ODA3 INSTITUTE

WHP — ADOPTION / PRACTITIONER FIELD GUIDE

Securing Physical AI

A Practitioner's Guide to PAI-SF™

Document ID: ODA3-2026-08-WHP-COM-014
Classification: Public — GEL v1.0

© ODA3 Pvt Ltd

ODA3 INSTITUTE

SECURING PHYSICAL AI

A Practitioner's Guide to PAI-SF™

Doc ID: ODA3-2026-08-WHP-COM-014
Document Type: WHP — Adoption / Practitioner Field Guide
Framework: PAI-SF™ v1.0 — Physical AI Security Framework
Classification: Public — GEL v1.0
Status: FINAL
Date: 12 August 2026
Audience: CISOs, Security Architects, AI Security Teams, OT/ICS Security Teams, Robotics and Autonomy Engineers, Safety Engineers, AI Governance Leads, Risk and Compliance Teams, Incident Responders, Product Security Teams, Assurance Teams

---

Publication Boundary

This is an informative practitioner field guide. It does not create, modify, replace, or supersede the controlled requirements of PAI-SF™ v1.0. The current canonical public structure is **12 domains and 41 controls**. Assessment scoring thresholds, detailed auditor procedures, certification decision logic, controlled test protocols, and other controlled assurance assets are outside this guide.

PAI-SF™ v1.0 is a Final Publication. Final Publication establishes the current controlled framework architecture and control catalogue; it does not imply universal sector calibration, regulator recognition, product approval, completed field validation, or certification availability.

This guide does not establish product-safety approval, functional-safety certification, regulatory compliance, deployment authorization, airworthiness, roadworthiness, medical-device approval, machinery conformity, or any equivalent sector determination.

Where this guide conflicts with a current controlled PAI-SF source, the controlled source governs.

Numbering note: Letter-suffixed sections (for example 44G–44I and 52A–52C) are stable publication identifiers introduced during controlled expansion of the guide. They are retained to preserve existing citations and do not indicate normative hierarchy or subordination to the preceding numbered section.

Methodology Note

This guide translates the PAI-SF™ v1.0 architecture into practitioner workflows for systems in which AI-enabled sensing, inference, autonomy, or command can create or materially influence physical consequence. It is grounded in the controlled/current PAI-SF source family, including ODA3-2026-07-NOR-PAISF-005 (Framework Standard), ODA3-2026-07-CAT-PAISF-006 (Control Catalogue), ODA3-2026-07-CSX-PAISF-004 (GAISSF Domain 9 to PAI-SF Control Crosswalk), ODA3-2026-07-QSG-PAISF-008 (Quick Start Guide), the current readiness artifacts, and ODA3-2026-07-SCP-PAISF-012 through -016 for robotics/industrial automation, autonomous vehicles/mobility, drones/uncrewed systems, medical robotics, and smart infrastructure/physical AI.

ODA3 Institute was founded in March 2026. This guide therefore does not claim proprietary deployment history, independently measured incident-reduction outcomes, assessor-consistency results, or field validation across all physical-AI sectors. Examples are illustrative unless explicitly identified otherwise.

Executive Summary

Physical AI changes the assurance problem because an AI output can become a physical command.

For a conventional software AI system, an unsafe or compromised output may expose information, mislead a user, trigger an unauthorized transaction, or corrupt a downstream workflow. Those consequences can be serious. In a physical-AI system, the same classes of failure can also propagate through motors, brakes, steering, manipulators, valves, pumps, medical instruments, access barriers, flight controls, mobile platforms, or other actuators.

The practitioner problem is therefore not merely whether a model is secure. It is whether the **complete pathway from perception to physical effect is controlled, observable, interruptible, and recoverable**.

This guide uses the following operational chain:

physical environment → sensing/perception → inference → autonomy boundary → command → actuation → physical effect → monitoring → intervention → recovery → revalidation

PAI-SF™ v1.0 structures that problem into **12 domains and 41 controls**. The domains address sensor and perception integrity; actuator and kinetic command safety; edge hardware and model attestation; autonomy boundaries; safe-state and degraded-mode behavior; human override and intervention; runtime monitoring and telemetry; incident response and physical recovery; supply-chain and update assurance; simulation, testing and validation; functional-safety interface; and the physical operating environment.

The framework does not replace functional safety. It provides an AI-security assurance layer that must interface with the safety, engineering, product, regulatory, operational and sector processes that already control the system.

The core practitioner rule

Do not assess the model in isolation when the risk travels through a physical consequence path.

For every material physical-AI function, practitioners should be able to answer:

What does the system perceive?

What assumptions make that perception trustworthy?

What decisions can the AI influence or make?

What autonomy is actually authorized?

What physical commands can result?

What independently constrains those commands?

What happens when perception, inference, communications or compute degrades?

Can a human intervene effectively under real operating conditions?

Can the system enter a credible safe or degraded state?

Is the complete sequence reconstructable from evidence?

Does an update change the physical safety/security case?

Who has authority to return the system to service?

How to Use This Guide

This guide is designed for task-based use; it does not need to be read sequentially.

If your immediate question is…	Start with…
Does PAI-SF apply to this system?	Sections 1–5 and Section 20
How do we scope a pilot?	Sections 18–21 and 44G
How do we architect the physical consequence path?	Sections 22–28
What evidence and telemetry do we need?	Sections 21 and 29–34
How do we respond to a physical-AI incident?	Sections 35–39
How do sector safety processes interact with PAI-SF?	Sections 4, 16, and 40–44
Which worksheets should we use?	Sections 47–52C
What does this guide explicitly not establish?	Section 58

PART I — UNDERSTAND THE PHYSICAL-AI ASSURANCE PROBLEM

1. What Counts as Physical AI?

For this guide, physical AI means an AI-enabled system in which sensing, inference, autonomy, recommendation, orchestration, or control can create or materially influence a physical-world consequence.

