ODA3 INSTITUTE

FRAMEWORK STANDARD

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

Framework Standard

ODA3-2026-07-NOR-PAISF-005 | Public-Facing — Normative Core Separation Draft

### Document Control

| | |
|---|---|
| **Doc ID** | ODA3-2026-07-NOR-PAISF-005 |
| **Doc Type** | NOR — Framework Standard |
| **Framework Tag** | PAI-SF™ v1.0 |
| **Audience** | CISOs, Security Architects, AI Governance Leads, Compliance Officers, Standards Body Participants |
| **Governing Licence** | GAISSF Ecosystem Licence (GEL v1.0) |
| **Access Classification** | Open / Public — normative core content; no Controlled-asset material in this document |
| **Publication State** | Publication Candidate — normative core separation draft, pending formal publish action |
| **Publish Date / Review Cycle** | 01 August 2026 (content finalized); publication date to be confirmed |
| **Classification** | Public-Facing — Normative Core Separation Draft |

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

### Framework Standard

## 1. Purpose and Position

This standard defines the normative core of PAI-SF™ v1.0: scope, terminology, principles, domain architecture, control taxonomy, lifecycle expectations, evidence model, and conformance boundary. It separates the framework standard from the broader Technical + Compliance Report narrative so future catalogues, guides, crosswalks, and sector notes can cite a stable normative source.

PAI-SF™ is the Physical AI Security Framework. It defines the AI-security assurance layer for systems that sense, decide, and physically act. It is part of the GAISSF Ecosystem alongside GAISSF™, UAIF™, and AI-IRF™.

## 2. Scope

PAI-SF™ applies to AI-enabled systems 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.

- **In scope:** robotics, autonomous vehicles, drones and uncrewed systems, AI-driven industrial automation, warehouse and logistics automation, medical robotics, smart infrastructure with AI-driven actuation, embodied AI agents, and AI-connected IoT with physical effectors.
- **Out of scope:** pure software AI risk unless the AI system output 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 yes, PAI-SF™ may apply.

## 3. 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.

Where security failures have safety consequences, that intersection is addressed through Domain 11, Functional Safety Interface. PAI-SF™ does not claim safety ownership.

## 4. Definition of Physical AI

Physical AI means an AI-enabled system whose model outputs directly or indirectly cause, guide, constrain, or fail to prevent physical action where erroneous, adversarial, or degraded operation could result in kinetic harm, infrastructure damage, operational disruption, or human injury.

## 5. Core Principles

- 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
- Kinetic Zero Trust

Kinetic Zero Trust is the 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.

## 6. Domain Architecture

PAI-SF™ v1.0 contains 12 domains and 41 controls. The domain architecture is canonical for v1.0 and should be used consistently in framework pages, control catalogues, evidence guides, sector notes, and assessment material.

| Domain | Name | Normative Coverage |
|---|---|---|
| 1 | Sensor and Perception Integrity | Cross-modal validation, calibration monitoring, and adversarial physical-world input detection. |
| 2 | Actuator and Kinetic Command Safety | Independent kinematic limits, command-path freshness, and actuator divergence monitoring. |

| Domain | Name | Normative Coverage |
|---|---|---|
| 3 | Edge Hardware and Model Attestation | Secure boot, hardware provenance, anti-tamper controls, and attestation key custody. |
| 4 | Autonomy Boundaries | Autonomy-level governance, operational design domain change control, graduated boundary enforcement, and handoff integrity. |
| 5 | Safe-State and Degraded-Mode Behavior | Safe-state definition, model-drift response, compromise-independent triggers, verified fallback behavior, and degraded-mode definition. |
| 6 | Human Override and Intervention | Independent override, operator readiness, and controlled reset after intervention. |
| 7 | Runtime Monitoring and Telemetry | Telemetry integrity, monitoring independence, coverage validation, and evidentiary retention. |
| 8 | Incident Response and Physical Recovery | Physical containment, volatile evidence preservation, and hardware-in-the-loop revalidation before return to service. |
| 9 | Supply Chain and Update Assurance | Component provenance, staged update rollout, physical regression testing, and third-party component assurance. |
| 10 | Simulation, Testing, and Validation | Hardware-in-the-loop testing, adversarial physical-world testing, and simulation-to-reality gap quantification. |
| 11 | Functional Safety Interface | Integration of AI-specific hazards into safety cases, control-to-safety-integrity mapping, standards gap tracking, and exception escalation. |
| 12 | Physical Operating Environment | Operating-environment specification, environmental change detection, and human proximity zone design. |

