ODA3 Institute

CERT-In AI Security Auditing Guidelines:

Evidence Readiness and Examination Matrix

An Independent Practitioner Guide to Source Traceability, Audit Evidence and Examination Limitations



EDITION 1 · VERSION 1.0

Publication metadata

Publication type	Practitioner research guide
Title	CERT-In AI Security Auditing Guidelines: Evidence Readiness and Examination Matrix
Institutional author / publisher	ODA3 Institute
Edition	Edition 1 · Version 1.0
Publication date	24 November 2026
Document identifier	ODA3-2026-11-RG-001
Audience	CISOs, security architects, AI governance leads, compliance professionals, independent auditors, standards participants and technical assessment practitioners
Coverage	41 illustrative examinations; selective coverage, not a line-by-line treatment of the source
Copyright	© 2026 ODA3 Pvt Ltd. All rights reserved.

Independent research notice.
This publication is independent ODA3 Institute practitioner research and is not issued, approved, certified or endorsed by CERT-In. Its technical review was founder-led within ODA3 Institute; no separate external independent technical review was performed. It is not a regulatory conformity report, legal opinion or third-party audit.

Executive proposition

An AI audit is not established by the presence of a policy, inventory, risk-tier label or completed checklist. The examination turns on a narrower question: which claim does each record support, for which AI system and version, under what operating conditions, and within what period? Evidence that a control was designed does not, by itself, show that the control was implemented or operated effectively.

CERT-In's Artificial Intelligence (AI) Security Auditing Guidelines (4 November 2026) provide the source context for this practitioner examination matrix. ODA3 Institute translates selected guideline topics into traceable source references, candidate records, examination questions and explicit evidence limitations. The matrix is an independent research aid: it does not determine legal applicability, confer audit authority, or establish conformity with a regulatory requirement.

This edition presents 41 selected evidence-readiness records, together with nine clearly labelled domain overviews. The questions are intended to help readers distinguish documentary assertions from examinable records and observed outcomes. Coverage is selective and is not a line-by-line treatment of every provision in the Guidelines.

Terminology used in this guide

AI risk and impact assessment (AIIA); corrective action plan (CAP); responsible, accountable, consulted and informed (RACI); access control list (ACL); vulnerability assessment and penetration testing (VAPT); security operations centre (SOC); AI bill of materials (AIBOM); software bill of materials (SBOM); information and communication technology (ICT); model risk management (MRM). These are working expansions, not substitutes for definitions in the cited primary sources.

Methodology and authority boundaries

Primary authority.
This guide takes regulatory wording and qualifiers from the dated CERT-In primary documents, not from third-party descriptions or ODA3 Institute's own matrix. The 2022 CERT-In Directions and their FAQs are separate source instruments; this guide does not treat the 2026 Guidelines as replacing them.

Source versus interpretation.
The source reference in each examination heading identifies the cited external provision. The examination question, candidate evidence and evidence limitation are ODA3 Institute's practitioner interpretations, not quotations from CERT-In or regulator-prescribed evidence lists.

Legal applicability is separate.
Statements using "shall," "should," "may," "recommended," or tier conditions within the guidance must not be converted mechanically into a conclusion about enforceable statutory duties, universal procurement dates or sanctions. Applicability may depend on operative law, sectoral instruments and engagement terms; qualified legal advice may be required.

No blanket assurance.
A source requirement, the presence of a control design, a deployment record and an operating observation are distinct evidentiary propositions. An inaccessible record is not evidence of either conformity or nonconformity by itself.

Scope of coverage. Section 6 of the 2026 Guidelines addresses framework architecture and guiding principles. This edition does not allocate a standalone examination group to §6; relevant governance ideas may be reflected in examinations attributed to other sections. Absence of a dedicated group must not be read as absence of the source provisions.

Distinct instruments. The April 2022 CERT-In Directions establish separate ICT incident-reporting and record-related requirements within their applicable scope. They should not be paraphrased as automatically requiring retention of every prompt, retrieval result or AI tool-call trace; the 2026 Guidelines supply a different examination context.

Framework interoperability.
The ODA3 Institute COM-019 methodology provides system-owner context, while COM-020 Public v1.1 provides a bounded self-assessment method. Neither grants CERT-In audit authority or independently establishes regulatory conformity.

How to use the examination guide

Step 1 — Establish applicability.
Inventory the actual AI systems, operators, deployments, vendors and material interfaces. Record the applicable scope and responsible legal entity. Avoid assuming a risk tier or source clause applies merely because it appears in a checklist.

Step 2 — Establish the source criterion.
Record the guideline version, section, tier applicability, exact modality and any separate legal or contractual basis. Preserve the rationale for exceptions or escalation; read tier entries in the column for the tier concerned.

Step 3 — Locate candidate records and their owners.
Use each candidate-evidence line to identify likely record-holding functions and route requests accordingly. Obtain only relevant and permitted records, documenting system version, creation time, provenance, custody, access limitations and transformations.

Step 4 — Form a bounded question.
Test the specific proposition with an appropriate method, including relevant alternate execution paths, negative observations and plausible non-control explanations. Do not equate a documented plan with operation.

Step 5 — Record limits and escalation.
Describe evidence that was inaccessible, refused, not identified, contradictory, checked-and-not-observed or observed negative, preserving source-specific distinctions. Do not turn “not checked” into “compliant.”

