Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
©2025 ESREL SRA-E 2025 Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P8469-cd

Navigating Security and Safety Challenges in Autonomous Vehicle Systems: A Risk-Based Assessment Framework

Aqsa Rahim

Department of Technology and Safety, UiT The Arctic University of Norway. E-mail: aqsa.rahim@uit.no

Babar Ali

National University of Science and Technology. E-mail: babar.ali792@hotmail.com

Javad Barabady

Department of Technology and Safety, UiT The Arctic University of Norway. E-mail: javad.barabady@uit.no

Abstract: In the past few years, the advancement and adoption of autonomous vehicles AVs have rapidly increased. As a result of which safety and security of Avs has become a big challenge. These vehicles are dependent on systems including various sensors, AI systems and connectivity systems These systems are vulnerable to security threats and safety risks. The autonomous driving system faces various security challenges including cyberattacks on the AV software, communication systems or on the cloud-based platforms. These security threats could affect control systems of the vehicle which could lead to catastrophic failures. In addition to this, system and sensor errors could also result in failure of the AI decision system, which pose risks not only to passengers but also to other road users. To the ensure the safety and reliability of AVs a comprehensive risk assessment framework is needed that will not only evaluate the impact of such incidents but also incorporate advance technologies for timely detection and prevention of such incidents. This paper will focus on developing a risk-based framework to assess and mitigate the challenges associated with AV technologies, emphasizing the safety and security dimensions. We will examine key safety standards and explore techniques such as real-time risk monitoring, machine learning-based threat detection, and resilient system design. Through a focus on risk management practices, we aim to establish guidelines for the secure and safe integration of autonomous vehicles, making the way for widespread adoption and public trust in this transformative technology.

Keywords: Risk-Assessment, Security Challenges, Autonomous Vehicle Systems.

1. Introduction

The self-driving cars mark a new advancement with the enhanced possibility of autonomous transport, lower traffic-related mishaps, and improved urban designs [1]. The adoption of self-driving vehicles into the urban environment, however, has several safety and privacy issues. AVs are different from traditional cars, because they utilize various algorithms, sensors, and cloud communications. The dependence creates an intricate environment for cyber security threats and complex safety dependencies [2]. With the more advancement in AV, solving security concerns and safety risks represents a challenge that must be addressed timely. The modern paradox AV ecosystems face is that it offers technologies for autonomy, yet those technologies are often subject to different risk factors [3]. Some of those factors include but are not limited to adversarial risks that threaten vehicle control, software exploitation, and remotely commandeering vehicle-to-everything (V2X) channels [4]. Likewise, the real world poses a completely different set of problems and risks stemming from unpredictable scenarios, sensor malfunctions due to bad weather, and algorithm biasness. Many edge-case scenarios are not even taken into consideration during training phases [5]. All these problems are further worsened by the lack of integrated legislation and no specific rules for AV risk assessments.

To bridge this gap, this paper proposes a risk-based assessment framework (as shown in Figure 1) designed to systematically evaluate and mitigate security and safety challenges in AV systems. The framework integrates three pillars: (1) threat identification through attack surface mapping and failure mode analysis, (2) risk quantification using probabilistic models to assess likelihood and impact, and (3) risk mitigation strategies via real-time monitoring and adaptive machine learning (ML) algorithms. By harmonizing safety standards (e.g., ISO 26262) with cybersecurity protocols (e.g., ISO/SAE 21434) [6], our approach enables holistic risk management tailored to the dynamic AV environment.

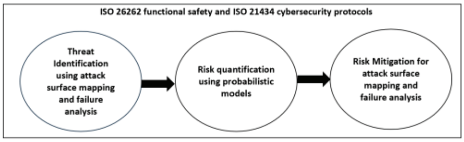
ISO 26262 functional safety and ISO 21434 cybersecurity protocols
Threat Identification using attack surface mapping and failure analysis
Risk quantification using probabilistic models
Risk Mitigation for attack surface mapping and failure analysis

Figure 1 Risk Assessment Framework

To contextualize risks, Table 1 provides a taxonomy of high-impact threats derived from empirical studies and industry incident reports.