# 7. Canonical Control Index

This section establishes the canonical v1.0 control identifiers and titles. Full 10-field control details should be maintained in the standalone Control Catalogue.

## D1. Sensor and Perception Integrity

| Control ID | Control Title |
|---|---|
| PAI-SF-SEN-001 | Multi-Modal Cross-Validation |
| PAI-SF-SEN-002 | Sensor Health and Calibration Monitoring |
| PAI-SF-SEN-003 | Adversarial Physical-World Input Detection |

## D2. Actuator and Kinetic Command Safety

| Control ID | Control Title |
|---|---|
| PAI-SF-ACT-001 | Independent Kinematic Bounding |
| PAI-SF-ACT-002 | Command Replay and Timing Attack Protection |
| PAI-SF-ACT-003 | Actuator Health and Divergence Monitoring |

## D3. Edge Hardware and Model Attestation

| Control ID | Control Title |
|---|---|
| PAI-SF-ATT-001 | Secure Boot and Model Attestation |
| PAI-SF-ATT-002 | Hardware Provenance and Anti-Tamper |
| PAI-SF-ATT-003 | Attestation Key Custody |

## D4. Autonomy Boundaries

| Control ID | Control Title |
|---|---|
| PAI-SF-AUT-001 | Autonomy Level Definition and Escalation Governance |
| PAI-SF-AUT-002 | Operational Design Domain as a Living Artifact |
| PAI-SF-AUT-003 | Graduated Boundary Enforcement |
| PAI-SF-AUT-004 | Autonomy Handoff Integrity |

## D5. Safe-State and Degraded-Mode Behavior

| Control ID | Control Title |
|---|---|
| PAI-SF-SAF-001 | Safe-State Definition for Credible Failure Modes |
| PAI-SF-SAF-002 | Model-Drift-Triggered Safe-State |
| PAI-SF-SAF-003 | Safe-State Trigger Independence Under Compute Compromise |
| PAI-SF-SAF-004 | Verified Fallback Control Law for Degraded Mode |
| PAI-SF-DEG-001 | Degraded Operating Mode Definition |

## D6. Human Override and Intervention

| Control ID | Control Title |
|---|---|
| PAI-SF-HOV-001 | Independent Physical Override |
| PAI-SF-HOV-002 | Override Familiarity and Operator Training |
| PAI-SF-HOV-003 | Post-Override Reset Integrity |

## D7. Runtime Monitoring and Telemetry

| Control ID | Control Title |
|---|---|
| PAI-SF-TEL-001 | Telemetry Integrity and Tamper-Evidence |
| PAI-SF-TEL-002 | Telemetry Independence from the Monitored System |
| PAI-SF-TEL-003 | Coverage Validation Against the Threat Model |
| PAI-SF-TEL-004 | Chain of Custody and Retention |

## D8. Incident Response and Physical Recovery

| Control ID | Control Title |
|---|---|
| PAI-SF-IRP-001 | Physical Containment Playbook |
| PAI-SF-IRP-002 | Volatile Physical Evidence Preservation |
| PAI-SF-IRP-003 | Hardware-in-the-Loop Revalidation Before Return-to-Service |

## D9. Supply Chain and Update Assurance

| Control ID | Control Title |
|---|---|
| PAI-SF-SUP-001 | Component Provenance and Bill of Materials |
| PAI-SF-SUP-002 | Staged Update Rollout with Physical Regression Testing |
| PAI-SF-SUP-003 | Third-Party Component Assurance Requirements |

