ODA3 INSTITUTE

CONTROL CATALOGUE

Physical AI Security Framework (PAI-SF™ v1.0)

Control Catalogue

ODA3-2026-07-CAT-PAISF-006 | Public-Facing — Standalone Catalogue Draft

Document Control

Doc ID	ODA3-2026-07-CAT-PAISF-006
Doc Type	CAT — Control Catalogue
Framework Tag	PAI-SF™ v1.0
Governing Licence	GAISSF Ecosystem Licence (GEL v1.0)
Access Classification	Open / Public — framework reference content; no Controlled-asset material (scoring, test procedures, auditor guidance, certification decision logic) in this document
Publication State	Publication Candidate — standalone catalogue draft, pending formal publish action
Publish Date / Review Cycle	01 August 2026 (content finalized); publication date to be confirmed
Classification	Public-Facing — Standalone Catalogue Draft

Physical AI Security Framework (PAI-SF™ v1.0)

Control Catalogue

1. Purpose

This catalogue extracts the 41 PAI-SF™ v1.0 controls into a standalone reference artifact. It is intended to make control IDs, titles, objectives, evidence expectations, dependencies, exclusions, and maturity relevance searchable and citable without requiring readers to use the broader Technical + Compliance Report.

The catalogue is a framework reference document. It is not an assessment methodology, scoring rubric, certification decision guide, or sector-specific implementation manual.

2. Catalogue Structure

- Each control record preserves the 10-field structure used in the PAI-SF™ v1.0 Technical + Compliance Report.
- Control IDs and titles are canonical for v1.0 unless superseded by a later formal release note.
- Evidence tags such as **[T2]**, **[T3]**, and **[T4]** indicate evidence-source expectations as carried from the source register; they do not create an independent certification threshold.
- Controlled assets such as scoring thresholds, test procedures, auditor guidance, and certification decision logic are outside this public catalogue boundary.

3. Control Count Verification

Domain	Name	Controls	Focus
D1	Sensor and Perception Integrity	3	Cross-modal validation, calibration monitoring, and adversarial physical-world input detection.

Domain	Name	Controls	Focus
D2	Actuator and Kinetic Command Safety	3	Independent kinematic limits, command-path freshness, and actuator divergence monitoring.
D3	Edge Hardware and Model Attestation	3	Secure boot, hardware provenance, anti-tamper controls, and attestation key custody.
D4	Autonomy Boundaries	4	Autonomy-level governance, operational design domain change control, graduated boundary enforcement, and handoff integrity.
D5	Safe-State and Degraded-Mode Behavior	5	Safe-state definition, model-drift response, compromise-independent triggers, verified fallback behavior, and degraded-mode definition.
D6	Human Override and Intervention	3	Independent override, operator readiness, and controlled reset after intervention.
D7	Runtime Monitoring and Telemetry	4	Telemetry integrity, monitoring independence, coverage validation, and evidentiary retention.
D8	Incident Response and Physical Recovery	3	Physical containment, volatile evidence preservation, and hardware-in-the-loop revalidation before return to service.
D9	Supply Chain and Update Assurance	3	Component provenance, staged update rollout, physical regression testing, and third-party component assurance.
D10	Simulation, Testing, and Validation	3	Hardware-in-the-loop testing, adversarial physical-world testing, and simulation-to-reality gap quantification.
D11	Functional Safety Interface	4	Integration of AI-specific hazards into safety cases, control-to-safety-integrity mapping, standards gap tracking, and exception escalation.
D12	Physical Operating Environment	3	Operating-environment specification, environmental change detection, and human

Domain	Name	Controls	Focus
			proximity zone design.

Total: 12 domains / 41 controls.

4. Control Catalogue

Domain 1. Sensor and Perception Integrity

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

Field	Content
Domain	Domain 1. Sensor and Perception Integrity
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 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
Domain	Domain 1. Sensor and Perception Integrity
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.

Field	Content
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
Domain	Domain 1. Sensor and Perception Integrity
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.
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
Domain	Domain 2. Actuator and Kinetic Command Safety
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
Domain	Domain 2. Actuator and Kinetic Command Safety
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.

Field	Content
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
Domain	Domain 2. Actuator and Kinetic Command Safety
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
Domain	Domain 3. Edge Hardware and Model Attestation
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

Field	Content
	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
Domain	Domain 3. Edge Hardware and Model Attestation
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.
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
Domain	Domain 3. Edge Hardware and Model Attestation
Objective	Protect attestation/signing keys from extraction or misuse.
Applicability	All systems relying on cryptographic attestation (ATT-001, TEL-001).

Field	Content
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
Domain	Domain 4. 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]**
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

Field	Content
	escalation is a suspected root cause.

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

Field	Content
Domain	Domain 4. 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
Domain	Domain 4. 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

Field	Content
	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
Domain	Domain 4. 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]
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
Domain	Domain 5. 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
Domain	Domain 5. 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

Field	Content
	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
Domain	Domain 5. 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]
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
Domain	Domain 5. 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 degradation) — another concrete instance of the framework's core justification.

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

Field	Content
Domain	Domain 5. 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

Field	Content
	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
Domain	Domain 6. Human Override and Intervention
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 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
Domain	Domain 6. Human Override and Intervention
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
Domain	Domain 6. Human Override and Intervention
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 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
Domain	Domain 7. 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
Domain	Domain 7. 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

Field	Content
	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
Domain	Domain 7. 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.
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
Domain	Domain 7. 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

Field	Content
	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
Domain	Domain 8. Incident Response and Physical Recovery
Objective	Provide predefined, tested procedures for physically containing an active incident, distinct from digital 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
Domain	Domain 8. Incident Response and Physical Recovery
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
Domain	Domain 8. Incident Response and Physical Recovery
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.

Field	Content
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
Domain	Domain 9. Supply Chain and Update Assurance
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
Domain	Domain 9. Supply Chain and Update Assurance
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.

Field	Content
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
Domain	Domain 9. Supply Chain and Update Assurance
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
Domain	Domain 10. Simulation, Testing, and Validation
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

Field	Content
	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
Domain	Domain 10. Simulation, Testing, and Validation
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 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
Domain	Domain 10. Simulation, Testing, and Validation
Objective	Quantify, rather than assume, how well simulated validation predicts real-world behavior.

Field	Content
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
Domain	Domain 11. Functional Safety Interface
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.
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
Domain	Domain 11. Functional Safety Interface

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
Domain	Domain 11. Functional Safety Interface
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
Domain	Domain 11. Functional Safety Interface

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
Domain	Domain 12. Physical Operating Environment
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 change what ENV-001 *describes* — see locked boundary paragraph above.

Field	Content
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
Domain	Domain 12. Physical Operating Environment
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
Domain	Domain 12. Physical Operating Environment
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.

Field	Content
Implementation Expectation	Physical zone design (signage, barriers, sensor 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.