ODA3 INSTITUTE

TCR — TECHNICAL & COMPLIANCE REPORT

Operational Complementarity Between Google's Secure AI Framework (SAIF) and the GAISSF™ Ecosystem

A Practitioner Methodology for Evidence-Bounded AI Assurance Across SAIF, GAISSF™, UAIF™, AI-IRF™, and PAI-SF™

Document ID: ODA3-2026-08-TCR-GAI-001
Classification: Public

Document ID	ODA3-2026-08-TCR-GAI-001 (paired with ODA3-2026-08-EXB-GAI-001)
Document Type	TCR — Technical & Compliance Report
Framework Tag	GAISSF™, UAIF™, AI-IRF™, PAI-SF™ (Google SAIF referenced as third-party source)
Regulatory Crosswalk	Independent analytical complementarity model — not an official SAIF crosswalk
Evidence Tier	T1–T4 inline (Sections 3–14)
Incident Provenance	Not applicable
Audience	Primary: Security Architects, AI Governance Leads, Cloud Security Teams — Secondary: CISOs, Compliance Officers, Independent Reviewers
Publish Date / Review Cycle	August 2026 — reassess on material change to SAIF guidance or the GAISSF™ Ecosystem

A Practitioner Methodology for Evidence-Bounded AI Assurance Across SAIF, GAISSF™, UAIF™, AI-IRF™, and PAI-SF™

Document ID: ODA3-2026-08-TCR-GAI-001 (paired with ODA3-2026-08-EXB-GAI-001) **Document Type:** TCR — Technical & Compliance Report **Framework Tags:** GAISSF™, UAIF™, AI-IRF™, PAI-SF™ (Google SAIF referenced as a third-party source, not an ODA3 framework) **Classification:** Public **Evidence Tier:** T1–T4 inline (Sections 3–14) **Audience:** Primary — Security Architects, AI Governance Leads, Cloud Security Teams. Secondary — CISOs, Compliance Officers, Independent Reviewers. **Publish Date / Review Cycle:** August 2026 — reassess on material change to publicly documented SAIF guidance, the GAISSF™ Ecosystem, or supporting evidence.

Executive Summary

Google's Secure AI Framework (SAIF) is a widely referenced practitioner framework for reasoning about AI-specific risk across an architecture — its four component areas and associated risks give organizations a vocabulary for identifying where AI systems introduce novel exposure. SAIF is deliberately not an accountability, evidence, incident-classification, or physical-effect assurance framework, and Google's own public materials do not present it as one. This report examines where that leaves a practical gap for any organization trying to move from "we have identified relevant AI risks using SAIF" to "we can demonstrate, with evidence, that a specific agentic AI system operates within its authorized boundary" — and proposes a complementarity method connecting SAIF's risk framing to the accountable-control and evidence discipline of GAISSF™, the incident-record structure of UAIF™, the governed response lifecycle of AI-IRF™, and — where an AI system can produce or materially influence a physical effect — the implementation and assessment requirements of PAI-SF™.

This report is independent analysis by ODA3 Institute. It is not sponsored, endorsed, approved, or validated by Google LLC, and it does not represent a partnership or affiliation with Google LLC. It does not establish an official or jointly approved SAIF–ODA3 crosswalk, and it does not certify any organization, product, model, or deployment against SAIF or any ODA3 framework.

The report's central proposition is narrow and evidence-bounded: SAIF and the GAISSF™ Ecosystem can occupy adjacent, reinforcing roles without either framework absorbing, validating, or certifying the other. Section 3 sets out how the four ODA3 frameworks relate to one another and to SAIF. Section 4 establishes the accountability and scoping discipline a pilot requires before any framework mapping begins. Sections 5 through 8 develop the evidence model, the agent accountability chain, incident and response integration, and an eight-phase operational assurance methodology. Section 9 translates that methodology into a proposed 90-day pilot plan. Sections 10 through 14 address assurance boundaries and governance, what this report explicitly does not establish, sector calibration considerations, independence and trademark obligations, and conclusions.

Readers should treat every claim in this report as tiered by evidence basis (Section 2) and should read Section 11, "Notably Absent," with the same weight as the report's positive claims — it is not a disclaimer appendix but a working part of the report's argument.

1. Purpose and Introduction

This report answers a specific, practical question: can an organization that has used Google's SAIF to reason about the risk of an AI system also build an accountable, evidence-based assurance case for that system using the GAISSF™ Ecosystem — without treating either framework as validating, certifying, or subsuming the other?

The question matters because SAIF and the GAISSF™ Ecosystem answer different parts of a shared problem. SAIF helps a practitioner ask *what could go wrong here, and where in the architecture*. GAISSF™, UAIF™, AI-IRF™, and PAI-SF™ answer a different set of questions: *who is accountable for this control, what evidence demonstrates it operates, how is a suspected failure classified and investigated, how is response governed, and — where relevant — how is a physical-effect capability specifically assured*. An organization that only does the first has a risk map without an evidence trail. An organization that jumps to control frameworks without SAIF-informed risk context risks building assurance around the wrong things.

Scope. This report provides analysis and a proposed operational methodology, including a Control Data Dictionary structure, an agent accountability model, a minimum agent test set, and a 90-day pilot plan. It is not a completed or exhaustive SAIF-to-GAISSF™ crosswalk, and Section 11 states explicitly what this report does not establish. Methodology elements are proposed by this report unless a source citation states otherwise — see the Capability State convention in Section 2.

Intended use. This report is written for organizations that already use, or are evaluating, Google Cloud and SAIF-informed practices and that operate or plan to operate agentic AI systems with meaningful tool access, delegated identity, or state-changing capability. It assumes no ODA3 framework has already been formally adopted, but it is also useful to an organization already using GAISSF™ that wants a structured way to bring SAIF-informed risk context into an existing control and evidence program.

