SRIDA IT PROFESSIONAL REFERENCE PUBLICATION

ISO/IEC 42001:2023

Artificial Intelligence Management System (AIMS)

Requirements, Implementation Guidance & Audit Testing Procedures

DOCUMENT ID	VERSION	STATUS
SRIDA-AIMS-42001-RTP-001	1.0	Final — Approved for Publication

Rajendra Bodda
Chief Information Security Officer (CISO)
Srida IT · 2026

Independent professional guidance.
Not an official ISO or IEC publication.

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT



ISO/IEC 42001:2023 — Artificial Intelligence Management System (AIMS)

Requirements, Implementation Guidance & Audit Testing Procedures

SAMPLE

Prepared by

Rajendra Bodda

Chief Information Security Officer (CISO)

Published by Srida IT
SRIDA-AIMS-42001-RTP-001 · Version 1.0 · October 2026

This publication is an independently developed professional guidance document prepared by Srida IT. It is not an official ISO or IEC publication, does not replace the licensed ISO/IEC 42001:2023 standard, and does not confer certification or accreditation.

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

Document Information and Control

Field	Value
Document title	ISO/IEC 42001:2023 — Artificial Intelligence Management System (AIMS)
Subtitle	Requirements, Implementation Guidance & Audit Testing Procedures
Document identifier	SRIDA-AIMS-42001-RTP-001
Version	1.0
Status	Final — Approved for Publication
Classification	Srida IT Professional Reference Publication
Author	Rajendra Bodda, Chief Information Security Officer (CISO)
Publisher	Srida IT
Publication year	2026
Issue date	October 2026
Reference standard	ISO/IEC 42001:2023 (Clauses 4–10; normative Annexes A and B; informative Annexes C and D)
Companion publication	SRIDA-ISMS-27001-RTP-001, ISO/IEC 27001:2022 Requirements, Implementation Guidance & Audit Testing Procedures
Next review	Annually, or on revision of the referenced standards

Version history

Version	Date	Author	Description of change
1.0	October 2026	Rajendra Bodda	First published edition, approved for publication.

Approval

Role	Name	Date
Approved for publication	Rajendra Bodda, Chief Information Security Officer (CISO)	25 October 2026

Distribution

Released for publication by Srida IT. Distribution and use are subject to the permitted-use terms in the Copyright, Legal and Usage Notices.

SAMPLE

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

Table of Contents

Part I Foundations 15

Chapter 1 Introduction to ISO/IEC 42001:2023 16
Chapter 2 Structure and Certification Overview 20
Chapter 3 AI Management System Principles 25
Chapter 4 ISO/IEC 42001 and Related Standards and Regulations 30
Chapter 5 Terminology, Definitions and Abbreviations 34
Chapter 6 How to Use This Manual 39
Chapter 7 Audit Methodology, Evidence and Sampling for AI Systems 45

Part II AIMS Requirements (Clauses 4–10) 51

Chapter 8 Clause 4 — Context of the organization 52
Clause 4.1 Understanding the organization and its context 53
Clause 4.2 Understanding the needs and expectations of interested parties 58
Clause 4.3 Determining the scope of the AI management system 62
Clause 4.4 AI management system 66
Chapter 9 Clause 5 — Leadership 70
Clause 5.1 Leadership and commitment 71
Clause 5.2 AI policy 75
Clause 5.3 Roles, responsibilities and authorities 79
Chapter 10 Clause 6 — Planning 83
Clause 6.1.1 Actions to address risks and opportunities: General 84
Clause 6.1.2 AI risk assessment 88
Clause 6.1.3 AI risk treatment 93
Clause 6.1.4 AI system impact assessment 98
Clause 6.2 AI objectives and planning to achieve them 103
Clause 6.3 Planning of changes 107
Chapter 11 Clause 7 — Support 112
Clause 7.1 Resources 113
Clause 7.2 Competence 117
Clause 7.3 Awareness 121
Clause 7.4 Communication 125
Clause 7.5 Documented information (7.5.1 General, 7.5.2 Creating and updating documented information, 7.5.3 Control of documented information) 129
Chapter 12 Clause 8 — Operation 134
Clause 8.1 Operational planning and control 135
Clause 8.2 AI risk assessment 140
Clause 8.3 AI risk treatment 144
Clause 8.4 AI system impact assessment 148
Chapter 13 Clause 9 — Performance evaluation 153
Clause 9.1 Monitoring, measurement, analysis and evaluation 154
Clause 9.2 Internal audit (9.2.1 General, 9.2.2 Internal audit programme) 158
Clause 9.3 Management review (9.3.1 General, 9.3.2 Management review inputs, 9.3.3 Management review results) 162
Chapter 14 Clause 10 — Improvement 166
Clause 10.1 Continual improvement 167
Clause 10.2 Nonconformity and corrective action 171