Materiality in this guide: “Material” refers to a change, event, dependency, condition, or evidence gap that could reasonably alter the physical consequence path, the safety/security case, the applicability or effectiveness of a control, or the evidence basis for an assurance or return-to-service decision. Organizations should define context-appropriate materiality thresholds according to system consequence, sector obligations, and their own governance processes.

Typical examples include:

industrial robots and collaborative robots;

autonomous mobile robots and warehouse automation;

autonomous or highly automated vehicles;

drones and uncrewed systems;

medical and assistive robotics;

AI-enabled machinery and production systems;

smart-building and infrastructure control;

AI-enabled access, traffic, utility, facility, or environmental control;

edge AI systems that issue or materially shape physical commands.

A system does not need to be fully autonomous to be in scope. Human approval can reduce or change risk, but a nominal human-in-the-loop does not remove the physical consequence path.

1.1 Physical consequence test

Ask four questions:

Question	If Yes
Can an AI output directly cause actuation?	Physical-AI scoping is indicated.
Can an AI recommendation predictably drive a human or automated controller into physical action?	Assess indirect command influence and meaningful human authority.
Can corrupted sensing or inference defeat a safety-relevant operating assumption?	Assess perception, environment and safe-state controls.
Can an AI/model/update change require revalidation of a physical safety or operational case?	Assess update, testing and functional-safety interfaces.

If all four are no, PAI-SF may not be the primary framework. Record the reasoning rather than assuming applicability from the presence of AI alone.

2. The Kinetic Gap

The kinetic gap is the assurance gap between an AI decision and its physical effect.

A technically valid inference can still become unsafe because:

the sensor input was spoofed, stale, occluded or miscalibrated;

the system operated outside its declared environment;

the autonomy level changed without authorization;

a command was replayed or arrived too late;

the actuator behaved differently from the commanded state;

the fallback mechanism depended on the same compromised compute path;

a human override existed but was inaccessible, slow, ambiguous or ineffective;

an update altered behavior without adequate physical regression testing;

telemetry captured the model output but not the command, actuator state or operator intervention.

Security therefore has to follow the complete physical consequence path.

3. PAI-SF™ and GAISSF™ Domain 9

GAISSF™ Domain 9 is retained for citation continuity. PAI-SF™ **supersedes and expands the physical-AI security-assurance layer** represented by D9. **Supersession applies to that assurance layer only; GAISSF™ Domain 9 remains retained for citation continuity under ODA3-2026-07-CSX-PAISF-004. This does not create two independent assurance obligations.** This is not a deletion of D9 and does not mean every historical GAISSF reference has already been rewritten.

Practitioner interpretation:

GAISSF™ remains the broader AI safety/security governance baseline where applicable.

GAISSF D9 remains a continuity reference for physical-AI considerations.

PAI-SF™ is the dedicated physical-AI framework for deeper scoping, control implementation and evidence.

UAIF™ can structure incident classification where an AI incident record is required.

AI-IRF™ can structure incident response.

Do not count GAISSF D9 and PAI-SF controls as two independent assurance obligations merely because both are referenced. Determine the applicable architecture and preserve the mapping.

Practitioner decision rule: For physical-AI security assurance, use **PAI-SF™ controls as the primary control set**. Reference **GAISSF™ Domain 9** only for citation continuity, historical mapping, or crosswalk purposes. Do not implement or assess both as independent physical-AI control obligations.

Domain-numbering note: GAISSF™ Domain 9 and **PAI-SF™ Domain 9** are different framework domains. GAISSF™ Domain 9 is the legacy physical-AI continuity reference discussed above. **PAI-SF™ Domain 9 is Supply Chain and Update Assurance.** Where “D9” is used in this guide, the framework name should be stated when ambiguity is possible.

4. PAI-SF™ Is Not Functional Safety

PAI-SF interfaces with functional safety; it does not replace it.

Functional-safety processes may determine acceptable hazards, safety integrity, safe states, validation requirements, product approvals, operational restrictions, and return-to-service authority. PAI-SF asks whether AI-security conditions can invalidate or undermine those assumptions.

Examples:

A safety case assumes a sensor is trustworthy; PAI-SF asks how spoofing, calibration drift and sensor disagreement are detected.

A machine safety function assumes a stop command is independent; PAI-SF asks whether compromise of shared compute can defeat the trigger.

A vehicle ODD defines weather and road conditions; PAI-SF asks how environmental changes are detected and how autonomy changes when the boundary is approached.

A medical robot has a validated procedure; PAI-SF asks whether a model, firmware or dependency update changes the behavior relevant to that validation.

Safety authority remains with the applicable sector process.

Standards and regulatory boundary: This guide does not establish conformity, equivalence, or compliance mapping between PAI-SF™ and any regulation, directive, standard, or sector assurance regime. Organizations subject to instruments such as AI regulation, cybersecurity standards, management-system standards, industrial-control standards, functional-safety standards, medical-device requirements, automotive safety regimes, aviation requirements, or other sector obligations should perform the applicable mapping and conformity analysis under the appropriate legal, engineering, safety, and standards expertise. PAI-SF's interface with those processes does not constitute conformity assessment against them.

5. Threat, Hazard and Failure Are Related but Different

Physical-AI teams should avoid collapsing three distinct concepts:

Threat: an actor, condition or attack mechanism capable of exploiting a weakness.

Hazard: a condition with potential to cause harm.

Failure: loss or degradation of intended function, whether malicious, accidental or systemic.

A spoofed camera target may be a threat-driven event. A calibration drift may be a non-malicious failure. Both can create the same hazardous physical state. PAI-SF implementation should therefore support security evidence without requiring malicious intent to be established before protective action.

PART II — THE 12-DOMAIN OPERATING MODEL

6. Domain 1 — Sensor and Perception Integrity

Control count: 3.

Purpose: establish confidence that physical-world inputs are sufficiently trustworthy for the decisions and actions that depend on them.

Practitioner focus:

cross-modal validation where a single sensor failure could cause unsafe action;

sensor health and calibration monitoring;

adversarial physical-world input detection where credible.

Evidence starters:

sensor inventory and trust assumptions;

calibration baselines and drift records;

disagreement thresholds;

degraded-perception behavior;

adversarial-input test results;

examples of sensor exclusion or confidence reduction;

mapping between perception confidence and allowed actuation.

