

ODA3 INSTITUTE

Technical + Compliance Report

AI-IRF™ Implementation Schema (AIIIS)

Machine-Readable Specification v1.0 — Companion Artifact to AI-IRF™ v1.0

79 controls · 19 domains · 8 SOAR-ready playbooks · 12 severity classes · 5 conformance profiles

DocID: ODA3-2026-06-TCR-HAI-003

Classification: Public

Additional Document Metadata

Doc Type	Technical + Compliance Report — Machine-Readable Companion Artifact (TCR)
Framework	AI-IRF™ v1.0 (part of the GAISSF Ecosystem: GAISSF™ UAIF™ AI-IRF™)
Related Publication	ODA3-2026-06-TCR-HAI-001 (normative source of truth — AIIS is a derived, machine-readable representation)
Evidence Tier	Not applicable — this is a schema/code artifact, not an evidentiary analysis; inline T1–T4 tags do not apply to JSON schema content
Audience	SOAR/XSOAR vendors, GRC platform engineers, assessment-tool developers, and self-assessing organisations
Publish Date / Review Cycle	1 June 2026; correction issued 19 July 2026 / Annual
Maintainer	ODA3 Institute (ODA3 Pvt Ltd)

Document ID	Version	Date	Classification
ODA3-2026-06-TCR-HAI-003	v1.0.1 document edition; JSON Schema v1.0.0	1 June 2026; corrected 19 July 2026	Public GEL v1.0

GAISSF ECOSYSTEM LICENCE (GEL v1.0) | ODA3-2026-06-TCR-HAI-003 | Public

Published by ODA3 Institute (ODA3 Pvt Ltd) under GEL v1.0, effective 1 June 2026. Classification: Public. Commercial Use of this schema in products or services requires a separate written commercial licence (GEL §3.4). AI training data use prohibited without licence (GEL §3.6). AI-IRF™, AI IRF CERT™, GAISSF™, UAIF™, and ODA3™ are trademarks of ODA3 Pvt Ltd, asserted on a use-in-commerce basis. Registration applications are pending. No statement in this document represents any mark as registered unless and until registration has been granted in the relevant jurisdiction (GEL §9.1). ODA3 Institute is the sole owner, proprietor, and governing body of all GAISSF Ecosystem frameworks in perpetuity (GEL §10.5). Governing law: Republic of India; DIAC arbitration, New Delhi seat (GEL §14). Attribution (GEL §6): AI-IRF Implementation Schema (AIIS) v1.0 © ODA3 Institute, GEL v1.0. Companion artifact to AI-IRF™ v1.0; structurally parallel to UAIS v1.0. Schema: <https://oda3.org/schemas/airf/aiis-v1.0.schema.json> | contact@oda3.org

SECTION 1 — PURPOSE AND RELATIONSHIP TO UAIS

AI-IRF™ v1.0 (AI Incident Response Framework v1.0) has, until this publication, existed only as a prose technical specification (ODA3-2026-06-TCR-HAI-001): 79 controls expressed as human-readable SHALL statements, auditable through documented assessment procedures but not directly consumable by software.

The AI-IRF Implementation Schema (AIIS) v1.0 closes that gap. It is the structured, machine-readable encoding of all 79 controls — canonical field names, typed values, and playbook definitions — designed as a structured interchange format for SOAR platforms, GRC tools, and assessment or assurance tooling; vendor-specific mapping, validation, and testing may be required. The schema reserves an optional conformance_test structure for embedded test vectors; this is planned for a future AIIS release and is not populated in v1.0.0.

RELATIONSHIP TO UAIS

AIIS is structurally parallel to UAIS v1.0 (oda3.org/schemas/uaif/uais-v1.0.schema.json), the existing machine-readable schema for UAIF™. Where UAIS encodes incident records and harm-weighted severity scoring for cross-framework reporting, AIIS encodes control conformance, response playbooks, and AI-IRF™-specific severity classification. A vendor already integrating UAIS will recognise the same design conventions in AIIS: canonical IDs, \$id namespace under oda3.org/schemas/, evidence tier tagging, and a severity score band structure intentionally aligned to UAIS's corrected level_4_high band (6.5–8.4).

What AIIS Makes Possible

- SOAR/XSOAR vendors can use PlaybookDefinition objects as structured inputs for integration; vendor-specific adapters, validation, security review, and testing may be required before operational use
- GRC platforms can map ControlRecord objects to existing compliance registers programmatically, rather than re-keying 79 controls by hand
- Future releases are planned to enable assessment-tool developers to build automated readiness scanners using ConformanceTest objects, supplementing rather than replacing human evidence review
- Self-assessing organisations can use the schema for scripted gap analysis before an independent assessment or any future certification engagement; results are readiness evidence, not certification or proof of legal compliance

What This Document Is Not

AIIS does not replace the Technical Report (TCR-HAI-001) as the normative source of truth. Where this schema and the Technical Report conflict, the Technical Report governs, consistent with GEL §10.2 (ODA3 sole version control authority). AIIS is a derived, machine-readable representation of the same 79 controls.

Methodology Note

METHODOLOGY

Control definitions, SHALL statements, and domain structure in this schema are drawn directly from ODA3-2026-06-TCR-HAI-001 [T1] and the four Option B normative extensions ODA3-ECO-011 (Content Safety), ODA3-ECO-013 (Federated Learning, Model Merge Defense, Quantization Security), ODA3-ECO-014 (Hallucination Detection), and ODA3-ECO-015 (Post-Quantum Cryptography) [T1]. Playbook step sequences and time-bound SLA values are derived from the same source documents [T1] with confidence thresholds set as illustrative defaults pending field calibration [T3]. Severity score values for the 12 incident classes are calibrated against the existing UAIF

severity_presentation_score model for cross-framework consistency [T2] and have not yet been validated against a production incident dataset [T3] — organisations adopting AIIS should treat the supplied thresholds as starting defaults, tunable per deployment.

Notably Absent

NOTABLY ABSENT

This schema does not include: (1) a reference implementation package (UAIF has oda3-uaif on PyPI; no equivalent oda3-airf package exists yet — see Section 6 Roadmap); (2) field-validated confidence thresholds for playbook trigger_condition objects — the values supplied are designed defaults, not empirically derived; (3) a paired Executive Brief — this is a technical artifact aimed at engineering and assessment-tooling audiences, not board-level readers; the relevant board-facing companion is the AI-IRF™ v1.0 Executive Brief (ODA3-2026-06-EXB-HAI-002); (4) PII-specific field handling guidance — organisations must apply their own data classification policy to any personal data captured in evidence_type log_record fields; (5) a formal OSCAL profile mapping — AIIS uses a custom JSON Schema rather than OSCAL natively; an OSCAL crosswalk is a candidate for v1.1.

CORRECTION HISTORY

This is v1.0 of AIIS, built against the corrected ODA3-2026-06-TCR-HAI-001 (revised to formally specify Domains 11–13 and restore HAL-4/HAL-5). Domains 11 (Detection Reliability), 12 (Agentic Threat Detection), and 13 (Severity Classification) were newly authored in the TCR-HAI-001 revision — these control IDs existed in earlier assessment artifacts, including a draft Statement of Applicability, prior to this revision but lacked a normative specification until now. CSS-2, CSS-4, CSS-6, and CSS-7 implementation detail was extended beyond the one-line summary in ODA3-ECO-011. PQC-1 was authored as a real, schedule-bound control operationalising the migration timeline in ODA3-ECO-015, rather than the unsourced present-tense control text that previously appeared in scheme documents. The total control count is 79, not 77 — the figure used in earlier ecosystem documents was based on an out-of-date Technical Report.

Document edition v1.0.1 (19 July 2026) corrects the CSS-2 mapping from Article 53(1)(c) to Article 53(1)(d), changes regulatory mappings from compliance assertions to traceability relationships, and clarifies that certification infrastructure, automated readiness tooling, and vendor integrations are not represented as operational or universally interoperable. JSON Schema version remains 1.0.0.

SECTION 2 — SCHEMA OVERVIEW

The schema defines five top-level object types. Each is detailed in this section with its complete field reference, followed by the full JSON Schema source in Appendix A.