Evidence classes and limits

Six classes distinguish what documentation, implementation records and observed outcomes can establish. Examples are illustrative throughout this publication.

Documented design

Examples: Policy, architecture, RACI, system card

Supported proposition: Intended rule, assigned ownership, planned boundary

Does not establish: that the rule was deployed or executed

Implementation

Examples: Versioned configuration, deployment approval, ACL state, model hash

Supported proposition: A specified configuration was represented or observed at a point in time

Does not establish: that all paths used that configuration

Operating outcome

Examples: Authenticated logs, tool invocations, approval and downstream commit records

Supported proposition: Specific decisions/actions occurred in an observed period

Does not establish: that uninstrumented paths or later commits are excluded

Negative observation

Examples: A defined search yielding no matching outcome

Supported proposition: The outcome was not observed in the stated search boundary

Does not establish: global absence without adequate coverage

Third-party representation

Examples: Provider statement, attestation, contractual warranty

Supported proposition: The provider represented a named condition

Does not establish: enterprise-side execution or unobservable internals

Simulation / rehearsal

Examples: Test transcript, tabletop record, simulation output

Supported proposition: The described test or rehearsal occurred

Does not establish: live deployment or actual incident reporting

Taxonomy distinction. An evidence class identifies the kind of material examined (for example, a documented design or a bounded negative observation). An evidence-result state records the outcome of the examination (for example, checked-and-not-observed or evidence inaccessible). One class may produce different states; neither taxonomy assigns a regulatory compliance grade.

Risk-tier interpretation

The following is an ODA3 Institute practitioner paraphrase of Table 11 (Composite Scores and Minimum Audit Requirements, printed p. 25) and Table 14 (Audit Intensity and Obligations by Risk Tier, printed p. 26) in the 2026 CERT-In Guidelines. The table is not a reproduction of source wording and does not independently determine statutory enforceability. Read the underlying columns and conditional phrases in the source before applying any entry. IMPORTANT — The four rows below are ODA3 practitioner paraphrases, not verbatim CERT-In table cells or a determination of legal enforceability. Audit periodicity, independent-auditor requirements and risk-trigger conditions must be checked against the corresponding source table and the assessed system before being used.

CERT-In system-risk tier	Audit mechanism and cadence — ODA3 source paraphrase	Governance, AIIA and testing — ODA3 source paraphrase
Tier 1 (Low)	Self-assessment or basic internal review with documented evidence (Table 11, p. 25).	Designated AI owner; AIIA optional (recommended); basic security assessment (Table 14, p. 26).
Tier 2 (Moderate)	Documented internal audit, plus independent audit by CERT-In-empanelled auditing organisations at least annually (Table 11, p. 25).	AI governance function; simplified AIIA described as mandatory; threat modelling and API/security testing (Table 14, p. 26).
Tier 3 (High)	Periodic independent third-party audit by CERT-In-empanelled auditing organisations; enhanced frequency where material changes, incidents or risk warrant (Table 11, p. 25). Table 14 describes six-month audit where risk or material change warrants (p. 26).	AI Governance Committee; comprehensive AIIA; validation independent of development; comprehensive AI security assessment and VAPT (Table 14, p. 26).
Tier 4 (Critical)	Independent audit by a CERT-In-empanelled auditing organisation with continuous monitoring (Table 11, p. 25); comprehensive audit, continuous monitoring and event-triggered reassessment (Table 14, p. 26).	Board/senior-management oversight; comprehensive AIIA with independent review where appropriate; enhanced independent validation and advanced adversarial/red-team testing (Table 14, p. 26).

The complete tier-specific treatment also varies by the nine core audit domains in §8. This public guide summarises selected overarching distinctions rather than reproducing all 36 internal domain-by-tier assessments. Examine the original CERT-In guidance before applying a tier-specific expectation to a particular system.

Terminology: CERT-In Tier 1–4 refers to system-risk tiers. Any ODA3 evidence or verification labels are separate classifications and must not be read as those system-risk tiers.

Interpretation boundary. CERT-In system-risk tier denotes the assessed impact/exposure category; organisational maturity describes the capability of governance and assurance processes; an audit assurance opinion expresses the examiner's conclusion for a stated scope and period. They are not interchangeable, and no one value proves that a control operated.

Nine core audit domains

The nine overviews below identify the audit-control domains in §8 of the 4 November 2026 CERT-In AI Security Auditing Guidelines. They summarise examination boundaries rather than reproducing the source's complete 65 control areas. Original guideline: CERT-In AI Security Auditing Guidelines (official PDF) These are ODA3 editorial summaries of source domains—not a reproduction of the 65 individual §8 audit-control cells or an official CERT-In examination checklist.

§8.1 AI Governance & Accountability (printed pp. 47–48)

Scope represented: AI Governance Policy, AI Governance Committee, Roles and Responsibilities, AI Risk and Impact Assessment, AI System Inventory, Policy Review and Update, External Requirements Management.

Readiness examination: A policy approved by the board establishes an authorised governance design, not that every production AI system is covered. Examine whether the inventory, ownership assignments, quarterly governance records where applicable, and material-change decisions are consistent with deployed systems.