## D10. Simulation, Testing, and Validation

| Control ID | Control Title |
|---|---|
| PAI-SF-TST-001 | Hardware-in-the-Loop Pre-Deployment Testing |
| PAI-SF-TST-002 | Adversarial Physical-World Testing |
| PAI-SF-TST-003 | Simulation-to-Reality Gap Quantification |

## D11. Functional Safety Interface

| Control ID | Control Title |
|---|---|
| PAI-SF-FSI-001 | AI-Specific Hazard Integration into Safety Case |
| PAI-SF-FSI-002 | Control-to-Safety-Integrity Mapping |
| PAI-SF-FSI-003 | Functional Safety Standards Gap Register |
| PAI-SF-FSI-004 | Safety Case Exception Escalation |

### D12. Physical Operating Environment

| Control ID | Control Title |
|---|---|
| PAI-SF-ENV-001 | Operating Environment Specification |
| PAI-SF-ENV-002 | Environmental Change Detection and Response |
| PAI-SF-ENV-003 | Human Proximity Zone Design and Validation |

## 8. 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 and update management → Third-party component changes → Decommissioning.

## 9. Evidence and Assurance Model

PAI-SF™ uses six evidence categories: design evidence, test evidence, runtime evidence, incident evidence, third-party evidence, and governance evidence. Evidence quality is assessed by relevance, independence, currency, coverage, and integrity.

Assurance under PAI-SF™ is evidence-based and continuous. It is not a one-time assertion of safety, security, regulatory compliance, or certification.

## 10. Conformance and Maturity Boundary

PAI-SF™ v1.0 uses a per-domain maturity model as the recommended direction. It does not establish an organization-wide certification tier in this standard, and ODA3 Institute does not currently claim certification authority for PAI-SF™.

Future scoring thresholds, maturity rubrics, test procedures, auditor guidance, scenario libraries, and certification decision logic are Controlled assets and must not be inferred from this public/normative core alone.

## 11. GAISSF Domain 9 Disposition

GAISSF™ Domain 9, Physical AI Safety, is to be retained for citation continuity and marked superseded/expanded by PAI-SF™ once the GAISSF Ecosystem documentation is updated. This standard does not claim that the live GAISSF page has already been changed.

This standard does not reproduce individual GAISSF D9 control names or statements here; the verified control-by-control mapping is published separately as **ODA3-2026-07-CSX-PAISF-004**.

## 12. Open and Controlled Asset Boundary

- **Open under GEL v1.0:** terminology, domain names and identifiers, control ID taxonomy, high-level control objectives, basic threat taxonomy, incident vocabulary alignment, and optional interoperability vocabulary.

- **Controlled:** official scoring thresholds, detailed assessment methodology, auditor guidance, red-team procedures, high-risk scenario libraries, and certification decision logic.
- Commercial use, implementation delivery, audit, assessment, and certification services remain subject to ODA3's licensing and partner-empanelment model.

## 13. Notably Absent

- No claim of regulatory recognition or endorsement by a regulator, standards body, laboratory, or certification authority.
- No claim that PAI-SF™ eliminates physical harm or provides a universal safety guarantee.
- No claim of proprietary physical-AI incident telemetry or client incident datasets.
- No claim that the 41 controls have been field-validated across all robotics, vehicle, drone, medical, industrial, or infrastructure contexts.
- No claim that PAI-SF™ replaces functional safety, product safety, hazard analysis, or sector-specific safety certification.

## Appendix A. Document Creation Backlog Position

This Framework Standard closes Tier 1 item 1 in the PAI-SF documentation maturity backlog. Remaining Tier 1 items are the standalone Executive Summary, Quick Start Guide, and Control Catalogue.

| Field | Value |
|---|---|
| Completed before this document | TCR-PAISF-001, EXB-PAISF-001, MEMO-PAISF-002, MEMO-PAISF-003, CSX-PAISF-004 |
| This document | NOR-PAISF-005 — Framework Standard |
| Recommended next document | CAT-PAISF-006 — Standalone Control Catalogue |