Object Type	Count	Purpose
ControlRecord	79	One record per AI-IRF™ v1.0 control — canonical ID, domain, SHALL statement, evidence type (conformance test vectors reserved for a future release, not populated in v1.0.0)
PlaybookDefinition	8	Structured, SOAR-ingestible incident response playbooks covering the automatable domains (MSR, APT, APA, MMD, SAI, MEX, PHY)
SeverityClassification	12	Incident class → severity score → regulatory notification routing, structurally parallel to UAIF's severity model
ConformanceProfile	5	Named adoption bundles (Foundation, Operational, Autonomous, SOAR_Integration, Regulatory_Evidence) mapping to required domain/control subsets
ConformanceTest	Future	Reserved structured test vectors; not populated or included in AIIS v1.0.0.

ControlRecord — Field Reference

Field	Type	Description
control_id	string	Canonical ID, e.g. 'MCP-1', 'SEV-CLS', 'PQC-1'
domain_code	enum (19)	One of the 19 normative domains
domain_name	string	Full domain name
minimum_tier	enum	T1+, T2+, T3, or T3+ — organisational tier at which control becomes mandatory
requirement_type	enum	preventive, detective, corrective, or governance
shall_statement	string	Normative SHALL requirement text from the Technical Report
evidence_type	enum	log_record, configuration_attestation, test_report, policy_document, audit_trail, or signed_artifact
time_bound_sla	object null	Present only for controls with explicit time requirements (e.g. MLC-2, MSR-5)
playbook_ref	string null	Reference to a PlaybookDefinition, where applicable
regulatory_mapping	array	List of {regulation, article} pairs to which this control is mapped for traceability; a mapping does not demonstrate

Field	Type	Description
		legal compliance
conformance_test	object	Embedded ConformanceTest (future release — reserved field, not populated in v1.0.0) — test_id, test_method, pass_criteria, evidence_tier

PlaybookDefinition — Field Reference

Field	Type	Description
playbook_id	string	Pattern 'PB-[A-Z]{2,4}-[0-9]{3}', e.g. 'PB-MSR-001'
trigger_domain	enum (6)	MSR, APT, APA, MMD, SAI, or MEX — the domains with automatable response logic
trigger_condition	object	control_id, condition_expression (boolean string), confidence_threshold
human_in_the_loop	object	required flag, blast_radius_threshold_users, financial_impact_threshold_usd — per MSR-4/APA-4
steps	array	Ordered list of {step_order, action, control_ref, max_duration_seconds}
severity_classification_ref	string	Links to a SeverityClassification incident_class

SECTION 3 — COMPLETE CONTROL RECORD REFERENCE (ALL 79 CONTROLS)

Domain 1 — MCP & A2A Protocol Security (7 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
MC P-1	T1+	preventive	MCP endpoints SHALL require mutual TLS (mTLS) or workload identity federation. No unauthenticated endpoints permitted.	configuration attestation	—
MC P-2	T2+	detective	All MCP message traffic SHALL be inspected at semantic level for injection patterns, malicious skill payloads, or protocol tampering.	log record	—
MC P-3	T2+	corrective	Unregistered MCP servers SHALL be automatically quarantined upon detection; notifications sent to orchestrator and platform owner within 5 minutes.	log record	5 minutes
MC P-4	T2+	preventive	MCP server manifests SHALL be registered with the agent orchestrator before endpoints go live; unregistered servers blocked at network layer.	configuration attestation	—
A2A -1	T1+	preventive	Inter-agent authorization SHALL enforce least-privilege action scopes. Policy-as-code enforcement mandatory.	policy document	—
A2A -2	T2+	preventive	All inter-agent messages SHALL be cryptographically signed (ECDSA) and integrity-verified before processing. Unsigned or tampered messages SHALL be rejected and logged with SIEM alert.	signed artifact	—
A2A -3	T2+	preventive	A2A communication SHALL enforce per-agent-pair rate limits and anomaly detection to prevent agent loop amplification attacks. Circuit-breaker SHALL trigger on >3 anomalies per hour, freezing the amplifying agent and alerting SOC.	configuration attestation	—

Domain 2 — Shadow AI & Continuous Visibility (5 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
SAI-1	T1+	detective	Organizations SHALL deploy instance-aware discovery across network edges, SaaS connectors, endpoints, and browser extensions.	log record	—
SAI-2	T1+	preventive	DLP controls SHALL inspect all AI data egress points for sensitive data exposure.	log record	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
SAI-3	T1+	detective	Shadow AI instances SHALL be automatically classified into risk tiers (sanctioned/shadow/rogue) within 1 hour of discovery and logged to CMDB.	log record	1 hour
SAI-4	T1+	corrective	Rogue AI instances SHALL trigger automated quarantine and incident declaration within 15 minutes of classification; AI Security Lead paged automatically.	log record	15 minutes
SAI-5	T1+	detective	Unauthorised AI deployments classified as rogue SHALL trigger automated SOC alert with risk classification within 5 minutes of rogue designation.	log record	5 minutes

Domain 3 — Adaptive Machine-Speed Response (5 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
MS R-1	T2+	detective	Response frameworks SHALL integrate LLM-as-Judge (LLM-J) metrics for real-time behavioral validation.	log record	—
MS R-2	T2+	corrective	Playbooks SHALL support conditional branching based on live telemetry deltas.	configuration attestation	—
MS R-3	T2+	corrective	Behavioral baselines SHALL trigger automated containment for high-confidence attacks (>95%) without human delay.	log record	—
MS R-4	T1+	governance	HITL SHALL be required for actions with blast radius >10 users, financial impact >\$10K, or safety-critical involvement.	audit trail	—
MS R-5	T2+	detective	MTTC (mean time to containment) for agentic threat events SHALL be 2 seconds or less.	log record	2 seconds

Domain 4 — Model Extraction & Distillation Defenses (4 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
MEX -1	T2+	detective	Behavioral distillation detection SHALL analyze query patterns for systematic extraction.	log record	—
MEX -2	T2+	preventive	Dynamic rate limiting SHALL use per-user/session quotas with semantic clustering.	configuration attestation	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
MEX-3	T3	preventive	Cryptographic output perturbation SHALL be applied for high-confidence suspected extraction (>85% similarity).	configuration attestation	—
MEX-4	T2+	corrective	API key revocation workflow SHALL execute within 5 minutes of confirmed extraction detection.	log record	5 minutes

Domain 5 — AI-BOM & Supply Chain Integrity (5 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
BO M-1	T2+	preventive	Every production AI system SHALL have a cryptographically signed manifest before deployment.	signed artifact	—
BO M-2	T2+	preventive	Training data and model weights SHALL be cryptographically hashed before ingestion.	log record	—
BO M-3	T2+	preventive	External data sources SHALL undergo source authenticity checks including vendor attestation.	policy document	—
BO M-4	T2+	governance	Incident playbooks SHALL include supply chain rollback and vendor notification SLA (<=24h).	policy document	24 hours
BO M-5	T2+	preventive	Runtime BOM validation SHALL occur on every model load; mismatch between loaded model and registered BOM blocks inference.	log record	—

Domain 6 — Agentic Permission Architecture (5 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
APA-1	T1+	preventive	Pre-execution validation gates SHALL enforce policy checks before any consequential action.	configuration attestation	—
APA-2	T1+	preventive	Agents SHALL operate under least-privilege, task-scoped permissions with auto-expiration (<=15 min).	configuration attestation	15 minutes
APA-3	T1+	governance	HITL SHALL be required for: data export >1K records, payment initiation, production delete, PII access, safety override.	audit trail	—
APA-4	T1+	governance	All permission grants, denials, and overrides SHALL be logged with immutable timestamps and cryptographic signing.	audit trail	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
APA-5	T1+	preventive	Tool use SHALL require a pre-execution approval gate, separate from the general APA-1 action validation gate, specific to tool invocation.	configuration attestation	—

Domain 7 — Regulatory & Compliance Operations (4 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
REG-1	T1+	governance	Incident playbooks SHALL include a regulatory notification decision tree.	policy document	—
REG-2	T1+	governance	Legal and Compliance SHALL be embedded as Consulted/Accountable roles in initial triage within 1 hour.	audit trail	1 hour
REG-3	T2+	governance	Organizations SHALL maintain incident-to-regulation mapping tables reviewed quarterly.	policy document	—
REG-4	T2+	governance	Audit-ready documentation SHALL be preserved with cryptographic integrity for minimum 3 years.	audit trail	—

Domain 8 — Physical AI & Safety Integration (4 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
PHY-1	T3	governance	Separate incident playbooks SHALL exist for physical-harm scenarios.	policy document	—
PHY-2	T3	governance	Safety and cybersecurity teams SHALL operate in integrated incident command structures.	policy document	—
PHY-3	T3	preventive	Systems SHALL mandate fail-safe fallback modes upon cyber anomaly detection.	test report	2 seconds
PHY-4	T3	preventive	Hardware-rooted trust SHALL default to human oversight if cryptographic attestation fails.	configuration attestation	—

