# Cross-Correlations and Challenges of Combined Artificial Intelligence, Functional Safety and Cybersecurity Development for Automotive, Aerospace, and Government Applications


Radu Iacob
Sr Manager R&D
Synopsys, Inc
Hillsboro, OR 97124
iacob@synopsys.com

Shivakumar Chonnad
Sr Architect
Synopsys, Inc
Sunnyvale, CA 94085
shiv@synopsys.com

Chris Clark
Sr Manager R&D
Synopsys, Inc
San Antonio , TX, 78148
clarkc@synopsys.com



***Abstract* – Artificial intelligence enabled transportation solutions are poised to take center stage across the automotive and aerospace industries creating a shift in customer expectations and industry practices. From sensor fusion and power optimization to adaptive cruising and robotaxis, today's automotive product development cross-correlates multiple disciplines, such as data science, machine learning, and neurocomputing, along with reliability, functional safety, and cybersecurity. Combined development delivers AI solutions for automotive systems that are safe, secure, and reliable, capable of performing in safety-critical environments while resisting cybersecurity threats. Similar considerations apply to the development of aerospace solutions. Artificial intelligence is becoming the driving force shaping today's technologies, global economies, the social and cultural environments around the world, hence the need to consider the impact, cross-correlations, and challenges between and within all related disciplines.**



***Keywords — AI; functional safety; cybersecurity; reliability; development lifecycle; mission; cross-correlation; trade-off; automotive; aerospace***


## 1. INTRODUCTION

Automotive product implementation encompasses automotive reliability, functional safety, cybersecurity and artificial intelligence aspects developed systematically within a quality management system framework. Like aerospace industry requirements, automotive reliability and functional safety are of primary importance for mitigating risks stemming from the operating mode and operational environment. For example, random hardware failures that may occur during mission mode due to random causes, such as a radiation burst, are of primary focus in automotive and aerospace applications. Cybersecurity risks, on the other hand, may arise and may be mitigated in relation to threats based on measurable risk. It presumes a potential target asset, a disruptive intent, and an attack path that must be protected to avoid access to the assets. Artificial intelligence can play multiple roles in the context of automotive and aerospace applications. It may implement a mission function, such as driving autonomously on a highway or flying unmanned to a destination, provide a safety mechanism such as obstacle identification and avoidance, or implement cybersecurity measures such as signature identification and dynamic response based on known or unknown cybersecurity attack patterns.

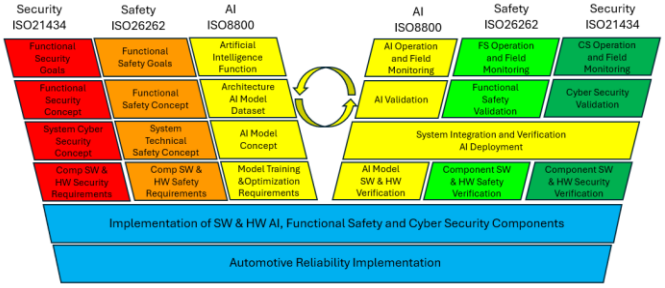


*Figure 1: V-Model of combined AI, Functional Safety and Cybersecurity development lifecycle for automotive development process*

Regardless of the scope, artificial intelligence development must be coordinated with functional safety and cybersecurity to identify common requirements, objectives and resources, and potential weaknesses that one aspect of the development may bring to the others. Figure 1 illustrates the automotive product development lifecycle, including the automotive reliability, functional safety, cybersecurity and artificial intelligence aspects. This paper is describing cross-correlations between relevant disciplines in automotive development, as well as potential challenges and mitigation trade-offs when combined aspects are part of the development.

## 2. AI SYSTEMS

An AI system is typically defined as a system comprised of a pre-processing stage, an AI model stage and a post-processing stage, receiving input data from various sources such as an array of sensors, and outputting data to consumers, for executing an AI function.

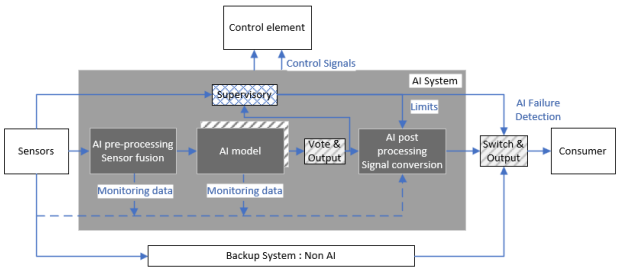


*Figure 2: Concept diagram of a fail-safe AI system, adapted from [8]*

Examples of AI systems include perception AI (classifiers, content interpreters, translation), generative AI (content creation), agentic AI (coding assistant, customer care, patient care), physical AI (self-driving cars, UAVs, general robotics, space and planetary exploration robots).