Notably absent: A second sensor is not automatically independent. Two sensors can share power, calibration logic, firmware, mounting, environmental blind spots or common preprocessing.

7. Domain 2 — Actuator and Kinetic Command Safety

Control count: 3.

Purpose: constrain the physical command path independently of the model's apparent correctness.

Practitioner focus:

independent kinematic or operational limits;

command freshness and replay/timing protection;

divergence monitoring between commanded and observed actuator state.

Questions:

What is the maximum command the AI can issue?

Where is that bound enforced?

Is enforcement independent of the AI application?

Can stale or replayed commands execute?

Does the system verify that the actuator did what was commanded?

What happens when command and physical state diverge?

A software limit implemented in the same compromised execution path is not automatically an independent safety/security bound.

8. Domain 3 — Edge Hardware and Model Attestation

Control count: 3.

Purpose: establish provenance and integrity of edge compute, firmware, models and trust anchors involved in physical behavior.

Evidence starters include secure-boot state, device identity, firmware provenance, model/package hashes, attestation results, anti-tamper events and key-custody records.

The practitioner question is not merely "is the model signed?" It is "can we establish what hardware, firmware, model, configuration and dependencies were actually executing when the physical behavior occurred?"

9. Domain 4 — Autonomy Boundaries

Control count: 4.

Purpose: define and enforce the authority the system is allowed to exercise.

The operating design domain or equivalent operating boundary should be treated as a living controlled artifact. Geography, weather envelope, task scope, speed, proximity, payload, route, procedure phase, operator state, communications state and other conditions may determine what autonomy is permitted.

A useful autonomy record contains:

authorized autonomy level;

allowed tasks/actions;

prohibited actions;

operating-domain constraints;

transition triggers;

reduced-autonomy modes;

handoff conditions;

revalidation triggers after boundary change.

10. Domain 5 — Safe-State and Degraded-Mode Behavior

Control count: 5.

Purpose: ensure that loss of confidence does not automatically become uncontrolled physical behavior.

“Stop” is not universally safe. A drone may need to land or loiter. A vehicle may need a minimum-risk maneuver. A surgical robot may need controlled withdrawal. An industrial process may require a sequenced shutdown rather than immediate power loss.

Practitioners should define:

credible safe states;

degraded modes;

entry conditions;

exit conditions;

independent triggers where required;

fallback behavior;

conditions where fallback itself introduces risk.

11. Domain 6 — Human Override and Intervention

Control count: 3.

Purpose: make human authority real rather than nominal.

Evidence of meaningful intervention can include:

physical or logically independent override mechanisms;

operator training and drills;

time-to-intervene measurements where meaningful;

intervention success/failure records;

override accessibility under degraded conditions;

controlled reset after intervention;

evidence that automation does not silently resume unsafe authority.

A human-in-the-loop claim is weak evidence if the operator cannot understand the situation, has insufficient time, lacks authority, or cannot physically interrupt the system.

12. Domain 7 — Runtime Monitoring and Telemetry

Control count: 4.

Purpose: make the physical consequence path observable and reconstructable.

Minimum useful telemetry frequently spans:

sensor state → perception/confidence → model/version → autonomy mode → command → actuator state → safety-controller state → operator action → physical outcome → recovery action

Telemetry should be evaluated for integrity, independence, coverage and evidentiary retention. Logging only the model output is insufficient when the incident question is what the machine actually did.

13. Domain 8 — Incident Response and Physical Recovery

Control count: 3.

Purpose: contain physical risk, preserve volatile evidence and revalidate before return to service.

Physical containment may include isolation, autonomy reduction, speed/power limitation, geofencing, controlled shutdown, manual control, removal from service or activation of a sector-defined safe state.

Evidence preservation should not override immediate human safety. Where safe to do so, preserve sensor state, command history, actuator state, model/configuration, safety-controller events, operator actions and relevant physical evidence before destructive reset or rebuild.

Return to service should require evidence, authority and revalidation appropriate to the system. For material physical-AI incidents, symptom disappearance is not sufficient proof of recovery.

14. Domain 9 — Supply Chain and Update Assurance (PAI-SF™)

Control count: 3.

Purpose: prevent component and update changes from silently altering the physical assurance case.

The supply chain includes more than the model provider. It can include sensors, controllers, firmware, edge hardware, autonomy software, maps, retrieval sources, plugins, libraries, gateways, integrators, remote-management services and safety components.

Every material update should answer:

What changed?

Which physical behaviors could change?

Which previous evidence is invalidated?

What regression tests are required?

Is staged rollout possible?

What is the rollback path?

What provider evidence is available?

What remains unknown?

15. Domain 10 — Simulation, Testing, and Validation

Control count: 3.

Purpose: test physical-AI behavior beyond ordinary software unit and integration testing.

Practitioner methods may include simulation, hardware-in-the-loop testing, adversarial physical-world testing, fault injection, sensor degradation, communications loss, boundary transitions and simulation-to-reality gap analysis.

Simulation evidence should not be treated as equivalent to field evidence without examining the assumptions that separate the simulated environment from the operating environment.

16. Domain 11 — Functional Safety Interface

Control count: 4.

Purpose: ensure that AI-security findings reach the safety processes that depend on the affected assumptions.

Key practices:

route AI-specific hazards and security findings into the applicable safety case;

map security controls to safety-relevant assumptions where appropriate;

track gaps between AI-security evidence and sector safety requirements;

escalate exceptions rather than silently accepting uncovered conditions.

PAI-SF evidence can inform a safety process. It does not itself grant safety approval.

17. Domain 12 — Physical Operating Environment

Control count: 3.

Purpose: make the environment part of the assurance boundary.

Relevant variables can include weather, lighting, floor condition, road condition, electromagnetic conditions, human proximity, obstacles, sterile-field conditions, site geometry, traffic, payload, vibration, temperature, communications availability and other context that affects sensing or action.