Domain 9 — Model Lifecycle Security (5 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
MLC-1	T2+	governance	Every production model SHALL have a unique immutable identifier and cryptographic signature.	configuration attestation	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
MLC-2	T3	corrective	Automated rollback capability SHALL execute within 15 minutes of compromise detection.	log record	15 minutes
MLC-3	T2+	governance	Recovery playbooks SHALL include rollback validation gates confirming integrity, performance, and adversarial robustness.	policy document	—
MLC-4	T2+	governance	Model version changes SHALL be logged in immutable audit trail with dual approval.	audit trail	—
MLC-5	T2+	governance	Drift detection thresholds SHALL trigger a dual-approval retraining workflow; retraining without dual approval is blocked.	audit trail	—

Domain 10 — Zero Trust Architecture for AI (ZTAI) (4 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
ZTA-1	T2+	preventive	AI infrastructure SHALL adopt explicit zero-trust design with documented architecture reviewed at least annually.	configuration attestation	—
ZTA-2	T2+	preventive	Continuous verification SHALL require re-authentication every 15 minutes or per high-risk action; no persistent sessions permitted.	configuration attestation	15 minutes
ZTA-3	T2+	preventive	Micro-segmentation SHALL isolate inference, vector storage, and orchestration layers with explicit allow-lists enforced in policy-as-code.	configuration attestation	—
ZTA-4	T2+	preventive	Runtime least-privilege enforcement SHALL include per-request authorization for tool execution with full audit trail.	configuration attestation	—

Domain 11 — Detection Reliability (3 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
REL-1	T2+	governance	Auto-containment SHALL require ≥ 3 ensemble signals, agreement ≥ 0.7	log record	—
REL-2	T2+	governance	Confidence scores SHALL decay 10% per consecutive similar alert	log record	—
REL-3	T2+	governance	Circuit-breaker SHALL disable automation after 3 FP within 1h	log record	—

Domain 12 — Agentic Threat Detection (3 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
APT-1	T2+	detective	Prompt injection detection >=90% across all channels	log record	—
APT-2	T1+	detective	Tool abuse prevention via pre-execution gate on every tool call	log record	—
APT-3	T2+	detective	Memory poisoning detection with automated quarantine <=60s	log record	—

Domain 13 — Severity Classification (2 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
SEV-CLS	T1+	governance	Deterministic severity scoring algorithm implemented and versioned	configuration attestation	—
SEV-RPT	T1+	governance	Reportability thresholds evaluated deterministically	configuration attestation	—

Domain 14 — Content Safety (7 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
CSS-1	T1+	preventive	Harmful content detection pipeline SHALL achieve >99.9% block rate for CSAM (zero-tolerance); alert and log all blocked outputs.	test report	—
CSS-2	T2+	governance	Maintain a sufficiently detailed summary of content used to train the GPAI model, published per EU AI Act Art. 53(1)(d) template requirements. Implement output filtering to detect verbatim or near-ver	policy document	—
CSS-3	T1+	preventive	AI-generated synthetic media SHALL carry machine-readable provenance marker (C2PA recommended) and human-facing disclosure label.	configuration attestation	—
CSS-4	T2+	preventive	Design review SHALL confirm the AI system does not employ subliminal techniques (content presented below the threshold of conscious perception) or techniques designed to exploit known vulnerabilities of specific groups to materially distort behaviour.	test report	—
CSS-5	T1+	detective	Monthly statistical analysis of AI outputs across protected demographic groups	test report	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
CSS-6	T2+	detective	Conduct a quarterly Demographic Impact Assessment (DIA) for all high-risk AI systems. The DIA	test report	—
CSS-7	T1+	preventive	Tier 1 + documented vulnerable population impact assessment, conducted alongside CSS-6's DIA, specifically evaluating outcomes for identified vulnerable groups rather than aggregate demographic category	test report	—

Domain 15 — Federated Learning Security (4 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
FL-1	T2+	preventive	Federated clients authenticated via mutual TLS with client-specific certificates; unregistered/revoked clients rejected.	configuration attestation	—
FL-2	T2+	detective	Byzantine-robust aggregation (FedAvg with clipping; Krum/Bulyan for high-threat); gradients >3 sigma from epoch mean flagged; clients with anomaly rate >5% rejected.	log record	—
FL-3	T2+	preventive	Differential privacy (DP-SGD) applied; epsilon <=1 for high-sensitivity data, epsilon <=8 maximum ceiling for lower-sensitivity data; epsilon=10 not recommended for production.	configuration attestation	—
FL-4	T2+	preventive	Global model signed with ECDSA (or CRYSTALS-Dilithium per PQC-1) after each aggregation round; signed checkpoint stored in AI-BOM registry.	signed artifact	—

Domain 16 — Model Merge Defense (3 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
MM D-1	T2+	preventive	Pre-merge source model vetting: AI-BOM provenance, absence of known malicious fine-tuning datasets, training data manifest completeness verified; incomplete provenance rejected.	signed artifact	—
MM D-2	T2+	detective	Post-merge backdoor scan (Neural Cleanse or ABS); semantic consistency testing across 500+ prompt templates; embedding space anomaly analysis.	test report	—
MM D-3	T2+	preventive	Model merge operations performed in isolated compute environment with no	configuration attestation	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
			internet access; all artifacts hash-verified entering and leaving.		

Domain 17 — Quantization Security (3 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
QBS-1	T2+	preventive	Pre-quantization behavioural baseline established on 1,000+ representative inputs; stored as signed artefact in AI-BOM.	signed artifact	—
QBS-2	T2+	detective	Post-quantization differential analysis against baseline; divergence >5% flagged; divergence >10% rejects model.	test report	—
QBS-3	T2+	preventive	Only trusted, audited quantization toolchains used (ONNX Runtime, TensorRT, PyTorch quantization); toolchain integrity hash-verified; third-party pre-quantized models screened via QBS-1/QBS-2.	signed artifact	—

Domain 18 — Hallucination Detection (5 controls)

ID	Tier	Type	SHALL Statement	Evidence	SLA
HAL-1	T1+	detective	RAG factual grounding verification: all factual claims grounded in retrieved source documents; citation extraction and source verification; ungrounded claims flagged for human review.	log record	—
HAL-2	T2+	governance	Model confidence scoring implemented; outputs below 0.7 confidence for consequential decisions require human review before action; uncertainty surfaced to end users.	configuration attestation	—
HAL-3	T2+	detective	For high-stakes outputs (medical, legal, financial, safety-critical): parallel inference with second model or rule-based verifier; inconsistencies >10% divergence flagged for human review.	log record	—
HAL-4	T1+	detective	Hallucination rate tracked via human correction feedback loop; alert at >3% of evaluated outputs; escalate to AI-IRF corrective action at >5% (trust_erosion_rate threshold).	log record	—
HAL-5	T2+	preventive	For regulated domains (medical diagnosis, legal analysis, financial advice): external fact-checking API or knowledge graph	configuration attestation	—

ID	Tier	Type	SHALL Statement	Evidence	SLA
			verification integrated for domain-specific claims.		

Domain 19 — Post-Quantum Cryptography (1 control)

ID	Tier	Type	SHALL Statement	Evidence	SLA
PQC-1	T3+	preventive	PQC readiness assessment SHALL be completed covering BOM-4, MCP-1, A2A-2, and REG-4 cryptographic use cases. Migration to CRYSTALS-Dilithium (BOM-4, A2A-2) at Tier 3 from Q1 2027; hybrid Kyber-1024 (MCP-1) from Q3 2027; hybrid ECDSA+Dilithium (REG-4, all tiers) from Q3 2027.	configuration attestation	—

SECTION 4 — SOAR-READY PLAYBOOK DEFINITIONS (8 PLAYBOOKS)

These 8 playbooks cover every AI-IRF™ v1.0 control with automatable, time-bound response logic. Each is expressed as a PlaybookDefinition object ready for ingestion into SOAR platforms per the OEM technology licensing model (ODA3-2026-06-STR-COM-001 Stage 6: ServiceNow, Splunk, Palo Alto XSOAR, Microsoft Sentinel, IBM QRadar/Resilient).

