ASIL E — Extending Automotive Functional Safety for Driver-Out Autonomy

Jherrod Thomas
Certified Functional Safety Expert
Jherrod.Thomas@aol.com
The Lion of Functional Safety™

***Abstract*—This article addressed the absence of human controllability in highly automated driving by proposing an extension to automotive functional safety that introduced a driver-out controllability class and required a uniform, one-level escalation of integrity targets. The aim was to establish a coherent basis for demonstrating safety without a human fallback by aligning hazard analysis, verifiable evidence, and lifecycle governance within a single framework. The study employed an extended hazard analysis and risk assessment that included an explicit driver-out decision with a corresponding escalation rule. It derived obligations for an integrity tier beyond current practice and integrated Safety of the Intended Functionality and the Underwriters Laboratories 4600 safety-case framework. Mandatory analyses comprised System-Theoretic Process Analysis for control-structure hazards, systematic identification of triggering conditions that degrade nominal performance, and construction of a structured safety case with traceable evidence. The approach was illustrated through a worked example on nighttime pedestrian non-detection to show requirement flow-down and a verification and validation plan. Results indicated that the driver-out classification elevated all hazards by one integrity level and produced an obligation set that exceeded prior thresholds. The framework specified higher diagnostic-coverage targets, architectural redundancy with fail-operational behavior, stricter latency and availability requirements, runtime monitoring with minimal-risk transitions, and post-deployment governance using telemetry, drift detection, incident response, and gated software updates. An evaluation workflow connected claims to evidence across development, testing, and operation, and the case study demonstrated measurable performance targets and auditable traceability. The proposed extension offered a transparent and reviewable route to establish acceptable safety for driver-out operation, while maintaining compatibility with established practice and enabling continuous assurance in service.**

***Index Terms*—Automotive Safety Integrity Level (ASIL E), ISO 21448 (SOTIF), driver-out autonomy (C4), hazard analysis and risk assessment (HARA), System-Theoretic Process Analysis (STPA), Safety of the Intended Functionality, Underwriters Laboratories UL 4600, verification and validation, runtime assurance, safety case, autonomous vehicles**

Manuscript received September 10, 2025; revised September 14, 2025.

I. Introduction

AUTONOMOUS driving at SAE Levels 4 and 5 removes the human driver from the operational loop, which unsettles the long-standing safety assumptions embedded in ISO 26262. In particular, hazard analysis and risk assessment have historically relied on human controllability to moderate risk and bound failure consequences. This paper addresses that gap by formalizing driver-out controllability as C4, applying a uniform one-level escalation of integrity targets, and defining ASIL E obligations for systems that must assure safety without human fallback. The approach integrates ISO 21448 for functional insufficiencies and UL 4600 for system-level safety cases, yielding a coherent framework that aligns process rigor, verification evidence, and runtime assurance for fully autonomous vehicles.

A. Context and Motivation

1) Historical foundation in ISO 26262: ISO 26262 has functioned as the definitive framework for automotive functional safety since its publication. It addresses electrical and electronic systems in road vehicles under 3000 kg and specifies Automotive Safety Integrity Levels (ASIL A to D), with ASIL D denoting the highest rigor. ASIL classification depends on three parameters: Severity (S), Exposure (E), and Controllability (C). The framework presumes that human controllability is a central safety premise [1]–[4].

The standard operationalizes safety through hazard analysis and risk assessment (HARA), which identifies vehicle-level hazards and assigns an appropriate ASIL. Each integrity level mandates corresponding development workflows, verification strategies, and safety mechanisms. Systems at ASIL C and D typically require semi-formal or formal verification to demonstrate compliance [1], [2], [5], [6].

2) Dependence of ASIL A to D on human controllability: Conventional ASIL determination assumes that a human driver remains available to intervene during faults. Controllability evaluates the driver's capacity to

avoid specified harm by reacting in time and with appropriate maneuvers. This assumption has been valid for conventional vehicles and for driver assistance at SAE Levels 0 to 2, where the human remains responsible for the dynamic driving task [7]–[11].

3) SAE Level 4 and Level 5 remove the human from the loop: Automation at SAE Levels 4 and 5 challenges the prior assumption set. At these levels, normal operation does not rely on human supervision. Level 4 systems function independently within a defined Operational Design Domain (ODD), whereas Level 5 extends that capability to all conditions [7]–[12].

This change is not merely incremental. It realigns safety responsibility from human oversight to technical means. Without a human fallback, the risk profile shifts, and the sufficiency of ISO 26262 processes alone becomes uncertain, prompting the need for augmented safety arguments and assurance evidence [4], [7], [13], [14].

B. Problem Statement

1) Loss of controllability in driver-out operation: Traditional controllability assumptions collapse in SAE Level 4 and Level 5 settings. In conventional analysis, controllability is central to ASIL assignment because a competent driver is expected to intervene and mitigate hazardous events. In driver-out operation, this human fallback is absent, so the assumed protective layer no longer exists [1], [4], [7], [8], [13].

2) Limits of ISO 26262 C0–C3 without a human driver: When no driver is present, the C0–C3 taxonomy cannot represent the true residual risk. The framework does not fully address cases in which:

- no human operator is available for immediate intervention [8], [13].
- failures must be contained entirely by automated safety mechanisms [15], [16].
- the vehicle must maintain safe operation or reach a safe state without assistance [6], [12].

These conditions exceed the intent of the current controllability classes and reveal a gap for automated systems [13].

3) Proposed extension: C4 = Driver-Out: To close this gap, define C4 as Driver-Out. C4 applies when human controllability is fundamentally unavailable and safety depends wholly on automated functions and fail-safe strategies [7], [13] C4 acknowledges that driver-out operation presents a qualitatively different risk profile that existing classes do not capture. It implies elevated assurance needs, stronger evidence, and augmented verification to compensate for the absence of human oversight [12]–[14], [17].

C. Thesis and Contribution

1) C4-augmented HARA with uniform +1 ASIL escalation: This work extends HARA to cover driver-out operation by introducing a C4 controllability class. C4 denotes scenarios in which human intervention is unavailable and safety relies entirely on automated detection, decision, and actuation. Building on this foundation, a uniform +1 ASIL escalation rule is proposed. For any hazard previously assessed under C0–C3, the absence of human controllability triggers a one-level increase in the target integrity. This policy yields consistent treatment of driver-out risks and removes ambiguity in allocation of safety requirements.

2) ASIL E obligations derived from escalation: Applying uniform +1 escalation to ASIL D produces a new obligation tier, ASIL E. This is the first formal extension beyond the A–D range and is tailored to the conditions faced by fully autonomous systems. ASIL E captures the elevated assurance burden that arises when human oversight is removed [7], [13].

- higher diagnostic coverage targets that exceed ASIL D thresholds [3], [18].
- reinforced redundancy with fail-operational behavior for critical paths [15], [19], [20].
- stricter performance and availability requirements for safety functions [13], [21].

3) Integration with ISO 21448 SOTIF and UL 4600: The ASIL E scheme is coupled with complementary standards to close gaps that functional safety alone does not address. ISO 21448 (SOTIF) covers hazards from specification weakness and performance limits rather than random hardware faults [4], [12], [22]. UL 4600 supplies a system-level safety case framework with criteria for validation of autonomous products [4], [7], [13]. The combined approach addresses:

- extended functional safety under ISO 26262 at ASIL E
- insufficiencies and performance limits under ISO 21448 SOTIF
- end-to-end safety argumentation and validation under UL 4600

This integration ensures coverage from component behavior to system evidence, which is essential when driver-out operation removes the human fallback [4], [7], [13].

4) Significance and expected impact: The proposed methodology establishes both a theoretical basis and a practical workflow for safety in fully automated driving. It codifies C4 within HARA, creates a consistent escalation rule, defines ASIL E obligations, and unifies them with SOTIF and UL 4600. Together these elements enable assurance commensurate with public-road

deployment while supporting regulatory compliance and transparent safety cases [7], [13], [14].

II. Extending HARA with C4

A. Baseline HARA Recap

1) Core parameters: S0–S3, E0–E4, C0–C3: ISO 26262 characterises risk using three parameters that jointly determine the Automotive Safety Integrity Level (ASIL): severity S, exposure E, and controllability C [23]–[25].

Severity (S). The scale ranges from S0 to S3 [24], [25]. S0 denotes absence of injury. S1 represents light to moderate injury that does not threaten life or cause permanent disability. S2 covers severe or potentially life-threatening injury with possible permanent impairment. S3 represents life-threatening or fatal outcomes with high risk of death.