AI systems include AI software components (AI models) and supporting conventional SW, and AI HW components (AI processors or accelerators) as well as supporting conventional HW such as host processors.

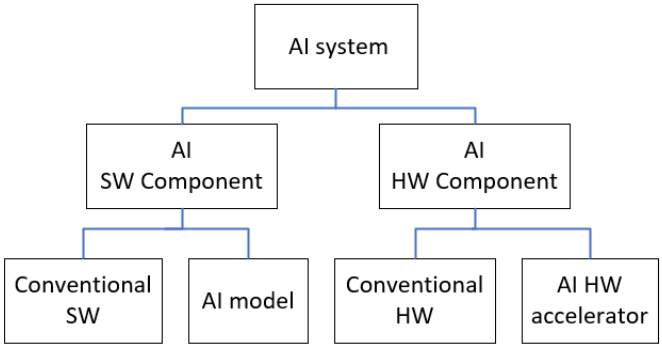


*Figure 3: Typical hierarchy of an AI system, adapted from [8]*

AI systems are developed using SW and HW AI technologies to implement the AI method itself, and to provide tools and procedures to create the trained AI model. The hardware infrastructure implementing AI systems are often referred to as hardware-in-the-closed-loop (HiL), whereas the outputs of the target AI HW components executing the target AI software components influence the HW inputs, and hardware-open-loop (HoL), whereas the outputs of the target AI HW components executing the target AI software components do not influence the HW inputs.

## 3. AI Systems Requirements

Governance initiatives are defining AI development requirements for safe, secure and trustworthy AI systems. For example, the "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" [1] provides guidance for the deployment and use of AI technologies in the US. In a similar manner European Union issued the "EU Artificial Intelligence Act" [2], intended to provide clear and up to date AI development and analyses requirements to EU constituents.

Outputs from these initiatives are reflected in development and release of normative standards such as NIST-AI-600-1[7] and ISO/PAS 8800 [8]. Normative standards such as those listed in Table 1 may influence and provide relevant guidance to industry on what and how to evaluate and collect data necessary to build and reinforce AI data models.

*Table 1: Relevant standards for governance compliant AI system development*

| Category | Relevant standards |
|---|---|
| Safety | IEC 61508[3], ISO 26262[4], ISO 21448[5], ARP4761[21] |
| Cybersecurity | ISO/IEC 27001[10], ISO/SAE 21434[6] |
| AI | ISO 42001[11], ISO 22989[9], ISO 23053[12], ISO 4804[13], ISO 5469[14], ISO 8800[8], NIST-AI-600-1[7] |
| Reliability | IATF 16949[16], AEC-Q100[17], DO-254[19] |

Risk identification and mitigation are accomplished by executing risk analyses such as Hazard Analysis and Risk Assessment (HARA) for functional safety, Threat Assessment and Remediation Analysis (TARA) for cybersecurity, AI Risk Assessment (AIRA) for implementation of the AI solution, and Reliability Analysis for reliability risk evaluation.

AI risks are stemming from development and operational considerations such as system design, ML/AI algorithms, the data utilized for model training, and the actual operating conditions and threats that may occur during operation. The most common causes are systematic issues stemming from incomplete requirements definition, unknown AI system limitations, unsuitable model selection and lack of model transparency, low data quality and relevance, data bias, model performance drift. Random causes such as unintended operation beyond technical capabilities and adversarial attacks shall also be considered. AI risk can be formulated as a combination of probabilistic risk and deterministic risk.

Functional safety risk associated with random harmful events is evaluated based on the probability of exposure and on considerations of severity and controllability, being quantified by Automotive Safety Integrity Levels [4].

Cybersecurity risk is evaluated based on the attack path analysis, attack feasibility rating and impact rating.

Reliability risk is based on the probability of failure in time of the AI hardware components comprised by the AI system infrastructure, under specified operating conditions.

Thus, a global risk function GR can be defined as a combination of AI risk (AIR), functional safety risk (FSR), cybersecurity risk (CSR) and reliability risk (RR):

$$GR = f(AIR, FSR, CSR, RR) \quad (1)$$

whereas f is a function of conditional probabilities that can be represented across the space of operating conditions for representative use cases.