PB-SAI-001 — Shadow AI Auto-Quarantine

Field	Value
Trigger domain	SAI
Trigger condition	unregistered_endpoint_detected == true AND classification_confidence >= 0.8
Confidence threshold	0.8
HITL required	No
Blast radius threshold	10 users / \$10,000
Severity class	shadow_ai_exposure
Steps	1. alert_soc (SAI-5, ≤300s) → 2. isolate (SAI-4, ≤900s) → 3. log_evidence (SEV-CLS, ≤60s)

PB-MSR-001 — Machine-Speed Agentic Containment

Field	Value
Trigger domain	MSR
Trigger condition	ensemble_agreement >= 0.7 AND signal_count >= 3
Confidence threshold	0.85
HITL required	Yes
Blast radius threshold	10 users / \$10,000
Severity class	agentic_runaway
Steps	1. rate_limit (MSR-1, ≤2s) → 2. isolate (MSR-3, ≤2s) → 3. escalate_human (MSR-4, ≤60s) → 4. log_evidence (SEV-RPT, ≤60s)

PB-MEX-001 — Model Extraction Response

Field	Value
Trigger domain	MEX
Trigger condition	output_divergence_score >= 0.6 AND query_pattern_anomalous ==

Field	Value
	true
Confidence threshold	0.75
HITL required	No
Blast radius threshold	0 users / \$0
Severity class	model_extraction
Steps	1. rate_limit (MEX-2, ≤30s) → 2. revoke_token (MEX-3, ≤60s) → 3. log_evidence (SEV-CLS, ≤60s)

PB-APA-001 — Agentic Permission Overreach Response

Field	Value
Trigger domain	APA
Trigger condition	permission_scope_exceeded == true
Confidence threshold	0.9
HITL required	Yes
Blast radius threshold	10 users / \$10,000
Severity class	credential_overreach
Steps	1. revoke_token (APA-3, ≤10s) → 2. escalate_human (APA-4, ≤60s) → 3. log_evidence (SEV-CLS, ≤60s)

PB-APT-001 — Agentic Threat Detection Response

Field	Value
Trigger domain	APT
Trigger condition	prompt_injection_score >= 0.9 OR tool_abuse_detected == true OR memory_poisoning_detected == true
Confidence threshold	0.9
HITL required	No
Blast radius threshold	10 users / \$10,000
Severity class	prompt_injection

Field	Value
Steps	1. isolate (APT-1, ≤5s) → 2. quarantine (APT-3, ≤60s) → 3. alert_soc (APT-2, ≤60s) → 4. log_evidence (SEV-RPT, ≤60s)

PB-MLC-001 — Compromised Model Rollback

Field	Value
Trigger domain	MMD
Trigger condition	model_compromise_confirmed == true OR drift_threshold_breached == true
Confidence threshold	0.85
HITL required	Yes
Blast radius threshold	10 users / \$10,000
Severity class	model_drift
Steps	1. escalate_human (MLC-4, ≤120s) → 2. rollback_model (MLC-2, ≤900s) → 3. notify_regulator (SEV-RPT, ≤3600s) → 4. log_evidence (MLC-5, ≤60s)

PB-MMD-001 — Model Merge Backdoor Response

Field	Value
Trigger domain	MMD
Trigger condition	backdoor_signature_detected == true
Confidence threshold	0.8
HITL required	Yes
Blast radius threshold	0 users / \$0
Severity class	supply_chain_compromise
Steps	1. quarantine (MMD-3, ≤60s) → 2. escalate_human (MMD-1, ≤300s) → 3. log_evidence (SEV-CLS, ≤60s)

PB-PHY-001 — Physical AI Safety Failsafe

Field	Value
Trigger domain	MSR
Trigger condition	physical_process_anomaly == true

Field	Value
Confidence threshold	0.95
HITL required	Yes
Blast radius threshold	0 users / \$10,000
Severity class	physical_safety_failure
Steps	1. isolate (PHY-1, ≤1s) → 2. escalate_human (PHY-3, ≤5s) → 3. notify_regulator (SEV-RPT, ≤3600s) → 4. log_evidence (PHY-2, ≤60s)

SECTION 5 — SEVERITY CLASSIFICATION (12 INCIDENT CLASSES)

AI-IRF™ v1.0 severity scoring follows the same 0–10 scale and SEV-1 through SEV-4 levels as UAIF. A score of 6.5 sets an internal regulatory-review routing flag for cross-framework consistency; it does not determine legal reportability, applicable law, authority, or notification deadline.

Incident Class	SEV Level	Score	Reg. Trigger	Notification Routing
prompt injection	SEV-3	6.8	Yes	ENISA/National CSIRT (24h); EBA (if financial) (4h)
model extraction	SEV-3	6.2	No	ENISA (72h)
agentic runaway	SEV-4	8.7	Yes	EBA (4h); National CSIRT (NIS2) (24h)
model drift	SEV-2	4.1	No	—
supply chain compromise	SEV-4	8.9	Yes	National CSIRT (NIS2) (24h); EBA (4h)
hallucination high stakes	SEV-3	6.5	Yes	Market Surveillance Authority (EU AI Act Art. 73) (360h)
federated poisoning	SEV-3	7.0	Yes	National DPA (GDPR) (72h)
quantization backdoor	SEV-3	6.6	Yes	National CSIRT (NIS2) (24h)
content safety violation	SEV-2	5.5	No	Market Surveillance Authority (360h)
physical safety failure	SEV-4	9.4	Yes	National CSIRT (NIS2) (24h); NCIIPC (6h)
shadow ai exposure	SEV-2	4.8	No	—
credential overreach	SEV-3	6.0	No	National DPA (GDPR, if personal data) (72h)

Severity Score Bands

Band	Score Range
Level 1 — Low	0.0 – 2.4
Level 2 — Moderate	2.5 – 4.9
Level 3 — Elevated	5.0 – 6.4

Band	Score Range
Level 4 — High	6.5 – 8.4
Level 4 — Critical	8.5 – 10.0

SECTION 6 — CONFORMANCE PROFILES AND IMPLEMENTATION ROADMAP

Profile	Tier Equivalent	Required Domains	Control Count
Foundation	Tier 1	MCP, A2A, SAI, APA, REG, SEV, CSS, HAL	24
Operational	Tier 2	MCP, A2A, SAI, MSR, REL, MEX, BOM, APA, APT, REG, MLC, ZTA, SEV, CSS, FL, MMD, QBS, HAL	72
Autonomous	Tier 3	MCP, A2A, SAI, MSR, REL, MEX, BOM, APA, APT, REG, PHY, MLC, ZTA, SEV, CSS, FL, MMD, QBS, HAL, PQC	79
SOAR Integration	N/A	SAI, MSR, MEX, APA, APT, MMD, PHY	26
Regulatory Evidence	N/A	REG, SEV, BOM, CSS	17

Reference Implementation Roadmap

UAIF's reference implementation ships as the oda3-uaif Python package. AIIS does not yet have an equivalent package. The following roadmap closes that gap:

Milestone	Target	Output
Schema publication (this document)	Document corrected; canonical JSON deployment to verify	TCR-HAI-003 v1.0.1 + validated JSON Schema v1.0.0; publish and checksum at the canonical URL
oda3-airf reference package (Python)	Q3 2026	pip-installable conformance validator + severity calculator, mirroring oda3-uaif structure
SOAR connector — Palo Alto XSOAR	Q3 2026	Native content pack importing all 8 PlaybookDefinitions
SOAR connector — Microsoft Sentinel	Q4 2026	Playbook templates + AIIS schema-validated incident records
Automated readiness-assessment scanner	Q4 2026	Reference tool implementing planned ConformanceTest objects for readiness self-assessment; not a certification or legal-compliance determination
OSCAL crosswalk (v1.1 candidate)	2027	Formal NIST OSCAL profile mapping for US federal AI procurement compatibility

APPENDIX A — COMPLETE JSON SCHEMA SOURCE

The JSON Schema (Draft-07) source for AIIS v1.0 is reproduced below for reference. Do not copy schema text from a rendered PDF for production use. The separately supplied, validation-tested JSON file is the authoritative machine-readable artifact for this corrected package. Publish it at the canonical URL, record its checksum, and verify JSON syntax and compatibility before implementation. Canonical URL: <https://oda3.org/schemas/airf/aiis-v1.0.schema.json>