Table 1 Risk matrix highlighting key threats to AV systems, ranked by likelihood and impact

Threat Category	Example Risks	Impact Severity
Cyber Security	GPS spoofing, CAN bus intrusion	Critical
Sensor Failure	LiDAR misclassification in fog	High
AI Decision erros	Edge-case scenario misjudgment	Critical
Cloud Vulnerabilities	Data exfiltration, DDoS attacks	High

By advancing a unified framework that prioritizes proactive risk mitigation, this work aims to catalyze the safe and secure deployment of AVs. The rest of the paper is organized as follows: Section 2 discusses threat identification and risk quantification, Section 3 discusses risk mitigation strategies, Section 4 discusses Policy and Regulatory Implications related to the framework and Section 5 discusses conclusion and future work.

2. Risk Assessment Framework: A Multi-Layered Approach

The framework is designed to address the complex relationship between safety and security issues faced in autonomous vehicle (AV). The framework combines threat modelling, probabilistic risk quantification, and mitigation, aligned with ISO 21434 (cybersecurity) and ISO 26262 (functional safety).

2.1. Threat Identification and Attack Surface Decomposition

Autonomous systems inherit risks from their interconnected hardware, software, and communication layers. We categorize threats into three domains as classified in Table 2:

- Cybersecurity: Exploitable vulnerabilities in AI models, V2X networks, or over-the-air (OTA) updates [7].
- Safety-Critical Failures: Sensor malfunctions (e.g., camera/LiDAR occlusion), algorithmic biases, or actuator errors [8].
- Operational Edge Cases: Unanticipated scenarios (e.g., extreme weather, ambiguous traffic scenarios) [9][10].

Table 2 Three different domains of Threats

	Case	Consequences
Attack Surface Mapping	Identify vulnerabilities in sensors (LiDAR, radar, Adversarial attacks on camera inputs) and perception algorithms (e.g., object detection, classification).	-False object identification (phantom objects) -Failure to detect critical obstacles -Misleading navigation decisions -Increased accident risks
	Map vulnerabilities in vehicle-	-Traffic disruption or congestion

(Cybersecurity)	vehicle (V2V) and vehicle-to-infrastructure (V2I) communication protocols, Man-in-the-middle (MITM) attacks on V2X messages.	- False emergency braking or acceleration -Incorrect Road hazard warnings -Compromised autonomous decision-making
	Analyze vulnerabilities in the operating system, middleware, and application layers.	-Unauthorized access to vehicle systems - Remote control of critical vehicle functions -System crashes and denial-of-service (DoS) - Data breaches and privacy violations
	Identify risks in electronic control units (ECUs), actuators, and power systems.	- Loss of vehicle control (braking/throttle manipulation) -Malfunctioning of safety-critical functions (e.g., ABS, steering) -Unexpected shutdowns or erratic behavior - Potential life-threatening situations
Failure Mode Analysis	Analyze failure modes under adverse conditions (e.g., fog, rain, snow).	- Failure to detect pedestrians, vehicles, or road obstacles - Incorrect object classification (e.g., mistaking snowbanks for solid objects) - Increased risk of accidents due to delayed or incorrect responses
	Identify biases in decision-making algorithms (e.g., favoring certain objects or scenarios).	-Unequal detection accuracy (e.g., prioritizing larger vehicles over pedestrians) - Unsafe driving decisions in underrepresented scenarios - Ethical concerns in risk prioritization (e.g., biases in collision avoidance)
Edge Cases	Identify scenarios not covered during training (e.g., rare road conditions or unexpected obstacles	- Increased accident risk due to failure to recognize or react to uncommon obstacles (e.g., fallen trees, animals, or sinkholes) - Misinterpretation of rare traffic situations (e.g., police hand signals, temporary road signs) leading to unsafe driving decisions - Poor generalization of perception models causing delays in obstacle avoidance - Unintended vehicle behavior (e.g., unnecessary stops or failure to yield) in edge cases not seen in training data - Ethical and legal concerns if the AI system cannot handle rare but critical scenarios (e.g., emergency vehicle interactions)

2.2. Risk Quantification and Prioritization

Risks are evaluated using a probability-impact matrix, combining likelihood estimates (derived from historical incident data and simulations) and severity scores (based on harm to humans, property, or trust). Machine learning models further refine these estimates by analysing real-world driving datasets. Table 3 gives a risk matrix for the threats identified in the previous section, for each threat case, its risk factor, Likelihood (L) of the risk, Impact of the risk (I) and its probability is given.

