Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ISO 26262 ASIL-B Compliance for AI-Integrated HMI in Commercial Vehicle VCUs

Mahesh Kumar Shanmugam

Kettering University, USA

ARTICLE INFO

Received: 02 Nov 2024

Revised: 18 Dec 2024

Accepted: 28 Dec 2024

ABSTRACT

The increasing deployment of artificial intelligence (AI) and machine learning (ML) within commercial vehicle human–machine interface (HMI) systems introduce functional safety challenges that conventional automotive safety standards do not yet fully resolve. Although ISO 26262 establishes a mature lifecycle framework for electrical and electronic systems, important questions remain regarding specification of learned behaviour, traceability between requirements and trained models, robustness under distribution shift, governance of training datasets, and assurance of probabilistic decision-making. These issues are particularly significant in heavy-duty commercial vehicles, where high vehicle mass, air-brake delay, trailer dynamics, and demanding duty cycles increase the operational consequences of delayed or misleading driver information. This paper proposes a practical ASIL-B compliance framework for AI-integrated HMI systems operating within commercial vehicle control units (VCUs). The framework combines ISO 26262 functional safety processes, ISO 21448 Safety of the Intended Functionality (SOTIF), and emerging AI governance guidance including ISO/IEC TR 5469, ISO/IEC 23894, ISO/IEC 42001, and the NIST AI Risk Management Framework. The proposed approach extends conventional automotive safety engineering through AI-aware hazard analysis and risk assessment (HARA), AI-extended design failure mode and effects analysis (DFMEA), fault tree analysis (FTA), structured verification and validation (V&V), and safety-case development tailored to bounded-authority AI-HMI architectures. The paper argues that ASIL-B is an appropriate integrity target when AI functionality remains advisory or supervisory and when deterministic safety mechanisms retain responsibility for critical warnings and fallback behaviour. The primary contribution is a systems-oriented ASIL-B compliance framework for AI-integrated HMI systems in heavy-duty commercial vehicle VCUs that incorporates AI failure modes, operational-domain assumptions, and bounded AI authority principles within an automotive functional safety context.

Keywords: ISO 26262, AI-Integrated HMI, Commercial Vehicle VCU, Functional Safety, ASIL-B Compliance

1. Introduction

Artificial intelligence and machine learning technologies are increasingly becoming part of commercial vehicle HMI architectures. Modern heavy-duty trucks now integrate AI-enabled driver-support functions capable of prioritising alerts, interpreting operational conditions, predicting subsystem anomalies, and adapting information presentation according to vehicle state and driver context. Within contemporary commercial vehicle electronic architectures, the vehicle control unit increasingly operates as a systems-integration platform that coordinates communication between braking systems, powertrain controllers, telematics modules, driver monitoring systems, and HMI subsystems.

This architectural evolution reflects broader industry movement toward software-defined vehicles and intelligent driver-assistance ecosystems. Commercial fleet operators increasingly demand predictive

diagnostics, operational efficiency, driver-support automation, and improved fleet safety analytics. Consequently, AI-enabled HMI systems are no longer limited to infotainment or convenience applications; they increasingly influence operational awareness and driver decision-making during safety-relevant conditions.

Despite these developments, functional safety assurance for AI-enabled automotive systems remains an unresolved engineering challenge. ISO 26262 is well established as the dominant framework for automotive functional safety and provides rigorous processes for hazards arising from malfunctioning electrical and electronic systems [1]. However, AI and machine learning introduce behaviours that are fundamentally different from deterministic software architectures traditionally addressed by the standard.

Several unresolved questions remain central to AI functional safety engineering. Conventional requirements engineering assumes deterministic system behaviour, whereas AI systems frequently rely on learned probabilistic inference derived from training datasets. Traceability between high-level safety requirements and trained model parameters remains difficult to establish. Robustness against distribution shift, environmental variation, or out-of-distribution inputs is difficult to verify exhaustively. Training datasets increasingly function as safety-critical engineering artifacts, requiring governance processes comparable to conventional software lifecycle controls. AI outputs may also appear plausible and technically valid while still producing unsafe operational outcomes because of insufficient contextual understanding.

Recent studies addressing machine-learning extensions to ISO 26262 argue that conventional automotive safety lifecycles require additional ML-specific lifecycle phases including data preparation, model training, and deployment assurance [9]. This observation reflects a broader industry concern that conventional automotive safety methods alone are insufficient for AI-enabled systems.

The commercial vehicle context intensifies these concerns significantly. Heavy-duty trucks operate under safety constraints that differ materially from passenger vehicles. These vehicles are subject to regulations such as FMVSS 121 for air-brake systems and FMVSS 136 for electronic stability control. High gross combination weight, long stopping distances, air-brake pneumatic delay, trailer articulation effects, and load-dependent stability margins create operational conditions where driver response quality becomes particularly important.

Commercial drivers frequently operate under long duty cycles, fatigue conditions, night-driving environments, adverse weather, work-zone traffic, and congested freight corridors, increasing operational exposure and safety complexity [12]. In such conditions, delayed or misleading AI-generated HMI outputs may influence driver decisions during situations where recovery capability is constrained by vehicle dynamics and operational inertia.

An important distinction must therefore be made between informational AI functionality and safety-critical decision authority. The central engineering question is not whether AI can support commercial vehicle HMIs, but rather how AI authority should be bounded within a defensible functional safety architecture.

Current industrial practice increasingly favours architectures where deterministic safety mechanisms remain responsible for critical safety functions while AI systems provide contextual interpretation, prioritisation, or advisory support. This architectural principle reflects a growing recognition that AI systems should operate within constrained authority boundaries rather than replacing deterministic safety supervision.

Currently, no comprehensive ASIL-B compliance framework specifically addresses AI-integrated HMI systems within heavy-duty commercial vehicle VCUs. Existing literature discusses AI assurance in

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

general automotive or autonomous-driving contexts but does not provide a practical integration framework combining ISO 26262, SOTIF, AI governance, and commercial vehicle operational assumptions.

This paper addresses that gap by proposing a systems-oriented ASIL-B compliance framework for AI-integrated HMI systems in commercial vehicle VCUs. The paper contributes:

1. A commercial-vehicle-oriented HARA methodology for AI-HMI systems.
2. A bounded-authority functional safety architecture.
3. An AI-extended DFMEA structure.
4. A three-stream verification and validation strategy.
5. A structured ASIL-B safety-case framework integrating AI governance and functional safety evidence.

The paper is intentionally framed from a systems architecture and engineering management perspective. In practice, AI functional safety is no longer solely a software-engineering issue; it is increasingly a systems-governance and integration-management problem involving architecture allocation, operational assumptions, update governance, lifecycle traceability, and organisational accountability.

2. Background

2.1 ISO 26262 Functional Safety

ISO 26262 remains the dominant automotive functional safety standard for electrical and electronic systems and provides a structured lifecycle-oriented framework for identifying hazards arising from malfunctioning system behaviour and deriving safety requirements proportional to operational risk [1]. The standard defines a comprehensive development process covering item definition, hazard analysis and risk assessment (HARA), safety-goal allocation, functional safety concept development, technical safety concept definition, hardware and software development, verification, validation, and safety-case construction.