Exposure (E). The frequency of encountering the hazardous scenario is graded E0 to E4 [23], [24]. E0 indicates very low probability during operation. E1 is low, E2 is medium, and E3 is high probability under typical driving. E4 denotes very high probability, often present in common operating conditions.

Controllability (C). Driver ability to avert harm is classified C0 to C3 [24], [25]. C0 means generally controllable, with more than 99% of drivers able to avoid harm. C1 is simply controllable, with more than 90% able to avoid harm. C2 is normally controllable, with more than 60% able to avoid harm. C3 is difficult or not controllable, with fewer than 60% able to avoid harm.

2) Baseline ASIL determination with C capped at 3: ASIL is assigned by intersecting S, E, and C in the standard determination matrix [23]–[25]. The resulting levels progress from QM to ASIL A, ASIL B, ASIL C, and ASIL D, reflecting increasing rigour in safety requirements [24], [26], [27]. QM applies when only quality management is warranted [24], [26]. ASIL A introduces basic functional-safety measures [24], [25]. ASIL B elevates verification and confirmation activities [24], [25]. ASIL C requires high assurance techniques, including more formal verification where applicable [24], [27]. ASIL D imposes the most stringent lifecycle controls and evidence expectations [24], [25], [27].

In the conventional framework, controllability is bounded at C3, which models the practical limit of human intervention in hazardous events. The cap embodies the assumption that some driver action remains possible, however often insufficient, in the most challenging conditions [24], [25].

B. C4 Definition and Escalation Rule

1) C4 = Driver Out: C4 designates situations with no human controllability. It targets SAE Level 4 and Level 5 operation, where a human is neither available nor expected to intervene in hazardous events [4], [28]–[31]. C4 is qualitatively distinct from C0–C3 because it captures zero human control rather than reduced capability. In C4 conditions, the share of drivers able to avoid harm is 0%, since no driver is present or engaged in the dynamic driving task [28]–[31].

2) Uniform +1 ASIL escalation: The uniform +1 escalation addresses the heightened risk introduced by driver-out operation. Removing human fallback increases the assurance burden on the automated stack, so each baseline ASIL advances by one level for hazards evaluated under C4 [28], [31].

Stepwise rationale. When a baseline analysis yields QM, C4 elevates the target to ASIL A, replacing quality-only controls with formal safety requirements and verification [24], [25]. If the baseline is ASIL A, the target becomes ASIL B, requiring stronger verification and architectural safety mechanisms [24], [25]. A baseline ASIL B advances to ASIL C, which typically calls for higher diagnostic coverage and, where applicable, formal methods [24], [27]. A baseline ASIL C advances to ASIL D, triggering the most stringent lifecycle evidence, fault detection, and fault handling measures recognized in the traditional framework [24], [27]. Finally, a baseline ASIL D advances to ASIL E, a level beyond the conventional matrix introduced to reflect driver-out risk [28], [31].

Consistency principle. The uniform shift preserves the relative risk ordering established by ISO 26262 while aligning the target integrity with the absence of human backup [28], [30].

C. Effects and Workflow Trigger

1) Universal escalation under autonomy: Applying C4 systematically elevates every identified hazard by one ASIL when driver-out conditions hold. This policy recognises that the removal of human oversight increases the criticality of all safety-relevant scenarios across the full range of S and E combinations [28]–[31]. The uniform approach avoids selective application that could create coverage gaps and preserves consistent treatment of hazards within the autonomous-vehicle risk space [28], [31].

2) From ASIL D to ASIL E: The shift from ASIL D to ASIL E is the principal consequence of the C4 extension, since it introduces requirements that exceed the traditional ISO 26262 matrix. ASIL E typically demands [28], [31]:

- diagnostic coverage beyond ASIL D targets, potentially exceeding 99% [25], [32];
- redundancy with fail-operational behaviour at the architectural level [19], [20];
- a safety case aligned with SOTIF for performance-limit and specification risks, and with UL 4600 for

systematic argumentation and validation planning [4], [28], [33], [34].

3) Activation of the driver-out workflow: The presence of C4 triggers a dedicated development and validation workflow tailored to autonomy [28], [29], [31].

Enhanced HARA. The analysis incorporates SOTIF-oriented hazard identification [4], [33], explicitly accounts for the removal of the human–machine interface as a control path [35], [36], and evaluates sensor limits and environmental boundaries within the operational context [37], [38].

Extended safety case. The argument structure follows UL 4600 principles [28], [34], includes quantitative validation targets where appropriate [39], [40], and defines as well as monitors the operational design domain with clear boundary conditions [41], [42].

Verification and validation. The programme prioritises scenario-based testing that targets edge cases and boundary conditions [43], [44], complements this with simulation at stated statistical confidence levels [9], [45], and supports it with field operational testing that enables comprehensive data capture and analysis [46], [47].

D. Non-C4 Branch (Explicit)

1) Decision rule: **If C4 does not apply, proceed with the normal ISO 26262 workflow.** The proposed scheme preserves full compatibility with established functional safety practice. For any hazard where driver-out conditions are not present, analysis and development continue under the unmodified ISO 26262 processes [4], [23]–[25].

2) Rationale and scope: Backward compatibility. Existing systems and organisational procedures remain valid. Teams may retain proven methods for conventional vehicles while invoking the enhanced requirements only where driver-out operation is in scope [24], [26].

Scope definition. The branch clearly marks the boundary between conventional and autonomous safety obligations. It prevents the unnecessary application of ASIL E measures when effective human controllability is available [4], [41].

3) Resource and regulatory alignment: Resource optimisation. Engineering effort and assurance evidence concentrate on genuine driver-out scenarios. Conventional systems follow the standard lifecycle with no additional burden [23], [26].

Regulatory alignment. The branch maintains conformance with current automotive safety regulations and standards while still providing a structured pathway for autonomous deployments [4], [24].

4) Process selection and auditability: The non-C4 branch corresponds to the No path in formal assessment flowcharts. This explicit decision point yields unambiguous process selection and simplifies audits, implementation planning, and regulatory review of the extended methodology [48], [49].

E. Supporting Artifact

1) Proposed ASIL Matrix: Driver Present vs. Driver Out: Figure 1 presents the proposed ASIL matrix, rendered in tabular form yet numbered and cited as a figure. It extends the ISO 26262 matrix by adding C4 and applies a uniform +1 escalation across all severity–exposure combinations. The figure functions as the central artifact for driver-out safety assessment by providing explicit targets for safety engineers and assessors [28], [31].

2) Escalation pattern and assurance principle: Under C4 conditions, traditional designations QM and ASIL A–D escalate to ASIL A–E respectively, so that no safety-critical case remains below ASIL A in autonomous applications. This mapping operationalises the core principle that fully automated systems require stronger safety assurance than human-supervised systems [28]–[31].

3) Traceability and use in practice: The artifact preserves traceability to ISO 26262 by retaining the established structure while adding the C4 branch. Organisations can apply the table to evaluate autonomous-vehicle hazards consistently and to derive matching requirements without ambiguity or uneven treatment across scenarios [23], [24], [28], [31].

III. ASIL E Obligations

Attaining ASIL E for driver-out autonomous operation requires extending classical functional safety into a coherent system-level assurance program. The obligation set includes construction of a structured safety case, alignment with complementary safety standards, completion of mandatory analyses, and fulfillment of an enlarged requirement portfolio that covers runtime assurance and post-deployment controls.

A. Safety Case

An ASIL E safety case is a structured and auditable argument that the autonomous system is acceptably safe for its intended functions without human supervision. The argument is organized into clearly connected claims, supporting rationale, and corroborating evidence to maintain end-to-end traceability from high-level goals to concrete artifacts and test outcomes.

- **Claims.** Top-level safety objectives state, for example, that the ADS maintains a safe state across all ODD conditions without driver intervention [28].

Fig. 1: Proposed ASIL Table with C4: Driver Present vs Driver Out

- **Arguments.** The justification explains how architecture, components, and processes together meet these objectives, typically structured with Goal Structuring Notation or an equivalent framework [34].
- **Evidence.** Corroborating materials include design specifications, verification reports, test and simulation outputs, and field operational test records that substantiate each argument element [28].