A deployment is not adequately scoped if the team cannot state the conditions under which the physical-AI system is expected to operate.

PART III — ADOPT PAI-SF™ WITHOUT CREATING A SECOND SAFETY SYSTEM

18. Adoption Pathway

Use four stages.

Stage 1 — Orientation

Understand the physical consequence path, PAI-SF domains, existing safety/security ownership and evidence sources.

Stage 2 — Pilot

Select one bounded system. Build the system boundary, consequence path, domain applicability record and starter evidence pack. Do not begin with an enterprise-wide control rollout.

Stage 3 — Operational Integration

Map PAI-SF activities into existing engineering, product security, OT security, change management, incident response, safety and GRC workflows.

Stage 4 — Governance Feedback

Use findings, incidents, test results and update evidence to improve GAISSF governance, PAI-SF implementation, sector safety interfaces and organizational decision rights.

19. Minimum Viable Adoption Package

Before a pilot, create at least:

Physical-AI System Scoping Record.

Physical Consequence Path diagram.

Domain Applicability Record.

Evidence Source Register.

Autonomy and Operating-Boundary Record.

Safe/Degraded-State Record.

Human Intervention Record.

Update/Revalidation Trigger Register.

Incident and Recovery Authority Record.

Sector/Safety Interface Record.

These artifacts are adoption aids, not new normative PAI-SF controls.

20. Ten-Question Physical-AI Scoping Record

#	Scoping question	Record
1	What physical system or process can the AI influence?	System/process boundary
2	What sensors or external observations feed the decision path?	Perception sources
3	What decisions can the AI recommend or make?	Decision authority
4	What physical commands can result directly or indirectly?	Command path
5	What is the authorized autonomy level and operating domain?	Autonomy/ODD
6	What independent limits constrain unsafe action?	Safety/security bounds
7	What safe or degraded states exist?	Fallback model

#	Scoping question	Record
8	How can a human intervene, and under what time/authority constraints?	Intervention model
9	What telemetry can reconstruct perception-to-effect behavior?	Evidence/observability
10	Which update, sector, safety or regulatory process controls change and return to service?	Governance interfaces

If the team cannot answer questions 1, 4, 5, 6, 7 or 10, the system is not ready for a confident PAI-SF implementation plan.

21. Evidence Quality

PAI-SF uses six evidence categories:

design evidence;

test evidence;

runtime evidence;

incident evidence;

third-party evidence;

governance evidence.

Evidence quality should be considered through relevance, independence, currency, coverage and integrity.

A useful practitioner evidence ladder is informative rather than normative:

Level	Example	Typical limitation
E1 Representation	policy statement, supplier claim, model card	says what should be true
E2 Configuration/Design Evidence	architecture, configuration export, signed manifest	shows intended/current state but not behavior
E3 Test Evidence	simulation, HIL test, adversarial test, failover exercise	bounded by test coverage and assumptions
E4 Runtime Evidence	tamper-evident telemetry, actuator/safety events, monitored intervention	strongest operational evidence but still context-dependent
E5 Independent Corroboration	independent test, separate safety controller, third-party evidence	independence must itself be established

Do not convert this ladder into an official PAI-SF scoring scale. It is a practitioner aid for evidence planning.

Practitioner note — evidence depth: The E1–E5 ladder is not an assessment score or mandatory evidence threshold. Evidence depth should be proportionate to system consequence, control objective, deployment context and the requirements of the controlled PAI-SF™ assessment

methodology. For higher-consequence systems, representation-only evidence should not ordinarily be treated as sufficient where test, runtime or independently corroborated evidence is reasonably obtainable.

PART IV — ARCHITECT THE PHYSICAL CONSEQUENCE PATH

22. Reference Architecture

Model the system as a chain of trust boundaries rather than a single AI component.

Environment
↓
Sensors / perception inputs
↓
Preprocessing / sensor fusion
↓
Model / inference
↓
Planner / autonomy logic
↓
Policy / command authorization
↓
Controller / actuator interface
↓
Physical actuator
↓
Physical effect

Parallel controls should include:

independent safety controller where applicable;

human override;

telemetry/evidence path;

identity and authorization;

update/provenance chain;

safe/degraded-state mechanism.

23. Trust-Boundary Review

At each boundary ask:

Can data be spoofed?

Can state become stale?

Can authorization be bypassed?

Can commands be replayed?

Can the component be replaced or updated?

Is the component independently observable?

Does failure propagate silently?

Is there a credible containment mechanism?

24. Perception Integrity Pattern

A perception assurance pattern should document:

sensor modalities;

calibration state;

expected environmental envelope;

fusion logic;

disagreement handling;

confidence thresholds;

spoofing/blinding assumptions;

degraded-perception response;

evidence retained when perception is challenged.

The goal is not perfect perception. The goal is to prevent uncertain perception from silently becoming unconstrained physical action.

25. Actuation-Bound Pattern

For every safety-relevant actuator, document:

command source;

allowed range;

rate limits;

freshness/timing requirements;

replay protection;

authorization boundary;

independent constraints;

actual-state feedback;

divergence response.

26. Autonomy-Boundary Pattern

Define autonomy as an explicit operating state, not a marketing label.

Record:

what the system may decide;

what it may execute;

when approval is required;

who can change autonomy;

what conditions force reduced autonomy;

how transitions are logged;

how the system behaves when handoff fails.

27. Meaningful Human Authority

A human override should be evaluated across five dimensions:

Awareness: does the person know intervention is required?

Comprehension: can the person understand the relevant system state?

Time: is there enough time to act?

Authority: is the person permitted to intervene?

Effectiveness: does the intervention actually constrain physical behavior?

Useful evidence includes drills, intervention timing, override success/failure, post-override reset behavior and operator competence records.

28. Safe-State Independence

A safe-state trigger should be reviewed for common-mode compromise. If the AI application, safety trigger and actuator authorization all depend on the same compute, identity, network or configuration path, the apparent fallback may not be independent.

Independence is a system property, not a diagram symbol.