A major strength of ISO 26262 is its emphasis on traceability and evidence-based development. Automotive Safety Integrity Levels (ASILs), ranging from ASIL A to ASIL D, are assigned using severity, exposure, and controllability metrics [1]. These integrity levels determine the rigor of required development and verification activities. ISO 26262 has therefore become foundational to modern automotive safety engineering because it establishes systematic processes for ensuring that electrical and electronic systems achieve acceptable residual risk levels.

Despite its importance, ISO 26262 was primarily developed around deterministic software and hardware architectures. Conventional automotive software can typically be specified through explicit rules, state transitions, and predictable execution behaviour. AI-enabled systems differ fundamentally because their behaviour emerges from training processes and data-driven optimisation rather than fully deterministic programming logic. Consequently, AI systems introduce assurance challenges associated with probabilistic inference, dataset dependency, uncertainty estimation, and operational-domain variability.

These differences create limitations when applying conventional ISO 26262 methodologies directly to AI-enabled systems. Functional safety engineers must increasingly evaluate robustness under unseen operating conditions, distribution-shift behaviour, confidence calibration, and dataset representativeness. Traditional verification approaches based on deterministic requirement traceability

may therefore become insufficient when system behaviour depends heavily on learned representations and statistical inference [3], [9].

2.2 ISO 21448 (SOTIF)

ISO 21448, commonly referred to as Safety of the Intended Functionality (SOTIF), addresses hazards arising from functional insufficiencies rather than direct hardware or software faults [2]. SOTIF is particularly relevant for systems that rely on perception algorithms, contextual interpretation, environmental understanding, and AI-enabled decision support.

In conventional functional safety analysis, hazards are generally associated with malfunctioning behaviour caused by faults within electrical or electronic systems. However, AI-enabled systems may generate unsafe outcomes even when no hardware or software fault exists. An AI model may behave according to its trained operational logic while still producing unsafe recommendations because the operational scenario exceeds the assumptions represented within the training data or because the intended functionality is insufficient for the encountered situation.

This distinction is especially important for AI-integrated HMI systems in commercial vehicles. For example, an AI-generated operational recommendation may appear technically valid but become unsafe under rare environmental conditions, degraded sensor states, or unusual trailer configurations. Similarly, contextual misunderstanding or insufficient scenario representation may influence driver decision-making even when deterministic safety mechanisms remain operational.

ISO 21448 therefore complements ISO 26262 by extending safety analysis beyond deterministic malfunctioning behaviour into performance limitations, scenario insufficiencies, and operational-domain constraints. In practice, AI-enabled automotive systems increasingly require integration of both ISO 26262 functional safety principles and SOTIF-oriented performance analysis [1], [2].

2.3 AI Safety Standards and Governance Frameworks

The rapid adoption of AI-enabled systems within safety-critical industries has resulted in the emergence of several governance and assurance frameworks relevant to automotive functional safety. These frameworks focus on AI-specific risks including uncertainty management, robustness assurance, transparency, operational governance, and lifecycle accountability.

ISO/IEC TR 5469 provides guidance regarding artificial intelligence in safety-related systems and discusses AI properties, lifecycle concerns, risk factors, and validation considerations [3]. The document recognises that AI-enabled systems introduce unique challenges associated with probabilistic behaviour, training-data dependence, and operational variability. Similarly, ISO/IEC 23894 introduces AI risk-management guidance focused on governance, transparency, organisational accountability, and lifecycle risk mitigation [4].

ISO/IEC 42001 further extends this discussion by defining an AI management-system framework emphasizing governance structures, continuous improvement, oversight responsibilities, and operational control mechanisms [5]. Unlike traditional software quality-management approaches, these standards recognise that AI assurance depends not only on technical correctness but also on organisational governance processes associated with datasets, model updates, monitoring activities, and operational assumptions.

The NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) also contributes an important governance-oriented perspective [6]. The framework organises AI assurance around four core functions: govern, map, measure, and manage. Although not automotive-specific, the NIST framework strongly influences current AI governance discussions because it emphasises lifecycle monitoring, organisational accountability, and continuous operational oversight.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The Institution of Engineering and Technology (IET) report *The Application of Artificial Intelligence in Functional Safety* further highlights practical engineering concerns associated with AI deployment in safety-critical environments [10]. The report identifies challenges including dataset bias, overfitting, inadequate scenario coverage, uncertainty regarding AI trustworthiness, and excessive human reliance on AI-generated outputs. Importantly, the report also emphasises that testing alone is insufficient for establishing AI trustworthiness and that analytical evidence, governance discipline, and operational supervision are equally necessary components of safety assurance.

2.4 Emerging Automotive AI Safety Guidance

Emerging automotive AI safety initiatives such as ISO/PAS 8800 attempt to address safety-related concerns associated with AI-enabled electrical and electronic systems in production road vehicles. These initiatives focus on insufficiencies related to dataset limitations, robustness, uncertainty, bias management, and operational-domain definition. However, such guidance should currently be interpreted as evolving directional support rather than fully mature standalone compliance frameworks.

As a result, practical automotive AI safety assurance still depends heavily on integrating multiple partially overlapping methodologies including ISO 26262, ISO 21448, AI governance frameworks, organisational safety processes, and application-specific engineering controls. This fragmented state reflects the broader challenge facing the automotive industry: AI-enabled systems evolve more rapidly than conventional safety standards, requiring engineering organisations to combine established functional safety principles with emerging AI assurance methodologies [1]–[6].

3. Related Work

The growing integration of artificial intelligence into automotive systems has generated increasing research interest in functional safety, AI governance, and machine-learning assurance. However, current literature remains fragmented between traditional automotive functional safety engineering and broader AI assurance research. Most existing studies focus either on autonomous driving systems or on generic AI governance frameworks, while comparatively limited work addresses AI-integrated HMI systems operating within commercial vehicle VCU architectures.

One of the most significant contributions in this area is the IEEE Access study titled *"A Systematic Approach to Enhancing ISO 26262 With Machine Learning-Specific Life Cycle Phases and Testing Methods"* [9]. The study argues that conventional ISO 26262 processes are insufficient for machine-learning-enabled systems because AI development introduces lifecycle activities not addressed within traditional automotive software engineering. These activities include dataset preparation, model training, retraining governance, operational-domain validation, and robustness assessment. The authors further emphasise that AI assurance cannot rely solely on conventional software testing because machine-learning behaviour is heavily dependent on training data quality and operational representativeness.

The Institution of Engineering and Technology (IET) report *"The Application of Artificial Intelligence in Functional Safety"* also provides an important engineering perspective regarding practical AI safety challenges [10]. Unlike purely theoretical AI discussions, the IET report focuses on the operational realities of deploying AI-enabled systems within safety-critical environments. The report identifies several major concerns including data quality, dataset bias, overfitting, insufficient scenario coverage, uncertainty regarding AI trustworthiness, and excessive human reliance on AI-assisted systems. These observations are especially relevant to AI-integrated HMI systems because the safety problem is not limited to algorithmic correctness alone; it also involves driver perception, trust calibration, and behavioural response to AI-generated information.

Another important contribution is IEEE 2846-2022, which introduces an assumptions-oriented framework for safety-related models used in automated driving systems [7]. Although the standard primarily targets automated driving technologies, its emphasis on operational assumptions and foreseeable scenario analysis is highly applicable to commercial vehicle HMI architectures. AI-enabled HMI systems similarly depend on assumptions regarding driver attentiveness, environmental conditions, communication timing, and operational-state interpretation. Incorrect assumptions regarding these factors may significantly influence system safety and controllability.