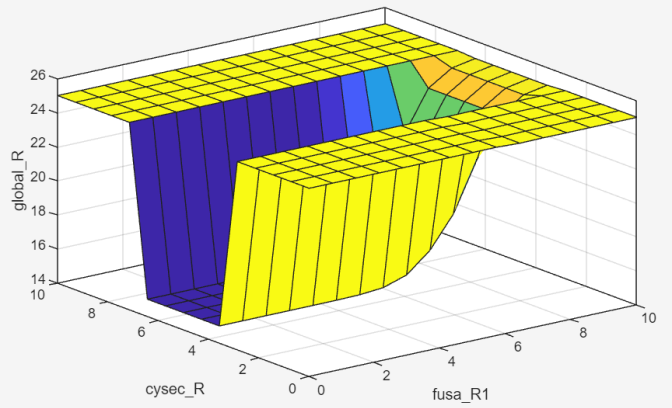


*Figure 4: Representation of a combined risk function*

Risk mitigation measures may lead to overlapping requirements, complementary requirements and/or contradictory requirements. Contradictions between functional safety and cybersecurity requirements may arise from:

- Risk assessment priorities (HARA vs TARA)
- Handling faults vs security breaches
- Transparent design vs obfuscation
- Static design vs dynamic updates

For example, a functional safety requirement may be that the system fails-safe in the event of a random hardware fault, typically moving to a safe state such as a power-down/up cycle, while the cybersecurity requirements are asking to avoid such states, as shutting down the system may be an opportunity for attackers to initiate a Denial-of-Service (DoS).

Such contradictions may be addressed by enhanced development methodologies like combined risk management, balancing availability, security and AI safety, aligning life-cycle processes, and by design optimization intended for risk reduction.

In a practical example, considering typical constraints of power, speed and precision for AI systems as shown in Table 2, a risk trade-off can be made by power management optimization, while preserving speed and precision.

*Table 2: Typical constraints for an AI System*

| Constraint | Safety | Security | AI |
|---|---|---|---|
| Power | M | L | H |
| Speed | H | M | H |
| Precision | H | M | H |

The global risk function after constraints trade-off and architectural optimization is illustrated in Figure 5.

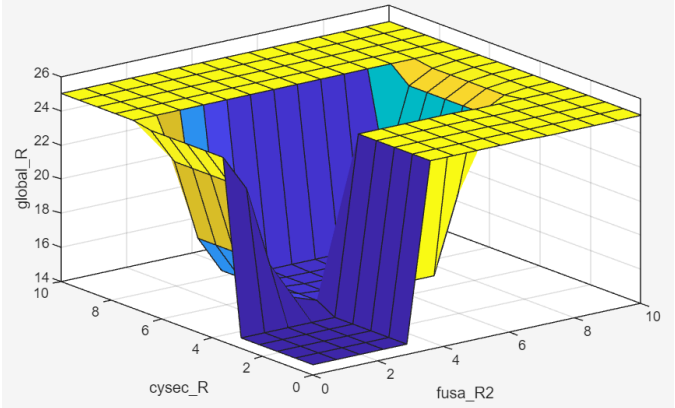


*Figure 5: Representation of the combined risk function after trade-off optimization*

## 4. DEVELOPING THE AI SOLUTION

AI systems are essential for executing complex missions and they are integral part of the implementation solutions as described in Table 3.

*Table 3:Representative AI applications*

| Application | Description |
|---|---|
| Autonomous solutions | Road vehicles, aircraft, space vehicles |
| UAV for disaster relief and recovery | Intervention in natural disasters aftermath: hurricanes, earthquakes, volcanic eruptions, tsunami, drought, wildfires |
| | Industrial or transportation accidents on ground, sea, air |
| | Humanitarian relief in affected areas |
| Planetary and space exploration | Ground and flying self-navigating vehicles for terrain exploration and mapping, geological and geophysical research etc. |
| | Exploration of space and other planets |
| Robotics | Industrial, military, medical and civil work robots |

The most employed AI functions for autonomous solutions are the following [13]:

- Image recognition: segmentation, labeling, object identification, feature and scene recognition
- Sensor fusion
- Simultaneous localization and mapping
- Predicting future behavior of relevant objects
- Route planning for safe and lawful driving
- Minimum risk maneuvering to execute driving plan
- Communication with other traffic partners
- Evaluation of nominal performance
- Safe mode transition and fail-safe degradation
- Payload delivery
- Adversarial training

The AI system development lifecycle is represented in Figure 6.