PART V — BUILD OBSERVABILITY FOR PHYSICAL AI

29. The Physical-AI Evidence Chain

For material actions, aim to reconstruct:

what the environment was → what the system sensed → what it inferred → what autonomy state applied → what command was authorized → what actuator did → what safety mechanism did → what the operator did → what physical outcome followed

30. Tiered Observability Baseline

Tier A — Minimum

system/model/version;

timestamp;

autonomy mode;

material command;

operator intervention;

major safety-state transition.

Tier B — Operational

Tier A plus sensor health, perception confidence, authorization context, actuator feedback, degraded-state transitions and update provenance.

Tier C — High-consequence

Tier B plus independent safety-controller events, synchronized multi-source telemetry, tamper-evident custody, detailed physical-state reconstruction and sector-required evidence.

These tiers are implementation guidance, not PAI-SF conformance levels.

31. Evidence Retention Questions

Which telemetry is ephemeral?

Are clocks synchronized?

Can logs be altered by the same administrator or component under investigation?

Are sensor and actuator records correlated?

Are safety-controller events retained separately?

Can provider-side evidence be requested after an event?

Does privacy law constrain retained sensor data?

Is video or biometric evidence involved?

What must be hashed, exported or preserved before reset?

PART VI — UPDATES, SUPPLY CHAIN AND CHANGE

32. The Update Is Part of the Physical Assurance Case

A model update can change perception. A firmware update can change timing. A map update can change route behavior. A dependency update can alter numerical behavior. A provider update can change outputs without changing your local application code.

Therefore, change management should classify updates by potential physical effect rather than by software package type alone.

33. Update Revalidation Decision

Ask:

Does the update change a model, sensor, fusion path, controller, autonomy rule, command path, safety mechanism or operating-domain assumption?

Could it change latency, confidence, fallback or actuator behavior?

Does it invalidate prior simulation/HIL evidence?

Is staged rollout possible?

Can the prior version be restored safely?

Does the applicable safety process require formal revalidation?

34. Third-Party Evidence Ladder

Third-party reliance should be recorded without pretending visibility exists where it does not.

Provider evidence state	Example	Practitioner treatment
Representation only	supplier statement/model card	preserve as provider representation
Assurance artifact	SOC report, safety artifact, SBOM, signed release note	validate scope and currency
Customer-verifiable	customer-run tests, signed hashes, change notice, telemetry	combine with enterprise evidence
Independently corroborated	independent lab/test evidence where applicable	record independence and scope
Unknown	provider-internal facts unavailable	record as unresolved; do not infer absence of issue

Interpretation note: A provider representation is an assertion made by the provider or supplier. An assurance artifact carries additional formal scope, provenance, verification, or independent-assessment context—for example, an independently issued assurance report, signed attestation, or controlled safety artifact. An assurance artifact is not automatically sufficient evidence: practitioners should still evaluate its scope, currency, independence, applicability, exclusions, and relationship to the system under review.

Operational closure may occur with provider unknowns, but the unknowns should remain visible and governed.

PART VII — INCIDENT RESPONSE AND RETURN TO SERVICE

35. First Principle: Human Safety Takes Priority

Evidence preservation is important. Human safety and the applicable engineered safe-state process take priority.

Where an immediate physical hazard exists, authorized teams should act to reduce danger even if the action destroys some evidence. Record what was done, why it was necessary, who authorized it and what evidence was lost or changed.

36. Physical-AI Incident Intake

Capture:

system and location;

current physical state;

people potentially exposed;

autonomy mode;

observed abnormal behavior;

sensor/perception anomalies;

commands and actuator behavior;

current safe/degraded state;

operator intervention;

provider involvement;

latest model/firmware/update;

evidence at risk of loss;

sector/safety authority engaged.

37. Containment Patterns

Containment may include:

reducing autonomy;

restricting speed, power, motion or payload;

geofencing;

isolating a sensor or input source;

disabling a connector or remote command path;

switching to manual control;

entering a validated degraded mode;

controlled shutdown;

removing a unit from service;

fleet-wide hold where shared dependency risk exists.

Do not assume shutdown is always the safest containment action.

38. Evidence Preservation Before Reset

Where safe and authorized, preserve:

sensor state and calibration;

perception/inference outputs;

model/version/hash;

autonomy mode;

command sequence;

actuator feedback;

safety-controller events;

operator actions;

network/identity events;

provider notices;

update history;

physical scene/equipment condition where relevant.

39. Return-to-Service Gate

Return to service should be treated as two linked but distinct decisions:

Track A — Security evidence closure: Has the security condition been contained, understood to the level required for the decision, remediated or explicitly accepted, and validated with appropriate evidence?

Track B — Safety / operational authorization: Has the person or process holding the applicable safety, product, clinical, aviation, road, machinery, infrastructure or operational authority authorized return where such authorization is required?

Security closure does not substitute for Track B, and safety authorization does not erase unresolved security evidence.

A return-to-service decision should answer:

What condition was contained or corrected?

What evidence supports that conclusion?

What remains unknown?

What simulation, HIL, physical or sector validation was performed?

Has the applicable safety authority approved the return where required?

Has change management recorded the remediation?

Is rollback available?

What enhanced monitoring applies?

What condition triggers immediate re-containment?

Who authorized the return?

Restart is not recovery.

PART VIII — SECTOR APPLICATION

40. Robotics and Industrial Automation

Priority themes include actuation bounds, autonomy/handoff, safe/degraded states, human override, telemetry, recovery and OT/safety integration.

Do not treat a robot cell as merely an endpoint. The relevant assurance boundary can include robot controller, PLC, safety PLC, vision system, model runtime, network, cell guarding, human proximity, production orchestration and remote vendor access.

Practitioner first calls

OT/industrial security;

robotics/controls engineering;

machinery/functional safety owner;

production/operations authority;

product/vendor owner where applicable.

Stop-work boundary

Do not bypass engineered safety functions, guarding, lockout/tagout, safety PLC logic or required machinery-safety procedures in order to preserve AI evidence.

41. Autonomous Vehicles and Mobility

Priority themes include perception, localization, ODD, autonomy transitions, steering/braking/throttle command bounds, minimum-risk behavior, telemetry, fleet updates and safety interface.