UL 4600 further contributes to the discussion by emphasising evidence-based safety cases for autonomous and AI-enabled systems [8]. Unlike traditional compliance-oriented standards, UL 4600 focuses on structured safety arguments supported through analytical and empirical evidence. This approach is particularly useful for AI-enabled systems because AI behaviour often cannot be completely specified through deterministic requirements alone. Instead, safety assurance increasingly depends on demonstrating bounded operational behaviour, governance discipline, and acceptable residual risk under defined operating conditions.

Research related to Safety of the Intended Functionality (SOTIF) has also become increasingly relevant for AI-enabled automotive systems. ISO 21448 addresses hazards resulting from functional insufficiencies rather than direct faults [2]. This distinction is important because AI-related hazards may arise even when the system behaves according to its trained operational logic. For example, a machine-learning model may produce technically valid outputs that nevertheless become unsafe because the operational scenario exceeds the assumptions represented within the training data. Consequently, SOTIF concepts have become central to AI safety discussions within the automotive industry.

Several studies additionally highlight the importance of uncertainty estimation and robustness analysis in machine-learning-enabled automotive systems. AI models operating within safety-related environments must demonstrate resilience against degraded sensor inputs, communication latency, environmental variation, and out-of-distribution (OOD) operational conditions. Robustness evaluation therefore becomes an essential complement to conventional software verification activities. Similarly, confidence calibration has emerged as a critical concern because overconfident incorrect AI outputs may encourage inappropriate driver trust and unsafe behavioural responses.

Despite these developments, current literature still lacks a practical and integrated ASIL-B compliance framework specifically addressing AI-enabled HMI systems within heavy-duty commercial vehicle VCUs. Existing work frequently concentrates on autonomous-driving perception systems, advanced driver-assistance systems (ADAS), or general AI governance methodologies rather than bounded-authority AI-HMI architectures. Commercial vehicle operational assumptions such as high vehicle mass, trailer dynamics, prolonged duty cycles, and air-brake response delays are also insufficiently represented within current AI functional safety discussions.

This paper therefore contributes a practical systems-oriented framework integrating ISO 26262 lifecycle discipline, ISO 21448 SOTIF principles, AI governance methodologies, and commercial vehicle operational considerations into a unified ASIL-B compliance strategy. The proposed framework specifically addresses bounded AI authority, AI-aware HARA, AI-extended DFMEA, integrated verification and validation, and governance-oriented safety-case development for AI-integrated HMI systems operating within commercial vehicle VCUs.

4. Item Definition and ASIL-B Rationale

The item considered in this paper is an AI-integrated human–machine interface (HMI) subsystem operating within a commercial vehicle vehicle control unit (VCU) architecture. The system functions as

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

an intelligent interface layer between the driver and multiple vehicle subsystems, supporting operational awareness, contextual interpretation, and safety-relevant information presentation. The AI-enabled HMI receives inputs from braking controllers, vehicle-dynamics systems, powertrain modules, telematics gateways, environmental sensors, driver-monitoring systems, and operational-status controllers distributed throughout the vehicle network. These inputs are processed through AI-enabled software components that generate contextual alerts, adaptive driver-support outputs, warning prioritisation, and operational recommendations intended to improve driver situational awareness and response quality during dynamic driving conditions.

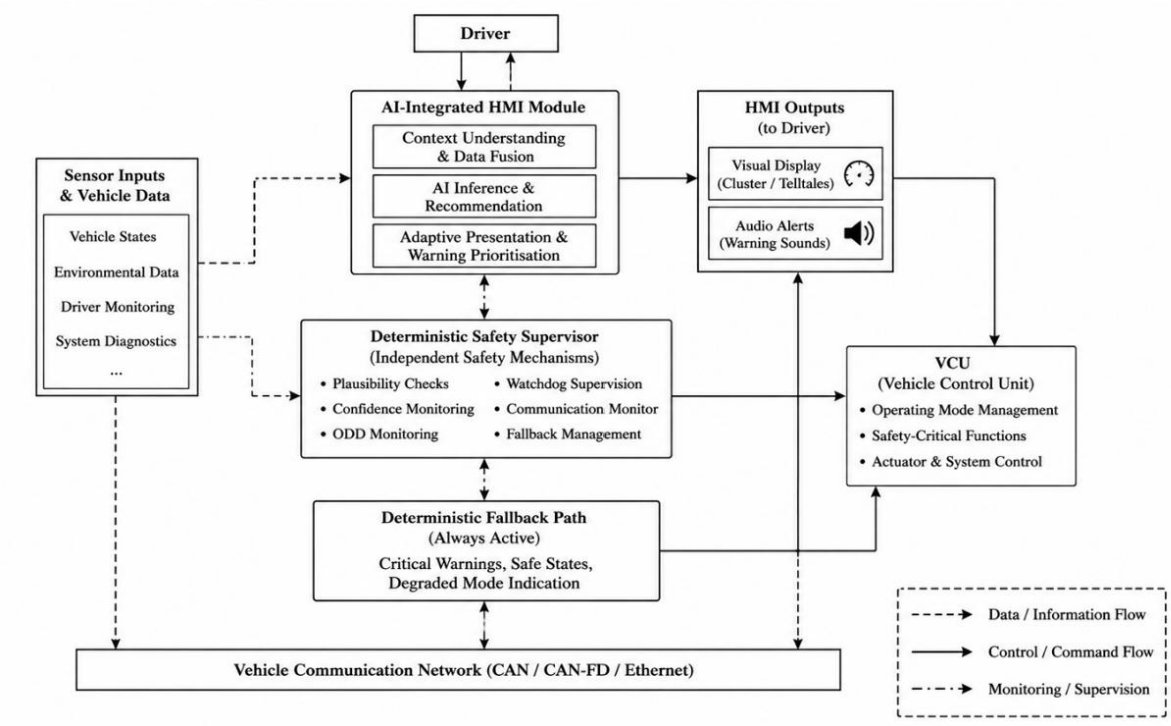

Figure 1: AI-Integrated HMI Architecture within Commercial Vehicle VCU

The HMI communicates with the driver through several integrated channels, including digital instrument clusters, centre-console displays, visual telltales, audio-warning systems, optional head-up displays, and driver-assistance notification interfaces. Within the proposed architecture, the AI functionality is intentionally restricted to advisory and supervisory authority. The AI subsystem may prioritise warnings, contextualise operational information, recommend driver actions, and personalise alert presentation according to operating conditions and driver context. However, the AI system does not directly control braking, steering, propulsion, or emergency actuation functions without independent deterministic supervision. This architectural separation is fundamental to the proposed safety strategy because it ensures that critical vehicle-control authority remains under deterministic safety mechanisms rather than probabilistic AI inference.

The assumed operator is a trained professional commercial driver functioning within normal duty-cycle limitations. The operational design domain (ODD) includes highway freight operation, urban logistics routes, stop-and-go traffic conditions, night driving, moderate adverse weather, work-zone environments, and long-duration freight duty cycles typical of commercial transportation operations. Vehicle platforms within scope include heavy-duty trucks operating with multiple powertrain variants, trailer configurations, and communication architectures based on CAN, CAN-FD, and Ethernet-enabled vehicle networks.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A critical aspect of the item definition is the distinction between informational, advisory, supervisory, and actuation authority. This distinction directly influences functional safety classification because the degree of AI authority determines how significantly the system can affect vehicle behaviour and driver decision-making. In the proposed architecture, ASIL-B is considered appropriate because hazards arise primarily through driver interpretation and behavioural influence rather than direct autonomous control of safety-critical actuators. The AI-HMI influences the driver's understanding of operational conditions, but independent deterministic systems remain responsible for essential safety functions and emergency intervention mechanisms.