The safety case must be reviewable by regulators and independent assessors. Every safety requirement is linked to the design elements and validation activities that satisfy it, recorded in a traceability matrix. This matrix demonstrates that each hazard identified by HARA, including C4 driver-out scenarios, is addressed by specific system mechanisms and verification methods, thereby enabling transparent and efficient safety assurance reviews [34].

B. Systematic Safety Standards to Integrate

ASIL E requires the coordinated application of ISO 21448 (SOTIF) and UL 4600 alongside ISO 26262. Together they address hazards from functional insufficiencies in nominal operation and impose lifecycle governance for deployed autonomy. The obligations below summarize what must be demonstrated for compliance at ASIL E.

1) ISO 21448 SOTIF: ISO 21448 addresses hazards that arise without hardware faults by focusing on functional limitations, performance shortfalls, and context-specific degradations in nominal operation [4]. The ASIL E obligations are:

- **Identify functional insufficiencies.** Determine how capability gaps can induce hazardous scenarios in ordinary use, for example perception ambiguities under adverse weather or challenging illumination [4], [50].
- **Analyze triggering conditions.** Define environmental and operational boundaries that precipitate

TABLE I: ASIL E Mandatory Analyses

Analysis Method	Scope/Purpose	Key Outputs	Evidence Deliverables	Safety Case Integration
System Theoretic Process Analysis (STPA)	Identify unsafe control actions & systemic hazards across ADS control structure	• Control action catalog • Unsafe control variations • Causal factor analysis • Safety constraints	• STPA worksheets • Control structure diagrams • Hazard analysis reports • Safety constraint specifications	System-level safety claims supported by STPA-derived constraints & verification evidence
SOTIF Analysis (ISO 21448)	Address functional insufficiencies & triggering conditions in nominal operation within ODD	• Triggering Condition ID (TCI) • Performance limitation catalog • ODD boundary definition • Coverage targets	• TCI databases • Scenario generation reports • Simulation test results • Field operational test data	Performance envelope claims backed by statistical validation & edge-case coverage evidence
UL 4600 Safety Case Framework	Establish structured safety argumentation & lifecycle governance for autonomous products	• Safety case skeleton (GSN) • Evidence pack definitions • OTA governance procedures • Lifecycle obligations	• Goal Structuring Notation diagrams • Evidence catalogs • OTA validation reports • Incident response records	Top-level safety argumentation structure linking all claims to organized evidence packs

degradation, such as heavy rain or tunnel entry, and show how these conditions are detected and managed [37].

- **Verify ODD coverage.** Demonstrate that the ODD is complete and that the ADS achieves adequate performance across it using simulation, track testing, and field operational tests (FOTs) [4], [42].

Lifecycle implication. Integration into ASIL E requires both proactive discovery of unsafe scenarios through systematic generation and edge-case exploration, and reactive monitoring in operation to detect perception or decision insufficiencies as they emerge [33], [51].

2) UL 4600: UL 4600 provides a safety-case framework tailored to autonomous products, emphasizing continuous validation, monitoring, and governance for systems in service [28]. The ASIL E obligations are:

- **Autonomy safety assurance.** Show robust ADS performance across all defined ODDs using diverse methods, including scenario-based testing, formal verification where applicable, and statistical assessment of edge-case coverage [28], [39].
- **Runtime monitoring with OTA governance.** Implement mechanisms to detect deviations from expected behavior during operation, including sensor degradation and model drift, and manage over-the-air updates to preserve or improve safety performance [28], [52].
- **Safety argument maintenance.** Keep the safety case current as field evidence, incidents, and software changes accrue, and ensure that OTA processes are vetted to prevent safety regressions [28].

This integration keeps the ASIL E safety case auditable and current across the vehicle's operational life by linking pre-deployment validation with post-deployment maintenance and oversight.

C. Mandatory Analyses

ASIL E requires a disciplined suite of analyses that expose system hazards, quantify performance limits, and bind evidence to safety claims across the lifecycle. The core activities are STPA for control-structure hazards, SOTIF-triggering condition identification for functional insufficiencies in nominal operation, and UL 4600 safety-case structuring to organize claims and evidence. Table I delineates the scope and deliverables of all mandated analyses for ASIL E systems. It shows how hazard identification, mitigation of functional insufficiencies, and lifecycle governance evidence are consolidated into a coherent safety case that supports compliance and enables continuous safety assurance.

1) System-Theoretic Process Analysis (STPA): STPA models the ADS as a hierarchical control structure that spans perception, planning, and actuation, then evaluates how control flaws can produce hazardous outcomes [51]. The ASIL E expectations are:

- **Unsafe control actions.** Enumerate control actions such as apply braking or maintain lane and characterize unsafe variations including braking too late or unjustified lateral maneuvers.

- **Causal analysis.** Trace causal factors across software, hardware, environment, and organizational processes, including sensor fusion failures and erroneous situation assessment [51].
- **Constraints and requirements.** Derive explicit safety constraints and convert them into requirements linked to design and verification plans, ensuring prevention or timely detection of each unsafe control action [51].

2) SOTIF Triggering Condition Identification: Building on ISO 21448, ASIL E obliges a rigorous program to identify Triggering Conditions for Insufficiencies that degrade nominal performance without faults [37]. The required outcomes are:

- **Scenario coverage.** Maintain comprehensive scenario sets that represent environmental variation, road surface states, and traffic interactions known to trigger performance limits [37].
- **Statistical targets and verification.** Set quantitative coverage targets, for example the 95th percentile of nighttime fog, and verify by a mix of simulation and real-world testing that performance remains within safe margins under these conditions [4], [33].

3) UL 4600 Safety-Case Skeleton and Evidence Packs: UL 4600 structures the safety argument for autonomous products so that claims, rationale, and corroborating evidence are coherent and auditable across deployment [28], [34]. ASIL E obligations include:

- **Claims and goals.** Define top-level claims and safety goals aligned with HARA outcomes and the ASIL E requirement set.
- **Evidence packs.** Organize diverse evidence types into curated packs for design, verification, validation, and operational monitoring, combining independent sources such as simulation logs, test reports, FOT data, and incident investigations [28], [34].

The resulting artifacts enforce consistency and completeness of justification, simplify regulatory review, and support continuous assessment as software and field evidence evolve.

D. Expanded Requirement Set

Meeting ASIL E requires extending the requirement stack beyond ISO 26262 Functional Safety Requirements (FSRs) and Technical Safety Requirements (TSRs) to include SOTIF obligations, runtime assurance, and in-service controls.

1) Retain FSR/TSR: Classical FSRs and TSRs remain the basis for controlling random hardware failures and systematic faults [24]. Examples include "the braking system shall detect wheel slip within 10 ms" and "system uptime greater than 99.9%." These requirements preserve continuity with established safety engineering practice while anchoring higher-layer obligations.

2) Add SOTIF Requirements: SOTIF supplements ISO 26262 by addressing hazards from functional insufficiencies and performance limits in nominal operation.

- **Performance envelope.** Specify minimum perception accuracy, path-planning reliability, and control latency guarantees across the defined ODDs, with verification strategies tied to each metric [42], [50].
- **Triggering-condition mitigation.** Incorporate design features, for example LiDAR redundancy or infrared sensing, and define fallbacks such as safe-stop maneuvers to manage identified limitations [4], [33].

3) Runtime Assurance (RTA): RTA continuously evaluates operational safety and enforces transitions to predefined safe states when confidence degrades [20], [29].

- **Online monitors.** Track real-time indicators such as perception confidence, model-uncertainty estimates, and sensor-health metrics [20], [29].
- **Confidence estimation.** Fuse indicators using Bayesian or Dempster-Shafer methods to compute a safety confidence metric that triggers mitigations when thresholds are crossed [20].
- **Minimal Risk Condition.** Execute autonomous disengagement to a Minimal Risk Condition, for example a controlled pull-over, when degradations are unrecoverable [19], [20].

4) Post-Deployment Requirements: In-service controls keep the safety case current throughout the vehicle lifecycle under UL 4600 governance [28], [34].

- **Telemetry and drift detection.** Collect operational data, including near-misses, RTA activations, and environmental factors. Apply drift-detection algorithms to identify model performance decay from distribution shift or sensor wear, prompting recalibration or retraining [28].
- **Incident reporting.** Maintain workflows that capture, analyze, and categorize field incidents and anomalies, and feed these outcomes into safety-case updates and product improvements [28], [52].
- **OTA governance.** Enforce verification and validation gates for software updates, with rollback mechanisms that prevent unsafe deployments and preserve prior safety performance [28], [34].