§8.2 Data Lifecycle, Quality & Privacy (printed pp. 48–50)

Scope represented: Data Authorisation and Purpose Control, Data Classification and Access Control, Data Provenance and Lineage, Data Integrity, Data Quality, Bias and Representation Assessment, Retention and Secure Disposal, Inference Data Separation.

Readiness examination: An approved data-source register establishes an intended purpose and lineage; it does not show that production access and transformation paths honoured those constraints. Reconcile source permissions, provenance, actual processing and unauthorised-change detection.

§8.3 Model Risk Management (MRM) (printed pp. 50–52)

Scope represented: Model Registry, Version Control, Model Risk Identification, Model Validation, Model Performance Monitoring, Model Decommissioning, Model Documentation, AI Bill of Materials (AIBOM).

Readiness examination: A model registry identifies a declared version and its approval history. It does not show that live production traffic used that artefact. Check deployment hashes, serving configuration, rollback events and model-change approvals against an identified observation period.

§8.4 AI Ethics, Bias & Fairness (printed pp. 52–53)

Scope represented: Ethics Policy, Pre-Deployment Bias Testing, Proxy Variable Screening, Bias Mitigation Techniques, Ongoing Fairness Monitoring, Adversarial Fairness Audit.

Readiness examination: A fairness test on a selected dataset cannot establish equitable outcomes for populations or periods never tested. Examine subgroup coverage, proxy-variable handling, thresholds, drift triggers and the independence of any challenge exercise.

§8.5 Explainability & Transparency (printed pp. 53–55)

Scope represented: Explainability Technique Selection, Decision Logging, User-Facing Explanations, Auditor Explanation Access, Transparency Reporting, Consistency Verification, Explanation Quality.

Readiness examination: An explanation may be available but not faithful, stable or accessible to the relevant person or auditor. Compare comparable decisions, investigate inconsistent explanations and preserve the distinction between model documentation and observed explanation quality.

§8.6 Security, Adversarial Resilience & Incident Reporting (printed pp. 55–57)

Scope represented: AI Threat Modelling, Adversarial Robustness Testing, AI-Specific VAPT, CERT-In Incident Reporting, Model Integrity Monitoring, Supply Chain Security, API Security, Infrastructure Hardening.

Readiness examination: A penetration-test report is bounded by the surfaces, configuration and period tested. Trace material threats through the model, API, retrieval, agent and infrastructure paths; do not infer global absence from an untested bypass or a vendor statement.

§8.7 Continuous Monitoring, Drift Management & Validation (printed pp. 57–58)

Scope represented: Performance Monitoring, Data Drift Detection, Concept Drift Detection, Fairness Drift Monitoring, Re-Validation Triggers, Retraining Governance.

Readiness examination: The absence of drift alerts says little about metrics, segments or periods not monitored. Verify configured baselines, alert thresholds, alert delivery, revalidation triggers and deployment decisions following retraining.

§8.8 Human Oversight & Decision Accountability (printed pp. 58–60)

Scope represented: Oversight Model, Approval Gates, Manual Override, Safe State, Reviewer Competence, Override Logging, Effectiveness Monitoring.

Readiness examination: A working override button is a design feature, not an effective intervention. Examine whether authorised reviewers had timely information and could alter consequential outcomes; low override rates may indicate either sound decisions or ineffective challenge.

§8.9 Third-Party and AI Supply-Chain Assurance (printed pp. 60–61)

Scope represented: Vendor Due Diligence, Model and Component Provenance, SBOM, Contractual Security Requirements, Provider Change Management, AI-as-a-Service Monitoring, Dependency and Vulnerability Monitoring, Exit, Portability and Fallback Strategy.

Readiness examination: Supplier due diligence and an exit plan document contractual intent, not that alternate models, data export, continuity or audit-evidence preservation will actually work. Examine dependencies, supplier change notifications and tested recovery paths.

Examination guide — 41 selected records

The 41 examinations constitute a purposive editorial selection of topics from the cited Guidelines. Selection favoured distinct source-to-evidence reasoning patterns, examinability with bounded records, source and tier distinctions, and the usefulness of explicit limitations. Selection was neither random, proportional nor designed to represent all source controls. The omissions and uneven section/domain coverage are disclosed in the coverage note below. Examinations are grouped by controlling source section and numbered E01–E41; the corresponding M identifiers preserve the historical research mapping.

Citation convention: "§" identifies a provision; "Table" identifies the cited source table; "p." and "pp." identify the printed source page(s), not the PDF viewer index. A citation combining several clauses is an ODA3 composite reference; consult each cited clause for its precise scope and wording. References to the 2022 Directions are explicitly marked "2022 PDF". CERT-In system-risk tiers (Tier 1–4) are distinct from ODA3 evidence classifications.

Each record asks what an examiner could establish from a bounded evidence set. Candidate records are illustrative, and limitations apply to the particular claim—not as generic legal exclusions.

Coverage and selection transparency

Counts below describe the primary editorial grouping of the 41 public examinations, not the number of CERT-In obligations, risk exposures or §8 control areas. Multiple examinations cross-reference other sections. This is not a statistically representative or complete coverage claim.

