GAP ANALYSIS: EXISTING AI INCIDENT TAXONOMY STANDARDS & GUIDELINES

Technical & Compliance Report | ODA³ Institute | Q2 2026

Audience: Technical Practitioners — CISOs, Security Architects, AI Governance Leads, Compliance Officers

HOW TO USE THIS DOCUMENT — See the navigation guide below to find your entry point.

Audience	Start Here	Key Sections
CISOs & Security Architects	Section 4 (Core Gaps) & Section 6 (UAIF Framework)	Severity matrix, incident/vulnerability split, SIEM integration mapping
AI Governance Leads	Section 2 (Methodology) & Section 5 (Operational Gaps)	Lifecycle modeling, chronic harm proxies, workflow field requirements
Compliance Officers	Executive Brief & Section 3 (Fragmentation Model)	EU AI Act readiness, cross-jurisdiction reporting matrix, regulatory timelines
Standards Bodies	Section 7 (Coverage Matrices) & Appendix	Normative control language, draft schema architecture, methodology notes

1. EXECUTIVE SUMMARY

As of Q2 2026, a single AI incident can cost your organization $2,625–$14,000 in manual crosswalk labor and create a direct path to an EU AI Act audit finding. Why? Because nine competing AI incident standards exist, but none provide end-to-end interoperability. This paper identifies the five critical gaps preventing consistent classification—including the absence of a CVSS-equivalent for AI—and provides the Unified AI Incident Framework (UAIF) to fix them before the August 2026 enforcement deadline.

The global taxonomy landscape is fragmented across three layers:

- Database Schemas (ITU-T, India TEC, OECD) → Define fields to collect; lack causal and harm depth.
- Security/Adversarial Taxonomies (MITRE ATLAS, NIST AML, AVID) → Classify attack techniques; exclude non-adversarial failures and socio-technical factors.
- Management-System Standards (ISO/IEC 42001, ENISA, OWASP) → Require incident processes; lack classification detail for cross-organizational aggregation.

Most Urgent Findings

! Critical

[Observed Gap] No Standardized Severity Classification

No standardized severity classification equivalent to CVSS exists for AI incidents across any current framework.

! Critical

[Observed Gap] Chronic/Systemic Harms Remain Invisible

Acute-incident-focused taxonomies cannot surface cumulative bias, trust erosion, or environmental impact.

! Critical

[Observed Gap] Causal vs. Effect Confusion

Classification layers conflate cause, mechanism, and consequence, blocking effective root-cause analysis.

◎ High

[Inferred Gap] Generative & Agentic AI Types Largely Unclassified

Most taxonomies predate generative AI or treat it as an afterthought, leaving non-adversarial hallucinations, RAG leakage, and agent escalation without classification paths.

! Critical

[Observed Gap] Mandatory Reporting Outpaces Taxonomy Maturity

EU AI Act Articles 61 and 62 and state-level mandates impose reporting obligations that existing taxonomies cannot satisfy in a standardized, defensible manner.

Recommended Immediate Action

SHALL

Adopt the UAIF hybrid model (OECD regulatory metadata → ATLAS adversary mapping → AVID risk surface → Custom AI Control Plane extensions). Implement causal-layer separation and stakeholder-weighted severity scoring for high-risk AI systems by Q3 2026.

⚠ *Limitation Statement: Financial estimates ($2,625–$14,000/incident) are modeled from compliance workflow benchmarks drawn from analogous cybersecurity standards adoption (e.g., manual CVE-to-STIX crosswalks), capped at the 25th percentile of historical analogues. These figures are intended for comparative risk ranking, not precise financial forecasting.*

2. METHODOLOGY

2.1 Standard Screening Criteria

Scope: Active or emerging AI incident, vulnerability, or management-system standards published or under draft review between 2023 and Q2 2026.

- Inclusion: Publicly documented schemas, regulatory guidelines, or consortium taxonomies with at least one incident classification field.
- Exclusion: Purely academic papers without operational field definitions; vendor-proprietary internal playbooks.

2.2 Gap Categorization

All findings are categorized at point of use to distinguish analysis from prescription:

- [Observed Gap] — Explicitly absent in screened standards, confirmed through direct documentation review.
- [Inferred Gap] — Not explicitly stated but logically deduced from structural omissions or scope boundaries.
- [Recommended Design] — Prescriptive direction based on practitioner need and regulatory trajectory.