This rationale aligns with current industrial practice. NXP documentation describing functional safety architectures for cockpit processors and HMI-domain controllers identifies ASIL-B as a common integrity target for safety-relevant HMI systems and display applications [11]. However, ASIL-B is only appropriate when several architectural and operational conditions are satisfied. AI authority must remain explicitly bounded, deterministic monitoring mechanisms must operate independently of the AI subsystem, driver controllability assumptions must remain valid under expected operating conditions, and safety-critical warnings must not be suppressible by AI behaviour alone.

If the AI-integrated HMI were capable of independently authorising hazardous VCU mode transitions, suppressing mandatory safety alerts, or misleading the driver during emergency conditions without deterministic supervision, the resulting hazards could justify higher integrity classifications such as ASIL-C or ASIL-D. Consequently, ASIL allocation should never be assumed solely from the presence of an HMI function or cockpit-domain architecture. Instead, ASIL determination must emerge from explicit hazard analysis and risk assessment (HARA) considering operational assumptions, controllability arguments, driver interaction, system architecture allocation, and the bounded nature of AI authority within the overall commercial vehicle safety concept [1], [2].

5. HARA for AI-Integrated HMI

Hazard analysis and risk assessment (HARA) for AI-integrated HMI systems requires a broader analytical perspective than conventional display-system safety analysis. Traditional HMI safety evaluations primarily focus on deterministic hardware and software failures such as display loss, communication interruption, processor malfunction, or corrupted signal transmission. However, AI-enabled systems introduce additional safety concerns associated with probabilistic inference, contextual misunderstanding, confidence miscalibration, uncertainty in operational interpretation, and human behavioural interaction with AI-generated outputs. Consequently, HARA for AI-integrated HMI systems must evaluate not only component malfunctions but also performance insufficiencies, operational-domain limitations, and the influence of AI behaviour on driver decision-making [1], [2].

One of the most critical hazard categories involves missing, delayed, or suppressed warnings. In commercial vehicle operations, failure to present safety-critical alerts within acceptable response time thresholds can directly reduce driver controllability during hazardous situations. Such hazards may include complete failure to display warnings, warning latency exceeding safe reaction margins, suppression of alerts because of AI prioritisation errors, or delayed escalation caused by incorrectly configured confidence thresholds. Since heavy-duty commercial vehicles operate with significant inertia and extended braking distances, even relatively small delays in warning presentation may materially affect operational safety.

Another major hazard category involves misleading or incorrect outputs generated by the AI-HMI subsystem. These hazards include false alerts causing unsafe driver reactions, incorrect operating-mode indications, misleading recommendations related to braking or powertrain conditions, and contextual misinterpretation leading to inappropriate driver decisions. Unlike deterministic software faults, these

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

hazards may occur even when the AI model is functioning according to its learned operational logic. In such situations, hazards arise because the operational environment exceeds the assumptions represented within the training data or because the AI model lacks sufficient contextual understanding [2], [10].

Display integrity failures remain equally important within AI-integrated HMI architectures. Representative hazards include display freeze, audio-warning loss, incorrect telltale presentation, symbol corruption, unit conversion errors, and misassociation of warnings with unrelated vehicle subsystems. Although these hazards resemble conventional HMI failures, their consequences may become more severe in AI-assisted systems because drivers increasingly depend on integrated contextual information during complex operational scenarios.

AI-specific hazards introduce additional complexity into the HARA process. These include operation under out-of-distribution (OOD) conditions, AI misclassification of operational state, overconfident incorrect inference, training-data imbalance, contextual misunderstanding, and excessive driver reliance on AI-generated recommendations. For example, an AI model trained primarily using daytime highway freight conditions may behave unpredictably during severe weather events, unusual work-zone traffic patterns, or poorly represented trailer configurations. Similarly, confidence miscalibration may result in uncertain outputs being presented with unjustified certainty, increasing the likelihood of unsafe driver trust and behavioural dependence [3], [4], [10].

Commercial vehicle operational assumptions significantly influence severity classification within the HARA framework. Heavy-duty trucks possess high vehicle mass, extended stopping distances, trailer articulation dynamics, and air-brake pneumatic delay characteristics that differ substantially from passenger vehicles. Consequently, hazards that might be classified as moderate severity in light vehicles can become severe hazards within commercial trucking environments because recovery capability is constrained by vehicle dynamics and operational inertia [12].

Exposure ratings are similarly elevated because commercial vehicles operate continuously across extended duty cycles, night-driving environments, adverse weather conditions, congested freight corridors, and work-zone traffic situations. Professional drivers may encounter hazardous operational scenarios repeatedly over prolonged operating periods, increasing the practical frequency and realism of exposure assumptions used during HARA evaluation.

Controllability assessment must additionally consider driver workload, reaction time availability, traffic density, vehicle inertia, road geometry, and the availability of independent fallback systems. Although trained commercial drivers generally possess higher situational awareness and operational discipline than consumer drivers, practical recovery capability remains constrained by braking distance, vehicle mass, trailer behaviour, and surrounding traffic conditions.

Representative HARA Table

Hazard	Severity	Exposure	Controllability	ASIL
AI suppresses brake-system warning	S3	E4	C2	ASIL-B
Incorrect downhill mode recommendation	S3	E3	C2	ASIL-B
False collision warning causing abrupt response	S2	E4	C2	ASIL-B
Display freeze during fault condition	S3	E3	C2	ASIL-B
AI confidence overstatement	S2	E4	C2	ASIL-B

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The HARA process for AI-integrated HMI systems should remain iterative throughout the complete development lifecycle. Unlike conventional deterministic systems, AI-related hazards may emerge progressively during simulation testing, usability evaluation, field validation, operational deployment, or post-deployment monitoring activities. Changes in operational environments, dataset composition, model updates, or driver interaction behaviour may reveal previously unidentified hazards or modify existing risk assumptions. Functional safety assurance for AI-enabled systems therefore requires continuous reassessment and operational feedback integration rather than relying solely on static upfront hazard analysis performed during early development phases [1], [2], [10].

6. Functional Safety Concept

The proposed functional safety concept for the AI-integrated HMI architecture is based on the principle of bounded AI authority. This principle reflects a fundamental assumption within modern automotive safety engineering: artificial intelligence may enhance operational awareness, contextual interpretation, and driver support, but deterministic safety mechanisms must remain responsible for essential hazard prevention and safety protection. In other words, AI functionality is permitted to support the driver and improve information management, yet critical vehicle safety decisions cannot rely solely on probabilistic AI inference without independent deterministic supervision [1], [10].

Within the proposed architecture, the AI subsystem primarily supports information prioritisation, contextual interpretation, adaptive information presentation, and driver-advisory functionality. The AI-enabled HMI may analyse operational conditions, interpret sensor-fusion data, prioritise warnings according to context, and provide recommendations intended to improve driver situational awareness. However, critical safety functions continue to operate through deterministic pathways that remain independent of AI behaviour. This architectural separation is essential because it limits the authority of AI-generated outputs and ensures that fundamental safety mechanisms remain predictable, verifiable, and independently supervised.