SAMPLE

© 2026 Srida IT · Version 1.0 · Independent guidance, not an official ISO/IEC publication

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

Part III Annex A Reference Controls 175

Chapter 15 **A.2 Policies related to AI 176**
A.2.2 AI policy 177
A.2.3 Alignment with other organizational policies 183
A.2.4 Review of the AI policy 188

Chapter 16 **A.3 Internal organization 193**
A.3.2 AI roles and responsibilities 194
A.3.3 Reporting of concerns 199

Chapter 17 **A.4 Resources for AI systems 204**
A.4.2 Resource documentation 205
A.4.3 Data resources 210
A.4.4 Tooling resources 215
A.4.5 System and computing resources 220
A.4.6 Human resources 225

Chapter 18 **A.5 Assessing impacts of AI systems 230**
A.5.2 AI system impact assessment process 231
A.5.3 Documentation of AI system impact assessments 237
A.5.4 Assessing AI system impact on individuals or groups of individuals 242
A.5.5 Assessing societal impacts of AI systems 247

Chapter 19 **A.6 AI system life cycle 252**
A.6.1.2 Objectives for responsible development of AI system 253
A.6.1.3 Processes for responsible AI system design and development 258
A.6.2.2 AI system requirements and specification 263
A.6.2.3 Documentation of AI system design and development 268
A.6.2.4 AI system verification and validation 273
A.6.2.5 AI system deployment 280
A.6.2.6 AI system operation and monitoring 286
A.6.2.7 AI system technical documentation 291
A.6.2.8 AI system recording of event logs 296

Chapter 20 **A.7 Data for AI systems 301**
A.7.2 Data for development and enhancement of AI system 302
A.7.3 Acquisition of data 307
A.7.4 Quality of data for AI systems 312
A.7.5 Data provenance 317
A.7.6 Data preparation 322

Chapter 21 **A.8 Information for interested parties of AI systems 327**
A.8.2 System documentation and information for users 328
A.8.3 External reporting 334
A.8.4 Communication of incidents 339
A.8.5 Information for interested parties 344

Chapter 22 **A.9 Use of AI systems 349**
A.9.2 Processes for responsible use of AI systems 350
A.9.3 Objectives for responsible use of AI system 355
A.9.4 Intended use of the AI system 360

Chapter 23 **A.10 Third-party and customer relationships 365**
A.10.2 Allocating responsibilities 366
A.10.3 Suppliers 371
A.10.4 Customers 376

Part IV AIMS Implementation Framework 381

SAMPLE

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

CHAPTER 1

Introduction to ISO/IEC 42001:2023

Artificial intelligence has moved from research teams into the operating core of organisations. Credit decisions, clinical triage, fraud screening, recruitment shortlists, customer-service chat, code generation and document summarisation are now routinely performed or shaped by AI systems. Many of these systems are not built in-house: they are embedded in software-as-a-service products, accessed through the application programming interfaces of foundation-model providers, or assembled by integrators from open-source models and commercial data sets. The result is that most organisations already depend on AI, whether or not they have decided to govern it.

ISO/IEC 42001:2023, *Information technology — Artificial intelligence — Management system*, is the first international management system standard written specifically for AI. It specifies requirements for establishing, implementing, maintaining and continually improving an artificial intelligence management system (AIMS) within the context of an organisation. This chapter explains why such a management system is needed, who the standard applies to, how it relates to regulation, and how this manual is intended to be used.