How to read this report. Section 3 should be read before any other technical section — it establishes how the four ODA3 frameworks relate to each other and where each one's authority begins and ends, which every later section assumes. Sections 4 through 9 are intended to be used sequentially by a pilot team. Section 10 through 14 govern how conclusions from that pilot may, and may not, be used or claimed publicly.

2. Evidence Methodology

Every material factual claim in this report carries an inline evidence tag. The tag states the basis for the claim, not its importance or a guarantee of correctness beyond what the cited source itself supports.

Tier	Basis	Typical use in this report
T1	Primary published source — an official Google or ODA3 publication, cited directly	Descriptions of SAIF, cited Google Cloud product documentation, descriptions of a released ODA3 framework
T2	Corroborated secondary or sector-context statement, not a primary citation on its own	Sector calibration observations; statements that require professional judgment applied to a primary source
T3	ODA3 analytical inference connecting two or more cited sources	Statements about how a SAIF concept and a GAISSF™ concept relate, where the relationship itself is ODA3's analysis rather than a documented fact
T4	ODA3 proposed or illustrative content, not yet independently field-validated	The Control Data Dictionary, the agent test set, the 90-day pilot plan, and all worked/illustrative examples

This report also uses a **Capability State** label wherever it describes a method, test, or record structure: *available in a released source* (cited directly), *proposed by this report* (ODA3's proposed methodology, not yet field-validated), *planned* (publicly stated future intent), or *not established* (no supporting basis found). Readers should not infer field validation, adoption, or effectiveness from a T4 tag or a "proposed by this report" label — those labels mean the opposite: that no such validation yet exists.

Framework mapping and complementarity statements throughout this report are analytical and evidence-bounded. They do not establish compliance, certification, regulatory approval, or that use of any framework named in this report would have prevented any specific incident or outcome.

3. The GAISSF™ Ecosystem and Its Relationship to SAIF

3.1 Four complementary frameworks

Following the publication of PAI-SF™, the GAISSF™ Ecosystem comprises four complementary ODA3 frameworks, each with a distinct and non-overlapping role. GAISSF™ is the enterprise AI safety, security, governance, and operational-assurance baseline. UAIF™ structures how an AI-related incident is recorded and classified. AI-IRF™ governs the response lifecycle once an incident is underway. PAI-SF™ provides the specialized implementation and assessment framework for AI systems that can produce or materially influence physical effects. None of the four frameworks is a substitute for SAIF, and SAIF is not a substitute for any of them — SAIF operates one layer earlier, informing the risk and architecture context that the ODA3 frameworks then convert into accountable controls, evidence, incident structure, and response governance. **[T3]**

3.2 A transition organizations should be aware of: GAISSF™ Domain 9 and PAI-SF™

Prior to PAI-SF™'s publication, GAISSF™'s Domain 9 was the framework's only location for physical-AI-relevant governance content. With PAI-SF™ published as Final Publication v1.0 on 3 August 2026, PAI-SF™ became the detailed implementation and assessment framework for systems capable of physical effects, organized across the twelve domains described in Section 3.5. **PAI-SF™ supersedes GAISSF™ Domain 9 for detailed physical-AI implementation and assessment guidance; it does not remove Domain 9 from GAISSF™**, which retains governance-level physical-AI requirements consistent with its role as the enterprise-wide baseline. **[T1]**

This distinction matters practically for any organization that consulted GAISSF™ materials before PAI-SF™'s publication: a physical-effect AI system should now be assessed against PAI-SF™ for implementation-level detail, with GAISSF™ Domain 9 continuing to apply at the governance level, rather than treating Domain 9 alone as sufficient for a system with meaningful physical-effect capability. Organizations should confirm they are working from the current post-transition framework structure — the counts and structure cited in this report's Source Register (Appendix G) reflect that current structure as independently checked against each framework's published page, though readers relying on this report for a live implementation should reconfirm current counts directly, since framework publications are versioned and may be revised. **[T3]**

> **FRAMEWORK RELATIONSHIP:**
>
> Following publication of PAI-SF™, the GAISSF™ Ecosystem comprises four complementary frameworks. GAISSF™ remains the enterprise AI safety, security, governance, and operational assurance baseline, including the governance-level Physical AI requirements retained in Domain 9. UAIF™ structures AI incident evidence. AI-IRF™ governs preparedness, containment, recovery, notification, learning, and post-incident validation. PAI-SF™ provides the specialized implementation framework for AI systems capable of producing or materially influencing physical effects. PAI-SF™ supersedes GAISSF™ Domain 9 for detailed Physical AI implementation and assessment guidance; it does not remove Domain 9 from GAISSF™.

GAISSF™ contributes: system scope and applicability; accountable control ownership; control objectives and implementation records; minimum evidence expectations; testing and exception management; management review; corrective and preventive action; and controlled assurance claims.