Source grouping	Selected examinations
§4	4
§5	11
§7	6
§8	5
§9	5
§10	2
§11	1
§12	3
§13	1
§14	1
2022 Directions	2

Coverage limitations. There is no standalone §6 group. The §8 domain overviews describe all nine domains, while only a subset receives a directly attributed public examination. The nine domain overviews are not a replacement for the source's 65 audit control areas. For source-level traceability, consult the individual examination headings and the concordance in Annex A.

Guideline §4 — source-linked examinations

E01 M01 | §4.10; p. 21

Examination question Could an auditor see from your inventory which AI systems, components and entities were in scope, which were excluded, and why?

Candidate evidence Versioned AI inventory incl. AI features inside purchased products; system boundary; criteria and applicability register

Evidence limitation An inventory does not prove completeness (embedded or interface-reached AI is commonly missed) and does not establish legal applicability.

E02 M31 | §§4.4–4.9; p. 16; p. 21

Examination question Have you recorded, per system, why it is in or out of scope, including any research-use exclusion?

Candidate evidence Per-system in/out-of-scope rationale; exclusion basis; entity mapping

Evidence limitation Source-based scope reasoning is not a legal determination. The applicability of the guideline to an entity without an Indian presence requires separate legal assessment.

E03 M36 | §§4.8, 8.1; §5.6; p. 19; p. 20; p. 47

Examination question Does each assessment name the exact version that was deployed, and was it repeated after each material change?

Candidate evidence Dated assessment; deployed version identifier; material-change log

Evidence limitation An assessment dated before deployment does not establish that it covered the deployed version.

E04 M44 | §4.8; p. 19; p. 20

Examination question For a retired system, can you show what happened to its models, embeddings and derived data, including copies held by suppliers?

Candidate evidence Lifecycle stage per system; stage-gate records; decommissioning records

Evidence limitation Decommissioning records do not prove deletion of copies held elsewhere.

Guideline §5 — source-linked examinations

E05 M02 | §5.3; pp. 23–24

Examination question Could a second person reproduce each of the eight factor scores from the recorded system version, facts and rationale?

Candidate evidence Versioned factor scores; source facts; misuse/failure scenarios; assumptions; scorer and date

Evidence limitation Factor scores and assumptions do not establish that the eight factors were independently corroborated, nor that the selected controls operated or reduced the scored exposure.

E06 M03 | §5.4; p. 25

Examination question Is each tier traceable to factor values, and is every upward reclassification or exception approved by a named authority with a date?

Candidate evidence Scoring record; factor-by-factor rationale; escalation or exception approval; reclassification history

Evidence limitation A tier label does not demonstrate effective implementation or determine legal enforcement.

E07 M04 | §§5.4–5.5; pp. 25–26

Examination question Does your audit plan for each Tier 2 system show an annual independent audit by an empanelled organisation and a documented internal audit?

Candidate evidence Source tables; per-system audit plan; empanelled-auditor engagement record if claimed

Evidence limitation An approved audit plan establishes intended cadence and assignment, but does not prove that an independent audit occurred, that an auditor had appropriate engagement authority, or that the audit scope and findings were adequate. Guideline wording alone does not determine legal enforceability, effective dates or penalties; applicability must be assessed separately.

E08 M32 | §5.7; p. 27; p. 28

Examination question Where you rely on proportionality relief, is each relying system documented as Tier 1 or 2?

Candidate evidence Entity-size basis; tier per system; scaled procedures used

Evidence limitation Organisation size does not lower a system's tier; relief does not reach Tier 3 or 4.

E09 M33 | §§5.4–5.5; pp. 25–26

Examination question For a Tier 1 system, can you produce a self-assessment linked to inspectable evidence and identify its reviewer?

Candidate evidence Self-assessment record; evidence index; reviewer

Evidence limitation A self-assessment without a retrievable evidence index may be impossible to substantiate; even an inspectable self-assessment is not proof that controls operated.

E10 M34 | §5.4 Table 11, p. 25 (Tier 2 annual; Tier 3 enhanced conditional); §5.5 Table 14, p. 26 (Tier 3 conditional six-month)

Examination question Is your audit frequency for each Tier 3 system justified against risk and material-change triggers?

Candidate evidence Audit plan and frequency rationale; enhanced-monitoring design; independent governance and fairness review

Evidence limitation A scheduled audit plan does not show that audits occurred or covered material changes.

E11 M35 | §5.4 Table 11, p. 25; §5.5 Table 14, p. 26 (audit intensity and reassessment context)

Examination question Can you trace the events that triggered audit reassessment to dated decisions and supporting evidence?

Candidate evidence Monitoring coverage map; reassessment trigger log; dated reassessment decisions

Evidence limitation Monitoring configuration does not establish continuous coverage, and a reassessment log does not show that every triggering event was detected.

E12 M38 | §§5.6, 8.9; pp. 27, 60–61

Examination question Would a provider-side model update be visible to you, and did it trigger revalidation?

Candidate evidence Provider change log; revalidation record

Evidence limitation Silent provider-side updates may not appear in customer change records.

E13 M49 | §5.5, Table 14; p. 26

Examination question For each Tier 2–4 system, can you present a dated AIIA at the depth stated for its assigned tier, linked to the relevant version and use case?

Candidate evidence AI risk/impact assessment, version and date, coverage of current system, assessed mitigations

Evidence limitation An AIIA records evaluated risks and proposed mitigations; it does not demonstrate that controls operated.