Table II presents the complete requirement set and specifies how each category supports comprehensive safety assurance.

Together, these obligations create an end-to-end assurance program that spans initial hazard analysis, structured safety cases, complementary standards, mandatory

TABLE II: Consolidated list of FSR, TSR, SOTIF, Runtime Assurance, and Post-Deployment requirements introduced by ASIL E

Requirement Category	Traditional ISO 26262	ASIL E Extensions/Additions	Key Performance Indicators
Functional Safety Requirements (FSR)	• Hardware fault detection • Software systematic fault prevention • Diagnostic coverage $\geq$90% (ASIL D)	• Enhanced diagnostic coverage $\geq$99% • Fail-operational architectures • Advanced redundancy mechanisms • Cross-domain monitoring	• Diagnostic coverage >99% • MTBF >10^9 hours • Fault detection time <50ms
Technical Safety Requirements (TSR)	• System availability targets • Performance specifications • Interface definitions	• Higher availability requirements • Stricter latency constraints • Enhanced cybersecurity measures • Multi-modal sensor fusion	• System uptime >99.99% • End-to-end latency <100ms • Sensor fusion confidence >95%
SOTIF Requirements	Not explicitly covered in ISO 26262	• Performance envelope definition • Triggering condition mitigation • ODD boundary monitoring • Specification completeness validation	• TCI coverage >95% • Performance degradation <5% • ODD violation detection <200ms
Runtime Assurance (RTA)	Limited runtime monitoring	• Continuous confidence estimation • Online safety monitors • Automated MRC transitions • Performance degradation detection	• Monitor update rate $\geq$50Hz • MRC activation time <2s • False positive rate <1%
Post-Deployment Requirements	Basic field monitoring	• Comprehensive telemetry collection • Statistical drift detection • Incident reporting workflows • OTA update validation gates	• Telemetry coverage >99% • Drift detection within 30 days • Incident response <48 hours

analyses, and lifecycle controls. The result is demonstrably safe driver-out autonomy without reliance on human intervention.

IV. Evaluation Process Aligned with the Safety Workflow

The ASIL E evaluation augments the ISO 26262 lifecycle with measures specific to driver-out autonomy. Across these stages, established functional-safety practices are combined with autonomy-focused analyses to ensure comprehensive hazard coverage and a defensible safety argument. The complete step-by-step process is illustrated in Figure 2, which highlights how each analysis block feeds into subsequent requirement derivation and V&V planning.

A. Mode and ODD Declaration

This stage specifies the operational modes and defines the Operational Design Domain (ODD) in which driver-out functionality (C4) applies. The ODD states environmental, geographic, and system-state limits, for example: urban roads at $\leq$ 50 km/h, daylight conditions, and absence of construction zones. Modes that permit driver-out operation are explicitly tagged C4 to activate the extended workflow, whereas other modes (for example, Manual Mode) remain governed by C0–C3 controllability classifications. A precise ODD boundary anchors hazard analysis and verification activities and clarifies the assurance claims for driver-out use [4], [9], [31], [41], [42].

B. HARA

With modes and ODD fixed, Hazard Analysis and Risk Assessment (HARA) is performed for behaviors within each mode. Using ISO 26262 parameters for severity (S0–S3), exposure (E0–E4), and controllability (C0–C3), each hazard receives a baseline ASIL. Controllability is conservatively capped at C3 at this stage to reflect the assumption of potential human intervention. The output is a traceable hazard register annotated with baseline ASIL levels (QM–D), which becomes the input to the subsequent C4 decision [23]–[25].

C. C4 Decision and Escalation

For hazards associated with a driver-out mode, apply a uniform one-level escalation to the baseline ASIL: QM→A; A→B; B→C; C→D; D→E. Thus, a baseline of D becomes ASIL E, while other levels rise by one step. Hazards outside driver-out modes follow the standard ISO 26262 path without change. This gate ensures elevated assurance whenever human controllability is absent [28], [31].

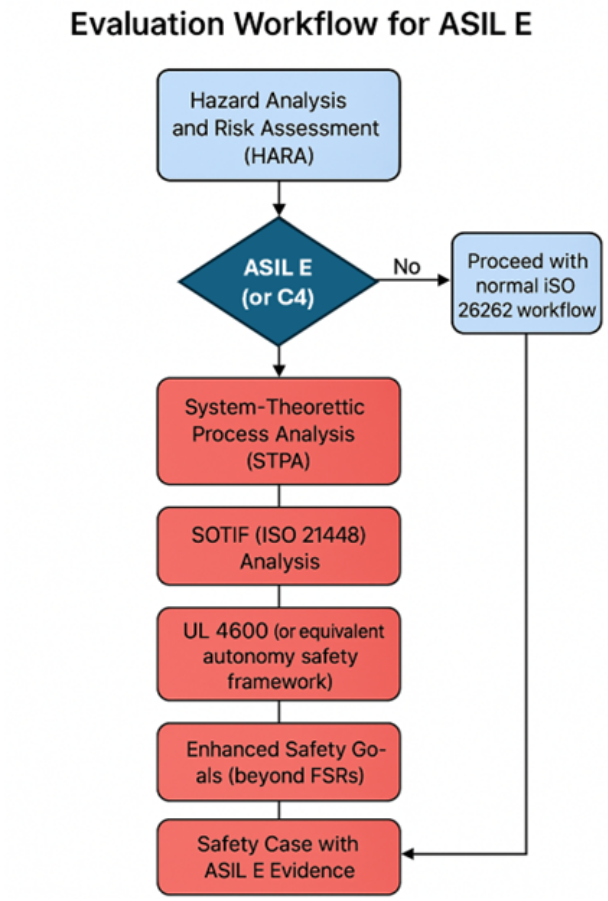

Fig. 2: Evaluation workflow for ASIL E, from Mode & ODD declaration through Operational Monitoring and Learning Loop

D. STPA

Following escalation, System-Theoretic Process Analysis (STPA) is executed to identify system-level hazards and unsafe control actions across the perception, planning, and actuation loops. The analysis models hierarchical control structures, enumerates control actions, and examines unsafe variants, for example, applying the brake too late or initiating an unintended lateral maneuver. The resulting safety constraints and requirements target interaction and emergent risks that traditional FMEA or FTA may miss, establishing the first critical analysis stage within the ASIL E flow [51].

E. SOTIF Analysis

The workflow next applies SOTIF (ISO 21448) to address functional insufficiencies and their triggers under nominal operation. The analysis proceeds through three focused activities:

- **Triggering conditions for insufficiencies.** Enumerate environmental and operational boundaries that can degrade perception, decision making, or control, for example low light, heavy rain, or road debris [22], [37].
- **Functional insufficiency assessment.** For each trigger, evaluate sensor limits, algorithmic bounds, and specification gaps, then derive SOTIF requirements, for example perception confidence $\geq 90\%$ under ECE R 115 rain levels [4], [33].
- **ODD mapping and coverage verification.** Confirm that all triggers lie within the declared ODD, and show sufficient performance through simulation, closed-track testing, and field operational tests [4], [42].

SOTIF expands coverage beyond random faults so that nominal performance limits receive equivalent safety attention in driver-out operation [4], [51]. Figure 3 maps key TCIs to ODD constraints and specifies the SOTIF-derived mitigation measures for low-light and adverse-weather scenarios.

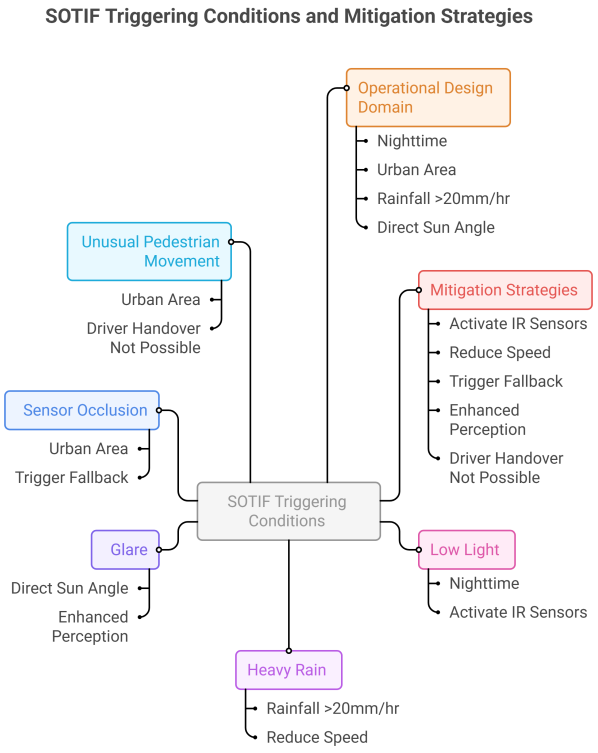

Fig. 3: Mapping of environmental and operational triggering conditions (TCIs) to ODD parameters and associated mitigation strategies

F. UL 4600 Alignment

UL 4600 supplies the safety-case structure and lifecycle obligations for autonomous products. The ASIL E

flow establishes:

- **Safety-case skeleton.** Formulate top-level claims from HARA, STPA, and SOTIF, and structure them in Goal Structuring Notation with explicit claim-argument-evidence links [28], [34].
- **Evidence pack definition.** Classify required evidence for each claim, including design records, verification reports, simulation logs, field-test data, and incident analyses [28], [34].
- **Lifecycle governance.** Define processes for over-the-air update governance, incident reporting, and continuous safety-case updates that incorporate field evidence and software changes [28], [52].