Table 3 Risk matrix for Threat cases

Cases	Risk Factor	L	I	Probability
Attack Surface Mapping	Adversarial attacks on camera inputs	H	C	Immediate Action
	Man-in-the-middle (MITM) attacks on V2X messages	M	C	Immediate Action
	Spoofing of V2X messages (e.g., fake traffic signals)	M	H	High Priority
	Denial-of-service (DoS) attacks on V2X communication channels	L	M	Medium Priority
	Physical tampering with ECUs	L	H	High priority
	Exploitation of unpatched software vulnerabilities in the AV's control system	H	C	Immediate Action
	Malware injection into the AV's operating system	M	H	High priority
Failure Mode Analysis	Power supply failures leading to system shutdown	L	M	Medium Priority
	Camera failure in low-light condition	M	H	High Priority
	LiDAR failure due to dirt or obstruction	M	H	High Priority
	Radar failure due to interference	L	M	Medium Priority
	Bugs in decision-making algorithms causing erratic behavior	L	H	High priority
	LiDAR performance degradation in heavy rain	M	H	High priority
	Biases in object classification (e.g., misclassifying vehicles)	L	M	Medium priority
	Radar interference from environmental noise	L	M	Medium priority
	Failure to detect pedestrians in low-light conditions	M	H	High Priority
Edge Cases	Unhandled scenarios (e.g., fallen tree on highway)	L	C	Immediate Action
	Rare weather conditions (e.g., black ice)	**L**	H	High Priority

The likelihood of each risk case given in Table 3 are based on evidence such as incidents from the real-world and evidence from theoretical as well as experimental data. Adversarial attacks on the camera input as well as exploiting the software vulnerabilities that have not been patched yet in the AV control system have high likelihoods from Cybersecurity research and cases that were reported in the past. These factors indicate the critical risks in autonomous systems. However, other factors such as Denial-of-service (DoS) attacks on V2X communication channels or V2X noise jamming radar from the background environmental noise are determined as low likelihoods because those are supported by evidence of lower quality in these contexts. The physical tampering with ECUs is possible but subjectively high concern and low complexity leads to a lower likelihood. Risks such as camera not functioning in low light scenarios and LiDAR sensor failure due to dirt or obstruction is moderate because there is significant supporting literature for the problems in automotive safety engineering research, but advances in technology have lowered the probability of their occurrence. Some biases within the algorithms while making decisions and failures within the system to detect pedestrians during low light scenarios are known risks which are adequately managed and mitigated through comprehensive testing and refinement which results in the algorithm's low probabilities of failure considering these risks. In extreme and edge case scenarios like automobile break down alongside the highways, dense fog with virtually no visibility or rare weather conditions including black ice, the possibility of these events occurring is still low mainly due to less occurrence of these events however, the risk associated with these events occurring is extremely high. The basis of evidence for these extreme edge scenarios is gathered from accident reports and meteorological phenomena which are outlined and underscore the gaps that still exist within autonomous vehicle systems. These judgments depict a multitude of risks along with clearly stating where there is lack of data which requires more validation and research along with different methods to neutralize these risks.

3. Mitigation Strategies: A Defence-in-Depth Architecture

A defense-in-depth architecture employs multiple layers of security controls to mitigate risks across the autonomous vehicle (AV) system. Below are tailored mitigation strategies for the risks identified in the Threat Identification phase.

3.1. Mitigation Strategies for Attack Surface Mapping in Autonomous Vehicles

Attack surface mapping involves identifying and analyzing potential vulnerabilities in a system that could be exploited by attackers. In the context of autonomous vehicles (AVs), these vulnerabilities exist across various components, including vehicle-to-everything (V2X) communication, software architecture, and electronic control units (ECUs). To mitigate risks, a multi-layered security approach is essential. The following strategies help minimize the attack surface and enhance the resilience of AV systems.