E14 M50 | §5.5, Table 14; p. 26

Examination question Can you identify the governance owner or body corresponding to each system tier and provide records of its decisions?

Candidate evidence Accountability register, committee charter, minutes, escalation records and independence arrangements

Evidence limitation A named owner, committee charter or board agenda does not establish effective oversight or operational intervention.

E15 M51 | §5.5, Table 14; p. 26

Examination question Does the test scope and depth match the tier and actual deployed system version, including relevant interfaces and threats?

Candidate evidence Test plans, asset/system scope, threat model, VAPT and red-team reports, retest records, monitoring coverage

Evidence limitation Testing conclusions are bounded to scenarios, environments, scope, and period examined; a report does not prove universal security.

Guideline §7 — source-linked examinations

E16 M05 | §7.7; pp. 41–42

Examination question Could anyone on the audit team be reviewing a control they designed, built, operated or advised on?

Candidate evidence Role history; engagement records; conflict declaration; independence rationale

Evidence limitation A declaration without underlying role evidence cannot establish independence.

E17 M06 | §7.8; p. 42; p. 43

Examination question Could you answer a pre-audit request with versioned, attributable records within the engagement window?

Candidate evidence Scoped document request; versions; provenance; completeness and access records

Evidence limitation Provided files alone do not establish operating truth or full-population coverage.

E18 M07 | §7.9; p. 43

Examination question Do your AI supplier contracts give audit and traceability access, and where they do not, is the gap recorded with an alternative procedure?

Candidate evidence Contract audit and traceability rights; record of inaccessible items; alternative procedure and result

Evidence limitation Unavailable or refused evidence does not establish conformity or failure. Contractual, statutory and other disclosure rights must be determined separately for the relevant parties and jurisdiction.

E19 M08 | §7.10, p. 44 (AI-assisted audit working-paper requirements)

Examination question If your auditor uses AI tools, could each AI-assisted finding be traced to original evidence and a named competent reviewer?

Candidate evidence Tool approval; working-paper log fields; validation record; reviewer sign-off; suspension record if reliability compromised

Evidence limitation AI-generated narrative is not independent evidence or an audit opinion.

E20 M39 | §7.6; p. 40; p. 41

Examination question For a Tier 3 or 4 system, can you demonstrate that the proposed audit team meets §7.6 competence and team-composition expectations, including the restriction on sole-auditor engagements?

Candidate evidence Team roster; competence evidence; certifications

Evidence limitation Credentials and a proposed roster do not establish the team's competence on this AI system or that required independent roles and personnel actually participated throughout the engagement.

E21 M41 | §7.10, pp. 43–45 (human auditor accountability; working-paper specifics p. 44)

Examination question Does the audit report name its limitations, and are key decisions attributed to named people?

Candidate evidence Decision log naming human decision-makers; limitation statements in report

Evidence limitation Reviewer sign-off does not show the reviewer examined the underlying evidence.

Guideline §8 — source-linked examinations

E22 M37 | §8.3; §8.9; pp. 51, 60–61

Examination question Does your AIBOM let an auditor confirm the model actually running matches the one recorded?

Candidate evidence AIBOM per system; hash records; plugin and interface list

Evidence limitation A BOM lists declared components; it does not prove there is no undeclared component, dynamic plugin or provider-side model swap.

E23 M40 | §8.3; §5.5; p. 51; p. 26

Examination question Does the person or function validating each model sit outside its development line to the degree its tier requires?

Candidate evidence Validation org chart; reporting lines; external engagement records

Evidence limitation Organisational separation on paper does not establish independence in practice.

E24 M43 | §8.6; p. 56

Examination question Did the last pen test cover AI-specific paths (prompt injection, tool abuse, retrieval leakage) on the current version?

Candidate evidence Dated pen-test report; scope; AI-specific test cases

Evidence limitation A dated report does not show it covered AI-specific attack paths or the current version.

E25 M45 | §8.8; pp. 58–60

Examination question Is the chosen oversight model justified against the system's autonomy and impact scores?

Candidate evidence Oversight-model rationale per system tied to tier

Evidence limitation A rationale shows the choice was considered, not that oversight works.

E26 M48 | §8.6, p. 56 (2026 Guidelines); Directions (i), (iv), pp. 1–2 (2022 PDF)

Examination question Can you produce timestamp and retention evidence for the ICT logs applicable under the 2022 CERT-In Directions, separately identifying any additional AI execution traces you rely on?

Candidate evidence Log inventory per AI component; time-source config; retention evidence

Evidence limitation Applicable ICT log retention does not establish complete AI execution provenance. The 2022 Directions must not be represented as universally mandating retention of every prompt, retrieval event or tool invocation.

Guideline §9 — source-linked examinations

E27 M18 | §9.6; p. 66

Examination question Can a sampled retrieval be traced to the authorisation boundary that should have applied to that user?

Candidate evidence Retrieval ACLs; tenant identity; query/retrieval traces; corpus versions

Evidence limitation One correct retrieval cannot establish isolation for all corpora or alternate paths.

E28 M19 | §9.7; pp. 66–67

Examination question Do your records show only that the gateway decided, or also what actually happened downstream?