1.1 Why an AI management system

Traditional information technology controls assume that a system behaves in the way its code specifies. AI systems, and machine-learning systems in particular, break that assumption in several ways. Their behaviour is learned from data rather than written explicitly, so the quality, representativeness and lawful origin of that data become control objects in their own right. Their outputs are probabilistic, so a system can be correct on average and still produce harmful results for particular individuals or groups. Their performance can degrade silently as the real world drifts away from the conditions represented in training data. And their reasoning can be difficult to explain, even to the people who built them.

SAMPLE

These characteristics create risks that conventional information security, quality or privacy management systems were not designed to address in full:

- **Harm to people outside the organisation.** An AI system can affect the rights, opportunities, safety or wellbeing of applicants, patients, customers, employees and members of the public who never interact with the organisation directly. These consequences are not always visible in an enterprise risk register that is framed around impact on the organisation.
- **Unfair or discriminatory outcomes.** Historical data can encode patterns of disadvantage that a model then reproduces at scale.
- **Opacity and loss of accountability.** Where no one can explain why a system produced a result, it becomes difficult to contest decisions, investigate incidents or assign responsibility.
- **Unreliable or unsafe behaviour.** Models can fail outside the conditions they were tested in, can be manipulated by adversarial inputs, and, in the case of generative AI, can produce plausible but false content.
- **Supply-chain dependency.** Organisations increasingly rely on third-party foundation models whose training data, evaluation results and change schedules they cannot directly inspect.
- **Regulatory exposure.** Legislators and regulators in many jurisdictions now expect organisations to demonstrate that AI is governed, documented, monitored and subject to human oversight.

A management system addresses these issues in a structured, repeatable way. Rather than relying on the good intentions of individual data scientists or the occasional ethics review, an AIMS establishes policy, accountability, risk and impact assessment, life cycle controls, monitoring, internal audit and management review, and links them in a cycle of continual improvement. It turns "responsible AI" from a set of principles into a set of auditable processes.

> **Management system, not product certification.**
>
> ISO/IEC 42001 certifies an organisation's management system for AI, not an individual model or product. A certificate demonstrates that the organisation has defined and operates processes to govern its AI systems within the certified scope. It does not, by itself, attest that any particular AI system is accurate, fair, safe or legally compliant.

© 2026 Srida IT · Version 1.0 · Independent guidance, not an official ISO/IEC publication

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

1.2 Who the standard applies to

ISO/IEC 42001 is intended to be applicable to any organisation, of any size, type or nature, that provides, develops or uses AI-based products or services. It does not limit itself to technology companies or to organisations that train their own models. In practice, the organisations that implement an AIMS fall into several broad patterns.

Table 1-1. Typical organisational profiles for an AIMS

Profile	Generic example	Typical AI role(s)	Where the AIMS effort concentrates
AI product developer	A software company that trains and sells a document-classification model	AI producer, AI provider	Life cycle controls (A.6), data quality and provenance (A.7), technical documentation and information for customers (A.8, A.10)
AI platform or service provider	A cloud provider offering hosted model inference and fine-tuning	AI provider, AI partner	Computing resources (A.4), event logging, responsibilities allocated between provider and customers (A.10)
Integrator	A SaaS company embedding a third-party large language model (LLM) in its customer-support product	AI provider, AI customer of the upstream model supplier	Supplier management, intended use, impact assessment, transparency to end users
Deploying organisation	A 900-employee digital lender using a credit-scoring model built by a vendor	AI customer (user)	Responsible use (A.9), impact assessment (A.5), human oversight, monitoring and supplier obligations
Regulated operator	A hospital group deploying a radiology triage model	AI customer, sometimes AI partner in validation	Clinical safety, impact on individuals, competence of human reviewers, incident communication
Public body	A municipal authority using a model to prioritise inspections	AI customer, sometimes AI producer	Societal impact, fairness, transparency to the public, accountability

An organisation is rarely in only one role. A bank may develop internal fraud models (producer), procure a vendor scoring model (customer) and expose a chatbot to customers (provider). One of the AI-specific additions in ISO/IEC 42001, compared with other management system standards, is that the organisation is expected to determine its role or roles with respect to the AI systems in scope, because those roles shape which risks and controls are relevant. Chapter 3 explains the role vocabulary drawn from ISO/IEC 22989.