3.3 UAIF™

UAIF™ contributes a structured incident record separating attack or failure mechanism, affected AI-system layer, confirmed and potential impact, severity indicators, causality, confidence, contextual modifiers, available evidence, and missing evidence — reducing the risk of turning an early hypothesis into an asserted fact. **[T1]**

3.4 AI-IRF™

AI-IRF™ contributes the governed response lifecycle: preparedness; detection and analysis; containment; recovery; notification where applicable; learning; and post-incident validation. Its role is not merely to close a ticket — it connects response decisions to evidence, approvals, recovery criteria, and recurrence prevention. **[T1]**

3.5 PAI-SF™

PAI-SF™ governs assurance where AI can receive physical-world input, determine or recommend action, issue or influence a kinetic command, or produce a physical effect, across a twelve-domain structure covering sensor and perception integrity, the physical operating environment, edge hardware and model attestation, autonomy boundaries, safe-state and degraded-mode behavior, actuator and kinetic-command safety, runtime monitoring, physical incident response and recovery, human override, supply-chain assurance, functional-safety interfaces, and simulation/testing/validation. PAI-SF™ v1.0 published as Final Publication v1.0 on 3 August 2026, confirming 12 domains and 41 controls, consistent with the counts stated in this report. **[T1]**

3.6 Complementarity operating sequence

SAIF risk and architecture framing → GAISSF™ control and evidence design → UAIF™ incident record → AI-IRF™ response and validation → PAI-SF™ physical-effect assurance where applicable → GAISSF™ management review and improvement. This sequence is an ODA3 analytical model. It is not a Google-supported integration. **[T4]**

4. Scope, Accountability, and Applicability

4.1 System boundary

A pilot begins with a system record, not a framework mapping. At minimum, identify: business purpose and material decisions supported; legal entity and accountable business owner; system owner and security-control owner; models, versions, and providers; prompts, system instructions, and policy components; retrieval sources and data stores; identities, service accounts, and delegated credentials; tools, APIs, plugins, and state-changing functions; deployment environments and network boundaries; monitoring, logging, and evidence repositories; external dependencies and sub-processors; users and affected persons; jurisdictions and retention restrictions; and any physical-world input, command, or effect. **[T4]**

An inventory entry should distinguish an AI assistant that proposes text from an agent that can create a payment, change a production system, send an external communication, modify an access rule, or control a physical device.

4.2 Accountable roles

Role	**Primary responsibility**	**Independence consideration**
Business owner	Purpose, materiality, risk acceptance, and operating authorization	Must not delegate residual-risk acceptance invisibly to the model or vendor
System owner	Architecture, lifecycle, availability, and change control	Must maintain an accurate system record
Security-control owner	Safeguard design, implementation, and operating evidence	Should not self-approve material exceptions
Model or AI owner	Model selection, evaluation, prompt/retrieval design, and model changes	Must preserve version and evaluation traceability
Data owner	Data authorization, quality, lineage, access, and retention	Must resolve rights and sensitive-data constraints
Incident commander	Response coordination and decision record	Must have authority to suspend or constrain operation
Independent reviewer	Test design, evidence challenge, and conclusion	Independence should be proportionate to impact and claim

4.3 Applicability decision

The statement of applicability should record: relevant SAIF risks or controls considered; applicable ODA3 control objectives; rationale for inclusion, tailoring, or exclusion; evidence source and test method; responsible owner; residual risk; exception expiry and compensating control; PAI-SF™ applicability decision; and approval and review date.

DISCIPLINE POINT:

“Not applicable” is a conclusion requiring evidence. It is not a device for reducing workload.

4.4 Physical-effect decision

Apply three questions:

- Can the system receive or infer information about the physical environment?
- Can it issue, select, recommend, or materially influence a command that changes physical state?
- Can failure, compromise, or misuse cause injury, unsafe motion, environmental damage, loss of physical control, or interruption of a safety function?

Any “yes” requires documented PAI-SF™ scoping. The degree of application may vary, but the decision cannot remain implicit.

5. Operational Assurance Evidence Model

The Control Data Dictionary is the minimum normalized record needed to translate a framework objective into an assessable control. It is not a new security platform and does not require every evidence item to be copied into an ODA3 repository.

Field	**Required content**
System and scope	Unique system identifier, environment, lifecycle stage, business process, and exclusions
Control record	Control identifier, objective, implementation statement, and status
SAIF context	Relevant SAIF source, risk, component, or control; exact source date
ODA3 relationship	Informs, supports, partially supports, no relationship established, or not assessed
Accountable roles	Business owner, system owner, control owner, evidence owner, and reviewer
Evidence source	System of record, query/report name, location, retention, and access conditions
Evidence period	Start and end date; event or population covered
Test method	Inquiry, inspection, observation, re-performance, sampling, or technical test
Population and sample	Complete population definition, sampling rationale, and selected items
Expected result	Objective pass condition and allowed tolerance
Result	Pass, partial, fail, not tested, or not applicable, with factual observations
Exception	Deviation, compensating control, owner, expiry, and approval
Decision	Reviewer conclusion, residual risk, and management action
Incident link	Related UAIF™ record and AI-IRF™ response record where applicable
Physical boundary	PAI-SF™ applicability, relevant domains, and physical-effect rationale

5.1 Evidence sufficiency

Evidence must be: relevant to the control objective; attributable to the correct system, environment, and owner; time-bounded to the assessed period; complete enough to support the conclusion; authentic and protected against unauthorized alteration; reproducible where a query or test is relied upon; lawfully retained and proportionate; and reviewable by a person with the required competence and independence.

