ODA3 INSTITUTE

ADOPTION GUIDE

The 10-Minute Introduction to

GAISSF™

Doc ID: ODA3-2026-08-WHP-COM-001
Classification: Public / Freely Distributable

Executive Summary

GAISSF™ (the Global AI Safety and Security Framework) gives organizations an evidence-driven operational structure for AI safety and security. Many AI governance programs do not yet answer one operational question clearly: what controls are actually in place, what evidence supports them, and where the uncertainty remains.

This guide is deliberately short. It will not make you a GAISSF™ practitioner. It will give you enough to decide whether the framework is relevant to your organization, and where to go next if it is.

> **METHODOLOGY NOTE**
> This guide is introductory and conceptual. It contains no client outcome data, no proprietary telemetry, and no incident statistics. The worked scenario in Section 4 is a labeled hypothetical, included to illustrate the framework's logic — not a case study or a claim of deployment history.

1. The AI Assurance Problem

Many organizations deploying AI today govern it using the same mechanisms they use for traditional software: policies, review processes, and risk registers. Traditional governance instruments may not fully address systems whose behavior and failure modes may be less predictable or less bounded than conventional software.

AI systems, particularly those with autonomy, tool access, or the ability to take action, can fail differently. They may drift. They may respond to inputs or contexts that were not anticipated during design or testing. They may be influenced or manipulated through inputs, context, retrieved data, or connected tools. And when something goes wrong, an organization without a structured evidence discipline may find it has no consistent way to show a regulator, an insurer, or its own board what actually happened, what changed, and what residual uncertainty remains.

This is the assurance gap: the distance between asserting that controls are adequate and demonstrating how those controls operate, what evidence supports them, and where uncertainty remains.

2. What GAISSF™ Is — and Is Not

What it is

- A control framework: 9 domains and a published architecture comprising 52 foundational controls, with 7 legacy physical-AI controls retained for continuity and cross-reference to PAI-SF™.
- An evidence discipline: controls are structured so their implementation can be supported by identifiable evidence, not just a policy you can describe.
- A common language designed for structured mapping against frameworks you may already use — NIST AI RMF, ISO/IEC 42001, the EU AI Act — so adopting it does not mean discarding existing work.

What it is not

- Not a certification product in which payment itself establishes conformance — any formal assurance conclusion must be supported by the applicable assessment criteria and evidence.
- Not a general cybersecurity framework relabeled for AI.

- Not a guarantee that an AI system is safe — it provides a structured basis for documenting controls, evidence, assessment conclusions, and residual limitations.

3. The Core Model

GAISSF™'s published architecture comprises 52 foundational controls across the core AI safety and security domains, plus 7 legacy physical-AI controls retained for continuity — PAI-SF™ now carries the authoritative expanded physical-AI requirements. Rather than treating every organization identically, GAISSF™ recognizes that a small team piloting a single AI assistant and a regulated enterprise running autonomous agents at scale face different starting points — which is why GAISSF™ Appendix P.1 defines a lightweight 9-control subset, drawn directly from the 52-control foundational scope, for use where its eligibility and scoping conditions are satisfied.

GAISSF™ at a Glance — The 9 Domains

- D1 — Model Integrity & Adversarial Robustness
- D2 — Runtime Security & Adversarial Defense
- D3 — Agentic Risk & Autonomous System Security
- D4 — Supply Chain & Third-Party AI Security
- D5 — Content Safety & Output Integrity
- D6 — Governance, Accountability & Human Oversight
- D7 — Human & Societal Harms
- D8 — Regulatory Alignment & Compliance
- D9 — Physical AI Safety (legacy) — retained for continuity; authoritative expanded treatment now in PAI-SF™

> **KEY PRINCIPLE**
> Assurance is demonstrated, not asserted. A control claim is only as credible as the evidence supporting its implementation and operation.

4. A Worked Scenario (Hypothetical)

Consider a hypothetical mid-size financial services firm piloting an AI assistant that drafts customer communications and has read access to account data. Its existing governance may include policy approval and vendor assurance, but may not yet provide system-specific operational evidence.

Applying GAISSF™'s foundational controls, the team would be expected to produce operational evidence such as: a documented data-access boundary for the assistant, a record of how outputs are reviewed before reaching a customer, a defined escalation path if the assistant produces something incorrect or unsafe, and logging or monitoring evidence supporting those controls in practice.

Where the Appendix P.1 eligibility and scoping conditions are satisfied, a limited pilot may begin with the lightweight subset before progressing to broader applicability analysis.

This scenario is illustrative only. It does not describe an ODA3 client, engagement, or outcome.

5. Who Should Use This

- CISOs and Security Architects extending existing security programs to cover AI systems.

- AI Governance Leads and Compliance Officers who need a control structure, not just a policy stance.
- Executives, boards, and General Counsel seeking a high-level, evidence-grounded understanding of how AI governance is being operationalized.
- Standards body participants evaluating GAISSF™ against other frameworks already in scope.

6. How GAISSF™ Relates to What You Already Have

GAISSF™ is designed as an augmentation layer, not a replacement. It is built to complement and be mapped against frameworks organizations may already use — NIST AI RMF, ISO/IEC 42001, SOC 2. The objective is to identify where existing governance and security controls already provide coverage, where GAISSF™ adds AI-security or operational-assurance depth, and where additional evidence or controls are needed — not to start over.

> **NOTABLY ABSENT**
>
> This guide does not include a control-by-control crosswalk, a formal assessment methodology, a regulatory-equivalence claim, or a certification conclusion. That level of detail belongs in the dedicated mapping and implementation guides, not in a 10-minute introduction.

7. Where Do I Start?

- 1. Identify one AI system or pilot to scope — not your entire AI estate.
- 2. If the system is suitable for lightweight scoping, compare it against the Appendix P.1 9-control subset; otherwise begin with full applicability scoping.
- 3. Read the GAISSF™ SMB Quick-Start Guide for a structured first self-review.
- 4. If your organization already runs NIST, ISO 42001, or SOC 2, read the relevant mapping guide before starting from scratch.

Your Next Step

Understand → Scope → Reuse Existing Controls → Implement → Assess

You are here: The 10-Minute Introduction to GAISSF™

- ↓ GAISSF™ SMB Quick-Start Guide
- ↓ Applicable mapping guide (NIST AI RMF / ISO 42001 / SOC 2)
- ↓ GAISSF™ implementation and evidence guides
- ↓ Assessment readiness

Next Reading

- GAISSF™ SMB Quick-Start Guide
- Mapping GAISSF to NIST AI RMF / ISO 42001 / SOC 2
- The AI Assurance Imperative: A Board Member's Guide to GAISSF™