3. THREE-LAYER FRAGMENTATION MODEL

LAYER 1 — Incident Database Schemas
• ITU-T J.AIID-IBC • India TEC 57090:2025 • OECD 27-Criteria Framework
• Gap: Schema fields exist but lack causal and harm depth

▼

LAYER 2 — AI Security / Adversarial Taxonomies
• NIST AI 100-2e2023/E2025 • MITRE ATLAS v5.4.0 • AVID SEP Domains
• Gap: Security-focused; excludes non-adversarial failures, chronic harms, socio-technical factors

▼

LAYER 3 — Management-System / Process Standards
• ISO/IEC 42001:2023 Annex A.10 • ENISA Ontology • OWASP AISVS (in development)
• Gap: Process requirements without classification detail; cannot aggregate incidents across organizations

Consequence: A single AI incident may require classification in 3+ incompatible formats — creating duplication, mismatched labels, and weak comparability

across telecom, critical infrastructure, cybersecurity, and broader AI governance contexts.

4. CORE GAP ANALYSIS (CONSOLIDATED)

1 CRITICAL

[Observed Gap] No Standardized Severity Classification

Problem: No widely accepted method for rating AI incident severity is comparable to CVSS. Frameworks use qualitative labels, numeric impact scales, or harm-specific measures without a shared weighting methodology.

Impact: Organizations cannot defensibly prioritize remediation. Insurance underwriters lack standardized metrics. Regulators view inconsistent scoring as governance weakness.

SHALL Implement the draft 5-Level AI Severity Matrix (see Section 6.2).

2 CRITICAL

[Observed Gap] Causal vs. Effect Confusion

Problem: Frameworks conflate cause, mechanism, and consequence in the same layer, preventing effective root-cause analysis and targeted remediation.

Required Separation: *Root Cause* → *Exploit Path* → *System Behavior* → *Incident Event* → *Realized Harm*

Impact: One causal chain produces multiple harms; one harm arises from multiple causes. Without separation, root-cause analysis and targeted remediation systematically fail.

3 CRITICAL

[Observed Gap] Chronic & Systemic Harms Are Invisible

Problem: Taxonomies focus on acute incidents. Chronic harms — cumulative bias, trust erosion, inference carbon footprint, model collapse — remain unclassified in every screened standard.

Impact: Long-term reputational, ESG, and systemic risk cannot be quantified, tracked, or reported to oversight bodies.

SHOULD Extend OECD metadata fields with chronic harm proxy metrics (see Section 6.3).

4 HIGH

[Inferred Gap] Incident vs. Vulnerability Boundary Undefined

Problem: AVID and ATLAS classify static vulnerabilities (e.g., "model vulnerable to prompt injection"), not actual incidents (e.g., "prompt injection occurred, causing X harm at Y severity"). SOC teams cannot track exposure duration, remediation status, or realized impact.

SHALL Implement dual-layer classification: Vulnerability Registry (static exposure) + Incident Schema (dynamic occurrence with timestamp, duration, affected stakeholders, remediation actions).

5 HIGH

[Inferred Gap] Generative & Agentic AI Incident Types Unclassified

Problem: Most taxonomies predate generative AI or treat it as an afterthought. Non-adversarial hallucinations, RAG data leakage, agent permission escalation, and multi-agent emergent behavior lack classification paths in any current standard.

SHOULD Publish non-adversarial incident extension aligned with AI Control Plane architecture before EU AI Act enforcement (August 2026).

5. OPERATIONAL & WORKFLOW GAPS

[Observed Gap] Reporting Workflow Underdeveloped

Current initiatives focus on incident classification, not operational intake. Without reporter_confidence, duplicate_detection, and remediation_status fields, incident databases become static archives rather than active management tools.

Example: Without deduplication, a SOC may receive 50 reports of the same model hallucination event, each logged as a separate incident. Without closure tracking, AI governance committees cannot verify that remediation was actually completed.

[Recommended Design] Minimum Viable Schema

The following core fields are required for operational utility in a production incident schema:

Field	Description
system_identity	Unique identifier for the AI system involved
sector_classification	NACE / NAICS sector code for cross-industry aggregation
geography	ISO 3166 country/region code
incident_timestamp	ISO 8601 datetime of incident occurrence
lifecycle_stage	Phase (design, deployment, operation, post-deployment)
cause_category	Root-cause classification aligned to causal layer
harm_category	Realized harm type from AVID SEP / OECD harm taxonomy
severity_score	1–5 level per the UAIF severity matrix
affected_stakeholders	Enumeration of impacted populations
response_actions	Remediation steps taken with closure status

6. THE UNIFIED AI INCIDENT FRAMEWORK (UAIF)

6.1 Architecture — Named Hybrid Model

To solve fragmentation, we propose the UAIF: a layered architecture mapping existing standards to functional roles, extended with AI Control Plane fields that address agentic AI and MCP-specific incident vectors absent from all current standards.