1.3 The first edition and its place in the AI standards family

ISO/IEC 42001:2023 was published in December 2023 as a first edition. It was developed by ISO/IEC JTC 1/SC 42, the joint technical subcommittee responsible for AI standardisation, which also maintains the AI concepts and terminology standard (ISO/IEC 22989), the AI risk management guidance (ISO/IEC 23894) and a growing family of related documents. ISO/IEC 22989 is cited as a normative reference, which means its terminology is indispensable for applying ISO/IEC 42001.

The standard follows the ISO Harmonized Structure used by ISO/IEC 27001, ISO 9001, ISO 14001 and other management system standards. Clauses 4 to 10 set out the requirements of the management system. Four annexes follow: Annex A (normative) lists reference controls; Annex B (normative) provides implementation guidance for those controls; Annex C (informative) describes potential AI-related organisational objectives and risk sources; and Annex D (informative) discusses use of the management system across domains and sectors. Chapter 2 describes this structure in detail.

At the time of writing, Srida IT is not aware of an amendment to ISO/IEC 42001 equivalent to the 2024 climate action amendments applied to ISO/IEC 27001 and other Harmonized Structure standards; readers should check the current status before relying on this.

1.4 Relationship to regulation

ISO/IEC 42001 is a voluntary international standard. It does not create legal obligations and conformity with it does not, by itself, demonstrate compliance with any law. It is nonetheless highly relevant to regulatory compliance, for three reasons.

1. **It provides the organisational machinery that regulation assumes.** Most AI-related regulation expects organisations to identify their AI systems, assess risks and impacts, maintain technical documentation, keep logs, ensure human oversight,

© 2026 Srida IT · Version 1.0 · Independent guidance, not an official ISO/IEC publication

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

CONTROL A.6.2.4

AI system verification and validation

Annex A reference control · applicability per AI risk treatment & SoA | WP-A6.2.4

1 Control Identification

Attribute	Detail
Control identifier	A.6.2.4
Control title	AI system verification and validation
Control objective area	A.6 AI system life cycle › A.6.2 AI system life cycle
Control purpose	To provide objective evidence, before release and whenever conditions change, that an AI system has been built to its specification (verification) and is fit for its intended purpose in its real context of use (validation), including its performance, robustness, fairness, safety and security.
Annex B implementation guidance	ISO/IEC 42001:2023, Annex B, B.6.2.4
Annex C objectives†	Robustness, Fairness, Safety, Security, Transparency and explainability, Availability and quality of training and test data, Accountability, Privacy
Annex C risk sources†	Risk sources related to machine learning, Complexity of environment, Level of automation, Technology readiness, System life cycle issues, Lack of transparency and explainability
AI life-cycle stages	Design and development, Verification and validation, Deployment, Continuous validation, Re-evaluation
Relevant AI roles	AI provider, AI producer, AI customer, AI partner
Relevant risks	• Models are released on the strength of optimistic metrics computed on leaked, unrepresentative or reused test data. • Systems that perform well on average fail for specific subgroups, edge cases or operating conditions. • Generative systems produce unsafe, false, biased or confidential outputs, or can be manipulated through prompt injection or jailbreaks, because adversarial testing was not performed. • Models degrade after deployment through data or concept drift, and no re-validation is triggered. • Validation is performed by the same people who built the model, without independent challenge, so blind spots persist. • Acceptance criteria are set or relaxed after results are known, undermining the integrity of release decisions.

† Mapping to ISO/IEC 42001:2023 Annex C (informative) as determined by Srida IT; verify against the licensed standard.

CONTROL UNDERSTANDING

2 Control Objective and Intended Outcome

The organisation defines, before testing begins, a verification and validation plan for each AI system with evaluation metrics and acceptance criteria traceable to requirements; evaluates the system on properly constructed, independent test data and in conditions representative of deployment; covers performance, subgroup fairness, robustness, safety, security and, for generative systems, adversarial red teaming; ensures independence of validation proportionate to risk; records results in an evaluation report used for an authorised release decision; and re-validates on defined triggers throughout operation.

3 Control Description

Verification asks whether the system was built right: that code, data pipelines, features, model behaviour and integrations meet the specification. Validation asks whether the right system was built: that it achieves its intended purpose for its users

© 2026 Srida IT · Version 1.0 · Independent guidance, not an official ISO/IEC publication

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

and AI subjects in the real context. For AI, both depend heavily on data and on statistical evidence rather than deterministic checks, so the plan must be explicit about what is measured, on what data and against what threshold.