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id": "https://oda3.org/schemas/airf/aiis-v1.0.schema.json",
  "title": "AI-IRF Implementation Schema (AIIS) v1.0",
  "version": "1.0.0",
  "description": "Published by ODA3 Institute under the GAISSEF Ecosystem Licence (GEL v1.0). Machine-readable implementation schema for AI-IRF™ v1.0 (AI Incident Response Framework). Defines structured control records, playbook definitions, and severity classification for the 79 controls across 19 domains. The schema reserves an optional conformance_test structure for embedded test vectors; this is planned for a future AIIS release and is not populated in v1.0.0. Machine-readable companion artifact to AI-IRF™ v1.0; structurally parallel to UAIS v1.0 (oda3.org/schemas/uaif/uais-v1.0.schema.json). Attribution: GAISSEF Ecosystem © ODA3 Institute, GEL v1.0 | https://oda3.org/airf",
  "type": "object",
  "definitions": {
    "ControlRecord": {
      "type": "object",
      "description": "Canonical machine-readable representation of a single AI-IRF™ v1.0 control. One ControlRecord exists for each of the 79 published controls. Embedded conformance_test objects are reserved for a future AIIS release and are not present in v1.0.0.",
      "required": [
        "control_id",
        "domain_code",
        "domain_name",
        "minimum_tier",
        "requirement_type",
        "shall_statement",
        "evidence_type"
      ],
      "properties": {
        "control_id": {
          "type": "string",
          "pattern": "^([A-Z][A-Z0-9]{1,3})(-[A-Z]+)?-?(RPT|CLS)?-[0-9]*$",
          "description": "Canonical control identifier, e.g. 'MCP-1', 'SEV-CLS', 'PQC-1'. Matches IDs published in ODA3-2026-06-TCR-HAI-001 and ECO-011/013/014/015."
        },
        "domain_code": {
          "type": "string",
          "enum": [
            "MCP",
            "A2A",
            "SAI",
            "MSR",
            "REL",
            "MEX",
            "BOM",
            "APA"
          ]
        }
      }
    }
  }
}
```

```

        "APT",
        "REG",
        "PHY",
        "MLC",
        "ZTA",
        "SEV",
        "CSS",
        "FL",
        "MMD",
        "QBS",
        "HAL",
        "PQC"
    ],
    "description": "One of the 19 normative AI-IRF™ v1.0 domains."
},
"domain_name": {
    "type": "string",
    "description": "Full domain name, e.g. 'Machine-Speed Response', 'Content Safety'."
},
"minimum_tier": {
    "type": "string",
    "enum": [
        "T1+",
        "T2+",
        "T3",
        "T3+"
    ],
    "description": "Minimum AI-IRF adoption-profile label associated with this control in v1.0. These labels support readiness scoping only and do not represent an issued certification, regulatory approval, or legal-compliance determination. T1+ denotes Tier 1 and above; T3 denotes the Tier 3 profile."
},
"requirement_type": {
    "type": "string",
    "enum": [
        "preventive",
        "detective",
        "corrective",
        "governance"
    ],
    "description": "Functional category of the control, aligned to NIST AI RMF Govern/Map/Measure/Manage where applicable."
},
"shall_statement": {
    "type": "string",
    "description": "The normative SHALL requirement text as published in the Technical Report."
},
"evidence_type": {
    "type": "string",
    "enum": [
        "log_record",
        "configuration_attestation",
        "test_report",
        "policy_document",

```

```
    "audit_trail",
    "signed_artifact"
  ],
  "description": "Primary evidence artifact type an assessor or automated readiness
tool may inspect when evaluating implementation of this control.",
},
"time_bound_sla": {
  "type": [
    "object",
    "null"
  ],
  "description": "Present only for controls with an explicit time requirement (e.g.
MLC-2 15-minute rollback, MSR-5 2-second MTTC).",
  "properties": {
    "metric": {
      "type": "string",
      "enum": [
        "MTTC",
        "MTTD",
        "rollback_sla",
        "response_sla",
        "notification_sla"
      ]
    },
    "threshold_seconds": {
      "type": "number"
    },
    "unit_display": {
      "type": "string"
    }
  }
},
"playbook_ref": {
  "type": [
    "string",
    "null"
  ],
  "description": "Reference to a PlaybookDefinition id, if this control is
operationalised as an automatable playbook step (primarily MSR, APT, APA, MMD domains).",
},
"regulatory_mapping": {
  "type": "array",
  "description": "List of regulation+article pairs mapped to this control for
traceability. A mapping does not demonstrate legal compliance, regulatory approval, or
satisfaction of the cited provision.",
  "items": {
    "type": "object",
    "properties": {
      "regulation": {
        "type": "string"
      },
      "article": {
        "type": "string"
      }
    }
  }
}
```

```
    }
  },
  "conformance_test": {
    "type": "object",
    "description": "Structured test vector for automated readiness evaluation by an
assessment tool or self-assessment script. NOT YET POPULATED in aiis-control-records-
v1.0.json (v1.0.0 release); reserved for a future AIIS release. Records validate without this
field in v1.0.0.",
    "required": [
      "test_id",
      "test_method",
      "pass_criteria"
    ],
    "properties": {
      "test_id": {
        "type": "string"
      },
      "test_method": {
        "type": "string",
        "enum": [
          "document_review",
          "log_inspection",
          "live_simulation",
          "configuration_scan",
          "interview"
        ]
      },
      "pass_criteria": {
        "type": "string"
      },
      "evidence_tier": {
        "type": "string",
        "enum": [
          "T1",
          "T2",
          "T3",
          "T4"
        ]
      }
    }
  }
},
"PlaybookDefinition": {
  "type": "object",
  "description": "A structured incident-response playbook interchange definition for
SOAR-platform integration. Vendor-specific transformation, validation, security review, and
testing may be required before operational use.",
  "required": [
    "playbook_id",
    "name",
    "trigger_domain",
    "trigger_condition",
    "steps"
  ],
}
```

```
"properties": {
  "playbook_id": {
    "type": "string",
    "pattern": "^PB-[A-Z]{2,4}-[0-9]{3}$"
  },
  "name": {
    "type": "string"
  },
  "trigger_domain": {
    "type": "string",
    "enum": [
      "MSR",
      "APT",
      "APA",
      "MMD",
      "SAI",
      "MEX"
    ],
    "description": "The domain whose control breach initiates this playbook. These six domains contain the AI-IRF™ v1.0 controls with automatable, time-bound response logic."
  },
  "trigger_condition": {
    "type": "object",
    "properties": {
      "control_id": {
        "type": "string"
      },
      "condition_expression": {
        "type": "string",
        "description": "Boolean expression evaluated against live telemetry, e.g. 'ensemble_agreement >= 0.7 AND signal_count >= 3'."
      },
      "confidence_threshold": {
        "type": "number",
        "minimum": 0,
        "maximum": 1
      }
    }
  },
  "human_in_the_loop": {
    "type": "object",
    "description": "HITL gate per MSR-4 / APA-4. Required for blast radius exceeding the defined thresholds.",
    "properties": {
      "required": {
        "type": "boolean"
      },
      "blast_radius_threshold_users": {
        "type": "integer"
      },
      "financial_impact_threshold_usd": {
        "type": "number"
      }
    }
  }
},
```