3.1.1. Securing V2X Communication

V2X communication (Vehicle-to-Vehicle and Vehicle-to-Infrastructure) allows instantaneous data transfer between AVs and surrounding infrastructure. Unfortunately, its security features are prone to interception, spoofing, and MITM attacks, which can interfere with crucial vehicle functions or even in worse case can manipulate them.

1. Encryption and Authentication: Implement end-to-end encryption (e.g., TLS, AES) and digital signatures to protect message integrity and prevent unauthorized interception.
2. Message Validation: Use cryptographic authentication mechanisms such as Message Authentication Codes (MACs) to verify the legitimacy of received data.
3. Intrusion Detection Systems (IDS): Deploy network-based IDS to detect anomalies in V2X communications, such as unauthorized data injection or unexpected message patterns.

3.1.2. Hardening Software and System Architecture

The operating system (OS), middleware, and application layers of AV software stack create a critical vulnerability zone for attackers. Attackers exploiting software weaknesses can take control of vehicle functions, block specific operations, or inject faulty code.

1. Regular Software Updates: Provide an over-the-air (OTA) update feature that ensures swift implementation of identified patches and security bugs to proactively minimize weaknesses.
2. Access Control and Sandboxing: Level for malware cross-contamination by limiting user permissions, controlling routine critical vehicle functions, or sandboxing malicious processes underneath user software layers.
3. Runtime Security Monitoring: Implement measures that enable real-time spotting of unusual actions within certain system processes and applications focusing on suspicious changes in the monitored software environment.

3.1.3. Protecting Electronic Control Units (ECUs) and Actuators

ECUs control fundamental vehicle operations such as acceleration, braking, and steering. Attacks on these components can be executed by physically manipulating ECUs as well as executing intrusion commands into the system.

1. Tamper-Resistant Hardware: Protect physical access of users to ECUs by means of secure hardware boxes and anti-tampering tools.
2. Anomaly Detection in Control Systems: Use intrusion detection systems that monitor and analyze ECU command patterns that is different from the known format.
3. Restricted Access to Diagnostic Ports: Access to any OBD-II and other diagnostic ports is secured with authentication checks to avoid unauthorized reprogramming or data manipulation.

3.2. Mitigation Strategies for Failure Mode Analysis

Failure mode analysis helps reveal and highlight potential weaknesses present in autonomous vehicles (AV) systems so that they can operate under a wide range of stress conditions in a dependable manner. It includes everything from sensor failures in harsh environments to algorithmic discrimination in coping with different scenarios. The following strategies outline effective mitigation approaches.

3.2.1. Sensor Failures

Cameras, radar, and LiDAR systems are used in autonomous vehicles to observe their environment. These technologies can face issues like poor visibility from fog, rain, or snow that can result in incorrect interpretations of hazards and objects present on the road.

1. Multi-Sensor Fusion: LiDAR, radar, and camera data can be integrated to increase perception accuracy during bad weather.
2. Weather-Adaptive Sensor Calibration: Calibrate environmental parameters in real time through noise adjustments, like rain LiDAR adjusting for heavy precipitation.
3. Self-Cleaning Mechanisms: Use automated wipers, air blowers, or hydrophobic coatings to prevent sensor obstructions caused by water, snow, or dirt.
4. AI-Based Signal Enhancement: Utilize machine learning algorithms to filter noise and enhance signal clarity in low-visibility conditions.

3.2.2. Addressing Algorithmic Biases

Decision-making algorithms in autonomous cars can develop biases that may result in dangerous activities such as identifying pedestrians as something else in dark environments or picking and choosing certain objects. These biases can contribute towards AV risks, and stem from inadequate training data and model overfitting.

1. Diverse and Representative Training Data: Train AI models on varied datasets that include different lighting conditions, weather scenarios, and pedestrian demographics to improve generalization.
2. Bias Detection and Correction: Employ fairness-checking to cut down biases in models used for detection and classification of objects.
3. Real-Time Performance Monitoring: Monitor decisions made from the algorithm in the field. Change the threshold of the decisions made from AI based on parameterised measured data.
4. Human-in-the-Loop Systems: Implement human controllers for algorithms which have the potential of causing safety issues.
5. Adaptive AI Learning: Employ reinforcement learning techniques where the AV models modify themselves in real time based on changing situations within the real world.