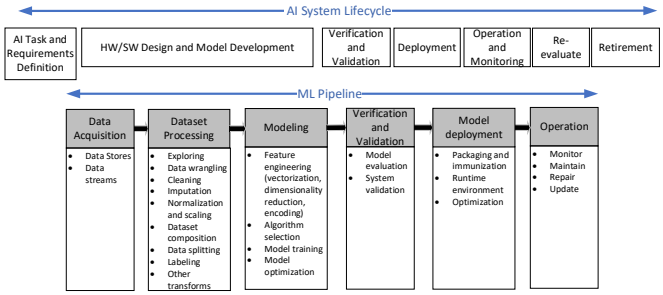


*Figure 6: AI system development lifecycle, adapted from [12]*

AI development starts by defining the application scope and requirements, identifying what will be modeled and for which purpose. The scope may include implementing a mission function, a safety or cybersecurity measure, improving reliability, optimizing an existing design, creating new functions, etc. AI development involves both software and hardware. The software development typically implements the AI algorithm and creates the AI model, which will execute the target tasks and may learn from new data.

Hardware resources may need to be developed or enhanced to support the execution algorithms. High-speed mission function solutions sometimes require a direct hardware implementation of the AI algorithms, including resources for neural network data structures and weight programmability, while involving software during the training phase. Both software and hardware resources employed by the AI solution are subject to functional safety and cybersecurity concerns. Hence, cross-correlations and challenges between these aspects need to be considered in the design.

Machine learning and AI are data-driven techniques. Therefore, data collection and preparation are of the highest importance for training, validation, and testing of the models. Thus, data science techniques are to be applied for the collection, analysis, and pre-processing of representative datasets. In most popular ML/AI applications, such as Large Language Models and Generative AI, the models are created and optimized based on contextual and relational information by processing exceptionally large volumes of data collected along significant periods of time. The training of such models, featuring hundreds of millions of parameters, takes time and resources, making them very expensive. Moreover, data may not be sufficiently relevant for the intended application.

In automotive and aerospace applications, in addition to ensuring relevant and unbiased data, the models can be enhanced by a process-driven approach which guarantees that laws of physics, as well as functional safety and cybersecurity requirements are embedded into the model.

In most of automotive and aerospace applications, processing speed is essential. For example, object detection in advanced driving assistance systems (ADAS) must be executed within a predefined interval to allow for the appropriate avoidance reaction, given the range of vehicle's speed and direction.

Therefore, various architectures were developed to best fit the functional requirements, including timing performance. For example, convolutional neural networks (CNN) are optimal for image processing, Bayesian regression may be best for classification, recursive neural networks (RNN) for time series and convolutional recursive neural networks (CRNN) for real-time scene detection. Architecture optimization is used to improve the overall performance by increasing precision and accuracy, while not overloading the computational effort so that to comply with the processing speed requirements. While there are well-known optimization measures, it is preferable to develop or adapt the optimization method such that to embed physical process laws and requirements specific to the application, for better control of results. These may involve functional safety and cybersecurity requirements.

Finally, the loss-function used for the convergence of the model toward the expected solution is typically designed based on a data-driven approach. At the same time, the loss function can be modified to embed physics information, or functional safety and cybersecurity aspects, leading to a faster and more efficient learning toward the desired enhanced model. Thus, the model can be trained with built-in mechanisms for error detection and cybersecurity features like intrusion detection, anomaly detection, and encryption.

Model cross-validation is performed to ensure that the model can handle safety-critical real-world scenarios, while being resistant to cyberthreats, such as injection of adversarial data. Penetration testing should be performed on the AI system to identify vulnerabilities in its ability to withstand unauthorized access to the model and datasets.

Finding the proper trade-off between the implementation of the artificial intelligence solution, and functional safety, cybersecurity and reliability measures, is a complex task that can be addressed through multidisciplinary risk analysis and regression techniques toward an optimum result.

## 5. Model Development, Deployment and MLOps

A typical automotive AI deployment model is made up of multiple layers, but in its basic form consists of data generation from sensors, in vehicle execution based on model inferencing, data transformation in a data center, and processed data delivery for AI consumption. The supervised ML process is represented in Figure 7.