A screenshot of a configuration page may show a point-in-time setting but not continuous operation. A policy may establish intent but not implementation. A log may show an event but not whether the event population is complete. **[T4]**

5.2 Evidence hierarchy for cloud records

A defensible evidence pack normally combines: approved design and policy; system inventory and configuration baseline; machine-generated operational records; change and exception records; control test results; incident and recovery records; and management review and corrective-action evidence. No single layer is sufficient for every objective. **[T4]**

5.3 Illustrative evidence sources described in Google Cloud

documentation

Official Google Cloud documentation describes logging, monitoring, security, and data services that may produce relevant records. Whether a particular record is enabled, complete, retained, accessible, and suitable for an ODA3 objective is deployment-specific. **[T1]**

Descriptive source	**Potential evidentiary use**	**Required caution**
Cloud Audit Logs	Administrative activity, data access where enabled, policy or resource changes	Confirm log type, coverage, exclusions, retention, and identity attribution
Vertex AI records	Model, endpoint, pipeline, evaluation, or deployment context where available	Verify service, region, feature, and record semantics; do not assume prompt or response retention
Security Command Center (Google Cloud)	Security posture or finding record	A finding is not proof of remediation or control effectiveness
Chronicle (part of Google Security Operations)	Detection, correlation, and investigation records	Validate source onboarding, parser behavior, time synchronization, and completeness
BigQuery datasets	Evidence aggregation, reproducible queries, and trend analysis	Govern dataset access, query version, transformation logic, and retention
Cloud Logging and Monitoring	Runtime events, metrics, and alert history	Verify monitored resources, alert routing, suppressed events, and observation gaps

Descriptive source	**Potential evidentiary use**	**Required caution**
Identity and access records	Principal, role, policy, and access-change evidence	Separate intended entitlement from actual use and delegated agent authority

BOUNDARY STATEMENT:

This table is illustrative and is not a product recommendation, implementation statement, or claim that the named service satisfies a GAISSF™ objective.

6. Agent Accountability Chain

6.1 Why agents change the assurance question

Google's public agent-security material emphasizes well-defined human controllers, carefully limited powers, and observable actions. An agent may combine ambiguous instructions, retrieved content, model inference, delegated identity, and external tools. The security question is therefore not limited to whether the model generated harmful text — it includes whether the system obtained authority, selected a tool, passed parameters, crossed a threshold, and changed state. **[T1]**

GAISSF™ provides the enterprise accountability and evidence wrapper; UAIF™ and AI-IRF™ provide incident and response discipline; PAI-SF™ applies if the action can produce physical effects. **[T4]**

6.2 Authority

Authority evidence should show: approved purpose and use case; named human and organizational owner; identities used by the agent; permitted resources and actions; prohibited actions; transaction, value, safety, or sensitivity thresholds; conditions requiring human approval; delegation and revocation method; and expiry or reauthorization conditions.

> **DISCIPLINE POINT:**
>
> The agent's technical ability is not its approved authority.

6.3 Constrained action

Constraints may include: tool allow-lists; least-privilege service identities; parameter validation; destination restrictions; environment separation; rate, value, or action limits; policy enforcement outside the model; multi-party approval for material actions; safe defaults; and deny-by-default handling of unknown tools or states. Prompt instructions alone are not treated as a complete authorization boundary for a material state-changing action.

6.4 Observation

The evidence design should preserve, where lawful and proportionate: initiating identity and request; applicable policy and system-instruction version; relevant prompt, retrieval, and tool context; model and deployment version; selected tool and parameters; authorization and approval result; external response; state change; timestamps and correlation identifiers; intervention or override activity; and evidence gaps. Observation is not synonymous with unrestricted content retention — privacy, confidentiality, privilege, security, and data-minimization requirements may restrict what can be stored.

6.5 Intervention

Intervention design should identify: alert and escalation conditions; human decision authority; workflow suspension; credential or token revocation; tool disablement; transaction cancellation where feasible; model or configuration rollback; safe-state transition for Physical AI; continuity arrangements; and criteria for resuming operation.

DISCIPLINE POINT:

An emergency stop that has never been exercised is a design claim, not operating evidence.

6.6 Reconstruction

A reviewer should be able to distinguish: confirmed events; technical observations; model-generated content; inferred intent or causality; missing records; disputed interpretation; confirmed and potential impact; and actions taken by people and systems. UAIF™ provides the structured incident record for this separation.

6.7 Validated improvement

Closure requires more than remediation assignment. The organization should show: causal and contributing-factor analysis; control or design change; owner and target date; regression test; evidence that the change operated; review of similar systems; updated risk and applicability records; and management acceptance of remaining risk.

6.8 Illustrative RACI for the accountability chain

A proposed starting allocation, not a universal organizational structure. Enterprises should adapt role names and preserve separation between implementation, residual-risk acceptance, and independent review. R = Responsible, A = Accountable, C = Consulted, I = Informed.

Chain stage	**Business owner**	**System owner**	**Security-control owner**	**Incident commander**	**Independent reviewer**
Authority	A	R	C	I	C
Constrained action	C	R	A	I	C
Observation	I	R	A	I	C
Intervention	C	R	A	R	I
Reconstruction	I	C	R	A	C
Validated improvement	A	C	R	R	C

DISCIPLINE POINT:

RACI allocation does not prove accountability. Each material decision must still produce a dated record identifying the decision-maker, evidence considered, exception or residual risk, and required follow-up.

6.9 Minimum agent test set

These are proposed tests, not a Google or SAIF test suite. All entries: Capability state — Proposed by this report.