The functional safety concept therefore establishes several key safety goals. First, the system must ensure timely and perceivable presentation of safety-critical warnings under all operational conditions. Second, the architecture must prevent misleading or incorrect AI-generated recommendations from causing unsafe driver behaviour. Third, the system must prevent unsafe VCU mode transitions influenced by incorrect AI interpretation or recommendation logic. Additional safety goals include maintaining accurate operational-state display, detecting HMI corruption or communication failure, preventing suppression of mandatory safety warnings by AI prioritisation logic, and ensuring safe degraded operation whenever AI confidence falls below acceptable thresholds or operational conditions exceed the validated operating domain.

To achieve these objectives, the proposed architecture incorporates multiple deterministic safety mechanisms designed to supervise and constrain AI behaviour. Representative mechanisms include deterministic fallback warning paths capable of bypassing AI processing during safety-critical events, watchdog supervision for detecting HMI freeze or processor failure, communication heartbeat monitoring between the HMI and VCU subsystems, plausibility checks comparing AI outputs against validated sensor data, AI confidence-threshold management, degraded-mode indicators informing the driver of AI limitations, driver confirmation requirements before safety-relevant mode changes, and timeout handling for stale or missing communication data. These mechanisms collectively ensure that AI functionality remains constrained within defined operational boundaries while deterministic safety systems maintain responsibility for critical protection functions.

The architecture additionally enforces separation between AI and non-AI software partitions to support freedom-from-interference objectives required by ISO 26262 [1]. Such partitioning prevents failures or

unexpected behaviour within AI components from propagating directly into deterministic safety mechanisms. This separation is particularly important in software-defined vehicle architectures where increasing computational integration may otherwise create unintended coupling between AI-enabled functionality and safety-critical control systems.

An important observation emerging from this architecture is that AI governance itself becomes part of the functional safety concept. In conventional deterministic systems, assurance depends primarily on software correctness, hardware reliability, and systematic development processes. In AI-enabled systems, however, assurance additionally depends on governance of training data, operational-domain assumptions, update management, robustness evaluation, runtime supervision, and behavioural consistency over time [3]–[6]. Consequently, AI assurance is no longer solely a software-engineering activity; it becomes a broader systems-governance and lifecycle-management challenge requiring integration between functional safety engineering, data governance, validation strategy, and operational oversight.

The IET report on artificial intelligence in functional safety specifically recommends constraining AI systems using conventional deterministic safety mechanisms wherever possible [10]. This recommendation strongly supports the bounded-authority architecture proposed in this paper. Rather than allowing AI systems to independently control safety-critical behaviour, the architecture intentionally limits AI authority while preserving deterministic supervision and fallback protection. Such an approach represents one of the most practical and defensible strategies for integrating AI-enabled HMI systems into commercial vehicle VCUs while maintaining compliance with established automotive functional safety principles [1], [2].

7. DFMEA Extended for AI

Design Failure Mode and Effects Analysis (DFMEA) remains one of the most valuable analytical methods within automotive functional safety engineering because it provides a structured mechanism for identifying potential failures, understanding their operational effects, and linking those failures to engineering controls and mitigation strategies. However, AI-enabled HMI systems require the DFMEA methodology to be extended beyond traditional hardware and software failure mechanisms. Conventional automotive DFMEA approaches are primarily oriented toward deterministic faults such as hardware degradation, communication errors, timing violations, and software logic defects. In contrast, AI-enabled systems introduce probabilistic behaviour, data dependency, operational uncertainty, and performance insufficiencies that cannot be fully addressed through traditional failure analysis alone [1], [3].

Within conventional HMI architectures, representative failure modes include display power loss, processor failure, graphics freeze, corrupted CAN messages, timing violations, communication interruption, incorrect telltale logic, and software state-machine errors. These failures typically arise from hardware faults, communication disturbances, software defects, or integration issues. Their behaviour is generally deterministic and can often be analysed using established diagnostic and fault-detection techniques defined within ISO 26262 development processes [1].

AI-integrated HMI systems introduce additional categories of failure associated with machine-learning lifecycle behaviour and operational uncertainty. These include training-data insufficiency, model uncertainty, out-of-distribution (OOD) operation, confidence miscalibration, biased datasets, model version mismatch, and post-deployment behavioural drift. Unlike conventional software defects, these issues may emerge even when the underlying implementation is technically correct because AI system behaviour depends heavily on training conditions, dataset representativeness, operational context, and environmental variability [3], [4], [10].

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Representative AI-specific failure modes include incorrect operational-context classification, overconfident but incorrect recommendations, insufficient scenario representation within training data, training/test data leakage, degraded sensor quality affecting inference reliability, and unintended behavioural changes resulting from uncontrolled model updates. For example, an AI model trained primarily using daytime highway freight scenarios may generate unsafe recommendations during rare work-zone conditions or severe weather situations not sufficiently represented within the training dataset. Similarly, confidence miscalibration may cause uncertain outputs to be presented to the driver with unjustified certainty, increasing the likelihood of inappropriate driver trust and unsafe operational decisions.

Representative DFMEA Extract

Cause	Failure Mode	Effect	Detection	Mitigation
CAN communication interruption	Warning unavailable	Driver misses safety alert	Heartbeat monitor	Deterministic fallback
OOD operational scenario	Incorrect AI recommendation	Unsafe driver action	OOD detector	Degraded-mode fallback
Confidence miscalibration	Incorrect high-confidence output	Driver overtrust	Calibration audit	Confidence thresholding
Uncontrolled model update	Behaviour deviation	Invalidated assumptions	Version management	Regression gating

An important advantage of DFMEA in AI-enabled systems is that it directly connects operational hazards to engineering controls and verification mechanisms. In conventional deterministic systems, mitigation strategies are typically centred on hardware diagnostics, communication redundancy, or software supervision. In AI-enabled systems, however, mitigation must additionally include dataset governance, robustness evaluation, confidence management, runtime monitoring, and controlled update-management processes [3]–[6].

Dataset governance becomes particularly important because training data increasingly functions as a safety-relevant engineering artifact. The representativeness, quality, and integrity of datasets significantly influence AI behaviour during operational deployment. Similarly, robustness evaluation becomes essential for assessing model behaviour under environmental variation, degraded sensor input, and operational-domain boundary conditions. Runtime monitoring mechanisms such as confidence-threshold management, plausibility checks, and OOD detection provide additional safeguards against unsafe AI behaviour during vehicle operation.

Controlled update-management and regression-testing processes are equally important because post-deployment model modifications may unintentionally alter safety-relevant system behaviour. Regression gating mechanisms therefore become necessary to ensure that retrained or updated models do not invalidate previously established safety assumptions or introduce unintended operational risks.

DFMEA therefore evolves into a broader systems-integration mechanism linking hazard analysis and risk assessment (HARA), technical safety concepts, AI governance processes, and verification evidence. Rather than functioning solely as a reliability-analysis tool, DFMEA in AI-enabled HMI systems becomes a central framework for integrating deterministic functional safety engineering with AI lifecycle governance and operational assurance practices [1], [4], [10].