Candidate evidence Agent identity; scoped delegation; gateway and tool-result traces

Evidence limitation Gateway denial alone cannot prove absence of prohibited execution elsewhere.

E29 M20 | §9.9; p. 67; p. 68

Examination question Can authority and action sequence be reconstructed across the whole agent chain?

Candidate evidence Inter-agent handoffs; actor identities; inter-agent authentication; delegated scope; tool execution records

Evidence limitation Missing handoffs or downstream systems limit end-to-end attribution.

E30 M21 | §9.10; p. 68; p. 69

Examination question For a cyber-physical AI system, what physical outcome did you observe during the defined interruption or fault test?

Candidate evidence Controller and actuator events; safety-function tests; simulation boundary

Evidence limitation Simulation, software cancellation or command receipt alone do not prove a physical safe state.

E31 M46 | §9.4, p. 65 (kill-switch checklist); §§9.9–9.10, pp. 67–69 (multi-agent and physical-system conditions)

Examination question When shutdown was tested, what happened to in-flight actions, queued tasks and delegated sub-agents?

Candidate evidence Shutdown design; test records; in-flight action handling

Evidence limitation A shutdown tested in staging does not establish that in-flight, queued or delegated actions halt in production; a documented kill switch is design evidence, not operating evidence.

Guideline §10 — source-linked examinations

E32 M22 | §10.5; p. 75; p. 76

Examination question Was an AI-specific event detected, triaged and escalated under examined conditions?

Candidate evidence Detection rule history; alert traces; triage outcomes; analyst training

Evidence limitation Records of SOC integration do not show coverage of AI events outside deployed detections.

E33 M23 | §10.6; p. 76; p. 77

Examination question Is each reporting decision traceable to the observed facts and to the time the incident was noticed or brought to notice?

Candidate evidence Incident timeline; classification rationale; reports and receipts

Evidence limitation A documented reportability decision does not establish that it was made within the applicable reporting window, or that any required report was delivered on time.

Guideline §11 — source-linked examinations

E34 M26 | §11.3; p. 80; p. 81

Examination question Does a sampled decision show that required intervention could occur and did occur?

Candidate evidence Review trigger; sampled decisions; override tests; actual reviewer interventions

Evidence limitation A documented oversight control does not show effective intervention across untested paths.

Guideline §12 — source-linked examinations

E35 M27 | §12.7; p. 87

Examination question What supports escalating a finding internally versus reporting an incident externally?

Candidate evidence Finding register; incident classification; escalation decision; legal criteria

Evidence limitation Finding severity is not by itself a six-hour reporting trigger.

E36 M28 | §§12.2–12.6; p. 83; p. 86

Examination question Do post-remediation records show the originally affected behaviour was re-examined?

Candidate evidence Finding evidence; corrective action; re-test result; updated records

Evidence limitation A closed remediation ticket does not alone establish verified control recovery.

E37 M42 | §12.6; p. 86

Examination question If an Unsatisfactory assurance opinion or two or more Critical findings triggered §12.6, can you show that the re-audit was completed within three months (or sooner if directed by a regulator) and covered the Critical and High findings and related domains?

Candidate evidence Re-audit report and completion date; original assurance opinion and findings; corrective action plan; scope and re-test evidence

Evidence limitation A re-audit booking or completion certificate does not establish that the affected controls were retested effectively or that the underlying causes of the original findings were removed.

Guideline §13 — source-linked examinations

E38 M29 | §13.4, pp. 92–93

Examination question Can a reported maturity level be reconstructed from evidence for each of the nine domains?

Candidate evidence Domain evidence; per-domain scores; rationale; exceptions

Evidence limitation A maturity level is not a risk tier and is not proof that a particular control worked.

Guideline §14 — source-linked examinations

E39 M30 | §14; p. 94

Examination question Is each referenced standard applicable through a controlling source or selected criteria?

Candidate evidence Applicable-criteria register; instrument status; source licence

Evidence limitation Being referenced does not make every standard mandatory.

Other source-linked examinations

E40 M24 | Direction (ii), p. 2; Annexure I, item (xx), p. 6 (2022 PDF)

Examination question Which reportable category was considered, and when was the incident first noticed or brought to notice?

Candidate evidence Noticing timeline; initial observations; category considered; submission receipts

Evidence limitation No legal determination for any specific incident; the 2022 instrument stays distinct from the 2026 guideline.

E41 M25 | Directions (i), (iv), pp. 1–3 (2022 PDF)

Examination question Can you demonstrate that the relevant ICT records used synchronised time sources for the examined period?

Candidate evidence ICT clock configuration; time-source settings; synchronisation or drift-monitoring records

Evidence limitation Time-synchronisation evidence does not establish log completeness, retention or full AI execution provenance.

High-consequence cross-references. Where the auditor’s tier judgment differs from the organisation’s, §7.3 requires the higher tier for audit scoping. Missing evidence is not, by itself, a basis to mark an applicable control “not applicable”; the Statement of Applicability and its justification should remain examinable. Planned material changes can trigger reassessment before deployment (§5.6). These are additional source distinctions, not new universal legal conclusions.