Conformance to UL 4600 keeps the safety case auditable and current from pre-deployment through in-service monitoring [28], [34].

G. Enhanced Safety Goals Beyond FSRs

After completing STPA, SOTIF, and UL 4600 alignment, revisit and extend the Safety Goals so that system-level hazards and insufficiencies inform the top tier:

- **SOTIF sub-goals.** Specify performance envelopes for perception, planning, and control under triggering conditions, for example lane-keeping error $\leq$ 0.2 m in low-sun conditions [22], [42].
- **Runtime assurance objectives.** Set online monitoring targets and safe disengagement criteria, for example initiate MRC within 2 s if perception confidence <70% [20], [29].
- **Operational monitoring goals.** Mandate telemetry collection and drift detection to sustain the safety case over time [28].

H. Requirements Derivation

The enhanced Safety Goals drive an integrated requirement set:

- **Design-time requirements.** Combine FSR and TSR for systematic faults and random hardware failures, for example diagnostic coverage > 90%, with SOTIF requirements that address functional insufficiencies and boundary conditions.
- **RTA requirements.** Define online monitors, confidence estimation methods, and MRC procedures that enforce the runtime objectives [20], [29].
- **Post-deployment obligations.** Specify telemetry, drift-detection triggers, incident reporting workflows, and OTA update gates to maintain safety claims in service [28], [34].

Each requirement traces to specific Safety Goals, ensuring full coverage of identified hazards and enabling structured V&V planning [24], [25], [34].

I. V&V Plan

The verification and validation plan for ASIL E must show that every requirement is satisfied across representative and challenging conditions.

- **Scenario and adversarial coverage testing.** Exercise the system in simulation and physical trials across all declared ODD conditions, including edge cases such as sensor spoofing and sudden obstacle appearance [39], [43], [53].
- **MRC demonstrations.** Validate end-to-end safe disengagement under degraded conditions detected by runtime monitors, including timing of MRC initiation and completion [19], [20].
- **OTA release gates.** Apply formal criteria for software approval, including regression testing, safety-case revalidation, and rollback preparedness to prevent unsafe deployment [28], [34].

The plan integrates evidence from simulation, closed-track experiments, laboratory analyses, and field operations to build a coherent, multi-modal body of proof.

J. Safety-Case Assembly and Independence

The final safety case consolidates claims, arguments, evidence, and explicit assumptions in a format consistent with UL 4600 and ASIL E.

- **Structured traceability.** Link each claim to its supporting evidence pack and to the underlying requirements to enable end-to-end argumentation [28], [34].
- **Assumption management.** Record system assumptions, for example calibration bounds or latency limits, together with the validation or monitoring measures that justify them [28], [34].
- **Independence requirements.** Meet ASIL E expectations for independent assessment, including at least one third-party review without development involvement and organizational separation among development, test, and safety-assessment roles, exceeding ASIL D practice [24], [25].

This governance improves objectivity and reduces bias in the evaluation of safety evidence.

K. Operational Monitoring and Learning Loop

After deployment, continuous assurance depends on systematic observation and rapid update cycles.

- **Telemetry collection.** Log runtime-assurance metrics, perception confidence, environmental conditions, and MRC triggers to support trend analysis and targeted investigation [28].
- **Drift detection.** Use automated methods to identify statistical shifts in model performance or sensor

behavior and trigger retraining, recalibration, or maintenance as needed [28].
- **Incident response and OTA governance.** For any incident or near miss, conduct root-cause analysis, update the safety case, and adjust requirements or safety goals as required; release validated improvements through controlled OTA pipelines and verify post-deployment effects [28], [34], [52].

This closed-loop operation keeps the ASIL E safety case current and responsive to real-world data, supporting a resilient lifecycle that adapts to evolving conditions and observed performance. Across these eleven stages, the workflow extends automotive functional safety to driver-out autonomy by integrating rigorous analysis, structured safety cases, and continuous operational oversight to achieve demonstrably safe fully autonomous operation.

V. Worked Example: Pedestrian Non-Detection at Night

This example applies the ASIL E methodology to a driver-out hazard in which the autonomous driving system (ADS) fails to detect a pedestrian under low-illumination nighttime conditions. The analysis covers HARA with C4 escalation, safety goals and requirements, and the Verification and Validation plan with links to the safety case.

A. HARA and C4 Escalation

Hazard definition. The ADS does not detect a pedestrian at night, creating collision risk. Baseline HARA (capped at C3).

- **Severity (S):** S3. Potential for life-threatening or fatal injury if a collision occurs [24] [25].
- **Exposure (E):** E2. Moderate likelihood of encountering pedestrians on urban roads at night [23].
- **Controllability (C):** C3. Human drivers would struggle to avoid harm given limited visibility and reaction time [24], [25].

The combination S3/E2/C3 maps to ASIL D in the ISO 26262 matrix, which is the highest traditional functional safety level [24], [25].

C4 decision and escalation. In driver-out operation within the declared ODD (urban roads, speed $\leq$ 50 km/h, nighttime), human intervention is unavailable, which sets controllability to C4. Applying the uniform +1 escalation rule elevates ASIL D to ASIL E. All subsequent derivations proceed under ASIL E obligations [28], [31].

B. Goals and Requirements

1) Safety Goals and SOTIF Sub-Goals: **Safety goal (SG-1).** The ADS shall detect and either avoid or come to a safe stop for all pedestrians within the declared ODD under nighttime conditions without human intervention.

SOTIF sub-goal (SG-1.1). Perception confidence for pedestrian detection shall be at least 95% under low-light conditions defined by ECE R115 nighttime illumination thresholds [4], [42].

2) Functional Safety Requirements (FSR) and Technical Safety Requirements (TSR):

- **FSR-1.** Diagnostic coverage for hardware and software faults affecting pedestrian detection shall exceed 99%, with detection of a latent fault within 50 ms [24], [25].
- **TSR-1.** End-to-end latency from image capture to control actuation shall be $\leq$ 100 ms under nominal conditions [24], [25].

3) SOTIF Requirements for Low-Light Conditions:

- **SOTIF-1.** The vision sensor suite, comprising visible camera and infrared sensing, shall sustain a minimum signal-to-noise ratio that enables pedestrian detection at 10 lux illumination [22], [37].
- **SOTIF-2.** Data-fusion confidence thresholds shall degrade gracefully, with automatic fallback to infrared-only detection when visible-light confidence is below 80% [4], [33].
- **SOTIF-3.** The ADS shall execute a safe-stop maneuver when pedestrian-detection confidence is below 70% for more than 200 ms in urban environments [4], [33].

4) Runtime Assurance (RTA) Requirements:

- **RTA-1.** Online monitors shall compute pedestrian-detection confidence and sensor-health metrics at a rate of at least 50 Hz [20], [29].
- **RTA-2.** If combined confidence is below 60%, the ADS shall initiate a Minimal Risk Condition within 2 s, consisting of controlled deceleration to a stop within the lane [19], [20].
- **RTA-3.** RTA monitors and trigger logic shall be validated to achieve a false-positive rate no greater than 1% to limit unnecessary MRC activations [20].