```

    "steps": {
      "type": "array",
      "items": {
        "type": "object",
        "required": [
          "step_order",
          "action",
          "control_ref"
        ],
        "properties": {
          "step_order": {
            "type": "integer"
          },
          "action": {
            "type": "string",
            "enum": [
              "isolate",
              "quarantine",
              "revoke_token",
              "rate_limit",
              "alert_soc",
              "rollback_model",
              "escalate_human",
              "log_evidence",
              "notify_regulator"
            ]
          },
          "control_ref": {
            "type": "string",
            "description": "control_id this step operationalises"
          },
          "max_duration_seconds": {
            "type": "number"
          }
        }
      }
    },
    "severity_classification_ref": {
      "type": "string",
      "description": "Reference to a SeverityClassification object (SEV-CLS) used to determine escalation path."
    },
    "SeverityClassification": {
      "type": "object",
      "description": "Machine-readable severity scoring object for SEV-CLS / SEV-RPT controls. Structurally parallel to UAIS severity_presentation_score but scoped to AI-IRF™ v1.0 incident classes rather than UAIF's cross-framework harm taxonomy.",
      "required": [
        "incident_class",
        "severity_level",
        "severity_score",
        "regulatory_trigger"
      ],
    },
  ],

```

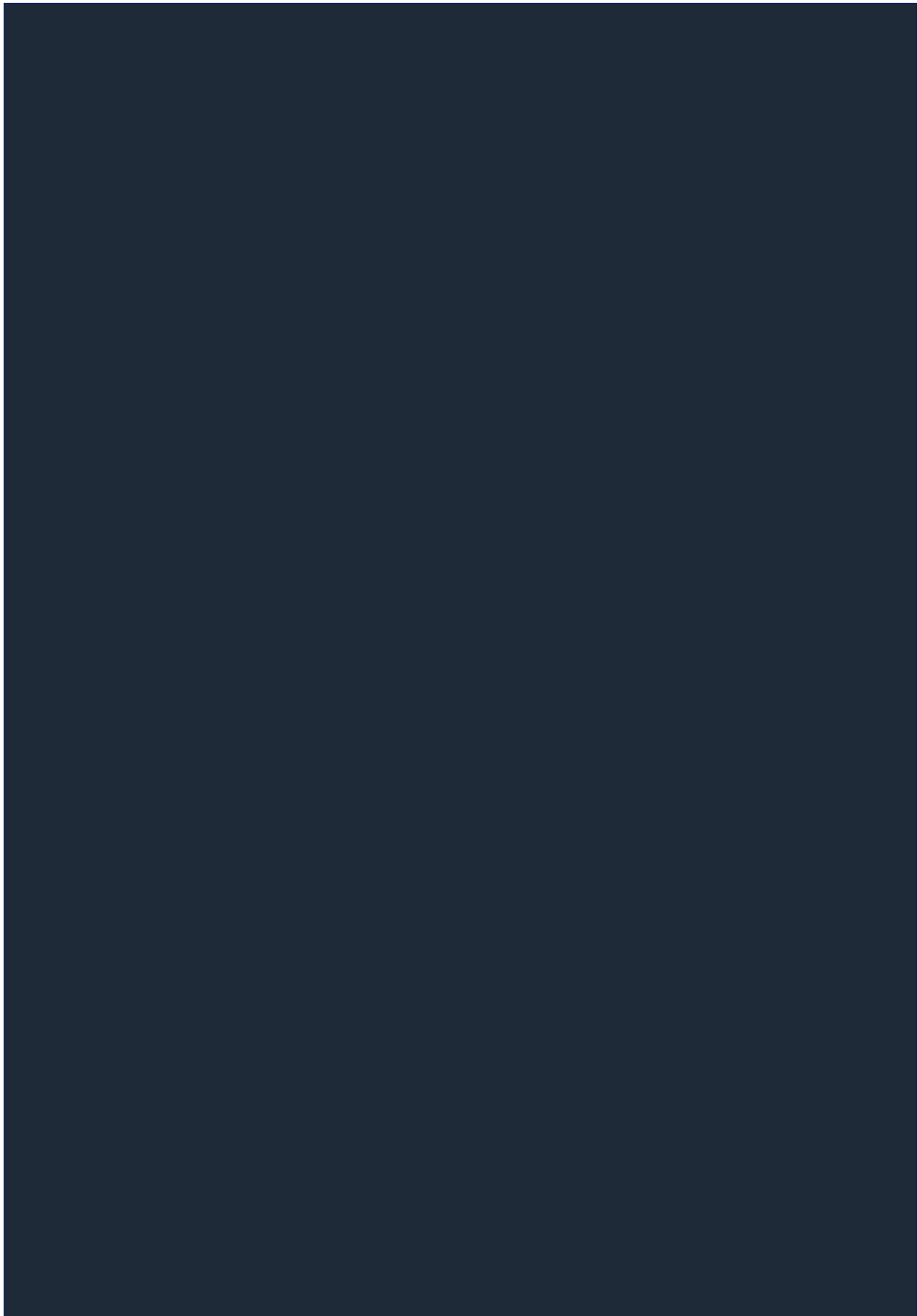
```
"properties": {
  "incident_class": {
    "type": "string",
    "enum": [
      "prompt_injection",
      "model_extraction",
      "agentic_runaway",
      "model_drift",
      "supply_chain_compromise",
      "hallucination_high_stakes",
      "federated_poisoning",
      "quantization_backdoor",
      "content_safety_violation",
      "physical_safety_failure",
      "shadow_ai_exposure",
      "credential_overreach"
    ]
  },
  "severity_level": {
    "type": "integer",
    "minimum": 1,
    "maximum": 4,
    "description": "SEV-1 (low) through SEV-4 (critical), per AI-IRF™ v1.0 §6."
  },
  "severity_score": {
    "type": "number",
    "minimum": 0,
    "maximum": 10
  },
  "score_bands": {
    "type": "object",
    "description": "AI-IRF™ v1.0 severity score bands, structurally aligned to UAIF's corrected level_4_high band (6.5-8.4) per UAIF TCR-STD-002 §4.8.",
    "properties": {
      "level_1_low": {
        "type": "string",
        "const": "0.0-2.4"
      },
      "level_2_moderate": {
        "type": "string",
        "const": "2.5-4.9"
      },
      "level_3_elevated": {
        "type": "string",
        "const": "5.0-6.4"
      },
      "level_4_high": {
        "type": "string",
        "const": "6.5-8.4"
      },
      "level_4_critical": {
        "type": "string",
        "const": "8.5-10.0"
      }
    }
  }
}
```

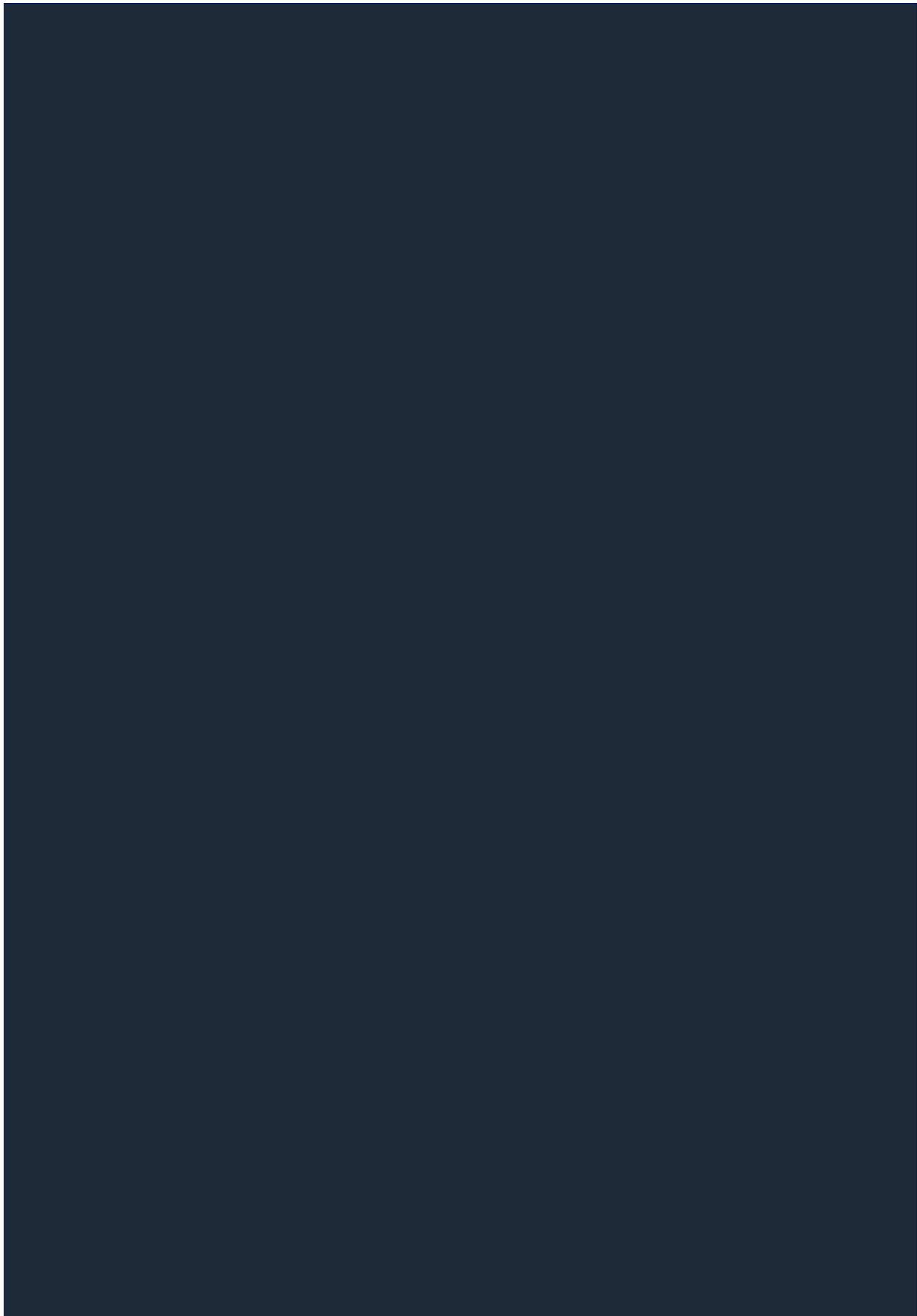
```