3.3. Mitigation Strategies for Edge Cases

Edge cases consist of obstacles an AV might face during its operation but hasn't been conditioned or trained for. Some edge cases would entail a tree blocking a freeway, or an erratic pedestrian activity. Such edge conditions can challenge the AV system's competence in decision making. To enhance the safety and reliability of AVs, these measures can be preferred:

3.3.1. Expanding and Enhancing Training Data

1. Diverse and Synthetic Datasets: Download ample real-world datasets and create training scenarios like AVs existing in the real world at set conditions to improve their performance for edge case handling.
2. Adversarial Training: Introduce challenging conditions (e.g., obstacles in unexpected locations) during model training to improve robustness.
3. Continual Learning: Implement machine learning models that can adapt, and update based on new data collected from real-world driving experiences.

3.3.2. Advanced Sensor Fusion and Perception

1. Multi-Modal Sensor Integration: Combine data from LiDAR, radar, and cameras to detect and classify unknown obstacles with higher accuracy.
2. AI-Based Anomaly Detection: Develop AI models capable of identifying unusual objects (e.g., a fallen tree) by comparing current sensor inputs with expected patterns.
3. Environmental Awareness Sensors: Integrate additional sensors (e.g., infrared, ultrasonic) to enhance AV perception in low-visibility conditions.

3.3.3. Continuous Real-World Testing and Validation

1. Edge Case Simulation Environments: Use high-fidelity simulation platforms to test AV behavior under rare and unpredictable conditions.
2. On-Road Data Collection and Feedback Loops: Deploy AVs in real-world conditions and use their experiences to refine perception models.
3. Crowdsourced Data Sharing: Utilize a shared AV network where vehicles contribute rare case scenarios to improve collective learning.

4. Safety Standards and Regulations for Autonomous Vehicles

The actual deployment of AVs is not only a matter of technology, but also one of creating appropriate AV regulation that will guarantee safety and security as well as trust from the public. In this part, we analyze how the proposed framework is integrated with the current standards, suggest modifications to existing ones, and provide steps that would enable the industry to accept these changes.

4.1. Alignment with Existing Standards

Developing an effective risk-based assessment system will give additional value to technologies and processes that are relevant to the current automotive standards without compromising them and at the same time focus on the AV systems singularity. Our approach allows for compliance with ISO 21434 and its lifecycle cybersecurity principles, consisting of threat and risk-based analysis and assessment, as well as mitigation of AV system threats. The framework's adaptive security policies and real-time monitoring mechanisms provide actionable measures to address evolving cybersecurity threats, as mandated by ISO 21434[11]. The framework integrates functional safety considerations by mapping safety-critical failures (e.g., sensor malfunctions, AI decision errors) to Automotive Safety Integrity Levels (ASIL). Our defence in-depth architecture ensures fail-operational capabilities, aligning with ISO 26262's requirements for redundancy and fault tolerance [12]. The use of digital twin simulations for scenario testing supports the standard's emphasis on rigorous validation and verification. Our framework's emphasis on dynamic risk assessment and continuous monitoring aligns with UNECE R155's requirements for cybersecurity management systems (CSMS). The inclusion of adversarial testing and forensic readiness mechanisms ensures compliance with the regulation's focus on proactive threat detection and incident response.

4.2. Recommendations for Regulatory Updates

While existing standards are useful to build upon, the development of AV technologies is unprecedented and require new approaches to be created to govern them. The makers of AV along with suppliers and regulators working together can strengthen the system of AV in sharing intelligence regarding threats. The certification procedures mostly make use of static risk evaluations that assume AV environments do not change. It is suggested to include the AV certification that includes continuous risk assessment and real-time threat response mitigation, so that the cars reevaluate the threat landscape while in motion. Another risk that remains largely unaddressed is adversarial attacks on AV perception systems such as LiDAR spoofing, adversarial patches etc. Authorities should demand some form of adversarial testing for AV manufacturers, so the vehicles can be validated safe against known and anticipated threats.