Test	**Objective**	**Minimum evidence**
Prohibited-tool test	Verify that the agent cannot invoke a tool outside its approved set	Test case, policy version, trace, and denial result
Delegated-identity test	Verify that identity and privileges match approved scope	Principal, roles, token conditions, attempted actions, and result
Prompt/instruction-boundary test	Determine whether untrusted content can alter authority or bypass approval	Payload, retrieved context, trace, decision, and observed result
Parameter-integrity test	Prevent unsafe or unauthorized tool parameters	Input, validation record, transformed parameters, and external response
Human-approval test	Verify that threshold actions cannot proceed without valid approval	Threshold, approval identity, timestamp, and action result
Kill/revocation test	Demonstrate timely suspension and credential invalidation	Trigger, revocation event, denied follow-on action, and elapsed time
Trace-completeness test	Determine whether a reviewer can reconstruct the action chain	Expected events, observed events, missing fields, and correlation result
Third-party dependency test	Identify failure or compromise behavior for external model/API/tool	Dependency record, scenario, fallback, observed behavior, and decision
Recovery test	Verify rollback, restoration, and safe resumption	Recovery plan, restored state, validation result, and approval
Physical-effect escalation test	Determine whether PAI-SF™ applies and whether safe-state controls operate	Capability analysis, PAI-SF™ scope, simulation/test record, and result

7. Incident and Response Integration

7.1 Detection is not classification

Detection initiates investigation; it does not conclude it. The first responsibility of an operational assurance process is to preserve uncertainty until sufficient evidence supports a defensible conclusion. This distinction reduces the risk of treating alerts, model output, or early indicators as established fact.

A detection record may show an anomalous request, prohibited tool attempt, model-output pattern, privilege change, or unexpected state transition. It does not automatically establish malicious intent, exploit success, causality, or impact. The first incident record should therefore preserve uncertainty. **[T4]**

7.2 UAIF™ incident record

For each material event, record at least: affected system and components; initiating event and detection source; confirmed facts; working hypotheses; attack or failure mechanism; affected AI-system layer; confirmed and potential impact; causality status; severity indicators; evidence sources; missing evidence; confidence; contextual modifiers; physical-world capability or effect; and links to related cases.

7.3 AI-IRF™ lifecycle

Lifecycle stage	**Assurance question**	**Required record**
Prepare	Were roles, escalation, evidence access, and response options established?	Plan, contacts, exercise result, evidence-access check
Detect	What signal triggered review and how reliable is it?	Alert, source, time, triage decision
Analyse	What is confirmed, inferred, and missing?	UAIF™ record, timeline, hypotheses, impact assessment
Contain	Which authority, tool, identity, model, or workflow must be constrained?	Decision, approval, containment action, verification
Recover	What state is safe and what validation is required?	Recovery criteria, restored state, tests, approval
Notify	What contractual, legal, regulatory, or stakeholder duties apply?	Decision basis, content, timing, recipients
Learn	What changed and how was effectiveness verified?	Corrective action, retest, recurrence review, management decision

7.4 Evidence preservation

Response actions can destroy or change evidence. The playbook should address: volatile trace capture; clock and identifier consistency; log-retention holds; model, prompt, retrieval, and policy versions; chain of custody where required; restricted or privileged materials; third-party evidence requests; integrity verification; and documented reasons where evidence cannot be retained.

7.5 Physical incident escalation

Where the event can produce physical effects, response must include PAI-SF™ considerations such as safe state, kinetic-command interruption, human override, physical recovery, environmental conditions, and functional-safety interfaces. Cyber containment alone may be insufficient if a device remains energized, moving, unstable, or capable of autonomous action. **[T4]**

8. Operational Assurance Methodology

Phase 1 — Scope and source control

A successful assurance engagement begins by establishing a stable assessment boundary. Every subsequent conclusion depends on the accuracy of this scope definition and the integrity of the source baseline.

Select one material agentic workload. Freeze the assessment boundary and architecture diagram. Establish the source baseline. Identify the applicable released versions of GAISSF™, UAIF™, AI-IRF™, and PAI-SF™. Assign accountable roles. Identify legal, privacy, sector, and contractual constraints.

> **EXIT CRITERION:**
> Approved system boundary, source register, and role record.

Phase 2 — SAIF-informed risk analysis

Use SAIF sources to identify relevant components, risks, and safeguards. Do not copy every item indiscriminately; record the exact source and explain why it informs the scoped architecture. **[T1]**

> **EXIT CRITERION:**
> Risk record with source-level traceability and no implied formal crosswalk.

Phase 3 — ODA3 applicability and control design

For each material risk: identify the GAISSF™ objective or other ODA3 requirement; assign relationship vocabulary; define implementation; identify evidence; define the test and expected result; assign owners and reviewer; decide whether PAI-SF™ applies; and record gaps and exceptions.

> **EXIT CRITERION:**
> Approved statement of applicability and Control Data Dictionary.

Phase 4 — Evidence-source validation

Validate record existence, system and environment attribution, coverage, retention, access, integrity, reproducibility, lawful use, and known blind spots. Do not label a product as an evidence source merely because documentation describes a relevant capability.

Phase 5 — Control and agent testing

Execute the minimum agent tests and risk-specific tests. Preserve the test case, environment, version, result, anomalies, and reviewer conclusion.

EXIT CRITERION:

Test pack with pass, partial, fail, or not-tested conclusions and a documented basis.

Phase 6 — Incident exercise

Run a scenario that requires detection, UAIF™ classification, AI-IRF™ containment, and recovery. Include an evidence gap and an ambiguous causal signal so the exercise tests uncertainty rather than a perfectly scripted event. If Physical AI is in scope, include safe-state and human-intervention steps.

EXIT CRITERION:

Exercised response record, lessons, and validated improvement actions.

Phase 7 — Independent review and management decision

Independent review provides assurance that conclusions are supported by evidence rather than implementation ownership or confirmation bias.

The independent reviewer should challenge: scope completeness; unsupported relationships; evidence sufficiency; test design; exception treatment; physical-effect scoping; incident conclusions; residual risk; and public or customer-facing claims. Management may authorize expansion, require remediation, accept bounded residual risk, or stop the pilot.

9. Ninety-Day Pilot Plan

Days 1–15: scope, roles, and current state

Deliverables: approved system boundary; owner and reviewer register; source baseline; architecture and dependency record; preliminary SAIF-informed risk analysis; applicable-law and data-constraint note; preliminary PAI-SF™ applicability decision; evidence-source inventory.