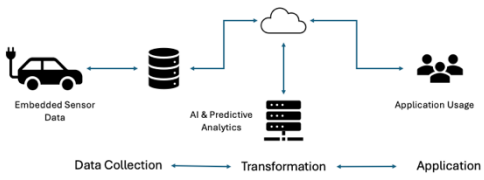


*Figure 7: A typical distributed automotive AI deployment model*

This complex environment has the potential to introduce a wide and varied threat surface to attackers. Therefore, special attention is paid to the model development, training and deployment on mobile AI systems, as well as to subsequent AI system monitoring, in the context of accomplishing the

cybersecurity requirements. This includes over-the-air (OTA) model updates resulted from continuous learning based on data collected from the AI systems operating in the field.

Model development takes into consideration specific aspects such as dataset quality and structure, development environment prepared to meet AI, safety and cybersecurity requirements, type of training along the model lifecycle (supervised, unsupervised, reinforcement learning), and the target endpoint constraints on the edge device where the model will be deployed for operation. Model development process flow and associated resources are represented in Fig. 8.

Model deployment on the endpoint takes into consideration the actual hardware resources of the target edge AI system, including computational power of the AI inference module, size of the RAM memory available for model parameters, parallel processing capabilities, target system speed and type of interfaces available for data input, output and within the inference module. Thus, the edge endpoint inference is typically much more restrictive than the model development environment therefore, model optimization is a necessary step for deployment.

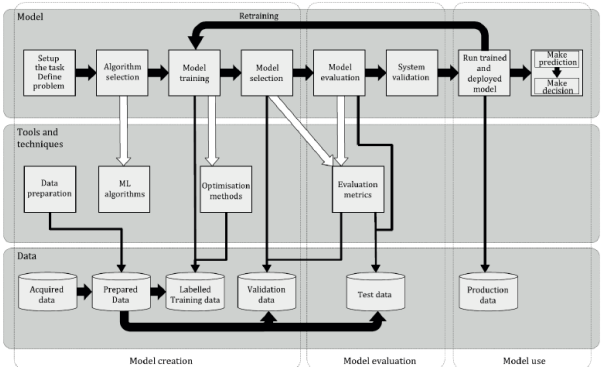


*Figure 8: A typical supervised ML model development flow, excerpted from [12]*

In a practical example, deploying a model for real-time applications on edge devices involves a recompilation of the model from a common development format such as ONNX, PyTorch or TensorFlow, to the target edge inference sub-system runtime environment. This recompilation may also include measures for increasing robustness to adversarial attacks and other additional cybersecurity features.

Deployment can be executed using containerization and docking techniques, which can further increase the immunity to external attacks. After deployment, the AI system continues to be monitored for performance and cybersecurity indicators defined during the development stage. When certain performance, safety or cybersecurity KPI thresholds are exceeded, updating or re-tuning of the edge AI system may be required. This is part of the continuous monitoring, maintenance and update activities as part of the MLOps stage in the AI system lifecycle.

## 6. AI INFERENCE

The model inference is executed on the AI edge system hardware to perform the AI tasks. The AI model inference sub-system represented between pre-processing and post-processing components in Figure 2, typically consists of an SoC that includes the host processor running a Real-Time-Operating-System (RTOS), the AI acceleration processor running the AI model and the associated NN and runtime libraries, the shared RAM and the system interface, as shown in Figure 9. The sub-system includes hardware and software AI and non-AI components that are subject to safety, cybersecurity, reliability and AI governance requirements.

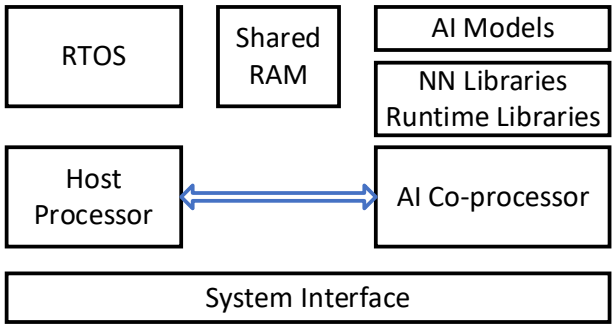


*Figure 9: AI model inference sub-system with associated software stacks*

Such a sub-system enables the edge AI device to run complex models such as LLM transformer, vision transformers, GenAI diffusion models and Multi-Modal-Models (MMM) that can process information from multiple sources of information simultaneously. For example, an MMM can process data from depth or stereo cameras, from LIDAR sensors, from microphones and from ultrasound sensors at the same time. Depending on the size of the shared memory and type of AI co-processor, an edge AI inference SoC can run models having between 1 billion and 7 billion parameters.