PAI-SF evidence can support security assurance. It does not grant type approval, roadworthiness or deployment permission.

42. Drones and Uncrewed Systems

Priority themes include sensor/navigation integrity, mission envelope, geofence enforcement, payload control, command freshness, return-to-home/landing behavior, communications loss, fleet/swarm coordination and air/ground/maritime operating constraints.

A hard stop may be impossible in flight. Safe-state logic must reflect the actual vehicle and operating environment.

43. Medical Robotics

Priority themes include patient-proximity perception, supervised autonomy, instrument/motion bounds, clinical workflow state, human intervention, update control, evidence preservation and medical-device/clinical safety interfaces.

This guide does not determine clinical effectiveness or medical-device approval.

44. Smart Infrastructure and Physical AI

Priority themes include distributed sensing, operator authority, bounded commands to infrastructure, degraded modes, edge update assurance, public-space telemetry and integration with facility, OT, emergency-management and public-safety processes.

PART VIII-B — PRACTITIONER ADOPTION SEQUENCE

44G. Minimum Viable PAI-SF Pilot Sequence

The framework suite is not sequential, but a PAI-SF implementation pilot still needs an execution order. The following sequence is **informative practitioner workflow**, not a new PAI-SF lifecycle:

Scope the system and consequence path. Complete the physical-consequence test and six-boundary scoping record.

Determine domain/control applicability. Use the controlled PAI-SF catalogue; do not infer applicability from this guide alone.

Name owners and decision authorities. Include security, engineering, safety, operations, provider and return-to-service authority.

Build the evidence plan. Identify what evidence should demonstrate implementation and operation, where it originates, its owner, retention, integrity protection and known gaps.

Test critical physical assumptions. Prioritize perception, command bounds, autonomy limits, safe/degraded behavior, intervention, telemetry and update/revalidation paths according to the system's consequence path.

Record findings and uncertainty. Separate observed fact, provider representation, analytical assessment, hypothesis, unknown and specifically checked-but-not-observed conditions.

Route remediation into existing engineering/safety processes. PAI-SF findings do not displace authorized safety, clinical, operational or regulatory processes.

Re-test affected controls and assumptions. Do not close a finding merely because a document was updated or a component was restarted.

Record residual gaps and management decisions. Identify unresolved provider dependencies, accepted operational constraints, deferred work and the responsible authority.

Set reassessment triggers. Material model, firmware, sensor, autonomy, environment, provider, safety-case or operating-scope changes should trigger targeted reconsideration.

Pilot output: a management-ready implementation baseline and evidence/gap record. It is not, by itself, a PAI-SF assessment conclusion, conformance determination or certification decision.

44H. Cross-Framework Routing During a PAI-SF Pilot

Framework applicability remains objective-driven and non-sequential.

Use **GAISSF™** where broader lifecycle governance, ownership and assurance controls apply.

Treat **GAISSF™ Domain 9** as a continuity and crosswalk reference only; see Section 3 for the governing D9/PAI-SF relationship.

Use **PAI-SF™** for the deeper physical-AI assurance layer.

If anomalous behavior is observed, preserve evidence and begin proportionate safety/security action without waiting for final classification.

Use **UAIF™** when a structured AI incident/vulnerability classification record is supportable. Do not invent an Observation record type.

Use **AI-IRF™** for incident-response preparation and execution. Classification and response may evolve together rather than operating as a rigid handoff.

44I. Evidence Status Vocabulary

For pilot records, use a small evidence-status vocabulary so teams do not collapse uncertainty:

Status	Meaning
Observed fact	Directly supported by enterprise-observed evidence
Provider representation	Asserted or supplied by a third party; provenance retained
Analytical assessment	Reasoned conclusion supported by identified evidence
Hypothesis	Plausible explanation requiring further evidence
Unknown	Evidence is presently insufficient
Checked — not observed	A defined condition was specifically checked and was not observed within the stated scope/window

Do not conflate evidence status with evidence category or depth. The vocabulary in this section records the **epistemic/provenance status of a claim**—for example, Observed fact, Provider representation, Analytical assessment, Hypothesis, or Unknown. The E1–E5 ladder in Section 21 is a separate practitioner aid describing the **kind and relative operational depth of evidence**. A single item may therefore have both an evidence category and an evidence status.

These labels are practitioner recordkeeping aids. They do not modify PAI-SF, UAIF or AI-IRF controlled terminology.

PART IX — PILOT DESIGN

45. Recommended Two-System Pilot

Use two systems with different consequence paths.

Pilot A — Bounded industrial/robotic system

Choose a system with known ownership, accessible telemetry and a defined safety process. Objective: validate perception-to-actuation mapping, evidence collection, autonomy boundaries and return-to-service governance.

Pilot B — Third-party or update-dependent physical-AI system

Choose a system where a supplier controls a material model, firmware, perception or autonomy component. Objective: validate supply-chain evidence, provider unknowns, update revalidation and contractual escalation.

46. Pilot Success Criteria

Do not define success as “all controls green.” Instead ask:

Was scope reproducible?

Were owners and decision rights clear?

Could the physical consequence path be reconstructed?

Were material evidence gaps visible?

Were provider representations separated from enterprise evidence?

Could teams identify safe/degraded states?

Was human intervention testable?

Were update/revalidation triggers clear?

Did PAI-SF integrate with existing safety/security processes without duplicating them?

Did the pilot produce actionable governance improvements?

PART X — PRACTITIONER WORKSHEETS

47. Physical Consequence Path Worksheet

System:
Owner:
Location/operating environment:
Model/provider/version:
Sensors:
Inference function:
Autonomy state(s):
Command authorization:
Actuators:
Independent limits:
Safe/degraded state:
Human override:
Telemetry:
Physical consequences:
Safety authority:
Return-to-service authority:

48. Domain Applicability Worksheet

Applicability instruction: All 41 PAI-SF™ controls are in scope for applicability consideration. Record each control as Applicable or Not Applicable, with rationale documented in accordance with the controlled Control Catalogue. The domain-level worksheet below is a summary aid and does not replace control-level applicability analysis.