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

8. Fault Tree Analysis

Fault Tree Analysis (FTA) remains an important analytical tool for evaluating combinations of failures capable of defeating the functional safety concept within AI-integrated HMI systems. FTA provides a structured method for decomposing top-level hazardous events into lower-level causal pathways involving hardware faults, software failures, communication disturbances, human interaction problems, and AI-related behavioural insufficiencies. Within automotive safety engineering, FTA is particularly valuable because it supports systematic reasoning regarding how multiple independent failures may combine to produce unsafe system behaviour [1].

For AI-integrated HMI systems operating within commercial vehicle VCUs, representative top-level undesired events include situations where the driver does not receive a required safety-critical warning, receives a misleading AI-generated recommendation, or observes an incorrect operational-mode display. Additional top-level hazards include unsafe VCU transitions influenced by AI-generated outputs, excessive driver workload or distraction caused by HMI behaviour, and failure of the system to communicate degraded AI functionality or operational limitations to the driver.

Hardware Faults

One major branch of the fault tree includes hardware-related failures such as display failure, processor faults, memory corruption, power-supply failure, and loss of audio-warning functionality. These failures may interrupt information presentation or prevent delivery of critical warnings to the driver. In commercial vehicle environments, where operational safety often depends on timely warning communication, such failures may significantly reduce driver controllability during hazardous conditions.

Software Faults

Software-related fault branches include stale data handling, priority-management errors, timing violations, and incorrect state-machine logic. These failures may result in delayed warnings, incorrect operational-state presentation, or improper prioritisation of safety-critical information. Since AI-integrated HMIs increasingly coordinate information from multiple vehicle subsystems, software integration complexity becomes a significant contributor to overall system risk.

AI Faults

AI-related fault branches introduce additional complexity because failures may arise from probabilistic inference behaviour rather than deterministic software defects. Representative AI faults include incorrect inference, confidence overstatement, poor edge-case generalisation, and dataset bias. Such hazards may emerge even when the AI implementation is technically correct because the operational scenario exceeds the assumptions represented within the training domain. Consequently, AI-related hazards often reflect performance insufficiency rather than conventional malfunctioning behaviour [2], [10].

Communication Faults

Communication-related fault branches include network latency, heartbeat loss, gateway routing errors, and corrupted CAN messages. Commercial vehicle VCUs increasingly rely on distributed electronic architectures involving multiple controllers, gateways, and communication networks. Delayed or corrupted communication between AI-HMI modules and deterministic safety systems may therefore compromise warning timing, operational-state synchronisation, or degraded-mode handling.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Human Interaction Faults

Human interaction faults represent another critical branch within the fault tree. These hazards include alert fatigue, ambiguous warning wording, poor visual placement of alerts, excessive driver distraction, and overtrust in AI-generated recommendations. AI-enabled HMIs influence not only technical system behaviour but also driver cognition and behavioural response. Consequently, human-factors considerations become integral to the overall safety analysis rather than secondary usability concerns.

Despite its usefulness, FTA has important limitations when applied to AI assurance. Conventional fault-tree methods assume largely deterministic relationships between component failures and hazardous outcomes. AI-related hazards, however, frequently arise from performance insufficiencies, operational complexity, probabilistic inference behaviour, and mismatch between the validated operating domain and real-world operational conditions [3], [10]. Such behaviour cannot always be represented effectively using purely Boolean fault relationships.

For this reason, FTA should not be treated as a standalone assurance method for AI-integrated HMI systems. Instead, it must operate alongside complementary analytical and validation approaches including DFMEA, SOTIF-oriented triggering-condition analysis, scenario-based validation, human-factors evaluation, and AI robustness assessment [2], [4], [10]. Together, these methods provide a more comprehensive understanding of both deterministic faults and AI-related performance limitations within commercial vehicle functional safety architectures.

9. Verification and Validation Plan

The proposed ASIL-B assurance strategy for AI-integrated HMI systems is structured around three parallel verification and validation (V&V) streams. This multi-layered approach reflects the reality that conventional automotive testing methods alone are insufficient for assuring AI-enabled systems operating within safety-relevant commercial vehicle architectures. Since AI-integrated HMIs combine deterministic software, probabilistic machine-learning behaviour, distributed vehicle communication, and human interaction, assurance activities must simultaneously address functional safety compliance, AI robustness, and driver usability. The proposed framework therefore integrates conventional ISO 26262 verification activities with AI-specific validation and HMI-focused human-factors evaluation [1], [2].

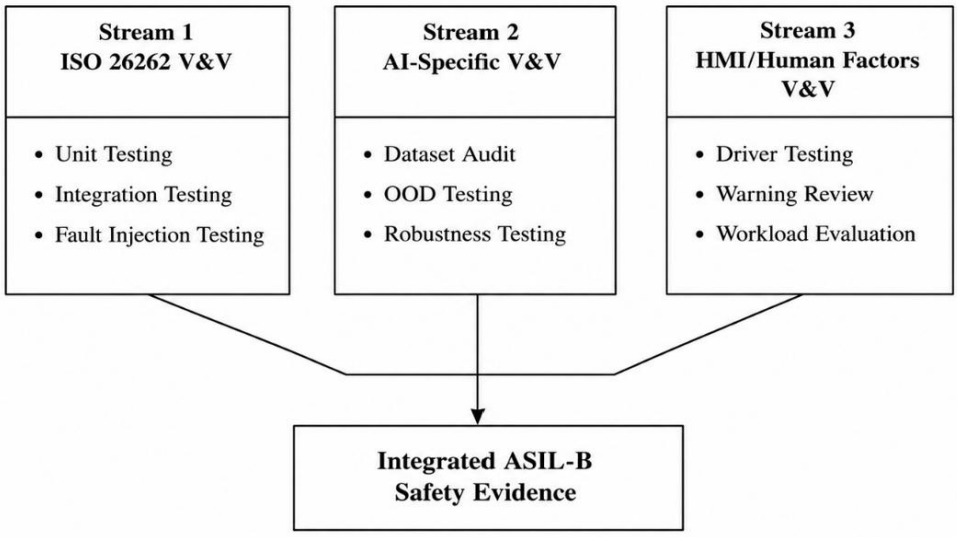

Figure 2: Three-Stream Verification and Validation Framework for AI-Integrated HMI

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Stream 1 — Conventional ISO 26262 Verification and Validation

The first verification stream focuses on conventional ISO 26262 functional safety activities associated with deterministic software and hardware behaviour. This stream includes requirements review, software and system architecture walkthroughs, unit testing, Modified Condition/Decision Coverage (MC/DC) analysis, software integration testing, system integration testing, fault injection testing, watchdog validation, timing analysis, communication-failure testing, and vehicle-level safety validation [1]. These activities remain essential because AI-integrated HMI systems still depend heavily on deterministic communication pathways, supervisory mechanisms, operating-state management, and safety-critical warning presentation logic.

Particular emphasis must be placed on freedom-from-interference verification between AI and non-AI software partitions. Since AI-enabled components may execute complex probabilistic inference tasks with varying computational behaviour, it is necessary to demonstrate that failures, delays, or resource contention within AI modules cannot compromise deterministic safety functions. Safety mechanisms must therefore demonstrate predictable and deterministic response behaviour under conditions including communication loss, corrupted data, timing violations, processor overload, and degraded operational states.

