

ODA3 INSTITUTE

PRACTITIONER CHEAT SHEET

UAIF Classification Cheat Sheet

Quick Reference for Record-Type and Root-Cause Classification

CHT-UAIF-083 | Public

© ODA3 Pvt Ltd

Document Control

Item	Value
Document Identifier	CHT-UAIF-083 (legacy identifier retained for citation continuity)
Doc Type	Practitioner Cheat Sheet - Classification
Framework Tag	UAIF™ v1.0
Canonical Schema	uaif-core-v1.0.json (JSON Schema Draft 2020-12)
Canonical SHA-256	55056e7cfa4fcf4b76cc4656833cb5b360a1b40fa90f23488d488409ccebeeca
Primary References	TCR-UAIF-014 Data Dictionary; TCR-UAIF-015 Validation Rules Catalogue; ODA3-2026-06-TCR-STD-003 Schema Specification
Companion Severity Aid	CHT-UAIF-084 - pending review; use TCR-STD-002 and TCR-UAIF-015 for authoritative current severity rules until reissued
Publication History	Issued 14 July 2026; updated 04 October 2026
Review Cycle	On every schema or classification-guidance correction; otherwise annual
Audience	Analysts and implementers classifying UAIF records
Licence	GAISSF Ecosystem Licence (GEL) v1.0
Contact	contact@oda3.org
Classification	Public

SCOPE BOUNDARY

This sheet is a practitioner reference for values already present in the frozen UAIF™ v1.0 schema. The schema enforces vocabularies and conditional fields; it does not enforce the human analytical reasoning used to choose between plausible root-cause labels. The close-call guidance below is therefore operational guidance, not an additional normative schema rule.

Why This Matters

Consistent classification supports aggregation, triage, trend analysis, and cross-team communication. This sheet helps analysts use the same `record_type`, `root_cause_primary_domain`, and `root_cause_specific_type` vocabulary while preserving the distinction between schema-enforced rules and analyst judgement.

1. Record Type - Pick First

Value	Use when...
Incident	Use when an event has manifested as an incident requiring the UAIF Incident severity bundle. In operational triage, this includes realised harmful effects and cases treated as incidents because harmful effect is credibly imminent. The schema itself enforces only the Incident/Vulnerability vocabulary and the Incident conditional fields.
Vulnerability	Use when a weakness or gap exists but has not manifested as an Incident.

NOT A UAIF v1.0 RECORD TYPE

Observation is not a normative UAIF™ v1.0 `record_type`. Keep unclassified observations or near-misses in upstream case/detection workflows until they are classified as Incident or Vulnerability.

2. Root Cause - Primary Domain

Value	Use when...
Human	A person's action, decision, error, misuse, or policy override is the primary causal domain.
System	The AI system or supporting infrastructure malfunctioned or behaved unexpectedly without an external actor being the primary causal domain.
Data	Training data, retrieval corpus, reference data, or input data is the primary defect vector.
External	An outside actor, supplier, dependency, or third-party system is the primary causal domain.

3. Root Cause - Specific Type

The schema permits 15 named values plus an x_-prefixed custom value. The 'typical primary domain' column below is a practitioner heuristic only; the schema does not enforce a fixed pairing between primary_domain and specific_type.

Typical primary domain	Specific type	Quick definition
External	Adversarial_Attack	Deliberate attacker action caused or contributed to the event. Requires exploit_path_description.
External / System	Prompt_Injection	Untrusted content altered model or agent behaviour through its input or context.
Data / External	Model_Poisoning	Training, fine-tuning, or feedback data was deliberately manipulated to corrupt model behaviour.
Data	Data_Poisoning	Non-model reference data or retrieval corpus was manipulated.
System	RAG_Leakage	Retrieval-augmented generation exposed content it should not have.
System	Hallucination	The system generated fabricated or ungrounded output without evidence of deliberate poisoning.
System / Human	Agent_Escalation	An autonomous or semi-autonomous agent exceeded intended authority or scope.
Human	Policy_Violation	The system or its operation violated an internal policy or guardrail without an external attack being the primary cause.
System	System_Failure	A conventional technical fault, outage, bug, or misconfiguration.
Human	Human_Error	A person's mistake caused or materially contributed to the event.
External	Supply_Chain_Compromise	A vendor, dependency, or third-party component was the source of compromise.
External / Human	Unauthorised_Access	Access occurred without authorisation, independent of whether AI behaviour was involved.
External / System	Data_Breach	Confirmed or suspected unauthorised disclosure of data.
External / System	Denial_of_Service	Availability was degraded or denied.
Human	Misuse	A legitimate user used the system in an unintended or prohibited way.
—	x_[custom]	Use only where none of the canonical values fit; the value must begin with x_ and satisfy the canonical schema pattern.