> **PRIMARY RISK:**
> Choosing an attractive demonstration system that is not materially representative.

Days 16–30: control and evidence design

Deliverables: statement of applicability; Control Data Dictionary; evidence-quality assessment; gap and exception register; agent authority record; test plan; incident scenario; data-retention and access decisions.

> **PRIMARY RISK:**
> Treating configured capability as evidence of operating effectiveness.

Days 31–60: configure evidence and test controls

Close critical evidence gaps; establish reproducible queries or reports; confirm correlation identifiers; execute identity, tool, approval, trace, and intervention tests; test recovery; validate third-party dependencies; execute applicable PAI-SF™ tests.

> **PRIMARY RISK:**
> Changing the environment so extensively that the pilot no longer tests the current operating state.

Days 61–75: incident exercise and corrective action

Execute the incident scenario; create the UAIF™ record; run the AI-IRF™ response lifecycle; preserve uncertainty and missing evidence; validate containment and recovery; assign corrective actions; retest high-priority changes.

> **PRIMARY RISK:**
> A scripted exercise that rewards process completion without testing decision quality.

Days 76–90: independent review and decision

Final pack: system and source baseline; applicability record; control implementation and evidence pack; agent test results; incident exercise; exceptions and residual risk; PAI-SF™ decision and results; independent review; management decision; expansion or closure plan.

Pilot success measures

Measure	**Evidence of progress**
Scope completeness	Named owners and approved boundary for every material component and dependency
Evidence coverage	Percentage of applicable controls with validated evidence source and test
Authority coverage	Percentage of state-changing tools with explicit owner, permitted scope, and approval rule
Trace completeness	Percentage of expected action-chain events reconstructable in test
Intervention readiness	Tested suspension, revocation, rollback, or safe-state result
Incident discipline	UAIF™ record distinguishes facts, inference, impact, and missing evidence
Recovery confidence	Recovery criteria met and independently reviewed
Corrective-action quality	High-priority actions retested for effectiveness
Retrieval effort	Time required to assemble the evidence pack from approved sources

CAUTION:

Metrics are diagnostic. A high percentage does not by itself establish adequate design or certification.

10. Assurance Boundaries, Claims, and Governance

10.1 Assurance conclusion

The pilot may support a conclusion about: the defined system and period; applicable objectives; evidence examined; tests performed; observed results; exceptions and limitations; and residual risk decision. It cannot support a universal statement that the organization, Google Cloud environment, model family, or agent platform is “SAIF certified” or “GAISSF™ compliant.”

10.2 Independence

Independence should be proportionate to the consequence of the claim. A control owner may perform routine self-testing, but a material external claim or certification decision requires a separately governed reviewer with appropriate competence and freedom from responsibility for the implementation being assessed.

10.3 Commercial-delivery boundary

Publication of this report does not authorize commercial implementation, assessment, training, or certification using ODA3 protected materials. Commercial delivery remains subject to the GAISSF™ Ecosystem License (GEL) and ODA3 Partner & Trust Ecosystem requirements, including empanelment where applicable. **[T1]**

10.4 Change governance

Reassessment triggers should include: material model or provider change; new tool or permission; new data or retrieval source; significant prompt or policy change; new deployment environment; material incident; new physical-world capability; relevant SAIF or ODA3 framework revision; legal or regulatory change; and evidence-source failure.

10.5 Publication governance

For future revisions of this publication: verify all ODA3 framework names, marks, counts, and statuses; periodically verify canonical framework URLs; recheck cited public sources when materially updated; perform technical and legal review before any future revision; verify evidence-tier tags; confirm Notably Absent statements; verify that tables do not imply product certification; test all links; and approve the final document and website metadata through the ODA3 publication process.

11. Notably Absent

The following were not established by this report:

- • Google endorsement, sponsorship, approval, adoption, or validation of GAISSF™, UAIF™, AI-IRF™, PAI-SF™, ODA3 Institute, or this report.
- • An official, jointly approved, or complete SAIF–ODA3 crosswalk.
- • A Google-supported GAISSF™ assurance overlay or software integration.
- • SAIF conformance or certification for any organization, product, model, service, or deployment.
- • GAISSF™ certification or assessment of Google Cloud, Vertex AI, Chronicle, Security Command Center, BigQuery, or another Google product.
- • Evidence that a named Google Cloud capability is enabled or effective in a particular customer environment.
- • A conclusion that public SAIF material describes Google's internal technical implementation.
- • A complete mapping of SAIF risks and controls to every GAISSF™, UAIF™, AI-IRF™, or PAI-SF™ requirement.
- • Legal or regulatory compliance.
- • Control effectiveness in a production system.
- • Independent validation of the proposed Control Data Dictionary, agent test set, or 90-day pilot.
- • A conclusion that logging all prompts, responses, retrieval content, or tool data is lawful, necessary, or proportionate.
- • A conclusion that information-security containment is sufficient for a Physical AI event.
- • Independent field validation of the proposed Control Data Dictionary, agent test set, complementarity model, or 90-day pilot methodology.
- • A conclusion that the proposed methodology will produce equivalent outcomes across sectors, architectures, cloud environments, or model providers.

> **NOTE:**
> These absences are constraints on interpretation, not defects concealed by the publication.

12. Sector Calibration Considerations

The complementarity method is sector-agnostic by design, but evidence, intervention, and notification requirements vary materially. The following are scoping prompts, not sector-specific guidance or legal conclusions; readers should consult ODA3 Institute's published sector guidance series for detailed treatment.