Srida IT recommends that every V&V plan address at least the following elements. **Metrics and acceptance criteria**: primary and secondary metrics chosen to match the consequence of errors (for example recall for a safety-critical detector, calibration for a probability used in pricing, precision at a fixed review capacity), with thresholds and confidence intervals fixed and approved before the final test run. **Test data integrity**: a holdout test set that is never used for training, tuning or model selection, constructed to prevent leakage (temporal splits for time-dependent data, group-aware splits so the same customer or patient is not in both training and test), representative of the deployment population, versioned and access-controlled; plus out-of-time and, where relevant, out-of-site test sets. **Subgroup and fairness evaluation**: disaggregated metrics for relevant groups and intersections, using fairness measures appropriate to the decision (for example differences in selection rate, false positive rate or false negative rate, or calibration within groups), with thresholds derived from impact assessment findings and documented reasons for metric choice, recognising that fairness definitions can conflict. **Robustness**: stress tests under distribution shift, noisy or missing inputs, boundary values, perturbations and, where relevant, adversarial examples, plus uncertainty and out-of-distribution detection behaviour. **Safety and security**: tests for failure modes identified in the threat model, such as data poisoning resilience, model extraction or membership inference exposure, and insecure output handling.

For **generative AI**, the plan should add task-specific quality evaluation (factual accuracy or groundedness against sources, relevance, refusal appropriateness), harmful content testing across defined categories, privacy leakage tests, and structured **red teaming** by people with adversarial skill and domain knowledge, covering jailbreaks, direct and indirect prompt injection (including via retrieved documents and tool outputs), data exfiltration, misuse for prohibited purposes and manipulation of vulnerable users, with findings triaged, fixed and regression-tested. Automated evaluation using another model as a judge may be used but should itself be validated against human ratings. **Validation in context** may include shadow deployment, pilot or A/B testing with human oversight, and user acceptance testing with representative users. **Independence**: for high-tier systems, validation should be performed or reviewed by a function independent of the development team (a model validation or model risk team, a second-line function, or an external assessor). **Re-validation triggers** should be defined: retraining, data source change, feature change, base model or supplier model version change, prompt or guardrail change, new deployment population or jurisdiction, drift metrics breaching thresholds, serious incidents, and periodic re-validation at a tier-based interval.

SAMPLE

4 Business, Ethical and Security Rationale

Verification and validation is the point at which the organisation's claims about an AI system meet evidence. Weak V&V is a leading cause of AI harm: data leakage inflates metrics, average-case results hide subgroup failures, and generative systems fail in ways that only adversarial testing reveals. Robust V&V supports the AI risk treatment required by Clause 6.1.3 and 8.3, provides evidence that impact mitigations work, and creates defensible records for customers, auditors and regulators. Several regimes expect AI systems to achieve appropriate levels of accuracy, robustness and cybersecurity, and expect testing against predefined metrics before placing on the market, including requirements for high-risk systems under the EU AI Act and model risk management expectations in financial services. The NIST AI RMF 1.0 Measure function similarly emphasises test, evaluation, verification and validation. Independent validation reduces conflicts of interest and is an established practice in model risk management. Re-validation recognises that AI performance is not static once deployed.

5 Risk and Harm Scenarios Addressed

- A credit model is evaluated on a random split that includes multiple applications from the same customer in training and test, overstating performance; in production approval accuracy is materially lower.
- A radiology triage model validated at a teaching hospital performs poorly on images from older portable machines at district sites, delaying urgent cases.
- A customer service assistant built on a third-party LLM follows instructions hidden in an uploaded document and discloses another customer's data.
- A hiring model meets its overall accuracy target but has a much higher false negative rate for applicants with disabilities, which was never measured.
- A supplier silently updates its foundation model version and the organisation's assistant begins producing confident hallucinations, with no re-validation triggered.
- The team lowers the acceptance threshold after the final test run to meet a launch date, with no recorded approval.