4. Common Close-Call Distinctions

This vs. that	How to distinguish them
Prompt_Injection vs. Misuse	Prompt_Injection: untrusted content manipulates model/agent context or instructions. Misuse: an authorised user deliberately uses the system outside intended or permitted use.
Hallucination vs. Model_Poisoning	Hallucination: ungrounded or fabricated output without evidence that training/fine-tuning/feedback data was deliberately manipulated. Model_Poisoning: deliberate manipulation of model-training or feedback data is part of the causal chain.
Agent_Escalation vs. Adversarial_Attack	Agent_Escalation describes the agent exceeding intended authority or scope. Adversarial_Attack identifies deliberate attacker action. If both mechanisms are present, choose the root-cause value that best represents the primary causal mechanism and document the other mechanism in the narrative/evidence record.
RAG_Leakage vs. Data_Breach	RAG_Leakage: the disclosure occurred through retrieval/generation behaviour. Data_Breach: broader unauthorised disclosure not specific to a RAG mechanism.
Human_Error vs. Misuse	Human_Error: an unintended mistake. Misuse: an authorised user knowingly acts outside intended or permitted use.

5. After Classifying - What It Triggers

- record_type = "Incident" -> require severity_analytical_score, severity_presentation_score, severity_level, and severity_rationale.
- record_type = "Incident" AND profile is Enterprise or Regulatory -> additionally require realized_harm_categories.
- profile = "Core" -> prohibit harm_acute, harm_chronic_proxy, realized_harm_categories, cumulative_bias_index, linked_vulnerability_id, and exposure_duration_hours.
- root_cause_specific_type = "Adversarial_Attack" -> require exploit_path_description.
- regulatory_trigger key present (true or false) -> require trigger_confidence and human_review_recommended.
- severity_rationale.calibration_status = "provisional" -> require weight_provisional_notice.

- atlas_technique_id present -> extensions must include SOC_Siem.
- Any gated AI Security L5 field present -> extensions must include AI_Security.

6. Classification Workflow

Step	Action	Result
1	Choose record_type	Incident or Vulnerability.
2	Choose profile	Core, Enterprise, or Regulatory.
3	Choose primary domain	Human, System, Data, or External.
4	Choose specific type	Use the canonical list; use x_[custom] only if no canonical value fits.
5	Document rationale	For close calls, record the facts that made one classification more appropriate than the alternatives.
6	Apply conditional fields	Complete all fields triggered by record type, profile, extensions, regulatory trigger, adversarial classification, or provisional calibration.
7	Validate	Run the completed JSON record through the canonical schema and the separate one-decimal score-precision check.

7. Notably Absent

- No Observation record type. Observation is an upstream workflow concept, not a normative UAIF™ v1.0 record.
- No schema-enforced mapping from root_cause_primary_domain to root_cause_specific_type. The domain groupings in this sheet are practitioner guidance.
- No claim that the cheat sheet resolves every multi-cause or ambiguous incident. Where more than one specific type plausibly fits, choose the best-supported primary causal classification and document the ambiguity/rationale.
- No dependency on an unpublished internal classification-rules document. The public schema, TCR-UAIF-014, and TCR-UAIF-015 define the current machine-enforced classification vocabulary and triggers.
- No assumption that CHT-UAIF-084 is current until it completes the same schema-alignment review. Use TCR-STD-002 and TCR-UAIF-015 for current severity requirements.
- No sector-specific x_extensions vocabulary is defined here; such vocabularies require separately validated and published sector artefacts.
- No certification, compliance, endorsement, legal, or regulatory status is created by following this sheet. Certification is not currently offered; any future programme would require separate published governance.

SAFE DESCRIPTION

A classification made consistently with this sheet may be described as "mapped to the UAIF™ v1.0 root-cause taxonomy." Do not describe it as "UAIF-certified", "UAIF-approved", or legally compliant solely because this sheet was followed.

8. Revision Note - 04 October 2026

This corrected reissue preserves legacy identifier CHT-UAIF-083 for citation continuity and aligns the cheat sheet to the frozen canonical uaif-core-v1.0.json schema (SHA-256 55056e7cfa4cf4b76cc4656833cb5b360a1b40fa90f23488d488409ccebceca). It removes Observation as a valid record_type; corrects the Incident/profile conditional rules; adds profile and extension-trigger boundaries; clarifies that primary-domain/specific-type pairings and close-call distinctions are practitioner guidance rather than schema-enforced mappings; removes dependency on unpublished internal classification guidance; qualifies CHT-UAIF-084 pending review; updates Public classification/contact wording; and aligns the certification boundary with the current ODA3 position. No new UAIF normative behavior is introduced.

GAISSF Ecosystem developed by ODA3 Institute, licensed under GEL v1.0.