12.1 Financial services

Agent authority and intervention (Sections 6.2 and 6.5) carry particular weight where an agent can initiate, approve, route, or materially influence a financial transaction. Pilot design should examine segregation of duties, transaction thresholds, delegated identity, fraud signals, reconciliation, reversibility, and sector-specific incident or outsourcing obligations. A model-generated recommendation may still be material even when a human performs the final transaction. **[T2]**

12.2 Healthcare and life sciences

Data provenance, privacy, clinical workflow boundaries, and human intervention require heightened attention where an agent processes health information or influences diagnosis, treatment, medication, prioritization, or device operation. This report does not establish clinical safety, medical-device compliance, or fitness for clinical use. Physical effects require PAI-SF™ scoping in addition to applicable clinical and functional-safety processes. **[T2]**

12.3 Public sector

Public-sector use may require additional transparency, procurement, records-retention, accessibility, administrative-law, and public-accountability analysis. Vendor-governance boundaries are especially important where the AI service is procured while the public body remains accountable for the decision or service delivered. **[T2]**

12.4 Critical infrastructure and industrial systems

Availability, safe state, degraded mode, physical operating environment, human override, and recovery may dominate the assurance case. A cybersecurity alert that appears low-impact in an information system may be material where delayed or incorrect action affects energy, transport, manufacturing, water, communications, or another essential service. PAI-SF™ applicability must be explicit. **[T2]**

12.5 Cross-sector rule

The common method may organize evidence, but it does not normalize away sector differences. Sector obligations, safety cases, professional duties, materiality thresholds, and notification timelines remain separate inputs to scope, test design, and conclusions.

> **POINTER:**
>
> ODA3 Institute publishes a dedicated sector guidance series (Healthcare, Financial Services, Insurance, Defence, Government, Critical Infrastructure, Education, Energy, Manufacturing, Retail, Telecom, Legal, Pharmaceutical, Aerospace, Automotive) at docs.oda3.org, developed independently of this report and not limited to SAIF-based deployments.

13. Independence, Trademark, and Attribution Notice

> **INDEPENDENT ANALYSIS NOTICE:**
>
> This report is an independent analysis by ODA3 Institute. It is not sponsored, endorsed, approved, or validated by Google LLC, and it does not represent a partnership with, or affiliation with, Google LLC.

Google, Google Cloud, Vertex AI, Chronicle, Security Command Center, BigQuery, and Cloud Logging are trademarks of Google LLC. SAIF refers to the Secure AI Framework published by Google. IAM is used here as a generic industry term for identity and access management, except where specifically identifying Google Cloud IAM as a named product. All third-party names, marks, and product names referenced herein are the property of their respective owners and are used solely for descriptive, nominative purposes to identify the products and frameworks analyzed; such use does not imply endorsement, affiliation, or sponsorship. This notice is a courtesy statement and is not a substitute for independent legal review.

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

14. Conclusions and Recommendations

SAIF and the ODA3 framework suite can occupy adjacent and reinforcing roles. SAIF helps practitioners understand AI-specific risk across an architecture and consider relevant safeguards. GAISSF™ can place accountable control objectives, evidence expectations, tests, and management decisions around that work. UAIF™ and AI-IRF™ can preserve incident uncertainty and govern response. PAI-SF™ can govern the additional boundary where an AI system can produce or influence physical effects.

The useful proposition is not that one framework completes or validates the other. It is that an enterprise can preserve each framework's independent role while building a traceable operating chain: risk context → accountable control → evidence → test → incident record → response → physical-effect assurance where applicable → validated improvement.

> **NOTE:**
> The example does not state that SAIF requires this exact implementation or that the implementation satisfies SAIF.

Appendix B: Evidence-Quality Checklist

- Is the evidence tied to the correct system and environment?
- Does it cover the required period?
- Is the underlying population defined?
- Can the query or report be reproduced?
- Are collection and transformation steps documented?
- Are time synchronization and correlation reliable?
- Are access and integrity protected?
- Are retention and deletion lawful?
- Does the evidence show design, operation, or both?
- Are missing records visible?
- Has contradictory evidence been considered?
- Is the reviewer competent and sufficiently independent?
- Does the conclusion stay within the evidence?

Appendix C: Evidence Boundary for Google Cloud and Multi-Cloud Environments

This report's evidence-source examples (Section 5.3) are Google Cloud–specific because the underlying complementarity proposition concerns SAIF. Most enterprises operate a mixed estate; this appendix separates what is illustrated here from what a multi-cloud or third-party-model deployment requires in addition.

- A provider-neutral evidence taxonomy so the same Control Data Dictionary fields can be populated regardless of which cloud, model provider, or agent framework produced the underlying record.

- Explicit mapping of each non-Google evidence source to the same fifteen dictionary fields defined in Section 5.
- A documented boundary statement for each third-party model or API distinguishing provider-supplied attestations from evidence the enterprise must independently generate.
- A single aggregation layer capable of reconciling evidence records with different native formats, retention rules, and access-control models without silently normalizing away material differences in evidentiary strength.

> **BOUNDARY STATEMENT:**
> This report does not provide a completed multi-cloud or third-party-model evidence pack. This is identified as a candidate future practitioner artefact, distinct from this report.