Domain	Applicable?	Rationale	Evidence owner	Key gap
D1 Sensor and Perception Integrity				
D2 Actuator and Kinetic Command Safety				
D3 Edge Hardware and Model Attestation				
D4 Autonomy Boundaries				

Domain	Applicable?	Rationale	Evidence owner	Key gap
D5 Safe-State and Degraded-Mode Behavior				
D6 Human Override and Intervention				
D7 Runtime Monitoring and Telemetry				
D8 Incident Response and Physical Recovery				
D9 Supply Chain and Update Assurance				
D10 Simulation, Testing, and Validation				
D11 Functional Safety Interface				
D12 Physical Operating Environment				

49. Evidence Source Register

Evidence	Source system	Owner	Retention	Integrity protection	Ephemeral?	Gap
Sensor state						
Model/version						
Autonomy state						
Command history						
Actuator feedback						
Safety events						
Operator actions						
Update provenance						
Provider evidence						

50. Human Intervention Test Record

Record scenario, trigger, operator role, information available, time available, intervention mechanism, observed response, success/failure, reset behavior and lessons.

51. Update/Revalidation Record

Record change identifier, changed component, potential physical effect, affected PAI-SF domains, previous evidence invalidated, tests required, staged rollout, rollback, safety review, deployment authority and monitoring conditions.

52. Physical-AI Recovery Gate

Record remediation, validation evidence, residual uncertainty, safety approval, change record, rollback, monitoring, re-containment trigger and return-to-service authority.

52A. Finding and Gap Register

For each material item record:

Field	Record
Finding/gap ID	Unique local reference
System / component	Affected scope
PAI-SF domain/control reference	From the controlled catalogue
Evidence status	Observed fact / provider representation / analytical assessment / hypothesis / unknown / checked-not-observed
Condition found	Concise statement
Physical consequence relevance	Why it matters to the consequence path
Existing safety/security control	Relevant current control or process
Owner	Accountable remediation owner
Action	Planned correction, validation or escalation
Due/review point	Organization-defined
Residual uncertainty	What remains unresolved
Closure evidence	Evidence supporting closure
Closure authority	Organization-defined decision owner

Artifact existence alone does not establish that a finding is closed or that a control is effective.

52B. Pilot Decision and Exception Record

Use this when the pilot encounters an unresolved provider dependency, temporary operational constraint, deferred remediation or exception.

Record the decision, affected system/control, evidence considered, alternatives, physical/safety consequence, compensating measures, owner, approving authority, review trigger, expiry/review point, and conditions requiring immediate reconsideration.

This record documents the organization’s decision process. It does not convert risk acceptance into PAI-SF conformance.

52C. Reassessment Trigger Register

At minimum consider triggers involving:

model, firmware, controller, sensor or actuator change;

autonomy-level or operating-envelope change;

new environment/site/route/clinical-use condition;

provider or dependency change;

safety-case or hazard-analysis change;

material incident, near miss or repeated anomaly;

loss of telemetry or evidence integrity;

new vulnerability affecting a consequence-path component;

changed legal, regulatory or sector authorization condition.

Record the trigger, affected scope, responsible owner, required review/test, and completion evidence.

PART XI — GOVERNANCE AND INTEGRATION

53. Role and Decision Rights

Role	Primary responsibility
CISO / Security leadership	Security governance, escalation and integration
Security/OT architect	Physical consequence path and trust boundaries
AI/model owner	Model behavior, version and inference evidence
Robotics/controls engineer	Command path, controller and actuator behavior
Safety owner	Safety-case authority and sector safety process
Operations owner	Operating conditions and service authority
SOC/CSIRT	Detection, triage, evidence and incident coordination
Platform/edge engineering	Runtime, identity, compute and update evidence
Vendor/procurement owner	Provider evidence and contractual escalation
Legal/privacy/compliance	Legal and regulatory determination
Executive risk authority	Material residual-risk decisions where assigned

One person may hold multiple roles in a small team. Role consolidation does not make missing authority or evidence disappear.

54. Existing-System Integration

PAI-SF adoption should reuse existing systems where possible:

CMDB/asset inventory for system identity;

GRC for control/evidence references;

SIEM/data platform for telemetry;

SOAR/case management for incident workflow;

PLM/ALM/change management for updates;

safety management system for hazard/safety evidence;

CI/CD or fleet-management system for deployment gates;

vendor management for provider evidence.

No proprietary ODA3 tooling is required by this guide.

55. Automation Boundary

Automation can support evidence collection, field validation, version comparison, telemetry correlation, routing, deterministic checks and test orchestration.

Automation should not by itself establish:

product safety;

legal reportability;

regulatory compliance;

root cause;

negligence or liability;

adequacy of a safe state;

meaningful human authority;

return-to-service approval;

PAI-SF certification or conformance.

PART XII — METRICS AND CONTINUOUS IMPROVEMENT

56. Metrics With Care

Useful metrics may include:

percentage of in-scope systems with documented physical consequence paths;

percentage with named safety and recovery authority;

telemetry coverage of perception-to-effect chain;

unresolved evidence gaps;

update changes requiring revalidation;

intervention-test completion;

time to establish safe/degraded state;

incidents closed with provider unknowns;

repeated failures by mechanism;

corrective actions completed and validated.

Avoid metrics that reward premature closure or equate low incident count with proven safety.

57. Governance Feedback Loop

finding/incident → affected assumption → affected PAI-SF control/process → owner → corrective action → validation → policy/runbook/control update → scheduled effectiveness review

Where the lesson affects broader AI governance, feed it into GAISSF™. Where an incident requires structured classification or response, use UAIF™ and AI-IRF™ according to their applicable boundaries.