5) Post-Deployment Requirements:

- **Telemetry (PD-1).** Record all occurrences of SOTIF-1 conditions and RTA-1 triggers, including environmental context, sensor readings, and any MRC execution, to support continuous analysis and traceability [28].
- **Performance drift (PD-2).** Apply statistical drift detection to pedestrian-detection performance. Initiate model retraining when degradation exceeds 5% over a rolling 30-day interval [28].
- **Incident handling and update control (PD-3, PD-4)**. Capture and categorize all pedestrian collisions and near-misses and integrate the findings into the safety case within 48 hours [28], [52].

Gate over-the-air updates that affect perception or RTA logic behind full regression V&V, including targeted nighttime and MRC scenario testing, prior to deployment [28], [34].

C. Verification, Validation, and Safety-Case Linkage

This section specifies the Verification and Validation plan and its contribution to the safety case in line with UL 4600. Evidence generated here demonstrates that requirements derived under ASIL E have been met and that the argument remains auditable and traceable throughout the lifecycle.

1) Adversarial Tests and Night Scenarios:

- **Simulation-based testing.** Construct physics-based adversarial cases, such as pedestrians in dark clothing at 5 lux with partial occlusion, and verify that pedestrian-detection confidence remains at or above the SG-1.1 threshold [39].
- **Closed-track testing.** Perform nighttime trials under illumination conditions prescribed by ECE R115, and measure field performance against SOTIF-1, SOTIF-2, and RTA-2 criteria [4].
- **Fault injection.** Introduce representative perception faults, including added camera latency and infrared dropout, to confirm FSR-1 diagnostic coverage and the reliability of RTA triggers [25], [53].

2) Minimal Risk Condition Demonstrations:

- **On-vehicle validation.** Induce abrupt drops in nighttime pedestrian-detection confidence to trigger RTA-2 and observe the safe-stop sequence. Confirm a stopping distance of no more than 5 m and verify closed-loop stability during deceleration [19], [20].

Figure 4 visualizes the GSN-style argument for the nighttime pedestrian detection hazard, linking requirements, V&V results, and post-deployment data.

3) UL 4600-Aligned Safety-Case Argument:

- **Claim 1. ADS detects pedestrians at night within the declared ODD.** Argument. Supported by simulation logs, closed-track reports, and field telemetry. Evidence. Adversarial test results meeting SG-1.1, illumination measurements from track tests, and performance statistics from field operational testing, organized as evidence packs A, B, and C [28], [34].
- **Claim 2. ADS executes a safe stop when perception confidence is insufficient.** Argument. Substantiated by fault-injection outcomes, RTA log analysis, and Minimal Risk Condition demonstrations. Evidence. Fault tree analysis confirming diagnostic coverage, RTA monitor performance metrics, and MRC maneuver reports including video records [19], [20].
- **Claim 3. The safety case remains valid through post-deployment change.** Argument. Maintained through controlled OTA processes, incident reporting, drift detection, and versioned safety-case updates. Evidence. OTA release-gate checklists, incident response timelines, drift-detection dashboards, and safety-case revision history [28], [34], [52].

Fig. 4: Goal-Structuring Notation (GSN) diagram for the pedestrian-not-detected scenario, showing top-level claims, arguments, and evidence sources

4) Independence and Auditability: An assessment team that is organizationally independent of the developers and primary testers reviews all evidence packs and the structured argument to reduce bias and confirm certification readiness. The safety case and V&V documentation provide full traceability from safety goals to requirements and then to evidence, enabling straightforward audit by regulators or third-party assessors [24], [25].

The pedestrian non-detection example shows a complete ASIL E workflow: HARA with C4 escalation, goals and requirements integrating SOTIF and runtime assurance, adversarial and track-based V&V with MRC demonstrations, and a UL 4600-aligned safety case. Continuous post-deployment monitoring and controlled updates preserve confidence in driver-out autonomy over time.

VI. Discussion

The proposed elevation of ISO 26262 to an ASIL E tier for driver-out autonomy marks a substantive shift in automotive functional safety for SAE Level 4 and Level 5 operation. The approach introduces a new controllability class, C4, together with a uniform +1 escalation across hazards, thereby reflecting the absence of human fallback and aligning safety analysis with operational conditions. The concept blends complementary standards

into one auditable structure. Benefits are clear, yet practical hurdles remain. The following subsections synthesise the advantages and outline an adoption pathway that is workable for industry and regulators.

A. Benefits

1) Recognition of zero-controllability conditions: Conventional ASIL A–D assessments presume a human can intervene to limit harm. ASIL E formalises the loss of this assumption in driver-out contexts by defining C4 and applying a consistent +1 increase in ASIL for each identified hazard. This adjustment prevents underspecification of safeguards, ensures that risk classification mirrors the operational envelope of autonomy, and grounds safety arguments in a vocabulary that matches real-world usage of fully automated vehicles [28], [31].

2) Integration of systematic safety standards: ASIL E binds ISO 26262, ISO 21448 (SOTIF), and UL 4600 into a single, coherent framework. SOTIF coverage ensures that performance limitations and specification gaps, such as perception weakness in low light or adverse weather, receive attention on par with random hardware faults. UL 4600 contributes a structured safety case with governance that spans pre-deployment validation, control of over-the-air updates, and post-deployment learning from incidents. The result is an end-to-end architecture with traceable claims and evidence, rather than siloed compliance activities [4], [28], [33], [34].

3) Regulatory clarity and streamlined approval: Building on the existing ASIL taxonomy avoids inventing a new regulatory scheme. The C4 category and the uniform +1 escalation give clear triggers for when ASIL E applies, which simplifies conformity assessment. Authorities can incorporate ASIL E through established mechanisms, including UNECE WP.29 GRVA processes, while preserving foundational safety principles. This improves consistency across jurisdictions and supports harmonised expectations for driver-out deployments [31], [41].

4) Public confidence and liability management: A rigorous, independent, and auditable safety case, backed by structured evidence packs, addresses public and legal scrutiny of autonomous operation. Continuous ingestion of operational data, including telemetry, incident reports, and drift detection signals, strengthens accountability and supports defensible liability analyses when events occur. Demonstrable alignment with recognised standards also improves communication with consumers, insurers, and policymakers, which can accelerate acceptance of driver-out technologies [24], [25].

B. Challenges

1) Increased development and verification costs: A uniform +1 ASIL escalation in driver-out operation raises safety targets across all hazards. Meeting ASIL E requires higher diagnostic coverage, additional redundancy, and broader verification and validation activities. Achieving near-ASIL-E diagnostic performance, for example 99.9% coverage, together with runtime assurance monitors and complete evidence packs, drives substantial expenditure. These burdens can limit market entry for startups and smaller OEMs and may shift leverage toward larger suppliers [25], [32].

2) Verification burden for machine-learning components: Autonomous stacks rely on machine learning for perception and planning, yet ISO 26262 and related standards were not conceived around data-driven modules. Deterministic diagnostic coverage and formal proofs are difficult because algorithms are probabilistic, high dimensional, and internally opaque. Constructing exhaustive tests for rare conditions, including adverse weather or adversarial occlusions, scales combinatorially. Progress depends on maturing V&V methods such as statistical coverage metrics, adversarial test generation, and explainable AI, which currently lack consistent regulatory acceptance [54].

3) Standards convergence and governance: Combining ISO 26262, ISO 21448, and UL 4600 requires alignment of terminology, scope, and lifecycle governance. Hardware-fault safety, functional insufficiency, and safety-case processes intersect, creating overlaps and potential conflicts. Working groups must resolve questions on evidence-pack granularity, cadence of safety-case updates, and thresholds for drift detection. Reaching consensus across ISO, SAE, UL, and regulators is slow and sensitive to regional positions, increasing the risk of fragmented ASIL E implementations across markets [22], [28], [34].

4) Organizational and cultural adaptation: Effective ASIL E deployment depends on cross-functional coordination across software, hardware, ML, safety, and compliance teams. New competencies are required, including STPA facilitation, SOTIF scenario generation, safety-case engineering, and governance of over-the-air changes. Teams accustomed to FMEA and FTA must adopt system-level practices that use iterative analysis and field telemetry. Institutions will need targeted training, process redesign, and in some cases reorganization to sustain a continuously maintained safety case [4], [51], [52].

C. Pragmatic Path to Adoption

A staged path can capture the safety gains of ASIL E while reducing adoption friction. The approach combines a transitional annex, targeted pilots within constrained operational design domains, and coordinated standards work that strengthens evidence generation and tool support.