Specialised systems and cross-cutting requirements

The original research register additionally covers generative-AI controls (§9.3), agentic and multi-agent controls (§§9.4, 9.7, 9.9), retrieval-augmented generation (§9.6), autonomous or cyber-physical systems (§9.10), infrastructure hardening (§10.3), SOC integration and reporting (§§10.5–10.6), human review (§11), findings and re-audit (§12), and maturity (§13). Section-specific cross-references are retained rather than treated as extra distinct requirements. The broader internal register also records mixed-instrument applicability questions outside the scope of this public edition; those require separate legal mapping.

For example, a gateway’s recorded denial can support a statement about its decision, while leaving unresolved whether an action executed along a different path. A model registry can support traceability of a declared version, but without execution records it does not prove that every request used that artefact. An AI supplier’s refusal to provide a specific item must be recorded distinctly from information simply not located in the available evidence basis.

Three worked evidence examinations

Example A — Consequential agent transaction

A system owner states that payment instructions require approval before execution. The investigator should distinguish (i) whether the policy applied to the transaction, (ii) whether an applicable approval existed, (iii) whether the gateway made a deny decision, and (iv) whether a prohibited downstream commit occurred in the defined examination boundary. A gateway denial does not settle (iv). A complete and timestamped search covering identified paths may support not observed within the searched boundary; it does not prove global absence.

Illustrative completed examination (not a real-system finding): deployment AG-01, version V3, 1–7 November 2026; the examiner obtained the approved policy, sampled gateway decisions, deployment routing records and downstream commit logs for named transaction IDs. The policy and gateway logs supported that the gateway denied a named unapproved request. A query of the specified downstream commit index, with the query and scope retained, returned zero matching commits for those IDs. Evidence-result state: “Checked-and-not-observed” within that defined index and period. The downstream event feed from a separate connector was inaccessible, so the wider execution-outcome proposition remains “Unresolved/inconclusive”. The two states apply to different propositions and must not be collapsed into a global “no execution occurred” conclusion.

Example B — Model update and provider-side drift

A model card and registry identify version V, but provider service updates are not under enterprise control. Ask separately what the enterprise deployed, what the supplier represented, what was observed for requests made during the relevant period, and which provider-side details were unavailable. An independent evaluation of observable outputs may be possible, but it does not reveal unexamined provider internals or establish that an unseen provider-controlled safeguard operated.

Example C — Human oversight in a high-impact workflow

A workflow advertises human-in-the-loop review and displays a working override button. Policies and training show intended arrangements; event records may show actual approval or intervention; independent observations may test whether those actions changed execution under specified conditions. Low override rates are ambiguous without reviewing the decision context, reviewer authority, and whether meaningful challenge occurred. If a cyber-physical actuator is involved, software acknowledgement is not sufficient by itself to establish safe physical transition.

Interpreting unresolved or negative evidence

The following eight states describe how evidence may appear in a bounded examination. They are descriptive record states, not regulatory grades or conformity findings. The category names follow COM-020 Public v1.1, §71; the descriptions below are explanatory practitioner paraphrases.

Evidence record state	How to interpret the state
Observed negative evidence	An examined record affirmatively shows a negative event or outcome within the stated boundary.
Contradictory evidence	Two or more sources support materially incompatible accounts requiring reconciliation.
Checked-and-not-observed	A specified search or test was performed and did not observe the target condition within its coverage.
Evidence not identified	The examined inventory or requests did not locate the required evidence; this does not establish its universal absence.
Evidence inaccessible	Evidence was identified but could not be examined because of technical, contractual or practical limits.
Evidence refused	Access or production was requested and declined; refusal is not proof of the underlying control outcome.
Unknown	The available evidence does not establish a reliable status for the proposition.
Unresolved / inconclusive verification result	An examination occurred but its result does not support a definite conclusion without further checks.

Annex A — Public examination concordance

This is a traceability index for the 41 public examinations, not an exhaustive source-control listing. “Primary grouping” denotes where the examination appears in this guide; several examinations cite secondary sections. Historical M identifiers are retained solely for cross-reference.

Public ID	Research ID	Grouping	Primary citation (abbreviated)
E01	M01	§4	§4.10
E02	M31	§4	§§4.4–4.9
E03	M36	§4	§§4.8, 8.1
E04	M44	§4	§4.8
E05	M02	§5	§5.3
E06	M03	§5	§5.4
E07	M04	§5	§§5.4–5.5
E08	M32	§5	§5.7
E09	M33	§5	§§5.4–5.5
E10	M34	§5	§5.4 Table 11, p. 25; §5.5 Table 14, p. 26
E11	M35	§5	§5.4 Table 11, p. 25; §5.5 Table 14, p. 26
E12	M38	§5	§§5.6, 8.9
E13	M49	§5	§5.5, Table 14
E14	M50	§5	§5.5, Table 14
E15	M51	§5	§5.5, Table 14
E16	M05	§7	§7.7
E17	M06	§7	§7.8
E18	M07	§7	§7.9
E19	M08	§7	§7.10, p. 44
E20	M39	§7	§7.6
E21	M41	§7	§7.10, pp. 43–45