PART XIII — NOTABLY ABSENT

58. What This Guide Does Not Establish

This guide does not establish or imply:

that PAI-SF™ eliminates physical harm;

universal safety guarantees;

regulator or standards-body recognition;

product certification;

PAI-SF certification availability;

regulatory compliance;

functional-safety certification;

airworthiness, roadworthiness or deployment authorization;

medical-device approval or clinical effectiveness;

machinery or infrastructure approval;

sector-specific scoring thresholds;

official assessment methodology or auditor procedure;

universal safe-state definitions;

universal autonomy levels;

universal human-intervention timing thresholds;

that a human-in-the-loop is automatically meaningful oversight;

that simulation proves field safety;

that a provider representation is independently verified evidence;

that absence of provider evidence proves absence of a provider-side issue;

that restarting a system proves recovery;

that a successful test proves control effectiveness in every operating condition;

that the 41 controls have been independently field-validated across all sectors;

proprietary incident telemetry or client outcome data;

guaranteed interoperability with any specific SIEM, SOAR, GRC, PLM, ALM, OT, safety or fleet-management product.

Acronym Reference

AI-IRF™ — AI Incident Response Framework **ALM** — Application Lifecycle Management **CMDB** — Configuration Management Database **CSIRT** — Computer Security Incident Response Team **GAISSF™** — Global AI Safety and Security Framework **GEL** — GAISSF Ecosystem License **GRC** — Governance, Risk and Compliance **HIL** — Hardware-in-the-Loop **ICS** — Industrial Control Systems **ODD** — Operational Design Domain **OT** — Operational Technology **PAI-SF™** — Physical AI Security Framework **PLC** — Programmable Logic Controller **PLM** — Product Lifecycle Management **SBOM** — Software Bill of Materials **SIEM** — Security Information and Event Management **SOAR** — Security Orchestration, Automation and Response **SOC** — Security Operations Center **UAIF™** — Unified AI Incident Framework

---

PART XIV — SOURCE AND TRACEABILITY BOUNDARY

59. Controlling and Supporting Source Hierarchy

This practitioner guide is informative. Where it conflicts with a controlled technical source, the controlled technical source governs.

Tier 1 — Controlling PAI-SF technical sources

1. **ODA3-2026-07-NOR-PAISF-005 — Physical AI Security Framework Standard v1.0.** Governs the canonical framework architecture, evidence/assurance boundary, GAISSF Domain 9 disposition, and open-versus-controlled asset boundary.

2. **ODA3-2026-07-CAT-PAISF-006 — PAI-SF™ Control Catalogue v1.0.** Governs the 41 control IDs, domain allocation, objectives, applicability, evidence requirements, dependencies and exclusions.

3. **ODA3-2026-07-CSX-PAISF-004 — GAISSF Domain 9 to PAI-SF Control Crosswalk.** Governs the D9 continuity/supersession mapping. The crosswalk's own source-title caveat remains applicable; this guide does not infer unverified historical D9 titles.

Tier 2 — PAI-SF implementation and sector sources

ODA3-2026-07-QSG-PAISF-008 — PAI-SF™ Quick Start Guide.

ODA3-2026-07-AIQ-PAISF-017 — PAI-SF™ Assessment Intake Questionnaire.

ODA3-2026-07-ERL-PAISF-018 — PAI-SF™ Evidence Request List.

ODA3-2026-07-ARC-PAISF-019 — PAI-SF™ Assessment Readiness Checklist.

ODA3-2026-07-SCP-PAISF-012 through -016 — sector-calibration packages for robotics/industrial automation, autonomous vehicles/mobility, drones/uncrewed systems, medical robotics, and smart infrastructure/physical AI.

These sources support scoping, readiness, evidence planning and sector calibration. They do not authorize this guide to disclose or reconstruct controlled scoring, auditor or certification logic.

Tier 3 — GAISSF Ecosystem practitioner integration sources

ODA3-2026-08-WHP-COM-007 — Which GAISSF Ecosystem Framework Do I Need?

ODA3-2026-08-WHP-COM-009 — The First 30 Days with GAISSF™.

ODA3-2026-08-WHP-COM-010 — Applying GAISSF™: Worked Scenarios.

ODA3-2026-08-WHP-COM-011 — When AI Breaks: The ODA3 Guide to AI Incident Response.

ODA3-2026-08-WHP-COM-012 — Classifying AI Incidents: A Practical Guide to UAIF™.

ODA3-2026-08-WHP-COM-013 — AI Incident Response Field Guide: Using UAIF™ and AI-IRF™ for Real-World AI Incidents.

These documents support practitioner routing and ecosystem integration. They do not override PAI-SF normative/control sources.

60. Source-State and Claims Boundary

Source status must not be inferred from document existence alone. In particular:

PAI-SF's canonical public structure remains **12 domains and 41 controls**.

The GAISSF™ Domain 9 / PAI-SF™ relationship is governed by Section 3 and ODA3-2026-07-CSX-PAISF-004.

PAI-SF interfaces with functional safety and sector assurance; it does not replace them.

Public practitioner material does not create official scoring thresholds, detailed assessment methodology, auditor procedures, controlled test protocols or certification decision logic.

The FINAL publication state of this guide applies to this practitioner guide only. It does not establish PAI-SF certification availability, regulator recognition, product approval, scoring thresholds, pass/fail logic, or field-validation claims.

61. GEL Attribution and Use Boundary

GAISSF™, UAIF™, AI-IRF™, and PAI-SF™ are frameworks of ODA3 Institute and are referenced under the GAISSF Ecosystem License (GEL) v1.0.

This attribution does not itself authorize unrestricted commercial delivery, assessment, certification, endorsement or software embedding. Applicable GEL terms and any controlled-service/partner rules govern those uses.

Final Publication Statement

This document is the **released FINAL edition** of ODA3-2026-08-WHP-COM-014. Its FINAL status establishes this practitioner guide as the released, version-controlled edition. Substantive changes require a controlled revision under the ODA3 document-control process. It does not establish PAI-SF certification availability, scoring thresholds or pass/fail logic, regulator recognition, product-safety approval, functional-safety equivalence, airworthiness, roadworthiness, clinical approval, sector-specific authorization, or field validation of all 41 controls.

© 2026 ODA3 Pvt Ltd.