SAI — SECURE AI

WHITEPAPER

AI Security & ISO 42001

A Field Guide

What the international standard for AI management asks of you, where AI systems most often fall short of it, and how to get to a certificate — a practical guide for enterprise and small-business leaders.

Author

Keith Patterson — Founder, SAI

Published

May 2026 · Edition 1

Contents

Executive summary

ISO/IEC 42001 is to artificial intelligence what ISO/IEC 27001 is to information security: the international, certifiable standard for managing it well. Published in December 2023, it is the first global standard for an AI management system — and it has moved from publication to serious enterprise adoption faster than almost anyone predicted. By early 2026, organizations including IBM, Anthropic, and Microsoft held certification, and the question “are you ISO 42001 certified?” had begun to appear in enterprise procurement and vendor security reviews.

For a private-sector company that builds or buys AI, ISO 42001 is becoming the cleanest available way to prove — to a customer, an auditor, an investor, or a regulator — that its AI is governed, not merely deployed. A self-declared “we take AI seriously” no longer carries weight. A certificate from an accredited body does.

This guide explains, in plain terms, what ISO 42001 is, what it requires, where AI systems most often fall short of it, and how an organization gets from where it is today to a certificate. It is written for the people who now carry AI governance — security and engineering leaders, and the founders doing both jobs at once.

> **What ISO 42001 is — and is not**
>
> ISO 42001 does not certify that a model is “safe” or accurate. It certifies that an organization runs a disciplined system for managing its AI — the process, not the model. It does not certify your car is fast; it certifies you run a proper garage. That distinction shapes everything in this guide.

1. Why ISO 42001, and why now

Most organizations adopted AI faster than they built the means to govern it. Models went into products, agents were given tools, staff started using AI to make or shape real decisions — and the policies, controls, and evidence that should have accompanied that lagged behind, often by a wide margin. For a while, that gap was invisible. It is no longer.

1.1 The people who used to take AI on trust have stopped

Enterprise customers now send vendor questionnaires that ask, specifically, how your AI is governed. Auditors expect AI systems to appear in the scope of a SOC 2 or ISO 27001 report. Investors performing technical due diligence ask to see AI risk documentation. And regulation is arriving — the EU AI Act's obligations for high-risk systems phase in through 2026. In every one of those conversations, the same thing is being asked for: credible, independent evidence that your AI is under control.

ISO 42001 is the recognized way to provide it. Because it is certifiable by an accredited body, it converts “trust us” into “here is our certificate” — the same move that SOC 2 made for SaaS security a decade ago. The trajectory is familiar: a few years ago, “are you SOC 2?” went from a sophisticated question to table stakes. ISO 42001 is on that same path for AI, and moving along it quickly.

1.2 The momentum is real

This is not a standard waiting for a market. By early 2026, IBM had certified its Granite models, Anthropic its Claude models, and Microsoft its 365 Copilot, alongside professional-services firms and infrastructure operators. Two forces are compounding that adoption: the EU AI Act's high-risk obligations approaching through 2026, and active work to adapt ISO 42001 into a European Norm — which will tie the standard still more tightly to the regulatory picture. The organizations certifying now are not early adopters chasing a badge. They are removing a question from every future sales and audit conversation.

2. What ISO 42001 is

ISO/IEC 42001 is the first international standard for an AI management system — an AIMS. The word management is doing real work in that phrase. The standard does not test a model or score an algorithm. It defines a system: a structured, repeatable way for an organization to set AI policy, assign responsibility, assess risk and impact, apply controls, audit itself, and improve over time.

2.1 A management-system standard — a familiar shape

ISO 42001 belongs to the same family as ISO 27001 (information security) and ISO 9001 (quality), and it shares their structure deliberately. Anyone who has been through an ISO 27001 certification will recognize the shape immediately: numbered clauses for the management system itself, an annex of selectable controls, a cycle of Plan-Do-Check-Act, and an emphasis on evidence over intention. An organization that already holds ISO 27001 is not starting from zero — the leadership, documentation, internal-audit, and improvement machinery can be extended to cover AI rather than rebuilt.

2.2 Certifiable — and that is the point

ISO 42001 can be certified. An accredited certification body audits your management system and, if it conforms, issues a certificate, maintained through periodic surveillance audits. This is what separates ISO 42001 from a voluntary framework: it produces an independent, recognized attestation a third party will accept. A framework helps you do the work; a certifiable standard lets you prove you did it.

2.3 Risk-based and proportionate

The standard is risk-based, which has a practical and welcome consequence: it scales. A ten-person company and a multinational bank are both held to the same standard, but the AI management system each must operate is sized to its actual AI risk. A small business does not need a bank's governance apparatus — it needs a system proportionate to what its AI actually does. ISO 42001 is explicitly built to allow that.

> **If you already have ISO 27001**
>
> If your organization already holds ISO 27001, treat ISO 42001 as an extension, not a second project. The hardest parts of a management system — leadership commitment, a documentation discipline, an internal-audit habit — already exist. The new work is AI-specific: the risk and impact assessments, the AI controls, and the scope.

3. What the standard requires

ISO 42001 has two halves. The numbered clauses define the management system itself. Annex A provides a catalogue of controls to apply within it. Both matter, and they work together.