CONTEXT LAYER — Custom AI Control Plane Extensions
- Agent permission scope • MCP endpoint IDs • OAuth pivot chains • RAG data lineage tags
- Purpose: Captures agentic and infrastructure-specific context absent from all existing taxonomies

▼

RISK LAYER — What — Vulnerability Surface & Failure Phase
- AVID SEP Domains (Security / Ethics / Performance) + CRISP-DM lifecycle mapping
- Purpose: Classifies the type of failure and the AI development phase in which it manifests

▼

TACTICAL LAYER — How — Adversary Techniques & TTPs
- MITRE ATLAS v5.4+ (adversary techniques and TTPs)
- Non-adversarial extension for hallucinations / emergent behavior (proposed)
- Purpose: Enables SIEM integration and threat intelligence correlation

▼

BASE LAYER — Schema — Regulatory Metadata & Reporting Alignment
- OECD 27-Criteria Framework (9 mandatory fields, 18 recommended)
- Purpose: Regulatory metadata, harm categories, and cross-jurisdiction reporting workflow alignment

6.2 AI Severity Matrix (Levels 1–5)

Level	Realized Harm Threshold	Operational Response	Regulatory Trigger
1 Info	No external harm; internal discovery only	Log and monitor	None
2 Low	Minor user confusion; < $10K response cost	Standard ticket; 30-day closure	Internal audit only
3 Med	Reputational harm; non-sensitive data exposure; $10K–$100K cost	Escalate to AI Governance; 14-day closure	Sectoral reporting if applicable

Level	Realized Harm Threshold	Operational Response	Regulatory Trigger
4 High	Physical safety impact; sensitive data exposure; > $100K cost	Emergency containment; < 72-hour regulatory notice	GDPR 72hr; EU AI Act Art. 61
5 Crit	Widespread societal harm; critical infrastructure; > €35M fine potential	Cross-agency response; immediate public disclosure	EU AI Act Art. 62; Federal/State mandates

Note: Stakeholder weights adjust the baseline level upward when societal or organizational impact exceeds individual harm thresholds.

6.3 Proxy Metrics for Chronic Harms

Harm Type	Proxy Metric	Tracking Window
Cumulative Bias	(Adverse demographic outcomes) ÷ (Total inference volume)	Rolling 90 days
Environmental Impact	gCO_2e per 1,000 inferences — trend line	Monthly aggregation
Trust Erosion	(Support tickets labeled "AI error") ÷ (Total AI interactions)	Quarterly review

7. COVERAGE MATRICES

Key: ☑ Fully covered ◍ Partially covered ✗ Not covered 🔄 In development

Matrix A — Coverage by Domain

Standard	Security Attacks	Non-Adversarial	Chronic/Systemic	Regulatory Alignment	Machine-Readable
India TEC 57090:2025	◍	◍	✗	◍	◍
ITU-T J.AIID-IBC	◍	◍	✗	◍	🔄
NIST AML Taxonomy	☑	✗	✗	✗	✗
ENISA Ontology	☑	◍	✗	☑	☑
AVID Taxonomy	☑	☑	◍	✗	☑
MITRE ATLAS	☑	✗	✗	✗	☑
OECD Guidelines	◍	◍	◍	☑	✗
ISO/IEC 42001:2023	✗	✗	✗	☑	✗

Matrix B — Coverage by Lifecycle Stage

Standard	Design	Deployment	Operation	Response/Mitigation	Audit/Closure
India TEC	✗	◍	◍	◍	◍
ITU-T	✗	◍	◍	◍	✗
NIST AML	✗	☑	☑	☑	◍
ENISA	◍	◍	◍	☑	◍
AVID	☑	☑	☑	☑	☑
MITRE ATLAS	◍	◍	◍	☑	✗
OECD	◍	◍	◍	☑	◍
ISO/IEC 42001	◍	◍	◍	☑	☑

EXECUTIVE BRIEF

ODA³ Institute | Board & Leadership Audience | Q2 2026

Cross-reference: Technical & Compliance Report — Sections 4–7 provide practitioner-level detail and normative controls for each finding below.