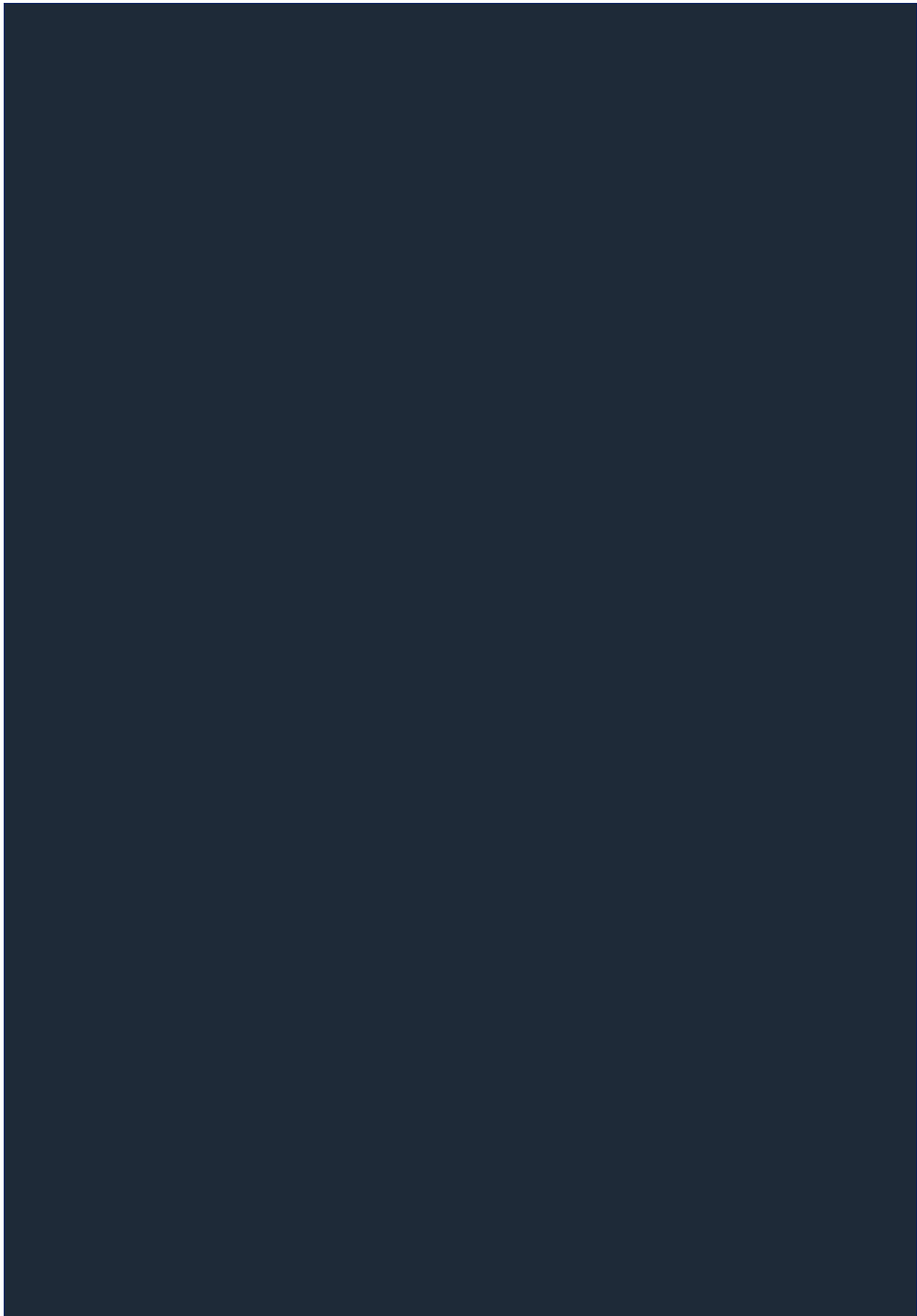
    },
    "regulatory_trigger": {
      "type": "boolean",
      "description": "Internal regulatory-review routing flag set when severity_score is
at least 6.5 for cross-framework consistency. This flag does not determine legal
reportability, applicable law, competent authority, or notification deadline; those require
role- and fact-specific review."
    },
    "notification_routing": {
      "type": "array",
      "items": {
        "type": "object",
        "properties": {
          "jurisdiction": {
            "type": "string"
          },
          "regulator": {
            "type": "string"
          },
          "notification_window_hours": {
            "type": "number"
          }
        }
      }
    },
    "description": "Candidate routing data populated from REG-1/REG-4 logic. Examples
such as DORA 4h, GDPR 72h, NIS2 24h/72h, and the EU AI Act Article 73 general 15-day outer
limit are contextual references only; applicability, trigger time, authority, and shorter
statutory cases must be verified for the incident."
  }
},
"ConformanceProfile": {
  "type": "object",
  "description": "Defines an adoption profile, structurally parallel to UAIF's 5 adoption
profiles (Core, Enterprise, Regulatory, SOC/SIEM, AI Security Extension).",
  "properties": {
    "profile_name": {
      "type": "string",
      "enum": [
        "Foundation",
        "Operational",
        "Autonomous",
        "SOAR_Integration",
        "Regulatory_Evidence"
      ]
    }
  },
  "tier_equivalent": {
    "type": "string",
    "enum": [
      "Tier 1",
      "Tier 2",
      "Tier 3",
      "N/A"
    ]
  }
},

```

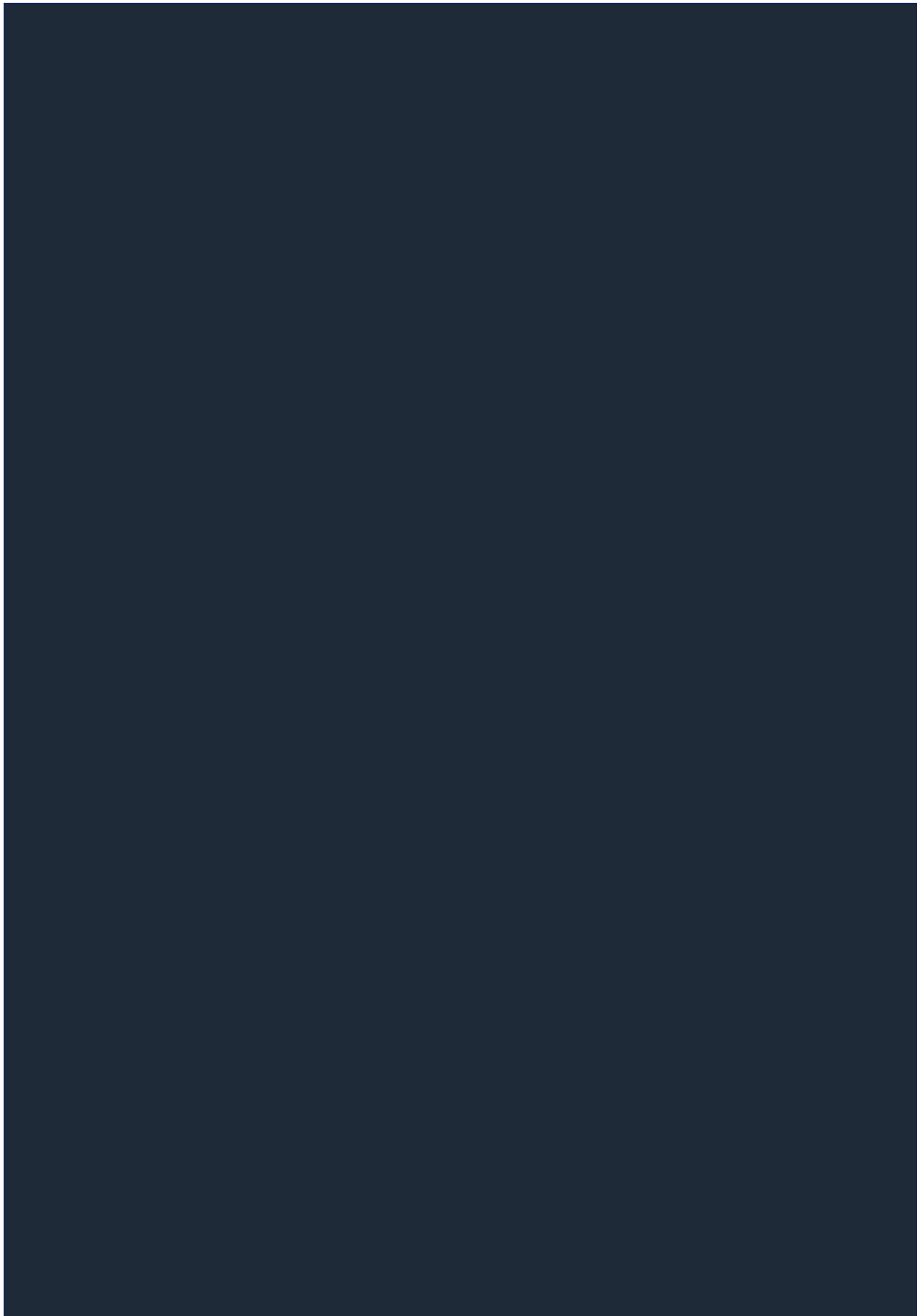
```
    "required_domains": {  
      "type": "array",  
      "items": {  
        "type": "string"  
      }  
    },  
    "required_control_count": {  
      "type": "integer"  
    }  
  }  
}  
}
```

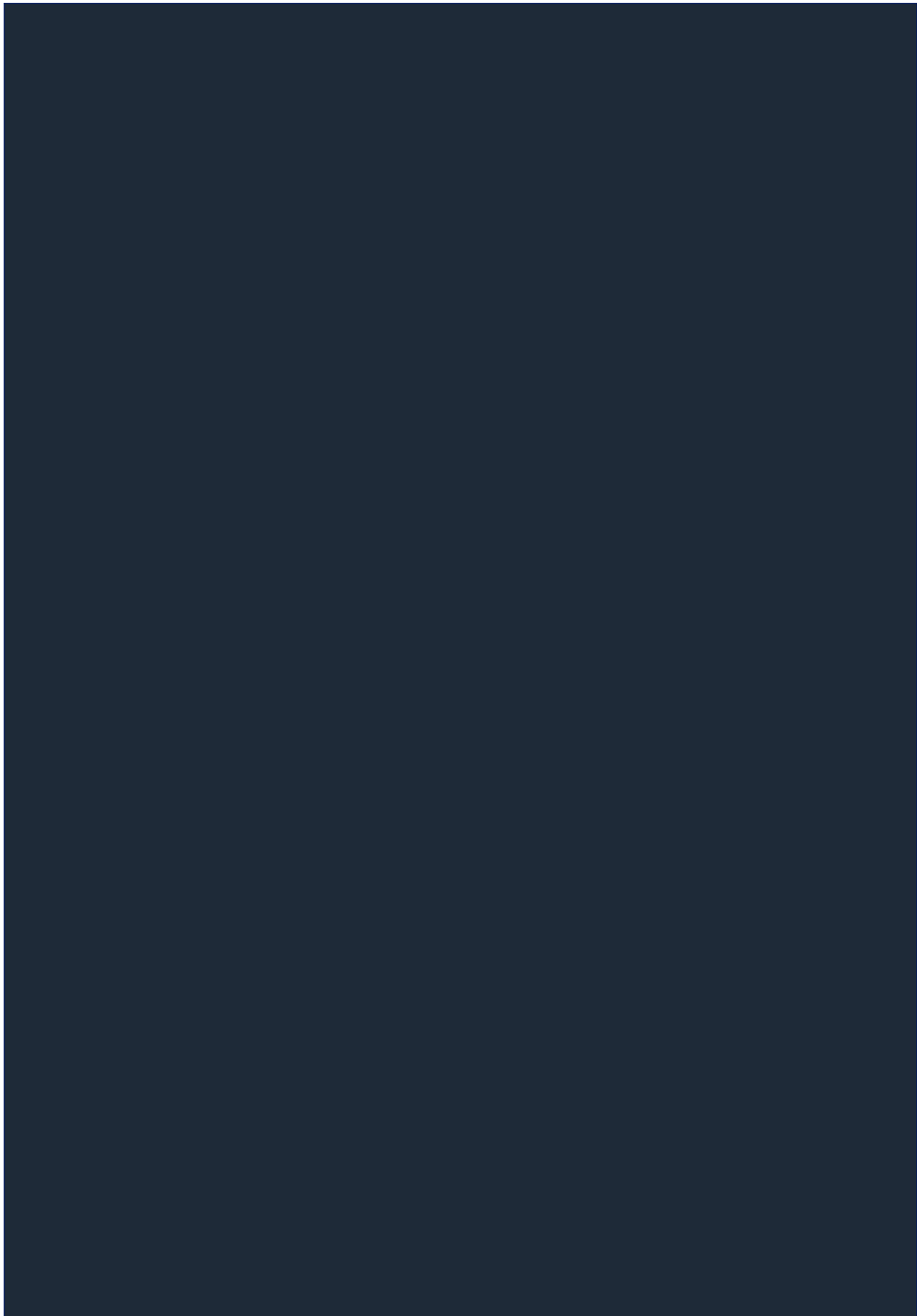














APPENDIX B — SAMPLE CONTROL RECORDS

Three representative ControlRecord instances showing the schema populated with real AI-IRF™ v1.0 control data.