© 2026 Srida IT · Version 1.0 · Independent guidance, not an official ISO/IEC publication

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

IMPLEMENTATION

6 Applicability and Scoping

Applicable to AI providers and producers for every AI system they develop or modify, and to AI customers for acceptance testing and validation in their own context of systems they procure or configure, since supplier evaluation cannot substitute for validation on the customer's population and use. AI partners performing evaluation services should operate to the same standard. Exclusion from the Statement of Applicability is very difficult to justify for any organisation that develops, configures or deploys AI systems in scope. Depth, independence and red-teaming effort may scale with tier.

7 Implementation Guidance

1. Publish an AI verification and validation standard defining required test categories by tier (functional verification, performance, subgroup fairness, robustness, security, generative quality and safety, red teaming, validation in context), independence requirements, evaluation report content and re-validation triggers and intervals.
2. Require a V&V plan for each system, approved before the final evaluation, that fixes metrics, acceptance thresholds with confidence intervals, test datasets and versions, subgroups, robustness scenarios and the release decision rule, each traceable to A.6.2.2 requirements.
3. Establish test data management: holdout sets constructed with leakage controls (temporal, group-aware splitting, deduplication), checked for representativeness against the deployment population, versioned, access-restricted from the development team where independence is required, and refreshed under control; record their provenance per A.7.5.
4. Implement disaggregated evaluation and fairness testing for person-affecting systems, selecting fairness metrics appropriate to the decision and documenting why; where protected attributes are unavailable, document proxy or synthetic methods and their limits.
5. Build robustness and security test suites: distribution-shift and stress scenarios, perturbation and adversarial tests, out-of-distribution detection, and tests derived from the AI threat model; integrate automated suites into CI/CD so that every candidate model or configuration change is evaluated.
6. For generative systems, run automated evaluation benchmarks (groundedness, harmful content categories, refusal behaviour, privacy leakage) and structured red-team exercises before release and after material change; track findings to remediation and add them to a regression suite; validate any model-as-judge evaluator against human ratings.
7. Assign independent validation for high-tier systems to a model validation or second-line function, or an external party, with authority to block release; for lower tiers require at least peer review by someone outside the immediate build team.
8. Validate in context through shadow mode, staged rollout or pilot with enhanced human oversight, and user acceptance testing with representative users, comparing outcomes with the non-AI baseline.
9. Record results in an evaluation report, attach it to the model registry entry, and require an authorised release decision that explicitly addresses any unmet criteria; prohibit post-hoc threshold changes without documented approval.
0. Operate re-validation: monitor drift and performance indicators in production, define triggers that automatically open re-validation, and perform periodic re-validation at tier-based intervals, recording outcomes and actions.

8 Roles and Responsibilities

Role	Responsibility
AI system owner	Accountable for an approved V&V plan, adequate evaluation and the release decision request.
Data scientist / ML engineer	Builds test suites and runs verification and evaluation; maintains reproducible evaluation notebooks and pipelines.
Independent model validation / model risk function	Challenges design, data and results, re-performs key tests and issues a validation opinion for high-tier systems.
Red team (internal or external)	Performs adversarial testing of generative and high-risk systems and reports findings by severity.
Model review board	Approves release based on the evaluation report and validation opinion; approves any accepted shortfalls.

9 Recommended Documented Information

- AI verification and validation standard
- V&V plans with pre-registered metrics and acceptance criteria
- Test dataset records (datasheets, version, split method, representativeness analysis)

SAMPLE

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

- Evaluation reports including disaggregated and robustness results
- Fairness and bias test results
- Red-team plans, reports and remediation tracker
- Independent validation reports and opinions
- Release approval records and accepted-shortfall decisions
- Re-validation trigger definitions, drift dashboards and re-validation records

ASSESSMENT AND TESTING

10 Audit Testing Procedures

Recommended Srida IT testing methodology; not prescribed by ISO/IEC 42001.