1) Transitional "ASIL D + ADS Annex": Where formal recognition of ASIL E is not yet feasible, authorities and industry can employ an "ASIL D + ADS Annex" model. ISO 26262 retains ASIL D as the nominal classification, while an annex specifies obligations that are equivalent to ASIL E for driver-out operation, including SOTIF integration, STPA, runtime assurance, and independence of the safety case. This preserves continuity with existing conformity processes while enforcing driver-out safeguards [28], [31].

- Annexed C4 definition. Define driver-out controllability as C4 and attach it to applicable ASIL D hazards within the annex.
- Annexed escalation. Require hazards under C4 to meet ASIL E safety-case, V&V, and operational monitoring obligations without renaming the underlying level.
- Regulatory mapping. Permit type-approval bodies to reference the annex so that systems satisfying "ASIL D + Annex" are treated as functionally equivalent to ASIL E and are eligible for driver-out deployment.

This transitional arrangement yields immediate clarity and higher rigour while broader standardisation progresses, enabling manufacturers to apply best practice and regulators to enforce elevated expectations within current certification structures.

2) Phased implementation and pilot programmes: Manufacturers should introduce the ASIL D + Annex obligations within limited ODDs, for example urban shuttles or campus delivery services. Controlled deployments enable collection of operational data and refinement of key workflows, including STPA facilitation, safety-case updates, and governance of over-the-air changes. Evidence from these pilots can inform iterative revisions to guidance, reduce uncertainty in ML verification methods, and demonstrate practical effectiveness of ASIL E obligations to authorities and the public.

3) Standards collaboration and tooling ecosystem: Industry consortia and standards bodies should coordinate implementation guidance, shared tooling, and open case studies to streamline ASIL E practices. Common libraries of technical safety concepts and interfaces, SOTIF scenario catalogues, STPA templates, and GSN patterns reduce duplicated effort and lower entry barriers. Joint research on ML V&V and runtime assurance will mature evidence-generation techniques, supporting consistent acceptance across regulators and accelerating broader adoption.

By recognising the qualitative shift in controllability, integrating complementary standards in a systematic manner, and using a transitional annex with pilots and shared tooling, the ASIL E methodology balances rigorous safety improvements with realistic delivery. Stakeholders can address cost, ML verification, and coordination challenges through collaborative pilots, targeted guidance, and phased policy updates, enabling safe and scalable driver-out deployments.

VII. Conclusion

This study addressed the absence of human controllability in Levels 4 and 5 by introducing a driver-out controllability class, prescribing a uniform one-level escalation of integrity targets, and deriving Automotive Safety Integrity Level E obligations. The framework unified ISO 26262 with Safety of the Intended Functionality and the UL 4600 safety-case approach. It was operationalized through explicit mode and operational design domain declarations, hazard analysis with a C4 decision, System-Theoretic Process Analysis, SOTIF triggering-condition analysis, structured safety-case assembly, and lifecycle monitoring. A worked example on nighttime pedestrian non-detection demonstrated requirement flow-down, verification and validation, and auditable traceability.

Positioned against prior literature, the work closes a recognized gap in ISO 26262 where controllability assumes human intervention. The explicit C4 definition, the uniform +1 escalation rule, and the coupling of SOTIF with a UL 4600-aligned safety case create a coherent basis for assuring driver-out operation. The inclusion of quantitative targets enhances evaluability: diagnostic coverage above 99 percent, system uptime above 99.99 percent, end-to-end latency under 100 ms, runtime-monitor update rates at or above 50 Hz, Minimal-Risk-Condition initiation within 2 s, telemetry coverage above 99 percent, drift-detection action within 30 days, and incident response within 48 hours.

The findings have theoretical and practical implications. The C4 escalation preserves ISO 26262 risk ordering while aligning integrity targets with zero-controllability conditions. The safety-case skeleton, curated evidence packs, and governance for over-the-air changes improve auditability and support certification. The worked example shows that scenario-based testing, adversarial simulation, fault injection, and on-vehicle Minimal-Risk-Condition demonstrations can be integrated into a single, reviewable argument for continuous assurance.

Limitations must be noted. Achieving near-ASIL-E diagnostic performance and fail-operational redundancy increases cost and design complexity. Verification of machine-learning components remains difficult because coverage metrics, adversarial robustness, and explainability are still developing. Harmonization across ISO 26262, ISO 21448, and UL 4600 requires sustained coordination to prevent duplicated evidence and to align drift-detection thresholds, evidence-pack granularity, and update cadences. Organizational readiness varies, and

generalization across diverse operational design domains will require larger field datasets.

Future work should focus on standardized scenario libraries and statistical coverage benchmarks for SOTIF validation, confidence-estimation and monitor-fusion methods with formal guarantees, reference patterns for fail-operational architectures with measurable recovery bounds, and shared tooling for evidence-pack curation and safety-case versioning. Policy mappings to UNECE and regional type-approval processes are also needed. Phased pilots in constrained operational design domains, supported by an interim "ASIL D plus ADS annex," can generate operational evidence and reduce adoption barriers.

The methodology provides a defensible and transparent route to demonstrate acceptable safety for driver-out autonomy, maintains backward compatibility for non-C4 operation, and enables continuous, data-driven assurance from pre-deployment through in-service monitoring. It offers industry and regulators a practical foundation for scaling safe driver-out deployments while advancing evidence standards necessary for public trust.

REFERENCES

[1] *Road Vehicles — Functional Safety — Part 1: Vocabulary*, Std. ISO 26 262-1:2018, 2018.

[2] R. Debouk, "Overview of the second edition of ISO 26262: Functional safety—road vehicles," *Journal of System Safety*, vol. 55, no. 1, pp. 13–21, 2019.

[3] F. Ferlini, L. O. Seman, and E. A. Bezerra, "Enabling ISO 26262 compliance with accelerated diagnostic coverage assessment," *Electronics*, vol. 9, no. 5, p. 732, 2020.

[4] K. Madala, C. Avalos-Gonzalez, and G. Krithivasan, "Workflow between ISO 26262 and ISO 21448 standards for autonomous vehicles," *Journal of System Safety*, vol. 57, no. 1, pp. 34–42, 2021.

[5] O. M. Kirovskii and V. A. Gorelov, "Driver assistance systems: Analysis, tests and the safety case. ISO 26262 and ISO PAS 21448," *IOP Conference Series: Materials Science and Engineering*, vol. 534, no. 1, p. 012019, 2019.

[6] *Road Vehicles — Functional Safety — Part 2: Management of Functional Safety*, Std. ISO 26 262-2:2018, 2018.

[7] R. Debouk, "Review of the latest developments in automotive safety standardization for driving automation systems," *Journal of System Safety*, vol. 58, no. 2, pp. 40–45, 2023.

[8] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, Std. SAE J3016_202 104, 2021.

[9] X. Li, A. C. A. Doss, B. A. Guvenc, and L. Guvenc, "Pre-deployment testing of low speed, urban road autonomous driving in a simulated environment," *SAE International Journal of Advances and Current Practices in Mobility*, vol. 2, no. 6, pp. 3301–3311, 2020.

[10] R. Salay, N. Kumar, and K. Czarnecki, "Enabling hazard analysis of ADAS and automated vehicles by characterizing driving scenarios," in *SAE Technical Paper*, no. 2018-01-1075. SAE International, 2018.

[11] H. Kang, Y. Lee, H. Jeong, G. Park, and I. Yun, "Applying the operational design domain concept to vehicles equipped with advanced driver assistance systems for enhanced safety," *Journal of Advanced Transportation*, vol. 2023, p. 4640069, 2023.

[12] *Road Vehicles — Safety of the Intended Functionality (SOTIF)*, Std. ISO 21 448:2022, 2022.

[13] *UL 4600: Standard for Evaluation of Autonomous Products*, Std., 2023.

[14] Y. Li, W. Liu, Q. Liu, X. Zheng, K. Sun, and C. Huang, "Complying with ISO 26262 and ISO/SAE 21434: A safety and security co-analysis method for intelligent connected vehicle," *Sensors*, vol. 24, no. 6, p. 1848, 2024.

[15] L. Rocha, P. Maciel, J. Cabral, and A. Costa, "Virtualized fault injection framework for ISO 26262-compliant digital component hardware faults," *Electronics*, vol. 13, no. 14, p. 2787, 2024.

[16] D. R. Biba, M. C. Ancuti, A. Ianovici, C. Sorandaru, and S. Muşuroi, "Power supply platform and functional safety concept proposals for a powertrain transmission electronic control unit," *Electronics*, vol. 9, no. 10, p. 1580, 2020.