3.1 The management system — clauses 4 to 10

The clauses follow the standard management-system structure. In plain terms, they ask the following:

Clause	Theme	What it asks of you
4	Context	Define the scope — which AI systems are covered — and your role for each: are you developing the AI, providing it, or using it?
5	Leadership	Establish an AI policy, assign clear responsibility for AI, and show genuine leadership commitment to it.
6	Planning	Run an AI risk assessment and an AI system impact assessment; set measurable AI objectives; plan how risks are treated.
7	Support	Provide the resources, competence, awareness, and documented information the system needs to function.
8	Operation	Actually carry out the risk and impact assessments and operate the controls — the system in motion, not on paper.
9	Performance evaluation	Monitor and measure the system; run internal audits; hold a management review.
10	Improvement	Act on findings — corrective action and continual improvement of the management system.

Two requirements inside Clause 6 deserve singling out, because they are where AI governance differs most from security governance. ISO 42001 asks for an AI risk assessment — the risks AI poses to the organization — and, separately, an AI system impact assessment — the effects an AI system has on the individuals and groups it touches. Many organizations do the first instinctively and have never done the second at all.

3.2 Annex A — the controls

Annex A is a catalogue of 38 controls grouped under control objectives. As with ISO 27001, not every control is mandatory: you select those that apply to your AI risk and record your reasoning in a Statement of Applicability. The control areas span the AI lifecycle:

Annex A control area	What it addresses
AI policy	A documented policy for the responsible development and use of AI, reviewed and kept current.
Internal organization	Defined AI roles, responsibilities, and reporting lines.

Annex A control area	What it addresses
Resources for AI systems	Knowing and documenting the data, tooling, compute, and human resources your AI systems depend on.
Assessing impacts of AI systems	A defined process for assessing an AI system's impact on individuals and society.
AI system life cycle	Responsible practices across design, development, verification, deployment, operation, and retirement.
Data for AI systems	Provenance, quality, and appropriate handling of the data used to build and run AI systems.
Information for interested parties	Giving users and affected parties the information they need — transparency obligations.
Use of AI systems	Responsible, intended use of AI systems once they are in operation.
Third-party relationships	Governing the AI you did not build — foundation models, APIs, and vendor-supplied AI.

The Statement of Applicability is not bureaucratic overhead. It is the document that, more than any other, tells an auditor — or a customer's security team — that you have actually thought about your AI risk rather than adopted a template.

4. Where AI systems most often fall short

Across real assessments, organizations miss ISO 42001 conformance in a small set of recurring ways. Knowing them in advance turns a daunting standard into a short list of things to fix.

- **They undercount their AI.** The single most common failure is at the scoping stage. The first list of AI systems is always too short — it misses the shadow AI staff use through public tools, the AI features embedded in third-party SaaS, and the AI components buried inside vendor platforms. The real scope is reliably larger than the first guess.
- **They never assess impact.** Organizations run a security risk assessment by reflex but have never assessed an AI system's impact on the people it affects. ISO 42001 requires both; the impact assessment is the half that is usually missing entirely.
- **Their data governance has gaps.** Training, fine-tuning, and retrieval data with no documented provenance, no classification, and no quality control. The model works; nobody can say what it learned from or whether that data should have been used.
- **They cannot see the AI they did not build.** Foundation models, hosted APIs, and vendor AI features sit outside any governance because they were never treated as part of the system. Annex A's third-party controls exist precisely for this blind spot.
- **Their governance is a document, not a system.** An “AI policy” PDF exists, but nobody operates it — no owner, no review cycle, no evidence it changes any decision. ISO 42001 certifies a working system; a policy that is only a file will not pass.
- **They have no evidence trail.** The system runs, but there is no record of the decisions behind it, no lineage, nothing an auditor can read. ISO 42001 conformance is demonstrated through evidence, and evidence has to be generated as the work is done — it cannot be back-filled the week before an audit.

5. The road to certification

Getting to an ISO 42001 certificate is a staged journey. The stages below are the same whether you are a startup or a large enterprise — only the weight of each one changes.

Stage 1 — Scope & discovery

Find every AI system you actually have. A practical first move is a focused, roughly one-week exercise: list every AI system in production, identify which ones touch customers or regulated decisions, and note where AI arrives indirectly — through staff tools and vendor platforms. Expect the list to grow as you go. This stage sets the scope for everything after it.

Stage 2 — Gap analysis

Map current practice against the clauses and Annex A. The output is two documents: a draft Statement of Applicability — which controls apply and why — and a prioritized gap list, showing what conformance will actually require.

Stage 3 — Build the management system

Stand up what is missing: the AI policy, the roles, the risk-assessment and impact-assessment processes, the selected controls, and the documentation that ties them together. For an organization with ISO 27001 already, this extends an existing system; for one without, it builds a first.

Stage 4 — Operate it

Run the system. A certification body cannot certify a management system with no operating history — it needs to see the system working: assessments performed, controls applied, decisions recorded. This stage is where evidence accumulates, and it cannot be rushed.

Stage 5 — Internal audit & management review

The standard requires both before certification. An internal audit checks the system against the standard; a management review puts the results in front of leadership. