ODA3 INSTITUTE

TECHNICAL + COMPLIANCE REPORT

PAI-SF™ v1.0 — Physical AI Security Framework

Technical + Compliance Report (TCR)

ODA3-2026-07-TCR-PAISF-001 | Internal Working Draft — Pre-Publication

Document Control

Doc ID	ODA3-2026-07-TCR-PAISF-001
Doc Type	Technical + Compliance Report (TCR)
Framework Tag	PAI-SF™ v1.0
Regulatory Crosswalk Tags	NIST AI RMF, ISO/IEC 42001, IEC 62443, ISO 26262, ISO 21448, IEC 61508, ISO 10218, ISO/TS 15066
Evidence Tier	T1–T4 (inline, Sections 4–13)
Incident Provenance	Not applicable — no ODA3 proprietary incident data used in this draft
Audience	CISOs, Security Architects, AI Governance Leads, Compliance Officers, Standards Body Participants (primary); Safety Engineers, OT Leaders, Robotics/Autonomy Assurance Teams (secondary)
Governing Licence	GAISSF Ecosystem Licence (GEL v1.0)
Access Classification	Open / Public core content; Controlled asset categories (scoring methodology, test procedures, auditor guidance, scenario libraries, certification decision logic) are referenced but not disclosed in this document
Publication State	Scheduled for Publication — 03 August 2026
Publish Date / Review Cycle	Scheduled for publication 03 August 2026; review cycle to be set upon publication approval
Classification	Public-Facing — Scheduled for Publication (03 August 2026)

PAI-SF™ v1.0 — Physical AI Security Framework

Technical + Compliance Report (TCR)

Methodology Note

> This report distinguishes confirmed facts from analytical interpretation throughout. Evidence tiers appear inline as **[T1]** Primary Verified, **[T2]** Secondary Verified, **[T3]** Controlled Simulation/Academic Proxy, or **[T4]** Anecdotal/Unverified, using ODA3's existing corpus-wide tiering — unchanged in meaning from GAISSF™/UAIF™/AI-IRF™ usage, not redefined here. As a framework published by a firm founded March 2026, PAI-SF™ makes no claim to proprietary telemetry or client incident data; threat and evidence content throughout is drawn from analysis of public disclosures, regulatory filings, and academic/controlled simulation, tiered accordingly.

1. Executive Summary

Physical AI — robotics, autonomous vehicles, drones, industrial automation, medical robotics, and embodied agents — introduces a category of AI risk that existing AI governance and security frameworks were not built to address: the translation of model inference into physical actuation. The Physical AI Security Framework (PAI-SF™ v1.0) is ODA3 Institute's fourth core framework, operating within the GAISSF Ecosystem alongside GAISSF™ (govern and assure), UAIF™ (classify), and AI-IRF™ (respond). PAI-SF™ does not replace functional safety, robotics safety, OT security, or sector-specific regulation — it defines the AI-security assurance layer for systems where model behavior, sensing, and actuation intersect with physical consequence.

This report documents PAI-SF™'s scope, principles, threat model, 12-domain architecture, 41-control register, lifecycle model, evidence and assessment methodology, standards mapping, and conformance pathway, together with an explicit account of what remains unproven or excluded.

2. Placement Within the GAISSF Ecosystem

Why a fourth core framework, not an extension of GAISSF™: GAISSF™'s 9 domains / 59 controls are built for AI security and safety as an information/model-behavior discipline. Domain 9, "Physical AI Safety," carries 7 of those 59 controls [T1 — published framework structure, confirmed at https://docs.oda3.org/frameworks/gaissf/index.html]. Kinematic safety bounds, actuator integrity, hardware-in-the-loop testing, and functional-safety alignment (ISO 26262, ISO 21448) do not map onto GAISSF's existing domain logic — when AI decides to move, apply force, or affect the physical world, the assurance problem changes in kind, not merely in degree.

Locked security/safety boundary: PAI-SF™ addresses physical-AI security assurance, including security conditions that may create or worsen safety consequences. It does not replace functional safety, product safety, safety engineering, hazard analysis, or sector-specific safety certification frameworks. GAISSF™ remains the top-level framework in the ecosystem whose name spans both safety and security; PAI-SF™ deliberately stays in its own lane — physical-AI security assurance — rather than extending into safety ownership. Where security failures have safety consequences, that intersection is addressed explicitly through Domain 11 (Functional Safety Interface), not by broadening PAI-SF™'s own claim.

Relationship table:

Framework	Primary Role	What PAI-SF™ Adds
GAISSF™ v1.0	Governs and assures AI security/safety controls	Domain-specific controls for sensor integrity, actuator bounding, kinetic safety, physical degradation, hardware attestation
UAIF™ v1.0	Classifies AI incidents	Physical consequence classification, kinetic impact taxonomy, sensor/actuator incident types
AI-IRF™ v1.0	Guides incident response	Physical containment procedures, safe-state management, HITL revalidation before return-to-service
PAI-SF™ v1.0	Governs physical-AI security, kinetic safety, sensor/actuator trust	—

GAISSF Domain 9 disposition (decided, not yet implemented on the published GAISSF page): GAISSF™'s 7 physical-AI controls (Domain 9) are to be marked **superseded/expanded by PAI-SF™**, not deleted. The decided plan is for Domain 9 to be retained as a stub with a citation-continuity pointer to this framework, preserving any existing reference to those control numbers while making PAI-SF™ the authoritative source going forward. As of this report's drafting, GAISSF™'s published documentation still shows Domain 9 in its original form — this disposition is a confirmed decision pending execution on the live GAISSF Ecosystem site, tracked as an open item. The control-by-control mapping itself is published as **ODA3-2026-07-CSX-PAISF-004** (see Appendix B).

Domain-naming crosswalk (collision check against GAISSF's published 9 domains): No exact name duplicates exist between GAISSF's 9 domains and PAI-SF™'s 12. Three pairs warrant explicit distinction to avoid citation ambiguity within the GAISSF Ecosystem:

GAISSF™ Domain	PAI-SF™ Domain	Distinction
D4 Supply Chain & Third-Party AI Security	PAI-SF D9 Supply Chain and Update Assurance	GAISSF D4 covers AI-specific supply chain risk in software/model pipelines; PAI-SF D9 covers physical hardware, firmware, and actuator component supply chains.
D2 Runtime Security & Adversarial Defense	PAI-SF D7 Runtime Monitoring and Telemetry	GAISSF D2 addresses runtime adversarial defense generally; PAI-SF D7 addresses physical telemetry integrity, independence, and evidentiary custody specifically.
D3 Agentic Risk & Autonomous System Security	PAI-SF D4 Autonomy Boundaries	GAISSF D3 addresses agentic software autonomy risk; PAI-SF D4 addresses physical autonomy-level authorization, ODD change control, and graduated boundary enforcement.

Licensing posture: PAI-SF™ is governed by the GAISSF Ecosystem Licence (GEL v1.0) directly, consistent with GAISSF™/UAIF™/AI-IRF™. GEL v1.0 already states its scope covers "all updates, versions, and future extensions published by ODA3 Institute" under the GAISSF Ecosystem — PAI-SF™, as the fourth framework within that ecosystem, is brought under this same instrument rather than governed by a separate, PAI-SF™-specific license. This corrects an earlier working assumption in this report's drafting process that GAISSF™/UAIF™/AI-IRF™'s Open assets were licensed under Apache 2.0; they are not — GEL v1.0 is itself a source-available, non-open-source, ODA3-controlled license (not OSI-approved, non-commercial by default, commercial use requiring a separate paid licence from ODA3 Institute). PAI-SF™'s Controlled assets (scoring methodology, test procedures, auditor guidance, scenario libraries, certification decision logic) follow the same Controlled-asset principle already applied across the ecosystem under GEL v1.0. Whether specific PAI-SF™-only amendments to GEL v1.0 are warranted (e.g., liability terms calibrated to physical-consequence risk, or a Standards Development Organisation carve-out extended to physical-safety-relevant bodies such as ISO/TC 299 or IEC TC 65) remains an open item for counsel review, not a decision made in this report.

3. Purpose, Scope, and Intended Users

Purpose: PAI-SF™ v1.0 provides a structured, evidence-based framework for governing, assuring, and evidencing the security and safety of AI-enabled systems that sense, decide, move, control, actuate, manipulate, or materially affect the physical world.

Scope inclusion: Robotics (industrial, collaborative, mobile, humanoid); autonomous vehicles (ground, aerial, maritime); drones and uncrewed aerial systems; AI-driven industrial automation and OT; warehouse/logistics automation; medical robotics; smart infrastructure with AI-driven actuation; embodied AI agents; AI-connected IoT with physical effectors.

Scope exclusion: Pure software AI risk (prompt injection, model theft, data privacy, algorithmic bias) unless it creates a direct physical, kinetic, operational, infrastructure, or human-safety consequence. Governing test: does the AI system's output cause, influence, or fail to prevent a physical action with potential for harm? If not, the concern belongs to GAISSF™/UAIF™/AI-IRF™.

Intended users: CISOs, Security Architects, AI Governance Leads, Compliance Officers, Standards Body Participants (primary, per ODA3's stated audience); Safety Engineers, OT Leaders, Robotics/Autonomy Assurance Teams, Risk Committees (secondary).

4. Definition of Physical AI

> *An AI-enabled system whose model outputs directly or indirectly cause, guide, constrain, or fail to prevent physical action — including movement, force application, environmental manipulation, or control of physical processes — where erroneous, adversarial, or degraded operation could result in kinetic harm, infrastructure damage, operational disruption, or human injury.*

Embodied AI, AI-enabled robotics, autonomous cyber-physical systems, and AI-connected OT receive full coverage. AI systems with only indirect physical consequences (e.g., a recommendation a human must act on) are covered where the causal chain is foreseeable and direct. Pure software AI with no physical effector path is out of scope.

5. Core Principles

Twelve principles ground PAI-SF™'s design: physical consequence awareness; sensor distrust and validation; bounded autonomy; actuator integrity; independent safety constraints; safe degradation; human override proportionality; evidence before assurance; lifecycle accountability; incident readiness; interoperability with existing standards; and **Kinetic Zero Trust**.

Kinetic Zero Trust is PAI-SF™'s defined core assurance principle: AI-generated physical actions should not be trusted solely because the model is authorized, confident, or operating within a known deployment context. Physical commands must be validated against independent safety boundaries, operational constraints, sensor confidence, environmental conditions, and human-override expectations before execution — enforced as a control-level constraint, not a model-level policy.

6. Threat and Hazard Model

Threats are tiered by maturity, not asserted uniformly:

Confirmed patterns [T2/T3 — demonstrated in published research and controlled testing, not asserted as widespread operational exploitation]: sensor spoofing and perception manipulation; adversarial physical-world inputs; actuator compromise; unsafe command execution; simulation-to-reality failure; runtime drift; environmental degradation; human proximity risk; emergency-stop failure.

Plausible analytical patterns [T3 — derived from systematic architecture analysis, feasible but not widely demonstrated operationally]: control-loop manipulation; edge hardware/firmware compromise; supply-chain compromise of AI-specific components; model tampering; unsafe autonomy escalation; embodied prompt injection with physical consequence; cloud-to-edge dependency failure; cascading infrastructure impact.

Maturity honesty statement: PAI-SF™ does not claim these threats have been observed in operational deployments at scale. Organizations should calibrate control investment to threat maturity and their own consequence profile, not to the length of this catalog.

7. Domain Architecture (12 Domains)

PAI-SF™ organizes assurance across 12 domains, stress-tested against a failure condition — any domain producing fewer than 3 substantive controls, or controls that reduce to "detect X, invoke another domain," is a merge candidate. All 12 passed.

1. **Sensor and Perception Integrity** — cross-modal validation, calibration monitoring, adversarial input detection.

2. **Actuator and Kinetic Command Safety** — independent kinematic bounding, replay/timing protection, divergence monitoring.
3. **Edge Hardware and Model Attestation** — secure boot, hardware provenance, key custody.
4. **Autonomy Boundaries** — autonomy-level escalation governance, ODD change control, graduated enforcement, handoff integrity.
5. **Safe-State and Degraded-Mode Behavior** — safe-state definition, model-drift-triggered safe-state, compute-compromise trigger independence, verified fallback control law, degraded-mode definition.
6. **Human Override and Intervention** — independent physical override, operator training, post-override reset integrity.
7. **Runtime Monitoring and Telemetry** — telemetry integrity, independence, coverage validation, chain of custody.
8. **Incident Response and Physical Recovery** — physical containment playbooks, volatile evidence preservation, HITL revalidation before return-to-service.
9. **Supply Chain and Update Assurance** — component provenance, staged rollout with regression testing, third-party assurance requirements.
10. **Simulation, Testing, and Validation** — HITL pre-deployment testing, adversarial physical-world testing, sim-to-real gap quantification.
11. **Functional Safety Interface** — AI-specific hazard integration, control-to-safety-integrity mapping, standards gap register, safety case exception escalation. *This domain interfaces with functional safety; it does not constitute or claim safety ownership — see the locked security/safety boundary in Section 2.*
12. **Physical Operating Environment** — environment specification, environmental change detection, human proximity zone design.

Domain 4 / Domain 12 boundary (locked): Domain 12 owns the physical operating environment as a validated description of where the system may operate. Domain 4 owns the authorization, autonomy limits, and change-control decisions that determine whether the system may act within, enter, leave, or expand that environment.

8. Control Register Summary

The register comprises **41 controls** across the 12 domains (3+3+3+4+5+3+4+3+3+3+4+3 = 41). Each control specifies: Control ID, Domain, Objective, Applicability, Implementation Expectation, Evidence Requirement, Assurance Expectation, Dependencies, Exclusions, and Maturity/Conformance Relevance. The complete 41-control register, at full field detail, is provided in Appendix A.

Illustrative controls:

PAI-SF-SAF-002 — Model-Drift-Triggered Safe-State [T3]: addresses gradual statistical divergence in learned-model behavior, a failure mode with no analog in discrete-fault-based functional safety standards (IEC 61508, ISO 26262) — the clearest concrete instance of PAI-SF™'s core justification for existing as a distinct framework.

PAI-SF-TEL-003 — Coverage Validation Against the Threat Model [T3]: periodically validates that telemetry would actually surface known threats (Section 6), rather than assuming coverage from data volume — the control that most directly demonstrates Domain 7 is an assurance function, not a passive logging service for other domains.

PAI-SF-FSI-004 — Safety Case Exception Escalation [T2]: requires AI-driven hazards not covered by applicable functional safety standards to be escalated to a named safety authority and

formally approved, rather than silently absorbed — the enforcement mechanism that gives Domain 11's mapping/gap-register controls (FSI-001–003) actual consequence.

9. Lifecycle Model

Concept and Intended Use → Physical Environment Definition → Hazard and Threat Modeling → Data and Simulation Design → Model Development → Hardware and Sensor Integration → Control-System Integration → Testing and Validation → Deployment Approval → Runtime Operation → Monitoring and Telemetry → Human Override (maintained throughout) → Incident Response → Recovery and Revalidation → Patch/Update Management → Third-Party Component Changes → Decommissioning.

Each stage carries a defined assurance objective, required evidence, characteristic failure modes, and a governance decision point. Detailed lifecycle-stage criteria should be maintained in a future implementation annex rather than implied as complete in this draft.

10. Evidence and Assurance Model

Six evidence categories: Design, Test, Runtime, Incident, Third-Party, Governance. Evidence quality is assessed on relevance, independence, currency, coverage, and integrity — not tier alone. Assurance is evidence-based and continuous, not assertion-based or point-in-time.

Honest limitation [T3/T4 disclosure]: the current public evidence base for physical-AI-specific security incidents, as distinct from general robotics/OT incidents, remains thin. Most of this framework's threat model rests on **[T3]** analytical extension from adjacent, better-documented domains (industrial control systems, autonomous vehicle safety testing) rather than **[T1]** physical-AI-specific incident data. This is stated plainly, not obscured.

11. Assessment Methodology

Architecture review, documentation review, simulation review, hardware-in-the-loop testing, controlled physical red-teaming, safety-boundary validation, telemetry/log review, change-control review, third-party component review, operator override validation, incident tabletop exercises, post-incident revalidation.

PAI-SF™ does not claim ODA3 currently operates testing labs, holds certification authority, or has proprietary incident datasets. Assessment provides evidence-informed confidence, not elimination of uncertainty.

12. Incident Classification and Response Alignment

PAI-SF™ extends UAIF™'s incident taxonomy with physical-consequence classification (kinetic impact, human safety exposure, infrastructure disruption, sensor/actuator anomaly types) and extends AI-IRF™'s response procedures with physical containment, kinetic evidence preservation, safe-state management during incident, and HITL revalidation before return-to-service — additions AI-IRF™'s general phase model (Prepare→Detect/Analyse→Contain→Recover→Learn) does not itself specify for physical systems.

13. Standards and Framework Mapping

All mappings below are **analytical**, not authoritative — PAI-SF™ does not claim endorsement, certification, or compliance under any external standard.

Standard	Analytical Relationship
ISO/IEC 42001	AI management-system governance extended to physical-AI-specific risk
ISO/IEC 23894	AI risk management guidance extended with physical-hazard specificity
NIST AI RMF	Physical-AI-specific control detail beneath Govern/Map/Measure/Manage
IEC 62443	OT/industrial security zone-conduit model referenced for actuation-layer integrity
ISO 26262	Automotive functional safety — PAI-SF™ identifies AI-specific gaps (discrete-fault model does not cover statistical drift)
ISO 21448 (SOTIF)	Safety of intended functionality — PAI-SF™'s runtime trust verification is a security-layer complement
IEC 61508	Safety lifecycle/SIL concepts — same AI-specific gap as ISO 26262
ISO 10218 / ISO/TS 15066	Robot and collaborative-robot safety — extended to AI-driven behavior beyond traditional programming

Open mapping questions: how AI-specific runtime drift integrates with deterministic functional safety lifecycles (IEC 61508) remains an identified gap, not a resolved mapping — tracked via Domain 11's gap register (FSI-003) and escalation control (FSI-004).

14. Conformance and Maturity Model

Recommended model: per-domain maturity scale (0–4), not organization-wide tiered attestation. A single organization-wide tier would imply cross-domain uniformity inappropriate for a first-version framework in a nascent evidence environment. Three illustrative levels — Foundational Physical-AI Attestation, Operational Physical-AI Assurance, High-Assurance Physical-AI Evaluation — are described as a **proposed pathway only**. ODA3 Institute does not currently hold or claim certification authority for PAI-SF™; any future certification decision logic remains hypothetical.

15. Open and Controlled Asset Boundary

Open (governed by GEL v1.0, the same instrument used across the GAISSF Ecosystem — not a separate PAI-SF™-specific license): terminology, domain names/identifiers, control ID taxonomy, high-level control objectives, basic threat taxonomy, incident vocabulary alignment with UAIF™, an optional telemetry/incident-vocabulary schema offered as interoperability vocabulary rather than a mandated format. Consistent with GEL v1.0's terms, "Open" here means available for citation, non-commercial use, and Internal Use under GEL v1.0 Section 4-5 — not open-source, and not free of restriction for Commercial Use, which requires a separate paid licence from ODA3 Institute under GEL v1.0 Section 5.3/17.

Controlled: official control-to-domain mappings, scoring/maturity thresholds, assessment methodology, auditor guidance, detailed red-team procedures, high-risk scenario libraries, certification decision logic (future/hypothetical pathway only).

A dependency/leakage audit of schema files precedes any open-license drafting, per standing ODA3 practice.

16. Notably Absent

PAI-SF™ v1.0 does not claim: regulatory recognition; certification authority; endorsement by any regulator or standards body; elimination of physical harm; a universal safety guarantee; proprietary incident-data conclusions; confirmed partnerships with UL, TÜV, ISO, or similar bodies; market adoption.

Locked position on safety (governing statement, restated from Section 2): PAI-SF™ addresses physical-AI security assurance, including security conditions that may create or worsen safety consequences. It does not replace functional safety, product safety, safety engineering, hazard analysis, or sector-specific safety certification frameworks (aviation, automotive, medical device, industrial, robotics). This is a deliberate boundary, not an evidence gap — PAI-SF™ is not merely under-resourced for safety ownership, it does not claim safety ownership as part of its scope.

Evidence gaps: limited operational physical-AI incident data at scale; immature simulation-to-reality validation methods; non-standardized adversarial physical-world testing; uncharacterized long-term degradation patterns; immature multi-agent/fleet assurance methods; incomplete supply-chain visibility; limited formal verification of learned components in physical contexts.

What this register has not yet proven [disclosed limitation, not resolved]: the 41 controls are distinct from each other on paper. That does not guarantee they remain distinct when mapped onto a real robot, drone, or autonomous vehicle architecture — for example, Domain 1's sensor health monitoring and Domain 7's runtime anomaly detection may draw on the same underlying telemetry stream in practice, even though they are architecturally separated here. Real-world pressure testing, not further framework-level design, is what will surface this. This register should be treated as revisable based on that experience, not as a final, field-validated instrument.

Appendix A: Complete Control Register (41 Controls, Full 10-Field Format)

Each control specifies: Objective, Applicability, Implementation Expectation, Evidence Requirement, Assurance Expectation, Dependencies, Exclusions, Maturity/Conformance Relevance. Control ID and Domain are given in each control's heading and section respectively.

Domain 1. Sensor and Perception Integrity

PAI-SF-SEN-001 — Multi-Modal Cross-Validation

Field	Content
Objective	Prevent single-sensor spoofing or blinding from causing unsafe action.
Applicability	Systems where a single perception failure could lead to unsafe physical action.
Implementation Expectation	Critical physical actions require corroboration from at least two independent sensor modalities; disagreement triggers a defined safety response.
Evidence Requirement	Sensor fusion architecture documentation; cross-validation test results; disagreement-response test records. [T2/T3]
Assurance Expectation	Test evidence covering representative disagreement scenarios, not only nominal-agreement operation.
Dependencies	PAI-SF-SAF-001 (safe-state response on unresolved

Field	Content
	disagreement).
Exclusions	Not applicable to non-safety-critical perception functions.
Maturity/Conformance Relevance	Expected at all conformance levels for safety-critical perception.
Ecosystem Relationship	Extends GAISSF™ model-input integrity to physical sensor diversity; provides sensor-manipulation incident categories for UAIF™; informs AI-IRF™ sensor-incident containment.

PAI-SF-SEN-002 — Sensor Health and Calibration Monitoring

Field	Content
Objective	Detect a single sensor's own degradation or miscalibration over time, independent of multi-sensor disagreement.
Applicability	All physical AI systems with sensors subject to drift, wear, or environmental degradation.
Implementation Expectation	Continuous self-diagnostic/calibration-drift monitoring per sensor, with a defined recalibration or exclusion threshold.
Evidence Requirement	Calibration drift logs; recalibration/exclusion event records. **[T2]**
Assurance Expectation	Evidence that drift thresholds are set below the point at which degraded input would affect safety, not only below sensor manufacturer tolerance.
Dependencies	PAI-SF-SAF-002 (model-drift-triggered safe-state) — sensor drift and model drift are related but distinct failure sources.
Exclusions	Not applicable to sensors with no meaningful drift/degradation mode (e.g., simple binary contact sensors).
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	Provides sensor-degradation incident categories for UAIF™, distinct from SEN-001's spoofing/disagreement categories.

PAI-SF-SEN-003 — Adversarial Physical-World Input Detection

Field	Content
Objective	Detect deliberate adversarial patterns designed to manipulate perception, distinct from passive degradation.
Applicability	Systems operating in environments where adversarial physical-world attack is a credible threat per the threat catalog.
Implementation Expectation	Auxiliary out-of-distribution/adversarial-pattern detection triggering reduced confidence or alert on detection.

Field	Content
Evidence Requirement	Adversarial test suite results; detection rate benchmarks. **[T3]**
Assurance Expectation	Test evidence against a defined, current set of known adversarial techniques, refreshed as the threat catalog updates.
Dependencies	Section 5 threat catalog (adversarial physical-world inputs entry).
Exclusions	Not applicable to fully enclosed/access-controlled operating environments where adversarial physical access is not credible.
Maturity/Conformance Relevance	Expected at High-Assurance level; Operational level may address only known, published attack patterns.
Ecosystem Relationship	Provides adversarial-perception incident categories for UAIF™; informs AI-IRF™ investigation of suspected deliberate manipulation.

Domain 2. Actuator and Kinetic Command Safety

PAI-SF-ACT-001 — Independent Kinematic Bounding

Field	Content
Objective	Ensure kinematic limits (force, torque, speed, position) are enforced independent of the AI decision path.
Applicability	Any system with powered actuation where exceeding kinematic limits could cause harm.
Implementation Expectation	Limits enforced at hardware/firmware level, architecturally independent from primary AI compute; configurable only through authenticated process; defaults to most restrictive limits on integrity failure.
Evidence Requirement	Architecture documentation showing independence; limit configuration change logs; enforcement test records including attempts to exceed limits via compromised AI path. [T2/T3]
Assurance Expectation	Test evidence that limits hold even when the AI model commands otherwise; formal verification expected for higher-consequence systems.
Dependencies	PAI-SF-ATT-001 (hardware root of trust for limit configuration integrity).
Exclusions	Not applicable to systems without powered actuation.
Maturity/Conformance Relevance	Expected at all conformance levels; depth of independence scales with consequence.
Ecosystem Relationship	Extends GAISSF™ access control/system integrity to physical actuation; provides actuator-incident categories for UAIF™; defines actuator containment for AI-IRF™.

PAI-SF-ACT-002 — Command Replay and Timing Attack Protection

Field	Content
Objective	Prevent replayed or timing-manipulated commands from reaching actuators.
Applicability	Systems with a networked or otherwise interceptable command path to actuators.
Implementation Expectation	Sequence numbering or timestamping with freshness checks on the command path.
Evidence Requirement	Replay-attack test logs. **[T3]**
Assurance Expectation	Test evidence of rejection under simulated replay and timing-manipulation attempts.
Dependencies	None beyond the general command-path architecture.
Exclusions	Not applicable to fully hardwired, non-networked command paths with no replay surface.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels for networked command paths.
Ecosystem Relationship	Provides command-injection incident categories for UAIF™, distinct from ACT-001's magnitude-bounding categories.

PAI-SF-ACT-003 — Actuator Health and Divergence Monitoring

Field	Content
Objective	Detect actuator behavior diverging from commanded behavior (mechanical wear, unexpected resistance) before it becomes unsafe.
Applicability	All systems with powered actuation subject to mechanical wear or environmental resistance changes.
Implementation Expectation	Closed-loop comparison of commanded vs. actual actuator response, flagging discrepancy beyond a defined threshold.
Evidence Requirement	Discrepancy logs; maintenance-trigger records. **[T2]**
Assurance Expectation	Evidence that discrepancy thresholds are tied to safety consequence, not only performance/efficiency concerns.
Dependencies	PAI-SF-SUP-002 (Domain 9 update/regression testing) where firmware updates could shift actuator response characteristics.
Exclusions	Not applicable to actuators with no meaningful wear-based failure mode.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	Provides actuator-anomaly incident categories for UAIF™, distinct from ACT-002's command-path-security categories.

Domain 3. Edge Hardware and Model Attestation

PAI-SF-ATT-001 — Secure Boot and Model Attestation

Field	Content
Objective	Verify the model/firmware executing at the edge is the authorized version.
Applicability	Physical AI systems where model or firmware compromise could lead to unsafe behavior.
Implementation Expectation	Secure boot with hardware root of trust; runtime attestation at configurable interval; defined safety response (not merely a log entry) on attestation failure.
Evidence Requirement	Secure boot architecture documentation; attestation success/failure logs; attestation-failure response test records. **[T2]**
Assurance Expectation	Architecture evidence plus test evidence of the failure response, not merely presence of the attestation mechanism.
Dependencies	PAI-SF-ATT-003 (key custody) — attestation is only as trustworthy as its key protection.
Exclusions	May be proportionally implemented for systems with negligible physical risk from model compromise.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels; Foundational expects at minimum boot-time integrity verification.
Ecosystem Relationship	Extends GAISSF™ model integrity/supply chain controls; provides attestation-failure incident categories for UAIF™; defines attestation failure as an AI-IRF™ trigger event.

PAI-SF-ATT-002 — Hardware Provenance and Anti-Tamper

Field	Content
Objective	Verify physical hardware components have not been substituted or tampered with post-manufacture.
Applicability	Systems where hardware substitution or tampering could affect physical safety.
Implementation Expectation	Hardware bill of materials with cryptographic provenance markers; physical tamper-evidence mechanisms.
Evidence Requirement	HBOM documentation; tamper-inspection records. **[T2]**
Assurance Expectation	Evidence of periodic physical inspection, not only design-time documentation.
Dependencies	PAI-SF-SUP-001 (Domain 9, component provenance) — closely related but ATT-002 is the physical/tamper-detection layer specifically.
Exclusions	Not applicable to components with no plausible physical-substitution attack surface.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.

Field	Content
Ecosystem Relationship	Provides hardware-tampering incident categories for UAIF™; defines re-provisioning procedures for AI-IRF™.

PAI-SF-ATT-003 — Attestation Key Custody

Field	Content
Objective	Protect attestation/signing keys from extraction or misuse.
Applicability	All systems relying on cryptographic attestation (ATT-001, TEL-001).
Implementation Expectation	Keys held in hardware security module or secure element, never exposed to general-purpose compute.
Evidence Requirement	HSM configuration audit; key-extraction test attempts. **[T2]**
Assurance Expectation	Test evidence of resistance to extraction attempts, not only configuration review.
Dependencies	Underlies PAI-SF-ATT-001 and PAI-SF-TEL-001 — both depend on key custody holding.
Exclusions	None — any system using cryptographic attestation depends on this control.
Maturity/Conformance Relevance	Expected at all conformance levels where attestation is used.
Ecosystem Relationship	A single point of failure for both Domain 3 attestation and Domain 7 telemetry integrity if not held — flagged as a cross-domain dependency worth explicit tracking.

Domain 4. Autonomy Boundaries

PAI-SF-AUT-001 — Autonomy Level Definition and Escalation Governance

Field	Content
Control ID	PAI-SF-AUT-001
Domain	Autonomy Boundaries
Objective	Ensure the system's authorized autonomy level is explicit and cannot be escalated without an authenticated, logged authorization process.
Applicability	All physical AI systems with more than one operating autonomy level (e.g., supervised vs. conditional vs. full autonomy). Not applicable to fixed single-mode systems with no escalation path.
Implementation Expectation	Discrete autonomy levels defined with explicit authorization criteria for each; escalation between levels requires an authenticated approval step distinct from routine operational commands.
Evidence Requirement	Autonomy level specification; escalation authorization logs; approval-role mapping. **[T2]**

Field	Content
Assurance Expectation	Design evidence plus authorization-log review demonstrating no undocumented escalation events occurred.
Dependencies	PAI-SF-AUT-002 (ODD definition, since autonomy level and validated operating domain are usually linked).
Exclusions	Does not apply to single-mode systems with no escalation capability.
Maturity/Conformance Relevance	Expected at all conformance levels for multi-level-autonomy systems.
Ecosystem Relationship	Extends GAISSF™ access control/authorization concepts to physical autonomy escalation. Provides autonomy-escalation incident categories for UAIF™. Informs AI-IRF™ investigation when unauthorized escalation is a suspected root cause.

PAI-SF-AUT-002 — Operational Design Domain as a Living Artifact

Field	Content
Control ID	PAI-SF-AUT-002
Domain	Autonomy Boundaries
Objective	Ensure the operational design domain (ODD) is maintained as a version-controlled specification with change-control, not a static one-time document.
Applicability	All physical AI systems operating under a defined ODD.
Implementation Expectation	Changes to the ODD (geography, weather envelope, task scope) trigger a defined revalidation process before the change takes effect; ODD maintained with version history.
Evidence Requirement	ODD version history; revalidation records tied to each ODD change. **[T2]**
Assurance Expectation	Documentation review confirming every ODD change has a corresponding revalidation record — no undocumented scope changes.
Dependencies	PAI-SF-ENV-001 (Domain 12) — see Open Items: this boundary needs explicit reconciliation before publication, since ENV-001 owns the physical/environmental description this control authorizes changes to.
Exclusions	Does not apply where the ODD is fixed by regulation and cannot be organizationally changed.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels; Foundational level expects at minimum a documented (if static) ODD.
Ecosystem Relationship	Extends GAISSF™ change-management controls to the physical operating boundary. Provides ODD-change incident categories for UAIF™.

PAI-SF-AUT-003 — Graduated Boundary Enforcement

Field	Content
Control ID	PAI-SF-AUT-003
Domain	Autonomy Boundaries
Objective	Ensure autonomy boundary enforcement is graduated (warning/restricted/prohibited zones) rather than a single binary safe/unsafe response.
Applicability	Systems where a single hard-stop response to any boundary approach would itself introduce risk (e.g., a moving vehicle, a drone in flight).
Implementation Expectation	At least three zones defined (warning, restricted, prohibited), each with a distinct response — operator alert, autonomy reduction, hard stop — rather than one shared response.
Evidence Requirement	Zone definition documentation; response-by-zone test records. **[T3]**
Assurance Expectation	Test evidence demonstrating each zone triggers its own distinct, correct response, not a shared fallback.
Dependencies	**This control terminates in a Domain 5 (Safe-State) response at its final zone — flagged honestly as the borderline case in the Domain 4 stress-test verdict below.** Also adjacent to PAI-SF-ENV-002 (Domain 12) — see Open Items.
Exclusions	Not applicable to systems where any boundary approach is itself unacceptable risk and only a single hard-stop response is appropriate.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels for systems where graduated response is safer than immediate stop.
Ecosystem Relationship	Provides graduated-boundary-violation incident categories for UAIF™; the final-zone response is the AI-IRF™/Domain 5 handoff point.

PAI-SF-AUT-004 — Autonomy Handoff Integrity

Field	Content
Control ID	PAI-SF-AUT-004
Domain	Autonomy Boundaries
Objective	Prevent ambiguous or contested control state during handoff between autonomous, human, and reduced-autonomy modes.
Applicability	All systems supporting more than one control mode with a handoff path between them.
Implementation Expectation	Handoff requires explicit, authenticated confirmation of which controller holds command authority at any instant, preventing a state where both or neither believe they are in control.
Evidence Requirement	Handoff protocol design; handoff test records covering simulated ambiguous-control scenarios. **[T3]**

Field	Content
Assurance Expectation	Test evidence specifically covering edge-case handoff timing (e.g., simultaneous claim, dropped claim) rather than only the nominal handoff path.
Dependencies	PAI-SF-HOV-001 (Domain 6) for the human-side end of the handoff.
Exclusions	Not applicable to fully autonomous systems with no human handoff path at all.
Maturity/Conformance Relevance	Expected at all conformance levels for systems supporting mode handoff.
Ecosystem Relationship	Provides handoff-failure incident categories for UAIF™; handoff ambiguity during an active incident is a AI-IRF™ investigation concern.

Domain 5. Safe-State and Degraded-Mode Behavior

PAI-SF-SAF-001 — Safe-State Definition for Credible Failure Modes

Field	Content
Control ID	PAI-SF-SAF-001
Domain	Safe-State and Degraded-Mode Behavior
Objective	Ensure the system has a defined safe state for each credible failure mode, triggerable independently of the primary AI system.
Applicability	All physical AI systems.
Implementation Expectation	Credible failure modes identified through hazard analysis; appropriate safe state defined per mode (halt, reduced capability, minimal-risk maneuver, human handover); at least one independent trigger path implemented.
Evidence Requirement	Hazard analysis identifying failure modes and safe states; safe-state architecture documentation; transition test records; physical-achievability validation (e.g., stopping-distance analysis). [T2/T3]
Assurance Expectation	Comprehensive hazard analysis plus validated transitions; higher-consequence systems expect fault-injection testing.
Dependencies	PAI-SF-ACT-001 (kinematic limits for enforcement); PAI-SF-HOV-001 (human override as an ultimate trigger).
Exclusions	None — all physical AI systems require safe-state definition; complexity scales with risk.
Maturity/Conformance Relevance	Expected at all conformance levels; sophistication scales with consequence severity.
Ecosystem Relationship	Extends GAISSF™ availability/resilience controls. Provides safe-state-failure categories for UAIF™. Safe-state activation is a primary AI-IRF™ response action; this control specifies the physical requirements.

PAI-SF-SAF-002 — Model-Drift-Triggered Safe-State

Field	Content
Control ID	PAI-SF-SAF-002
Domain	Safe-State and Degraded-Mode Behavior
Objective	Ensure safe-state logic accounts for gradual AI model drift, not only discrete hardware/software faults.
Applicability	Systems using learned models whose behavior can drift from validated performance over time or operating conditions.
Implementation Expectation	Monitoring for gradual divergence between expected and observed model behavior; defined safe-state or degraded-mode response to sustained drift, distinct from discrete-fault triggers.
Evidence Requirement	Drift-detection methodology; drift-triggered transition test records. **[T3]**
Assurance Expectation	Test evidence demonstrating drift detection at a meaningful threshold before drift becomes unsafe, not only after failure.
Dependencies	PAI-SF-SEN-002 (Domain 1, sensor calibration drift) for sensor-side drift; distinct from model-side drift addressed here.
Exclusions	Not applicable to systems using only deterministic, non-learned control logic with no drift-capable component.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels for learned-model systems.
Ecosystem Relationship	This is the clearest AI-specific gap PAI-SF™ fills relative to IEC 61508/ISO 26262, which model discrete faults, not statistical drift — the framework's core justification made concrete at control level.

PAI-SF-SAF-003 — Safe-State Trigger Independence Under Compute Compromise

Field	Content
Control ID	PAI-SF-SAF-003
Domain	Safe-State and Degraded-Mode Behavior
Objective	Ensure the safe-state trigger remains functional even if the AI compute/inference stack is adversarially compromised, not merely faulty.
Applicability	Systems where compute compromise (not just random failure) is a credible threat per the Section 5 threat catalog.
Implementation Expectation	Safe-state activation path tested specifically under simulated adversarial compute compromise, verifying the trigger fires independent of an actively hostile primary system.
Evidence Requirement	Adversarial fault-injection test records distinguishing compromise scenarios from random-failure scenarios. [T3/T4]

Field	Content
Assurance Expectation	Test evidence specifically covering the adversarial case — evidence of resilience to random failure alone is insufficient for this control.
Dependencies	PAI-SF-ATT-001 (Domain 3, attestation) — a compromised system that fails attestation should itself be a trigger condition.
Exclusions	Not applicable to air-gapped systems with no plausible compute-compromise threat model.
Maturity/Conformance Relevance	Expected at High-Assurance level; Foundational/Operational may address only random-fault independence.
Ecosystem Relationship	This is a security concern layered onto a safety mechanism — the clearest example in the register of PAI-SF™'s stated boundary between AI security and functional safety. Relevant to AI-IRF™ investigation of compromise-related incidents.

PAI-SF-SAF-004 — Verified Fallback Control Law for Degraded Mode

Field	Content
Control ID	PAI-SF-SAF-004
Domain	Safe-State and Degraded-Mode Behavior
Objective	Ensure degraded-mode operation substitutes a simple, independently verifiable control law rather than continuing to run the full learned policy at reduced parameters.
Applicability	Systems where immediate full stop is not the safest response and some continued controlled operation is required in degraded mode.
Implementation Expectation	Degraded mode substitutes a non-learned, formally simpler control function (e.g., a bounded deterministic motion profile) for the learned policy, rather than merely throttling the same model's outputs.
Evidence Requirement	Fallback control law specification and independent verification; test records comparing learned-policy-throttled vs. verified-fallback behavior. **[T3]**
Assurance Expectation	Verification evidence for the fallback control law itself (ideally formal verification for higher-consequence systems), not just behavioral test results.
Dependencies	PAI-SF-DEG-001 (degraded-mode definition, entry/exit criteria).
Exclusions	Not applicable to systems where degraded mode is defined as immediate full stop with no continued operation.
Maturity/Conformance Relevance	Expected at High-Assurance level; Operational level may accept throttled-learned-policy degraded mode as an interim approach.
Ecosystem Relationship	Addresses a question with no analog in non-AI functional safety (what replaces the AI during

Field	Content
	degradation) — another concrete instance of the framework's core justification.

PAI-SF-DEG-001 — Degraded Operating Mode Definition

Field	Content
Control ID	PAI-SF-DEG-001
Domain	Safe-State and Degraded-Mode Behavior
Objective	Ensure the system can operate in a defined degraded mode with reduced physical capability and risk when full capability is not assured.
Applicability	Systems where immediate safe-stop is not always the safest response (vehicles in motion, drones in flight, collaborative robots mid-task).
Implementation Expectation	At least one degraded mode defined that reduces physical risk (lower speed/force/reach/autonomy), is independently enforceable, has defined entry/exit conditions, and does not rely on the compromised component for its own safe operation.
Evidence Requirement	Degraded-mode specification and entry/exit criteria; architecture for independent enforcement; transition test records; validation that degraded mode is safe from worst-case entry states. [T2/T3]
Assurance Expectation	Validated transitions from representative and worst-case operational states, not only nominal-state transitions.
Dependencies	PAI-SF-SAF-001 (degraded mode is a type of safe state); PAI-SF-ACT-001 (kinematic limits for enforcement).
Exclusions	Not applicable where any degradation creates unacceptable risk and only immediate safe-stop is appropriate.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels; Foundational level expects at minimum a documented degradation strategy.
Ecosystem Relationship	Extends GAISSF™ resilience controls to physical degradation. Provides degraded-mode incident classification for UAIF™. Defines degraded-mode management during AI-IRF™ incident response.

Domain 6. Human Override and Intervention

PAI-SF-HOV-001 — Independent Physical Override

Field	Content
Objective	Ensure a physically accessible override exists, independent of the AI compute/software stack.
Applicability	All systems where humans are present in the operational environment or remote intervention is

Field	Content
	feasible.
Implementation Expectation	Hardwired where kinetic risk is significant; logged, non-repudiable activation; authenticated reset process.
Evidence Requirement	Architecture independence analysis; wiring diagrams; override activation/reset logs; test records under simulated AI compute failure. [T2/T3]
Assurance Expectation	Physical demonstration of override function under failure conditions, not only architecture review.
Dependencies	PAI-SF-SAF-001 (override triggers safe state).
Exclusions	For systems with no human access, override may be remote with equivalent independence requirements.
Maturity/Conformance Relevance	Expected at all conformance levels for systems with human proximity.
Ecosystem Relationship	Extends GAISSF™ human-in-the-loop controls; provides override-failure incident categories for UAIF™; defines override as primary AI-IRF™ response mechanism.

PAI-SF-HOV-002 — Override Familiarity and Operator Training

Field	Content
Objective	Ensure operators can actually use the override effectively under time pressure, not merely that it exists.
Applicability	All systems with a human-operated override mechanism.
Implementation Expectation	Periodic training and drills with measured response time.
Evidence Requirement	Training records; drill response-time logs. **[T2]**
Assurance Expectation	Evidence that response times are measured under realistic conditions, not only in ideal training scenarios.
Dependencies	PAI-SF-HOV-001 (the mechanism being trained on).
Exclusions	Not applicable to fully remote/automated override with no human-operator-dependent activation.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	A human-factors control distinct from HOV-001's architectural independence — provides operator-readiness context for AI-IRF™ incident review.

PAI-SF-HOV-003 — Post-Override Reset Integrity

Field	Content
Objective	Prevent silent resumption of operation after override without authenticated reset.
Applicability	All systems with an override mechanism.
Implementation Expectation	Reset requires explicit authenticated action and a

Field	Content
	diagnostic pass before resuming operation.
Evidence Requirement	Reset logs; diagnostic pass records. [T2]
Assurance Expectation	Evidence that reset cannot occur without the diagnostic pass, not merely that a diagnostic step exists in documentation.
Dependencies	PAI-SF-IRP-003 (Domain 8, HITL revalidation) where the override followed an actual incident.
Exclusions	Not applicable to override mechanisms with no distinct reset state (e.g., momentary override with automatic resumption by design).
Maturity/Conformance Relevance	Expected at all conformance levels.
Ecosystem Relationship	Governs the return-to-operation moment specifically — a different point in time from HOV-001's override event itself.

Domain 7. Runtime Monitoring and Telemetry

PAI-SF-TEL-001 — Telemetry Integrity and Tamper-Evidence

Field	Content
Control ID	PAI-SF-TEL-001
Domain	Runtime Monitoring and Telemetry
Objective	Ensure telemetry cannot be silently altered after collection, including by a compromised primary system attempting to conceal its own anomalous behavior.
Applicability	All physical AI systems where telemetry may be relied upon for incident investigation or assurance.
Implementation Expectation	Telemetry cryptographically signed or hash-chained at the point of collection; independent verification of the chain available during investigation.
Evidence Requirement	Telemetry signing architecture; chain-verification test records. [T2]
Assurance Expectation	Test evidence demonstrating tamper detection under simulated post-hoc alteration attempts.
Dependencies	PAI-SF-ATT-003 (key custody) — signing keys must themselves be protected for this control to hold.
Exclusions	Not applicable to non-safety-relevant telemetry with no investigative or evidentiary purpose.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	Telemetry integrity is foundational evidence for UAIF™ classification and AI-IRF™ investigation — a compromised telemetry chain undermines both.

PAI-SF-TEL-002 — Telemetry Independence from the Monitored System

Field	Content
Control ID	PAI-SF-TEL-002
Domain	Runtime Monitoring and Telemetry
Objective	Ensure a compromise of the primary AI/control stack cannot simultaneously blind the monitoring system observing it.
Applicability	Systems where monitoring is relied upon for safety or security assurance, particularly where compute compromise is a credible threat.
Implementation Expectation	Monitoring runs on architecturally separate compute (and power domain, where feasible) from the system it observes.
Evidence Requirement	Architecture documentation showing separation; test records demonstrating monitoring survives simulated primary-system compromise. **[T3]**
Assurance Expectation	Test evidence specifically covering the compromise scenario, not only independent-power-loss scenarios.
Dependencies	PAI-SF-SAF-003 (compute-compromise independence) — same architectural pattern applied to observation rather than action.
Exclusions	May be proportionally simplified for systems with negligible consequence from a blinded-monitoring scenario.
Maturity/Conformance Relevance	Expected at High-Assurance level; Operational level may accept logical (not physical) separation.
Ecosystem Relationship	Directly supports AI-IRF™'s ability to investigate an incident where the primary system itself is the suspected compromised party.

PAI-SF-TEL-003 — Coverage Validation Against the Threat Model

Field	Content
Control ID	PAI-SF-TEL-003
Domain	Runtime Monitoring and Telemetry
Objective	Ensure telemetry actually captures the signals needed to detect the specific threats in the Section 5 threat catalog, rather than collecting data volume without detection relevance.
Applicability	All physical AI systems with a defined threat model.
Implementation Expectation	Periodic validation injecting known threat patterns (simulated sensor spoofing, simulated actuator anomaly) and confirming telemetry would have surfaced them.
Evidence Requirement	Coverage test records mapped to specific threat-catalog entries. **[T3]**
Assurance Expectation	Evidence that coverage validation is repeated as the threat catalog is updated, not a one-time exercise.

Field	Content
Dependencies	Section 5 threat and hazard catalog (this control is only as good as the catalog it's validated against).
Exclusions	Not applicable to threat categories explicitly out of scope for the deployed system class.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	This is the control that most directly proves Domain 7 is not merely a service function for other domains — it is an audit of the collection system's own fitness for purpose.

PAI-SF-TEL-004 — Chain of Custody and Retention

Field	Content
Control ID	PAI-SF-TEL-004
Domain	Runtime Monitoring and Telemetry
Objective	Ensure telemetry is retained and access-controlled sufficiently to support post-incident investigation and potential regulatory or legal proceedings.
Applicability	All physical AI systems, with retention period scaled to consequence severity and applicable jurisdiction.
Implementation Expectation	Defined retention period, access logging, and custody documentation sufficient to establish evidentiary integrity if telemetry is later relied upon outside the organization.
Evidence Requirement	Retention policy; access control logs; custody documentation. **[T2]**
Assurance Expectation	Documentation review confirming retention practice matches policy in actual operation, not only on paper.
Dependencies	PAI-SF-TEL-001 (integrity) — custody is only meaningful if the underlying telemetry is itself tamper-evident.
Exclusions	Retention period requirements may not apply where no jurisdiction imposes a minimum and organizational policy is the only driver.
Maturity/Conformance Relevance	Expected at all conformance levels; sophistication of custody documentation scales with consequence.
Ecosystem Relationship	Supports AI-IRF™'s post-incident investigation and any downstream regulatory or legal process referencing the incident.

Domain 8. Incident Response and Physical Recovery

PAI-SF-IRP-001 — Physical Containment Playbook

Field	Content
Objective	Provide predefined, tested procedures for physically containing an active incident, distinct from digital

Field	Content
	containment.
Applicability	All physical AI systems with a plausible physical-consequence incident scenario.
Implementation Expectation	Playbooks per system class with defined roles and physical isolation steps (e.g., actuator isolation, area securing).
Evidence Requirement	Playbook documentation; tabletop exercise records. **[T2]**
Assurance Expectation	Evidence of tabletop exercises actually exercising the physical steps, not only reviewing the document.
Dependencies	PAI-SF-SAF-001 (safe-state as the first containment action).
Exclusions	Not applicable where physical containment is fully automated with no procedural/human step.
Maturity/Conformance Relevance	Expected at all conformance levels.
Ecosystem Relationship	Extends AI-IRF™'s general response procedures with physical-AI-specific containment; PAI-SF™ does not duplicate AI-IRF™'s general incident management.

PAI-SF-IRP-002 — Volatile Physical Evidence Preservation

Field	Content
Objective	Preserve perishable physical evidence (sensor state, actuator position, environmental conditions at incident time) before it is lost or altered.
Applicability	All physical AI systems where post-incident investigation may be required.
Implementation Expectation	Automated snapshot/freeze of relevant telemetry and physical state upon incident declaration.
Evidence Requirement	Preservation procedure documentation; sample preserved evidence packages. **[T2]**
Assurance Expectation	Test evidence that the snapshot trigger fires reliably at incident declaration, not only in a controlled demonstration.
Dependencies	PAI-SF-TEL-001, PAI-SF-TEL-004 (Domain 7 telemetry integrity and custody) — this control captures the moment; those govern the evidence afterward.
Exclusions	Not applicable to non-safety-relevant telemetry with no investigative value.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	Time-critical capture distinct from Domain 7's ongoing telemetry assurance function.

PAI-SF-IRP-003 — Hardware-in-the-Loop Revalidation Before Return-to-Service

Field	Content
Objective	Ensure systems involved in an incident undergo physical HITL revalidation, not merely software review, before returning to operation.
Applicability	Any physical AI system involved in an incident requiring containment or recovery.
Implementation Expectation	Mandatory HITL test pass tied to root-cause-specific scenarios before reactivation.
Evidence Requirement	Revalidation test records; root-cause traceability. [T2/T3]
Assurance Expectation	Evidence that revalidation scenarios specifically target the identified root cause, not a generic pre-deployment test suite.
Dependencies	PAI-SF-TST-001 (Domain 10, HITL testing methodology) — same testing capability, incident-driven trigger and scope here rather than pre-deployment.
Exclusions	Not applicable to incidents with no plausible physical root cause (e.g., purely administrative/governance incidents).
Maturity/Conformance Relevance	Expected at all conformance levels.
Ecosystem Relationship	The physical-return-to-service requirement AI-IRF™ does not specify on its own — PAI-SF™'s core addition to incident recovery.

Domain 9. Supply Chain and Update Assurance

PAI-SF-SUP-001 — Component Provenance and Bill of Materials

Field	Content
Objective	Establish provenance of safety-relevant hardware, software, and model components.
Applicability	All physical AI systems with safety-relevant components sourced from third parties.
Implementation Expectation	Maintained HBOM/SBOM/AI-BOM with vendor attestations.
Evidence Requirement	BOM records; vendor attestation documents. **[T2]**
Assurance Expectation	Evidence the BOM is maintained current with actual deployed components, not only at initial integration.
Dependencies	PAI-SF-ATT-002 (hardware anti-tamper) — provenance and tamper-detection are complementary.
Exclusions	Not applicable to fully in-house-manufactured components with no third-party sourcing.
Maturity/Conformance Relevance	Expected at all conformance levels.
Ecosystem Relationship	Extends GAISSF™ supply chain controls to physical AI components; provides supply-chain incident categories for UAIF™.

PAI-SF-SUP-002 — Staged Update Rollout with Physical Regression Testing

Field	Content
Objective	Prevent updates (model, firmware) from introducing unsafe physical behavior.
Applicability	All physical AI systems receiving model, firmware, or software updates.
Implementation Expectation	Staged rollout with mandatory physical safety regression testing before fleet-wide deployment.
Evidence Requirement	Regression test results; staged rollout records. [T2/T3]
Assurance Expectation	Evidence that regression testing specifically covers safety-critical behaviors, not only functional correctness.
Dependencies	PAI-SF-ACT-003 (actuator divergence monitoring) may detect update-induced behavior shifts post-deployment as a backstop.
Exclusions	Not applicable to fully air-gapped systems with no update mechanism.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	Extends GAISSF™ change management to physical safety regression; provides update-related incident categories for UAIF™.

PAI-SF-SUP-003 — Third-Party Component Assurance Requirements

Field	Content
Objective	Ensure third-party sensors, actuators, and models meet PAI-SF™ assurance expectations, not only internally developed components.
Applicability	All physical AI systems incorporating third-party safety-relevant components.
Implementation Expectation	Contractual/assurance requirements for suppliers, including vulnerability-notification obligations.
Evidence Requirement	Supplier assurance documentation; contractual clauses. **[T2]**
Assurance Expectation	Evidence suppliers have actually exercised notification obligations at least once (where applicable), not only that the clause exists.
Dependencies	PAI-SF-SUP-001 (provenance) — assurance requirements presuppose known provenance.
Exclusions	Not applicable where no third-party safety-relevant components are used.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	Governs external parties and contractual obligation, distinct from ATT-002's internal hardware-provenance control.

Domain 10. Simulation, Testing, and Validation

PAI-SF-TST-001 — Hardware-in-the-Loop Pre-Deployment Testing

Field	Content
Objective	Ensure safety-critical behaviors undergo HITL testing before deployment and after significant changes.
Applicability	Systems where simulation-trained or simulation-validated behavior has physical safety implications.
Implementation Expectation	HITL testing includes actual deployment hardware; covers safety-critical behaviors from hazard analysis, boundary conditions, and sim-to-real-transfer-challenging scenarios; repeated after significant changes.
Evidence Requirement	HITL test platform documentation; test scenario specifications traced to hazard analysis; test execution records; sim-to-real gap analysis. [T2/T3]
Assurance Expectation	Test coverage analysis demonstrating safety-critical behaviors are tested in HITL configuration, not only simulation.
Dependencies	PAI-SF-SAF-001 (safe-state behaviors are test priorities).
Exclusions	For systems with negligible physical consequence, simulation-only validation may be proportionally acceptable.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels; Foundational expects a documented testing strategy at minimum.
Ecosystem Relationship	Extends GAISSF™ testing controls to physical hardware validation; informs AI-IRF™ post-incident revalidation.

PAI-SF-TST-002 — Adversarial Physical-World Testing

Field	Content
Objective	Test the system against deliberately adversarial physical inputs, not only random/environmental variation.
Applicability	Systems operating in environments where adversarial physical attack is credible.
Implementation Expectation	Controlled red-team physical tests using known techniques from the threat catalog.
Evidence Requirement	Red-team test reports. **[T3]**
Assurance Expectation	Evidence the red-team scope reflects the current threat catalog, refreshed as it updates.
Dependencies	Section 5 threat catalog; PAI-SF-SEN-003 (adversarial input detection is what's being tested here).
Exclusions	Not applicable to fully enclosed, access-controlled environments with no adversarial physical access.
Maturity/Conformance Relevance	Expected at High-Assurance level; Operational level

Field	Content
	may address a narrower published-technique subset.
Ecosystem Relationship	Provides the test evidence underlying UAIF™'s adversarial-perception incident categories before any real incident occurs.

PAI-SF-TST-003 — Simulation-to-Reality Gap Quantification

Field	Content
Objective	Quantify, rather than assume, how well simulated validation predicts real-world behavior.
Applicability	All systems relying on simulation as part of their validation approach.
Implementation Expectation	Systematic comparison of simulated vs. real test outcomes for a sample of scenarios, tracked over time.
Evidence Requirement	Sim-to-real correlation reports. **[T3]**
Assurance Expectation	Evidence the comparison is repeated periodically, not a one-time validation exercise at initial deployment.
Dependencies	PAI-SF-TST-001 (the HITL results being compared against simulation).
Exclusions	Not applicable to systems validated entirely through physical testing with no simulation component.
Maturity/Conformance Relevance	Expected at High-Assurance level.
Ecosystem Relationship	Measures the validity of the testing method itself, distinct from TST-001/002 which test the system.

Domain 11. Functional Safety Interface

PAI-SF-FSI-001 — AI-Specific Hazard Integration into Safety Case

Field	Content
Objective	Ensure existing functional safety hazard analysis explicitly includes AI-specific failure modes (drift, adversarial input, sim-to-real gap), not only traditional hardware faults.
Applicability	All physical AI systems with an existing functional safety case or hazard analysis process.
Implementation Expectation	Joint safety/AI-governance review of hazard analysis against an AI-specific hazard checklist.
Evidence Requirement	Updated hazard analysis; joint review sign-off. **[T2]**
Assurance Expectation	Evidence the checklist is applied at each hazard analysis update, not only once at initial certification.
Dependencies	Section 5 threat and hazard catalog.
Exclusions	Not applicable to systems with no existing functional safety case to integrate with.
Maturity/Conformance Relevance	Expected at all conformance levels where a functional safety case exists.

Field	Content
Ecosystem Relationship	The bridge between GAISSF™ general AI safety and sector-specific functional safety — this domain's core purpose.

PAI-SF-FSI-002 — Control-to-Safety-Integrity Mapping

Field	Content
Objective	Map PAI-SF™ controls to relevant safety integrity level (SIL/ASIL) requirements where applicable, identifying gaps.
Applicability	Systems operating under a SIL/ASIL-based functional safety regime.
Implementation Expectation	Documented mapping exercise per deployed system class.
Evidence Requirement	Mapping documentation; gap register. **[T2]**
Assurance Expectation	Evidence the mapping is reviewed when either the PAI-SF™ control set or the applicable safety standard is updated.
Dependencies	PAI-SF-FSI-001.
Exclusions	Not applicable to systems with no SIL/ASIL-based regime in their sector.
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels for regulated sectors.
Ecosystem Relationship	Legitimate mapping-type control for an interface domain — not a substitute for FSI-004's enforcement mechanism, a complement to it.

PAI-SF-FSI-003 — Functional Safety Standards Gap Register

Field	Content
Objective	Maintain a living record of where functional safety standards do not yet address AI-driven physical behavior.
Applicability	All physical AI systems, particularly those in sectors with mature but AI-naive functional safety regimes.
Implementation Expectation	Gap register updated as new gaps are identified through operation or incident.
Evidence Requirement	Gap register with entries and status. **[T2]**
Assurance Expectation	Evidence the register is actually updated on a defined cadence, not a static document.
Dependencies	PAI-SF-FSI-001, PAI-SF-FSI-002.
Exclusions	None.
Maturity/Conformance Relevance	Expected at all conformance levels.
Ecosystem Relationship	Feeds potential future standards-body input — consistent with ODA3's standards-body-participant audience.

PAI-SF-FSI-004 — Safety Case Exception Escalation

Field	Content
Objective	Ensure AI-driven hazards not covered by the applicable functional-safety standard are escalated, documented, and approved before deployment or continued operation — not silently absorbed as an unaddressed gap.
Applicability	Any system where FSI-003's gap register identifies a hazard not covered by the applicable standard.
Implementation Expectation	Uncovered AI-driven hazards must be escalated to a named safety authority, documented with compensating controls, and formally approved before deployment or continued operation — this is the enforcement action the gap register (FSI-003) exists to trigger.
Evidence Requirement	Escalation records; compensating-control documentation; named-authority approval records. **[T2]**
Assurance Expectation	Evidence that every gap-register entry has either a closed escalation record or an open, tracked status — no gap sits unaddressed without a decision.
Dependencies	PAI-SF-FSI-003 (the gap register that triggers this control).
Exclusions	Not applicable to gaps assessed as having no credible physical consequence.
Maturity/Conformance Relevance	Expected at all conformance levels — this is the enforcement mechanism that makes Domain 11's mapping controls consequential rather than purely documentary.
Ecosystem Relationship	Turns FSI-001/002/003's mapping work into an actual governance decision point — the control that answers "does this domain do anything besides document," per the Founder's requested sharpening.

Domain 12. Physical Operating Environment

PAI-SF-ENV-001 — Operating Environment Specification

Field	Content
Objective	Explicitly define the physical environment (geography, weather, lighting, human traffic patterns) the system is validated to operate within — as a description, not an authorization.
Applicability	All physical AI systems with a defined operating environment.
Implementation Expectation	Documented environment specification.
Evidence Requirement	Environment specification document. **[T2]**
Assurance Expectation	Evidence the specification reflects actual deployment conditions, not only design-intent assumptions.
Dependencies	PAI-SF-AUT-002 (Domain 4) governs *authorization* to

Field	Content
	change what ENV-001 *describes* — see locked boundary paragraph above.
Exclusions	Not applicable to systems with no defined environmental boundary (rare, but possible for fully general-purpose systems).
Maturity/Conformance Relevance	Expected at all conformance levels.
Ecosystem Relationship	Provides the environmental baseline against which Domain 1 (perception) and Domain 4 (autonomy boundaries) validate their own assumptions.

PAI-SF-ENV-002 — Environmental Change Detection and Response

Field	Content
Objective	Detect when actual operating conditions exceed the validated environment specification and respond appropriately.
Applicability	All physical AI systems operating in variable or uncontrolled environments.
Implementation Expectation	Environmental sensing tied to defined thresholds, triggering alert or operational restriction when exceeded.
Evidence Requirement	Environmental monitoring logs; threshold-breach response records. [T2/T3]
Assurance Expectation	Evidence the response is appropriate to the specific environmental factor exceeded, not a single generic response.
Dependencies	PAI-SF-AUT-003 (Domain 4) — the graduated response ladder is owned there; ENV-002 owns detecting the environmental trigger that feeds it.
Exclusions	Not applicable to fully environmentally-controlled operating contexts (e.g., a fixed indoor facility with no variability).
Maturity/Conformance Relevance	Expected at Operational and High-Assurance levels.
Ecosystem Relationship	The detection half of a detection-plus-response pair whose response half lives in Domain 4 — an explicit, intentional cross-domain link, not an accidental overlap, per the locked boundary paragraph.

PAI-SF-ENV-003 — Human Proximity Zone Design and Validation

Field	Content
Objective	Ensure zones where humans may be present are explicitly designed and validated as part of the environment, not incidentally covered by kinematic limits alone.
Applicability	All physical AI systems operating where human presence is possible.
Implementation Expectation	Physical zone design (signage, barriers, sensor

Field	Content
	coverage) validated against actual human traffic patterns.
Evidence Requirement	Zone design documentation; validation walk-through records. [T2]
Assurance Expectation	Evidence the validation reflects observed traffic patterns, not only initial design assumptions.
Dependencies	PAI-SF-ACT-001 (Domain 2, kinematic bounding) — zone design and force/speed limits are complementary, not substitutes for each other.
Exclusions	Not applicable to systems with no possibility of human presence in the operating environment.
Maturity/Conformance Relevance	Expected at all conformance levels for human-accessible environments.
Ecosystem Relationship	A genuinely separate, physical/spatial design question from Domain 2's force/speed bounding — no overlap.

Appendix B: GAISSF™ Domain 9 Supersession Crosswalk

GAISSF™ Domain 9 "Physical AI Safety" (7 controls) is to be retained as a stub in GAISSF™'s published structure, marked superseded/expanded by PAI-SF™ v1.0 — this is a confirmed decision, not yet executed on the live GAISSF Ecosystem documentation as of this report's drafting. Existing citations to GAISSF D9 control numbers will remain valid as historical references once implemented; PAI-SF™'s 12-domain, 41-control register is the authoritative source for physical-AI security content going forward. The control-by-control crosswalk (GAISSF D9.1–D9.7 → specific PAI-SF™ controls) is published separately as **ODA3-2026-07-CSX-PAISF-004**, using the actual GAISSF D9 control statements.