The AI inference sub-system requirements for automotive, aerospace and government applications, as resulting from governance measures, should consider the following standards, among others:

- Safety of Intended Functionality: ISO 21448[5]
- Functional safety to mitigate random hardware faults: ISO 26262[4], ARP4761[21]
- Cybersecurity: ISO 21434[6], DO-178C[18]
- Reliability: IATF 16949[16], AEC-Q100[17], DO-254[19]
- AI safety, security and trustworthiness: ISO 42001[11], ISO 8800[8], NIST 600-AI-600-1[7]

These requirements are accomplished by implementing specific measures such as safety mechanisms, cybersecurity mechanisms, AI solutions for enhanced accuracy, precision, and recall, for model transparency and non-bias, and mechanisms for increased robustness to adversarial attacks. Implementation trade-offs for conflicting measures are presented in Table 4.

*Table 4: Cross-correlations and trade-off examples*

| Measure | Safety | Cyber-security | AI | Trade-off |
|---|---|---|---|---|
| Dual-core lockstep | (+) Increased diagnostic coverage (ASIL D) | (-) Double the number of attack paths | (-) Increasing consumed power | Duplicate blocks instead of entire core |
| Increase encryption length from 64-bit to 128-bit | (-) Increased area causes higher block failure rate | (+) Harder to break | (+) Higher resilience to adversarial attack | Implement BIST at start-up |
| Power-down safe state | (+) Start-up enables running latent faults safety mechanism | (-) Power down/ up may enable DoS attack | (-) Long start-up sequence for loading the model | Implement a reset safe-state instead of power-down |

## 7. CONCLUSIONS

Combining AI, functional safety, cybersecurity and reliability requirements in automotive, aerospace and government applications development is of critical importance for creating robust, safe and secure AI-driven automotive systems. Cybersecurity measures are needed for preventing unauthorized access or attacks that could compromise the safety and functionality of both AI models and vehicle AI systems. Datasets and models are cybersecurity assets that must be protected using cybersecurity mechanisms such as encryption keys, typically generated by root-of-trust or true random number generators. From the functional safety and reliability perspective, the hardware supporting the AI models and algorithms may be subject to random failures therefore, safety measures and reliability enhancements need to be considered for addressing the associated risks. The more complex the model and more hardware resources employed, the higher the probability of random hardware faults, leading to a decrease in the functional safety performance of the system. At the same time, adding more safety mechanisms may increase the vulnerability to cybersecurity attacks. On the other hand, cybersecurity mechanisms may often challenge the safety requirements, for example by executing OTA software updates on software units that are safety relevant. Such updates must be verified for functional safety compliance prior to being deployed on the edge devices. AI systems include all these safety and cybersecurity aspects, as well as other specific verifications such as those related to the stability of the model in time.

In conclusion, AI development involves implementation solutions that often require trade-offs and optimizations to cope with systematic challenges and with a wide range of unexpected situations during field operation. The complexity of such development is addressed by a combination of development processes working in synergy and by using special tools and methodologies.

## REFERENCES


[1] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence : 2023

[2] EU Artificial Intelligence Act : 2024

[3] IEC 61508 — Functional safety of electrical/electronic/programmable electronic safety-related systems –

[4] ISO 26262 Second edition : 2018 Road vehicles — Functional safety

[5] ISO 21448:2022 Road vehicles — Safety of the intended functionality

[6] ISO/SAE 21434:2021 - Road vehicles — Cybersecurity engineering

[7] NIST-AI-600-1:2024 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile

[8] ISO 8800:2024 Road vehicles — Safety and artificial intelligence

[9] ISO 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

[10] ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements

[11] ISO 42001:2023 Information technology — Artificial intelligence — Management system

[12] ISO 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)

[13] ISO 4804:2020 Road vehicles — Safety and cybersecurity for automated driving systems — Design, verification and validation

[14] ISO 5469:2024 Artificial intelligence — Functional safety and AI systems

[15] ISO 9001:2015 Quality management systems — Requirements

[16] IATF 16949:2016 — Automotive Quality Management System

[17] AEC-Q100:2023 — Failure Mechanism Based Stress Test Qualification For Integrated Circuits (base document)

[18] DO-178C — Software Compliance For Aerospace & Defence

[19] DO-254 — Design Assurance for Airborne Electronic HW

[20] ARP4754A — Guidelines for Development of Civil Aircraft and Systems

[21] ARP4761 — Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment

[22] ISO/TR 12204:2012 — Road vehicles — Ergonomic aspects of transport information and control systems — Introduction to integrating safety critical and time critical warning signals