[17] *Considerations for ISO 26262 ASIL Hazard Classification*, Std. SAE J2980_202 310, 2023.

[18] W. M. Goble and A. C. Brombacher, "Using a failure modes, effects and diagnostic analysis (FMEDA) to measure diagnostic coverage in programmable electronic systems," *Reliability Engineering & System Safety*, vol. 66, no. 2, pp. 145–148, 1999.

[19] T. Stolte, S. Ackermann, R. Graubohm, I. Jatzkowski, B. Klamann, H. Winner, and M. Maurer, "A Taxonomy to Unify Fault Tolerance Regimes for Automotive Systems: Defining Fail-Operational, Fail-Degraded, and Fail-Safe," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 251–262, Jun. 2022.

[20] T. Schmid, S. Schraufstetter, J. Fritzsch, D. Hellhake, G. Koelln, and S. Wagner, "Formal Verification of a Fail-Operational Automotive Driving System," Jan. 2021.

[21] T. Vidano and F. Assadian, "Control performance requirements for automated driving systems," *Electronics*, vol. 13, no. 5, p. 902, 2024.

[22] M. D. Menekşe, O. Özçetin, T. E. Ercan, and K. F. Doğan, "Safety of the Intended Functionality (SOTIF) based on System Theoretic Process Analysis (STPA): Study for Specific Control Action in Blind Spot Detection (BSD)," in *2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, Dec. 2024, pp. 1–5.

[23] "A study of automatic allocation of automotive safety requirements in two modes: Components and failure modes," in *The Role of ISO 26262*, 2020, pp. 83–97.

[24] J. N. Gowda, "ECU Inter-processor data communication End to End verification in Autosar for achieving Functional Safety Goals," *INCOSE International Symposium*, vol. 29, no. S1, pp. 443–453, 2019.

[25] F. Ferlini, L. O. Seman, and E. A. Bezerra, "Enabling ISO 26262 Compliance with Accelerated Diagnostic Coverage Assessment," *Electronics*, vol. 9, no. 5, p. 732, May 2020.

[26] C. Robinson-Mallet, J. Wegener, H. Heers, and P. Liggesmeyer, "Integration und Validation von Produktlinien fuer Fahrerassistenzsysteme im Kontext der ISO 26262 / Integration an validation of driver assistance product-lines in the context of ISO 26262," 2010.

[27] G. Bahig and A. El-Kadi, "Formal Verification of Automotive Design in Compliance With ISO 26262 Design Verification Guidelines," *IEEE Access*, vol. 5, pp. 4505–4516, 2017.

[28] P. Koopman, "UL 4600: What to Include in an Autonomous Vehicle Safety Case," *Computer*, vol. 56, no. 5, pp. 101–104, May 2023.

[29] P. Koopman and W. Widen, "Redefining Safety for Autonomous Vehicles," Aug. 2024.

[30] D. Jackson, V. Richmond, M. Wang, J. Chow, U. Guajardo, S. Kong, S. Campos, G. Litt, and N. Arechiga, "Certified Control: An Architecture for Verifiable Safety of Autonomous Vehicles," Mar. 2021.

[31] M. Wagner and C. Carlan, "The Open Autonomy Safety Case Framework," Apr. 2024.

[32] F. A. da Silva, A. Cagri Bagbaba, S. Hamdioui, and C. Sauer, "An automated formal-based approach for reducing undetected faults in ISO 26262 hardware compliant designs," in *2021 IEEE International Test Conference (ITC)*, Oct. 2021, pp. 329–333.

[33] M. Patel, R. Jung, and M. Khatun, "A Systematic Literature

Review on Safety of the Intended Functionality for Automated Driving Systems," Apr. 2025, pp. 2025–01–5030.

[34] D. Ratiu, T. Rohlinger, T. Stolte, and S. Wagner, "Towards an Argument Pattern for the Use of Safety Performance Indicators," Oct. 2024.

[35] M. Okada and B. Gallina, "Safety of the Intended Functionality of External Human Interfaces: Gaps and Research Agenda," in *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2024, pp. 578–583.

[36] A. Collin, A. Bilka, S. Pendleton, and R. D. Tebbens, "Safety of the Intended Driving Behavior Using Rulebooks," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, Oct. 2020, pp. 136–143.

[37] X. Xing, T. Jia, J. Chen, L. Xiong, and Z. Yu, "An Ontology-based Method to Identify Triggering Conditions for Perception Insufficiency of Autonomous Vehicles," Oct. 2022.

[38] M. Conrad and G. Schildbach, "Analysis of Functional Insufficiencies and Triggering Conditions to Improve the SOTIF of an MPC-based Trajectory Planner," Jul. 2024.

[39] L. Putze, L. Westhofen, T. Koopmann, E. Böde, and C. Neurohr, "On Quantification for SOTIF Validation of Automated Driving Systems," Apr. 2023.

[40] L. Peng, B. Li, W. Yu, K. Yang, W. Shao, and H. Wang, "SOTIF Entropy: Online SOTIF Risk Quantification and Mitigation for Autonomous Driving," Nov. 2022.

[41] J. Betz, M. Lutwitzi, and S. Peters, "A new Taxonomy for Automated Driving: Structuring Applications based on their Operational Design Domain, Level of Automation and Automation Readiness," Apr. 2024.

[42] V. Mohan, R. Harradi, and W. Hardt, "Enhancing SOTIF Analysis Using Model-Based Systems Engineering and Virtual Validation With Focus on Responsibility-Sensitive Safety," in *2024 International Symposium on Computer Science and Educational Technology (ISCSET)*, Jul. 2024, pp. 1–6.

[43] T. Menzel, G. Bagschik, and M. Maurer, "Scenarios for Development, Test and Validation of Automated Vehicles," Apr. 2018.

[44] V. J. E. Jiménez, H. Martin, C. Schwarzl, G. Macher, and E. Brenner, "Triggering Conditions Analysis and Use Case for Validation of ADAS/ADS Functions," 2022, vol. 13415, pp. 11–22.

[45] C. Reichenbächer, J. Hipp, and O. Bringmann, "Statistical Modelling of Driving Scenarios in Road Traffic using Fleet Data of Production Vehicles," 2024, pp. 185–196.

[46] S.-S. Shin, H.-J. Kang, and S.-J. Kwon, "A Study on Data Analysis for Improving Driving Safety in Field Operational Test (FOT) of Autonomous Vehicles," *Machines*, vol. 10, no. 9, p. 784, Sep. 2022.

[47] N. F. Salem, T. Kirschbaum, M. Nolte, C. Lalitsch-Schneider, R. Graubohm, J. Reich, and M. Maurer, "Risk Management Core – Towards an Explicit Representation of Risk in Automated Driving," *IEEE Access*, vol. 12, pp. 33 200–33 217, 2024.

[48] K. S. Kushal, M. Nanda, and J. Jayanthi, "Architecture Level Safety Analyses for Safety-Critical Systems," *International Journal of Aerospace Engineering*, vol. 2017, no. 1, p. 6143727, 2017.

[49] K. Bos, M. J. van der Laan, J. Groeneweg, G. J. Kamps, D. A. Legemate, I. Leistikow, and D. A. Dongelmans, "Grading recommendations for enhanced patient safety in sentinel event analysis: The recommendation improvement matrix." *BMJ open quality*, vol. 13, no. 2, Apr. 2024.

[50] M. D. Menekşe, O. Özçetin, T. E. Ercan, and K. F. Doğan, "Safety of the Intended Functionality (SOTIF) based on System Theoretic Process Analysis (STPA): Study for Specific Control Action in Blind Spot Detection (BSD)," in *2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, Dec. 2024, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10845230

[51] O. Özçetin and P. Brudke, "Comparison of System Theoretic Process Analysis and Cause Tree Analysis Applied on an Autonomous Parking System from Safety of the Intended Functionality Perspective," in *2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, Dec. 2024, pp. 1–5.

[52] D. Wichner, J. Wishart, J. Sergent, and S. Swaminathan, "Developing a Safety Management System for the Autonomous Vehicle Industry," Nov. 2024.

[53] O. Ahlgren, "Spice up apqp," 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:86562260

[54] K. Radlak, M. Szczepankiewicz, T. Jones, and P. Serwa, "Organization of machine learning based product development as per ISO 26262 and ISO/PAS 21448," in *2020 IEEE 25th Pacific Rim International Symposium on Dependable Computing (PRDC)*. IEEE, 2020, pp. 110–119.