9	$14K	Aug 2026
Competing Standards — Zero Interoperate	Max Per-Incident Manual Labor Cost	EU AI Act Enforcement Deadline

1. Bottom Line Up Front

> ***Fragmentation costs money and invites audits. Without a unified classification approach, your organization faces $2.6K–$14K per incident in manual reporting labor, duplicated compliance efforts, and direct exposure to EU AI Act penalties starting August 2026.***

Nine standards exist; zero interoperate. This brief provides a 90-day adoption path for the Unified AI Incident Framework (UAIF) to achieve compliance, reduce operational friction, and position your organization as a standards influencer in an emerging regulatory space.

2. Three Board-Level Risks

Risk	Why It Matters
▨ No Severity Standard	You cannot prove risk-based prioritization to regulators or insurers without a documented 1–5 severity rubric.
▨ Chronic Harm Blind Spots	Long-term trust, ESG, and systemic risk metrics are invisible to acute-incident-focused taxonomies.
▨ Regulatory Deadline Pressure	EU AI Act serious incident reporting begins August 2026. No mandated taxonomy exists yet; waiting guarantees inconsistent reporting and audit exposure.

3. Financial & Compliance Exposure

Cost Component	Estimated Range	Basis
Manual crosswalk labor	$2,625–$14,000 / incident	15–40 hrs × $175–$350/hr blended rate
Regulatory penalty risk	Variable (up to 7% global turnover)	EU AI Act Art. 71; HIPAA Tier 4

Cost Component	Estimated Range	Basis
Response inefficiency	20–40% time overhead	SOC workflow modeling

Limitation: Modeled from analogous cybersecurity standards adoption benchmarks, capped at the 25th percentile of historical analogues. Use for comparative risk ranking — not precise financial forecasting.

4. Board Action Summary

Risk to Monitor	Questions to Ask Management	Acceptable Evidence
Inconsistent incident classification	"How are we scoring AI incident severity, and can we defend it to an auditor?"	Documented 1–5 severity rubric applied to the last 3 incidents
Cross-border reporting duplication	"Do we map a single incident to one unified schema before jurisdictional translation?"	UAIF crosswalk log showing single-entry → multi-output
Chronic/systemic risk exposure	"Are we tracking cumulative bias, trust erosion, or environmental impact for high-impact models?"	90-day rolling proxy metric dashboard

5. Next 90 Days — SHALL Actions

SHALL Approve UAIF hybrid framework adoption for all high-risk AI systems.

SHALL Allocate 0.2–0.5 FTE to engage NIST AI RMF working group and ITU-T draft review process.

SHALL Implement causal-layer separation and 5-level severity scoring in incident intake workflows.

SHALL Budget $50K–$150K for machine-readable UAIF schema integration with existing SIEM and GRC platforms.

GLOSSARY OF KEY TERMS & ACRONYMS

Term	Definition
AVID	AI Vulnerability Database; classifies AI risk surfaces using Security/Ethical/Performance (SEP) domains
ATLAS	MITRE Adversarial Threat Landscape for AI Systems; tactic/technique matrix for AI-targeted attacks

Term	Definition
CVSS	Common Vulnerability Scoring System; the cybersecurity severity standard (scale 1.0–10.0)
MCP	Model Context Protocol; framework for AI agent authentication and data pipeline routing
NACE/NAICS	Industry classification systems used for sector mapping in regulatory reporting
OECD 27-Criteria	Cross-economy incident metadata framework (9 mandatory fields, 18 recommended)
SOC / SIEM	Security Operations Center / Security Information & Event Management systems
STIX 2.1	Structured Threat Information Expression; machine-readable cyber threat data format
UAIF	Unified AI Incident Framework; proposed layered hybrid model for cross-standard interoperability
CISO	Chief Information Security Officer
AI Control Plane	Authoritative enforcement layer for identity-bound execution, policy-constrained action authorization, and verifiable observability across AI-initiated operations

APPENDIX: NOTES & LIMITATIONS

Regulatory Timeline

EU AI Act enforcement dates and reporting requirements reflect current delegated act roadmaps as of Q2 2026. Dates are subject to market surveillance authority readiness and final publication of implementing acts. Organizations should monitor official EUR-Lex publications for updates.

Scope Boundaries

- Standards screened cover publicly available documentation only. Vendor-proprietary or
- classified government classifications are outside scope.
- Financial impact ranges are modeled estimates derived from analogous standards adoption workflows and are not AI-incident-specific empirical data. They are suitable for directional
- risk ranking only.
- UAIF schema field definitions are draft recommendations intended for standards body review
- and practitioner validation; they do not constitute a finalized interoperability specification.