results

Question #6

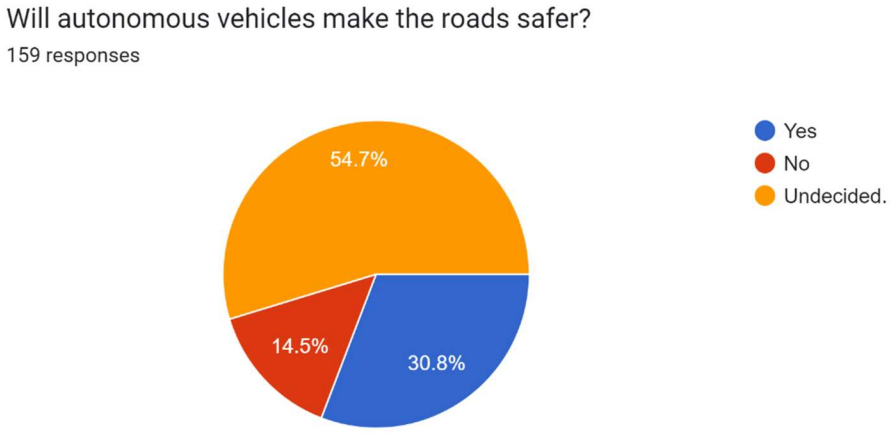

Figure A-26: Question 6 Results

Question #7

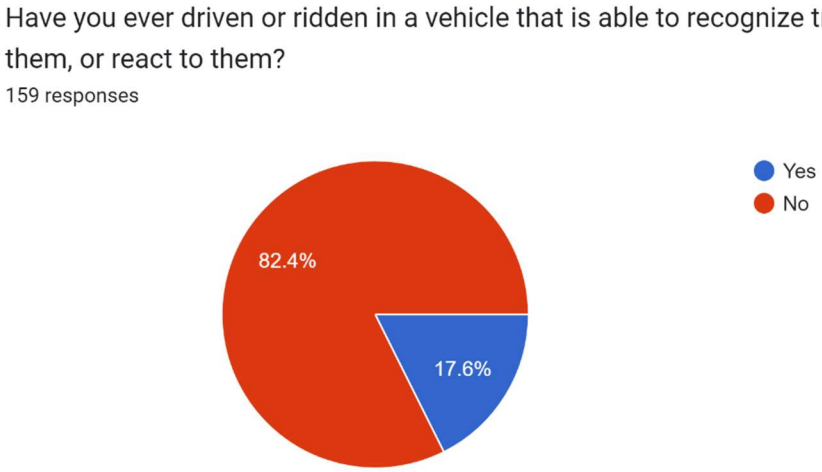

Figure A-27: Question 7 Results

Question #8 (Triggered by those who answered “No” or “Somewhat” to question 3.)

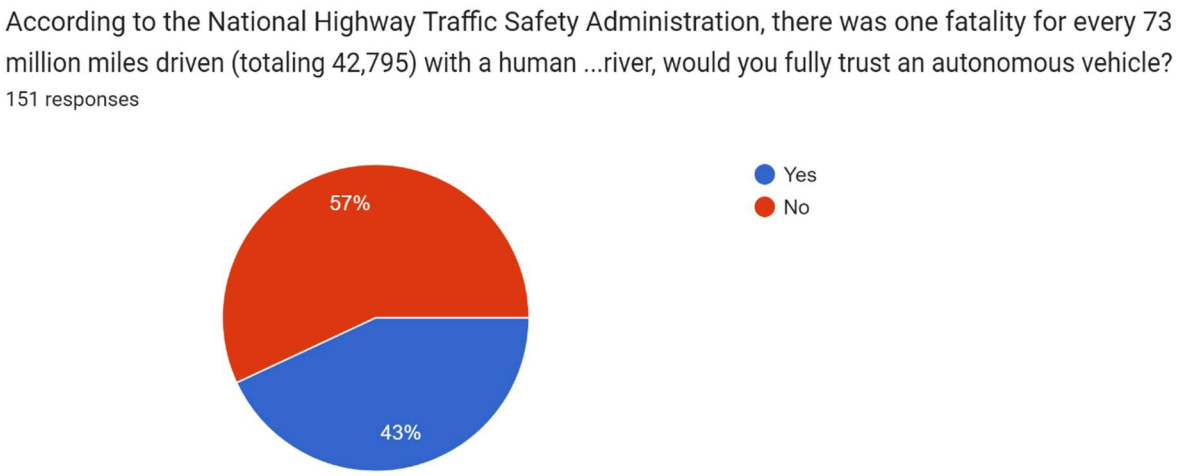

Figure A-28: Question 8 Results

Question #9 (Triggered by those who answered “Yes” to question 7.)