Fault injection testing is especially important because it validates whether the safety architecture can correctly detect and respond to abnormal operating conditions. Timing analysis similarly becomes critical in commercial vehicle systems where delayed warning presentation may materially reduce driver reaction capability because of vehicle inertia and long stopping distances.

Stream 2 — AI-Specific Verification and Validation

The second verification stream focuses specifically on AI-related assurance activities. AI-enabled systems require additional forms of verification because their behaviour depends not only on software correctness but also on dataset quality, operational-domain representation, model robustness, and lifecycle governance processes. Consequently, AI-specific assurance extends beyond traditional automotive software testing into broader evaluation of data integrity, training methodology, and behavioural consistency [3], [4].

This stream includes dataset-quality auditing, dataset provenance documentation, verification of independence between training, validation, and test datasets, scenario-coverage assessment, robustness testing, perturbation analysis, out-of-distribution (OOD) testing, confidence calibration evaluation, and model regression testing following updates or retraining activities. These activities collectively aim to demonstrate that the AI system performs consistently and predictably across representative operational conditions while appropriately identifying situations outside its validated operating domain.

Robustness testing is particularly important because AI models may encounter degraded sensor inputs, environmental variation, communication latency, or operational scenarios not represented during training. Perturbation analysis therefore evaluates how the model behaves under noisy, incomplete, or degraded input conditions. OOD testing further assesses whether the AI system can recognise unfamiliar operational situations and transition safely into degraded modes rather than generating confidently incorrect outputs [10].

The IET report on artificial intelligence in functional safety correctly notes that testing alone cannot establish AI trustworthiness [10]. Analytical evidence regarding dataset governance, process discipline, operational assumptions, and lifecycle management must supplement conventional testing evidence. As a result, AI assurance increasingly depends on governance quality as much as technical performance

metrics. Confidence calibration evaluation is similarly important because overconfident incorrect recommendations may encourage inappropriate driver trust and unsafe operational decisions.

Regression testing also becomes essential for maintaining behavioural consistency following model updates, retraining, recalibration, or deployment modifications. Since AI behaviour may change significantly following updates, controlled update governance and regression validation are necessary to prevent introduction of unintended safety-relevant behavioural changes into deployed vehicle systems.

Stream 3 — HMI-Specific Verification and Validation

The third verification stream addresses HMI-specific validation and human interaction assessment. Because AI-integrated HMIs directly influence driver perception, situational awareness, and behavioural response, functional safety assurance must include comprehensive evaluation of human factors and operational usability under realistic commercial vehicle conditions.

HMI-focused validation activities include driver-in-the-loop evaluation, warning salience testing, workload analysis, comprehension testing, mode-awareness studies, nuisance-alert analysis, degraded-mode evaluation, and long-duration operational assessment. These activities are intended to determine whether drivers correctly perceive, interpret, and respond to AI-generated warnings and operational information under realistic operational conditions.

Commercial vehicle testing must include representative freight duty cycles, multiple truck configurations, day and night operation, adverse weather environments, work-zone traffic conditions, and prolonged operational exposure. Heavy-duty commercial vehicles operate under conditions substantially different from passenger vehicles, and validation activities must therefore reflect the operational realities of freight transportation environments [12].

Alert-fatigue assessment becomes especially important within AI-enabled HMI systems because excessive, repetitive, or misleading warnings may gradually reduce driver responsiveness and trust calibration over time. An HMI system that generates frequent nuisance alerts may unintentionally encourage drivers to ignore critical warnings during genuinely hazardous situations. Consequently, validation must assess not only technical warning delivery but also long-term behavioural effects associated with warning frequency, prioritisation strategy, and driver workload.

An important validation objective is demonstrating that deterministic safety-critical warnings remain active, perceivable, and operational even when AI functionality becomes unavailable, degraded, or uncertain. This requirement directly supports the bounded-AI-authority principle proposed throughout this paper, ensuring that AI behaviour cannot compromise fundamental safety communication pathways within the commercial vehicle safety architecture [1], [10].

10. Safety Case Structure

The proposed ASIL-B safety case for AI-integrated HMI systems is organised into three integrated categories of evidence. Unlike conventional automotive safety cases that focus primarily on deterministic software correctness and hardware reliability, AI-enabled systems require additional governance-oriented evidence associated with datasets, model behaviour, operational assumptions, robustness management, and lifecycle control. Consequently, the safety case becomes both a technical assurance argument and a systems-governance argument demonstrating that AI functionality remains appropriately bounded and operationally supervised throughout the vehicle lifecycle [1], [8].

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

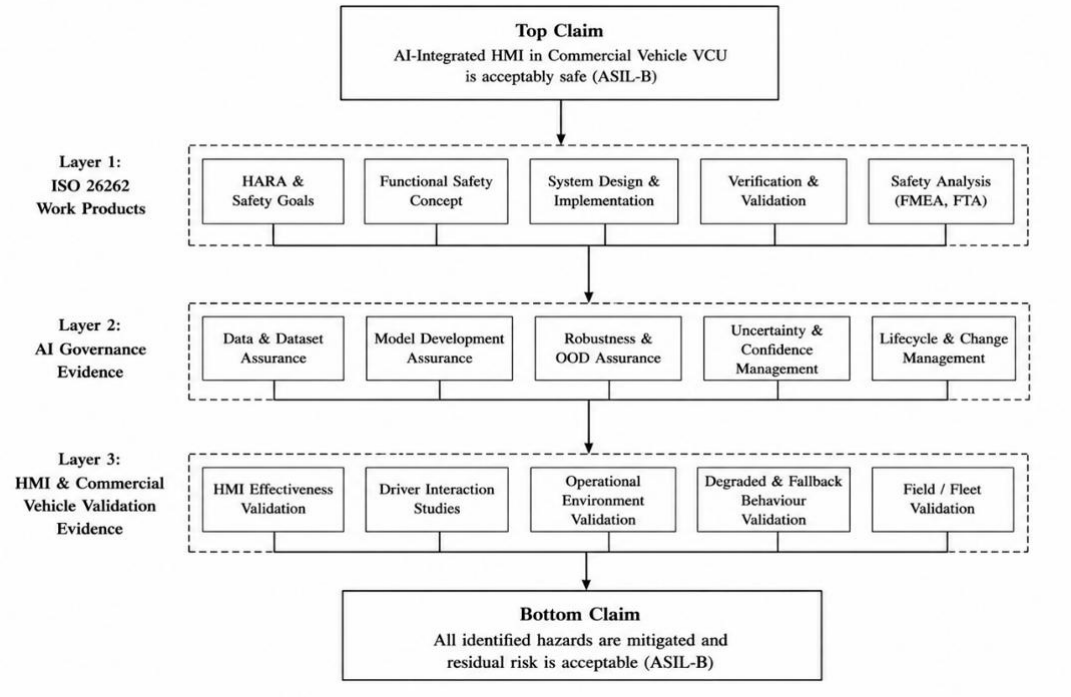

Figure 3: Safety Case Structure for AI-Integrated HMI

10.1 ISO 26262 Work Products

The first category consists of conventional ISO 26262 work products forming the core functional safety evidence package. This evidence includes the item definition, safety plan, hazard analysis and risk assessment (HARA), safety goals, functional safety concept, technical safety concept, architecture specifications, DFMEA, Fault Tree Analysis (FTA), dependent failure analysis, verification reports, validation reports, confirmation reviews, and safety assessments [1]. These work products collectively establish the deterministic functional safety foundation of the AI-integrated HMI system and demonstrate compliance with established automotive safety lifecycle processes.