Table 19-6. Audit testing procedures, Control A.6.2.4

Test ID	Method	Testing procedure	Result
TP-A.6.2.4-01	EXAMINE	Examine the AI verification and validation standard and verify that it defines required test categories by tier, test data integrity rules, fairness and robustness expectations, red-teaming requirements for generative systems, independence requirements, evaluation report content and re-validation triggers and intervals.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-02	SAMPLE	Select a sample of AI system releases and verify that each had a V&V plan approved before the final evaluation run (compare approval timestamps with evaluation run timestamps), with metrics and thresholds traceable to the requirements specification.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-03	INSPECT	For each sampled system, inspect the test dataset records and verify that the holdout set was excluded from training and tuning, that the split method addresses leakage (temporal or group-aware as appropriate), and that representativeness against the deployment population was assessed.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-04	RE-PERFORM	For at least one high-tier sampled model, re-run the evaluation notebook or pipeline on the registered holdout set using the registered model artefact, and verify that primary metrics and at least two subgroup metrics reproduce the figures in the evaluation report within tolerance; check for overlap between training and test record identifiers.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-05	EXAMINE	Examine evaluation reports for sampled systems and verify that disaggregated fairness, robustness and security results are reported against predefined thresholds, that any unmet criterion is explicitly addressed, and that the release decision was made by an authorised role.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-06	EXAMINE	For sampled generative systems, examine red-team plans, reports and remediation trackers and verify that testing covered jailbreaks, direct and indirect prompt injection, data leakage and harmful content categories, that findings were triaged by severity, and that high-severity findings were fixed and regression-tested before release or formally accepted.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-07	EXAMINE	For sampled high-tier systems, examine the independent validation report and verify that the validator is organisationally independent of the development team, had access to data and code, re-performed key tests and issued an opinion before release.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-08	SAMPLE	Select a sample of re-validation trigger events in the period (re-training, supplier model version change, prompt or guardrail change, drift alert breach, serious incident) from registry, change and monitoring records, and verify that each resulted in re-validation completed before or within the defined time of the change.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A
TP-A.6.2.4-09	INTERVIEW	Interview the data scientist, the independent validator and a model review board member for one sampled system and verify consistent understanding of the acceptance criteria, the known limitations and the conditions that would trigger re-validation.	☐ In place ☐ In place w/ obs. ☐ Not in place ☐ N/A

WRITE

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT

11 Sampling Guidance

Population: all production releases of AI systems in the audit period (new and changed), all independent validation reports, all red-team exercises, and all re-validation trigger events (identified from model registry, change management, supplier notifications, drift alerts and incident records). Release sample: 5 to 10 releases stratified by tier, with all high-tier releases if five or fewer; include at least one generative system and one procured or supplier-model-based system. Re-performance: at least one high-tier model, chosen where evaluation data are accessible in a controlled environment. Re-validation triggers: 5 to 8 events weighted to supplier model updates and drift breaches, which are the most commonly missed. Period: the audit period; for periodic re-validation, the most recent cycle.

12 Interview Questions

Interviewee	Question
Data scientist	How did you make sure no information from the test set leaked into training or model selection, and when were the acceptance thresholds fixed?
Independent model validator	What did you re-perform yourself, what did you find, and have you ever blocked a release?
Red team lead	Which attack categories did you test for the assistant, what were the most serious findings, and how did you confirm they were fixed?
AI system owner	What would cause your model to be re-validated before its next scheduled review, and how are you notified?
Model review board member	Describe a release where a criterion was not met. What was decided and who approved it?

13 Technical Verification

- Re-run the evaluation notebook on the registered holdout set and compare metrics with the evaluation report.
- Join training and test datasets on record or entity identifiers to detect overlap and leakage.
- Compare V&V plan approval timestamps with evaluation run timestamps in the experiment tracker to confirm criteria were pre-registered.
- Inspect model registry approvals and confirm each production version has an attached evaluation report and, for high-tier, validation opinion.
- Query drift dashboards and alert history and reconcile threshold breaches with re-validation records.
- Replay a sample of red-team prompts from the regression suite against the production generative system and confirm mitigations hold.

SAMPLE

14 Expected Audit Evidence

- Approved V&V standard and system-level V&V plans
- Test dataset datasheets and split documentation
- Evaluation reports with disaggregated, robustness and security results
- Red-team reports and remediation trackers
- Independent validation reports
- Model registry release approvals
- Drift monitoring and re-validation records
- Auditor re-performance worksheet

Free sample pages · The complete manual is available at sridait.com/manuals/rtp-iso42001 · © 2026 Srida IT