```
[
  {
    "control_id": "MCP-1",
    "domain_code": "MCP",
    "domain_name": "MCP & A2A Protocol Security",
    "minimum_tier": "T1+",
    "requirement_type": "preventive",
    "shall_statement": "MCP endpoints SHALL require mutual TLS (mTLS) or workload identity federation. No unauthenticated endpoints permitted.",
    "evidence_type": "configuration_attestation",
    "time_bound_sla": null,
    "playbook_ref": null,
    "regulatory_mapping": [
      {
        "regulation": "DORA",
        "article": "Art. 9"
      },
      {
        "regulation": "NIS2",
        "article": "Art. 21(2)(a)"
      }
    ]
  },
  {
    "control_id": "MCP-2",
    "domain_code": "MCP",
    "domain_name": "MCP & A2A Protocol Security",
    "minimum_tier": "T2+",
    "requirement_type": "detective",
    "shall_statement": "All MCP message traffic SHALL be inspected at semantic level for injection patterns, malicious skill payloads, or protocol tampering.",
    "evidence_type": "log_record",
    "time_bound_sla": null,
    "playbook_ref": null,
    "regulatory_mapping": [
      {
        "regulation": "EU AI Act",
        "article": "Art. 15"
      }
    ]
  },
  {
    "control_id": "MCP-3",
    "domain_code": "MCP",
    "domain_name": "MCP & A2A Protocol Security",
    "minimum_tier": "T2+",
    "requirement_type": "corrective",
    "shall_statement": "Unregistered MCP servers SHALL be automatically quarantined upon detection; notifications sent to orchestrator and platform owner within 5 minutes.",
    "evidence_type": "log_record",
    "time_bound_sla": {
```

```
    "metric": "response_sla",
    "threshold_seconds": 300,
    "unit_display": "5 minutes"
  },
  "playbook_ref": null,
  "regulatory_mapping": [
    {
      "regulation": "NIS2",
      "article": "Art. 21(2)(a)"
    }
  ]
}
```

APPENDIX C — SAMPLE PLAYBOOK DEFINITION

PB-MSR-001 (Machine-Speed Agentic Containment) as an illustrative PlaybookDefinition instance. Validate the instance and apply vendor-specific mapping, security review, and testing before importing it into a SOAR content pack.

```
{
  "playbook_id": "PB-MSR-001",
  "name": "Machine-Speed Agentic Containment",
  "trigger_domain": "MSR",
  "trigger_condition": {
    "control_id": "MSR-3",
    "condition_expression": "ensemble_agreement >= 0.7 AND signal_count >= 3",
    "confidence_threshold": 0.85
  },
  "human_in_the_loop": {
    "required": true,
    "blast_radius_threshold_users": 10,
    "financial_impact_threshold_usd": 10000
  },
  "steps": [
    {
      "step_order": 1,
      "action": "rate_limit",
      "control_ref": "MSR-1",
      "max_duration_seconds": 2
    },
    {
      "step_order": 2,
      "action": "isolate",
      "control_ref": "MSR-3",
      "max_duration_seconds": 2
    },
    {
      "step_order": 3,
      "action": "escalate_human",
      "control_ref": "MSR-4",
      "max_duration_seconds": 60
    },
    {
      "step_order": 4,
      "action": "log_evidence",
      "control_ref": "SEV-RPT",
      "max_duration_seconds": 60
    }
  ],
  "severity_classification_ref": "agentic_runaway"
}
```