5. Conclusion and Future Work

The rapid development of autonomous vehicles (AVs) calls for efficient policies that highlights the exceptional challenges to safety and security. In this paper, a framework was presented for risk-based assessment with the components of threat modelling, probabilistic risk quantification and mitigation strategies for the secure and safe operation of AV systems. The swift acquiring rate of autonomous vehicles (AVs) calls for efficient policies that highlight the exceptional challenges regarding safety and security.

In the future, several directions for studying this topic can be pursued. First, AV systems may be extended with quantum resistant cryptography for further security against the emerging advanced threats. Second, federated learning poses a new dimension in collaborative threat detection across AV fleets while ensuring data privacy. Finally, combining the AV systems with digital twin technologies for real time risk monitoring and predictive maintenance would make AV ecosystems more resilient.

References

1. Levinson, J., J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, et al. “Towards Fully Autonomous Driving: Systems and Algorithms.” In 2011 IEEE Intelligent Vehicles Symposium (IV), 163–68. IEEE, June 2011.
2. Liu, L., S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, and W. Shi. “Computing Systems for Autonomous Driving: State of the Art and Challenges.” IEEE Internet of Things Journal 8, no. 8 (2020): 6469–86.
3. Ren, K., Q. Wang, C. Wang, Z. Qin, and X. Lin. “The Security of Autonomous Driving: Threats, Defenses, and Future Directions.” Proceedings of the IEEE 108, no. 2 (2019): 357–72.
4. Hobert, L., A. Festag, I. Llatser, L. Altomare, F. Visintainer, and A. Kovacs. 2015. “Enhancements of V2X Communication in Support of Cooperative Autonomous Driving.” IEEE Communications Magazine 53 (12): 64–70.
5. Gao, X., and X. Bian. “Autonomous Driving of Vehicles Based on Artificial Intelligence.” Journal of Intelligent & Fuzzy Systems 41, no. 4 (2021): 4955–64.
6. Schmittner, C., & Macher, G. (2019). Automotive cybersecurity standards-relation and overview. In *Computer Safety, Reliability, and Security: SAFECOMP 2019 Workshops, ASSURE, DECSoS, SASSUR, STRIVE, and WAISE, Turku, Finland, September 10, 2019, Proceedings 38* (pp. 153-165). Springer International Publishing.
7. Garakani, H. G., B. Moshiri, and S. Safavi-Naeini. “Cybersecurity Challenges in Autonomous Vehicles: Their Impact on RF Sensors and Wireless Technologies.” In 2018 18th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM), 1–3. IEEE, August 2018.
8. Safavi, S., M. A. Safavi, H. Hamid, and S. Fallah. “Multi-Sensor Fault Detection, Identification, Isolation and Health Forecasting for Autonomous Vehicles.” Sensors 21, no. 7 (2021): 2547.
9. Jing, S., Y. Zhao, X. Zhao, F. Hui, and A. J. Khattak. “An Efficient High-Risk Lane-Changing Scenario Edge Cases Generation Method for Autonomous Vehicle Safety Testing.” IEEE Transactions on Intelligent Vehicles, 2024.
10. Bouchelaghem, S., A. Bouabdallah, and M. Omar. “Autonomous Vehicle Security: Literature Review of Real Attack Experiments.” In Risks and Security of Internet and Systems: 15th International Conference, CRiSIS 2020, Paris, France, November 4–6, 2020, Revised Selected Papers, vol. 15, 255–72. Springer International Publishing, 2021.
11. Costantino, G., M. De Vincenzi, and I. Matteucci. “In-Depth Exploration of ISO/SAE 21434 and Its Correlations with Existing Standards.” IEEE Communications Standards Magazine 6, no. 1 (2022): 84–92.
12. Taylor, W., G. Krithivasan, and J. J. Nelson. “System Safety and ISO 26262 Compliance for Automotive Lithium-Ion Batteries.” In 2012 IEEE Symposium on Product Compliance Engineering Proceedings, 1–6. IEEE, November 2012.