For survey respondents who had ridden in a vehicle that used TSR, we posed the following open-ended question, “What is the make of the vehicle that you have used traffic sign recognition? (i.e., Ford, GM, Tesla, Honda, Toyota, Waymo, Cruise).” We summarized the answers in the below chart, with Ford being the most used, followed by Tesla and GM.

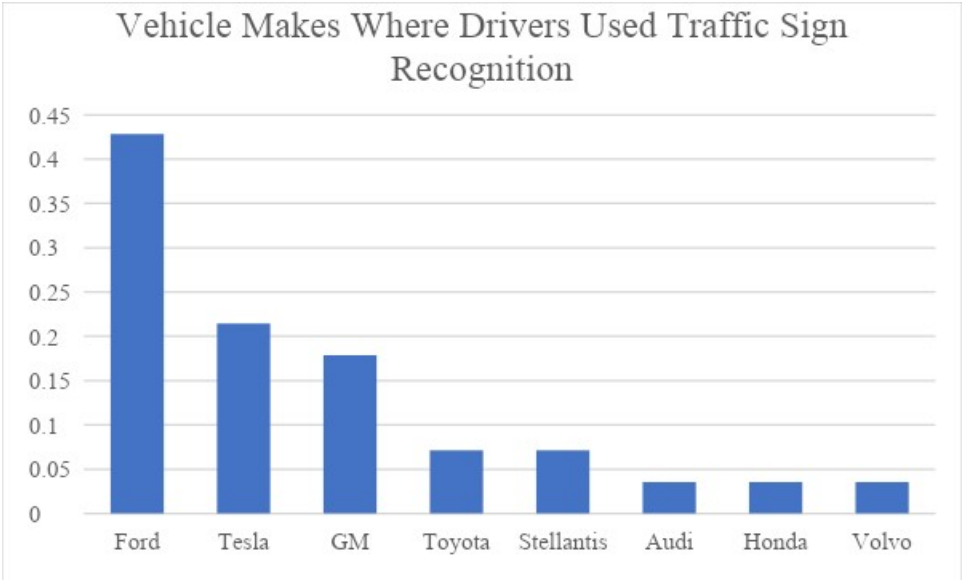

Figure A-29: Question 9 Results

Question #10 (Triggered by those who answered “Yes” to question 7.)

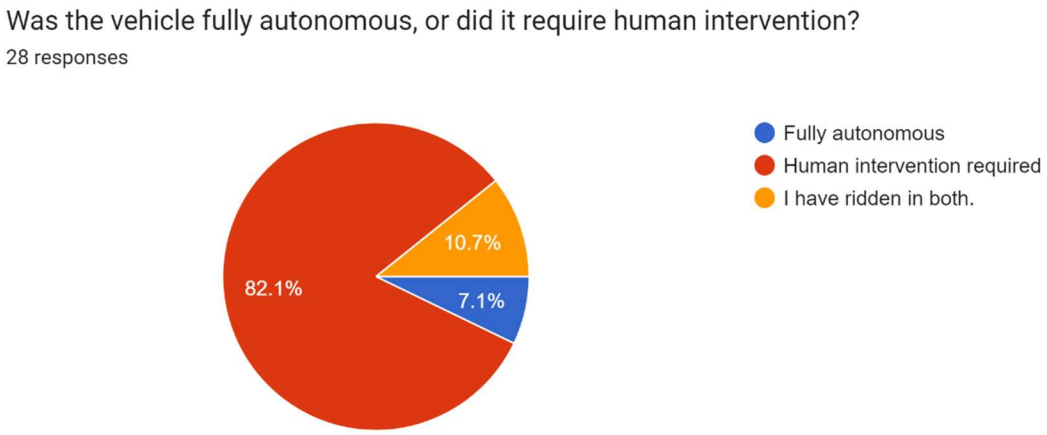

Figure A-30: Question 10 Results