Appendix D: Glossary

Term	**Meaning in this report**
SAIF	the Secure AI Framework published by Google, referenced from Google's public practitioner materials
GAISSF™	ODA3 Institute's Global AI Safety and Security Framework; the enterprise control and assurance baseline described in Section 3
UAIF™	ODA3 Institute's Unified AI Incident Framework for evidence-bounded incident records
AI-IRF™	ODA3 Institute's AI Incident Response Framework for governed preparation, response, recovery, and learning
PAI-SF™	ODA3 Institute's Physical AI Security Framework for systems that can perceive, influence, command, or produce physical effects
Agent	Software that can obtain information, invoke tools or services, or produce state-changing outcomes with some degree of autonomy
Control Data Dictionary	Maintained record connecting scope, objective, implementation, owner, evidence, test, result, and exception
Statement of applicability	Record of applicable, tailored, and excluded control objectives with rationale and approval
Capability state	Whether a capability is available in a released source, proposed by this report, planned, or not established
Evidence tier	T1–T4 classification defined in Section 2; a statement

Term	**Meaning in this report**
	about source basis, not automatic control effectiveness
Notably Absent	Explicit record of conclusions the available evidence does not establish
Intervention exercise	Test of the ability to suspend, revoke, constrain, roll back, or place a system in a safe state
Trace-completeness test	Reconstruction of a representative action from initiating request through tool or external effect
Residual risk	Risk remaining after controls, requiring an accountable and recorded decision

Appendix E: Illustrative UAIF™ Case Record Template

A worked, illustrative example of the UAIF™ fields referenced in Section 7, for a hypothetical suspected prompt-injection event. No actual incident data is used. Capability state: Proposed by this report.

UAIF™ field	**Illustrative entry**
Attack or failure mechanism	Suspected prompt injection via retrieved document content
Affected AI-system layer	Retrieval-augmented generation pipeline; agent tool-invocation layer
Confirmed impact	None confirmed at time of initial record
Potential impact	Unauthorized tool invocation; possible data exposure via downstream API call
Severity indicators	Elevated — agent has write-capable tool access in scope
Contextual modifiers	Detected in non-production canary environment; production traffic unaffected
Causal status and confidence	Suspected, unconfirmed — pending trace-completeness test
Available evidence	Retrieval-source log; agent planning trace (partial); tool-invocation log
Material evidence gaps	Prompt-level reasoning trace beyond the planning summary not currently retained
Physical effect	None identified; PAI-SF™ applicability reviewed and documented

Appendix F: Illustrative AI-IRF™ Response Log Template

A worked, illustrative example of an AI-IRF™ lifecycle log corresponding to the Appendix E case record. Capability state: Proposed by this report.

Lifecycle stage	**Illustrative log entry**
Prepare	Canary-environment playbook and named incident commander were pre-defined prior to detection
Detect	Retrieval-source anomaly alert triaged by on-call security-control owner within defined SLA
Analyse	Trace-completeness test initiated; causal status recorded as suspected, not confirmed
Contain	Agent tool access suspended in canary environment pending analysis; production unaffected, confirmed by scope record
Recover	Retrieval source allow-list tightened; agent tool access restored after re-test
Notify	Assessed against contractual and customer notification thresholds; determined not to meet notification criteria; decision recorded with rationale
Validate	Corrective action (allow-list change) re-tested after 14 days; no recurrence observed in canary monitoring; lesson added to control-design backlog

Appendix G: Source Register

Google, SAIF: Google's Guide to Secure AI, saif.google/, reviewed 31 July 2026. **[T1]**

Google, SAIF Risk Map Components, saif.google/secure-ai-framework/components, reviewed 31 July 2026. **[T1]**

Google Cloud, Use AI securely and responsibly, Google Cloud Architecture Center, last reviewed 5 February 2025, accessed 31 July 2026. **[T1]**

Google Research, Google's Approach for Secure AI Agents, 2025, accessed 31 July 2026. **[T1]**

Google Research, Securing the AI Software Supply Chain, 2024, accessed 31 July 2026. **[T1]**

ODA3 Institute, GAISSF™ v1.0, Global AI Safety and Security Framework, Final Publication v1.0. Verified live at docs.oda3.org/frameworks/gaissf/, 31 July 2026 (confirms current published structure: 9 domains, 59 controls, 52 foundational + 7 physical-AI in D9 — see Section 3.2 transition-note caution below).

ODA3 Institute, UAIF™, Final Publication. Verified live at docs.oda3.org/frameworks/uaif/, 2 August 2026 (confirms 11 sections, 57 fields/tests, 7-layer architecture, 5 conformance profiles).

ODA3 Institute, AI-IRF™, Final Publication. Verified live at docs.oda3.org/frameworks/airf/, 31 July 2026 (confirms 19 domains, 79 controls, and that the AIIS machine-readable control register is the authoritative structured source).

ODA3 Institute, PAI-SF™, Final Publication. Verified live at docs.oda3.org/frameworks/pai-sf/, 3 August 2026 (confirms 12 domains, 41 controls, published as a Final Publication).

ODA3 Institute, GAISSF™ Ecosystem License v1.0 and Partner & Trust Ecosystem rules. Verified live at oda3.org/legal/gel-v1-0/, 31 July 2026 (confirms Effective 1 May 2026, source-available/non-commercial-by-default classification, and that GAISSF™/UAIF™/AI-IRF™/ODA3 marks remain registered-trademark applications, not granted registrations — consistent with this report's ™-only convention).

Appendix H: Evidence-Tier and Claim Audit

Claim class	**Required treatment**
Description of SAIF	Cite official Google source and tag T1
Description of a Google Cloud capability	Cite official product documentation and tag T1; never claim customer implementation
Description of an ODA3 framework	Cite released ODA3 publication and tag T1
Deployment observation	Use a retained organizational record; normally tag T3
Complementarity relationship	Label as ODA3 analysis; tag T2 or T4 as appropriate
Proposed test or implementation method	Label proposed; tag T3 or T4 as appropriate
Legal conclusion	Excluded unless separately provided by qualified counsel
Certification or conformance	Prohibited unless supported by a governed scheme and assessment