Concordance — continued

Public ID	Research ID	Grouping	Primary citation (abbreviated)
E22	M37	§8	§8.3
E23	M40	§8	§8.3
E24	M43	§8	§8.6
E25	M45	§8	§8.8
E26	M48	§8	§8.6, p. 56 (2026 Guidelines); Directions (i), (iv), pp. 1–2 (2022 PDF)
E27	M18	§9	§9.6
E28	M19	§9	§9.7
E29	M20	§9	§9.9
E30	M21	§9	§9.10
E31	M46	§9	§9.4, p. 65; §§9.9–9.10, pp. 67–69
E32	M22	§10	§10.5
E33	M23	§10	§10.6
E34	M26	§11	§11.3
E35	M27	§12	§12.7
E36	M28	§12	§§12.2–12.6
E37	M42	§12	§12.6
E38	M29	§13	§13.4, pp. 92–93
E39	M30	§14	§14
E40	M24	2022 Directions	Direction (ii), p. 2
E41	M25	2022 Directions	Directions (i), (iv), pp. 1–3 (2022 PDF)

Annex B — Bounded evidence-examination protocol

Minimum source record. A candidate evidence item should identify system and deployment boundary, artefact owner, version or configuration, source and collection method, event or record time with time-source basis, relevant period, integrity indicator where available, and known exclusions or access restrictions. These are suggested fields for examination, not CERT-In-prescribed record formats.

Sampling. Define the population before selecting a sample: affected systems, versions, tenants, transactions, time range, paths and excluded strata. Record the selection rationale and known uninstrumented paths. A sample supports only the covered population and stated method; no finite sample alone establishes that unobserved execution paths are safe.

Negative-observation example (illustrative). An examiner reviews gateway decision records for one named production deployment, version and seven-day period, using a declared query for unauthorised tool-commit outcomes and recording the query, index scope, row count, access permissions, retention coverage and clock-quality limitations. A zero-result search supports “checked-and-not-observed within the named gateway dataset”; it does not prove that no bypass occurred via an unlogged path or downstream service.

Integrity and custody. Where relevant, compare collection-time digests to later copies, verify export scope, retain transformation logs and capture chain-of-custody transitions. Assess whether clocks were synchronised closely enough to correlate events. A hash can detect changes to a copy but does not independently establish completeness, source truth or event attribution.

Annex C — Applicability questions for qualified counsel

This public practitioner triage lists questions to refer to qualified legal counsel; it is not an internal approval form or a legal opinion. Record the relevant legal entities and their locations, AI systems and service recipients, whether the AI function is production or public-facing or confined to qualifying research, applicable sector instruments, incident classification and notice times, engagement terms and source versions. Identify unanswered questions and the responsible authority. Only qualified counsel can determine legal applicability or an exemption in a particular case.

Annex D — Corrections and document control

Edition 1 · Version 1.0 — 24 November 2026. This is the initial public-edition record. Subsequent substantive corrections will be issued through a documented revision, replacement or withdrawal decision, with a record of the affected edition and the date of that action.

Readers may submit source-citation errors and corrections via the ODA3 Institute website (https://oda3.org/). Receipt of feedback does not imply a finding is accepted; corrections are subject to internal verification and controlled issue handling.

Notably Absent — boundaries of this research

No real AI system, production evidence basis, incident or assessment client was examined.

No claim that every AI malfunction, hallucination, prompt-injection attempt, drift signal or audit finding is a six-hour reportable security incident.

No conclusion that all guidance statements independently establish statutory offences, uniform effective dates or universal audit procurement timelines.

No control-effectiveness finding, approved risk acceptance, functional safety determination, insurer acceptance, certificate or regulatory endorsement.

No assumption that supplier inaccessibility proves a control failed—or that an unrebutted supplier representation proves a control worked.

No claim that the nine illustrative domain overviews or the 41 selected questions constitute full, line-level coverage of the 107-page guideline. The examination selection does not establish full coverage of every source control area, subcondition or applicability exception.

No grant of CERT-In audit authority, ODA3 certification authority, or permission to reproduce controlled proprietary assessment methods.

Publication and commercial boundaries

This guide is an independent research resource. It is not a CERT-In-empanelled audit, an official CERT-In report form, a statutory audit, an auditor credential, a certification scheme, a regulatory equivalence map or individualised legal guidance. Candidate records do not guarantee acceptance by a regulator, auditor, insurer or court. This guide does not authorise an ODA3-branded assessment or certification; the larger internal research workbook is not included in this edition.

References

CERT-In. Artificial Intelligence (AI) Security Auditing Guidelines, version 1.0, 4 November 2026. https://www.cert-in.org.in/PDF/AI_Security_Auditing_Guidelines.pdf

CERT-In. Directions under Section 70B(6), 28 April 2022, including Annexure I. https://www.cert-in.org.in/PDF/CERT-In_Directions_70B_28.04.2022.pdf

CERT-In. FAQs on Cyber Security Directions, May 2022. https://www.cert-in.org.in/PDF/FAQs_on_CyberSecurityDirections_May2022.pdf

ODA3 Institute. How to Assess an AI System (COM-020), Public v1.1, §71. https://docs.oda3.org/documents/com-020/

ODA3 Institute. COM-019 — framework methodology. Methodology referenced for context; access and status are governed by ODA3 Institute.

This research is not an official CERT-In interpretation, statutory audit, specific-system assessment or certification. © 2026 ODA3 Pvt Ltd. All rights reserved. For permission to reproduce or commercially reuse original ODA3 Institute content, contact ODA3 Institute.