Architecture specifications are particularly important because they define allocation of safety responsibilities between deterministic safety mechanisms and AI-enabled advisory functions. Verification and validation reports provide evidence that safety requirements have been correctly implemented and validated across hardware, software, communication, and integration layers. Dependent failure analysis further demonstrates that failures within AI-related components cannot propagate uncontrollably into deterministic safety pathways.

10.2 AI-Specific Evidence

The second category of the safety case includes AI-specific assurance evidence. Since AI system behaviour depends heavily on training conditions, operational assumptions, and dataset composition, the safety case must include documentation demonstrating appropriate lifecycle governance and behavioural assurance [3], [4].

Representative AI-related evidence includes data requirements documentation, dataset provenance records, coverage rationale for operational-domain representation, labelling-governance procedures, model architecture rationale, training-configuration records, robustness evaluation results, out-of-distribution testing evidence, confidence-calibration analysis, model version-management documentation, regression-testing evidence, and documented operational limitations.

Dataset provenance documentation is especially important because training data increasingly functions as a safety-relevant engineering artifact. The safety case must therefore demonstrate how training datasets were collected, validated, labelled, and managed throughout the lifecycle. Similarly, robustness evaluation evidence is necessary to show how the AI system behaves under degraded environmental conditions, uncertain operational states, and operational-domain boundary scenarios.

Documented operational limitations also become a critical safety-case element because AI systems cannot realistically guarantee safe performance under all possible real-world conditions. Instead, the safety argument must explicitly define the operational assumptions, validated conditions, and known limitations under which the AI-HMI functionality remains acceptable for deployment.

10.3 HMI and Commercial Vehicle Evidence

The third category includes HMI-specific and commercial-vehicle-specific evidence associated with driver interaction, operational usability, and safety communication effectiveness. Representative evidence includes alert-priority specifications, timing validation results, audio and visual warning verification, driver comprehension studies, mode-awareness evaluation, nuisance-alert analysis, degraded-mode testing, and confirmation that deterministic safety warnings cannot be suppressed by AI-generated outputs.

Timing validation is especially important within heavy-duty commercial vehicles because delayed warning presentation may significantly reduce controllability due to long braking distances and vehicle inertia. Driver comprehension studies additionally provide evidence that commercial drivers correctly interpret warning messages, operational-state information, and degraded-mode indicators under realistic driving conditions.

The overall safety argument must explicitly demonstrate that AI authority remains bounded, deterministic supervision remains continuously active, degraded operational behaviour is clearly defined, operational assumptions are documented, and residual risk is supported through both analytical and empirical evidence. In practice, the safety case must show not only that the system functions correctly under normal conditions, but also that the architecture remains operationally safe when AI uncertainty, communication faults, degraded sensor inputs, or unexpected operational conditions occur.

An important observation emerging from AI-enabled automotive systems is that safety cases increasingly resemble governance arguments rather than purely deterministic correctness arguments. Assurance depends not only on what the AI system currently does, but also on how the organisation governs datasets, model updates, operational assumptions, behavioural consistency, and long-term lifecycle control [4]–[6]. Consequently, AI-integrated functional safety assurance extends beyond traditional software verification into broader organisational governance and systems-engineering management practices.

11. Conclusion

The integration of AI-enabled HMI functionality into commercial vehicle VCUs represents an important transition in automotive systems engineering. AI systems provide benefits such as contextual awareness, adaptive information management, predictive diagnostics, and improved driver-support capability. However, these advantages also introduce new safety-assurance challenges that extend beyond conventional deterministic automotive software engineering.

The first conclusion of this paper is that ISO 26262 remains necessary but insufficient for AI-integrated HMI systems. While the standard provides essential lifecycle discipline and verification rigor, AI-enabled systems additionally require SOTIF-oriented performance-insufficiency analysis, AI

governance, dataset assurance, robustness evaluation, uncertainty management, and human-factors validation.

The second conclusion is that bounded AI authority is essential for practical automotive AI safety assurance. AI systems may support contextual interpretation, information prioritisation, adaptive presentation, and driver assistance, but deterministic safety mechanisms must continue to retain responsibility for critical warnings, fallback behaviour, unsafe mode prevention, and operational-state integrity. Where AI influences safety-relevant driver behaviour, safeguards such as plausibility checks, confidence thresholds, ODD constraints, runtime supervision, and regression-controlled updates become necessary.

The third conclusion is that ASIL-B is appropriate when AI-HMI systems remain advisory or supervisory and when controllability remains with the driver and independent deterministic vehicle systems. However, ASIL allocation must always result from explicit HARA analysis incorporating commercial vehicle operational assumptions such as vehicle mass, stopping distance, trailer dynamics, and prolonged duty cycles.

The primary contribution of this paper is the development of a practical systems-oriented ASIL-B compliance framework for AI-integrated HMI systems in heavy-duty commercial vehicle VCUs. The framework integrates ISO 26262, ISO 21448, and emerging AI governance guidance while addressing AI failure modes, bounded AI authority, commercial vehicle operational assumptions, and integrated assurance methodology within the 2024 automotive functional safety landscape.

References

[1] International Organization for Standardization, *ISO 26262: Road Vehicles – Functional Safety*, ISO, Geneva, Switzerland, 2018.

[2] International Organization for Standardization, *ISO 21448: Road Vehicles – Safety of the Intended Functionality (SOTIF)*, ISO, Geneva, Switzerland, 2022.

[3] ISO/IEC, *ISO/IEC TR 5469: Artificial Intelligence – Functional Safety and AI Systems*, ISO/IEC, Geneva, Switzerland, 2024.

[4] ISO/IEC, *ISO/IEC 23894: Information Technology – Artificial Intelligence – Guidance on Risk Management*, ISO/IEC, Geneva, Switzerland, 2023.

[5] ISO/IEC, *ISO/IEC 42001: Information Technology – Artificial Intelligence Management System*, ISO/IEC, Geneva, Switzerland, 2023.

[6] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, Gaithersburg, MD, USA, 2023.

[7] IEEE Standards Association, *IEEE Std 2846-2022: Assumptions for Models in Safety-Related Automated Driving Systems*, IEEE, New York, NY, USA, 2022.

[8] Underwriters Laboratories, *UL 4600: Standard for Safety for the Evaluation of Autonomous Products*, 2nd ed., UL Standards, Northbrook, IL, USA, 2022.

[9] M. A. Rahman, S. Azam, M. M. Hasan, and J. Jonkman, "A Systematic Approach to Enhancing ISO 26262 With Machine Learning-Specific Life Cycle Phases and Testing Methods," *IEEE Access*, vol. 12, pp. 45789–45815, 2024.

[10] Institution of Engineering and Technology, *The Application of Artificial Intelligence in Functional Safety*, IET, London, U.K., 2024.

[11] NXP Semiconductors, *Driving Functional Safety in ADAS and Autonomous Vehicles*, NXP White Paper, Eindhoven, Netherlands, 2024.

[12] National Highway Traffic Safety Administration and Federal Motor Carrier Safety Administration, "Heavy Vehicle Automatic Emergency Braking Systems," *Federal Register*, vol. 88, no. 128, pp. 43502–43615, 2023.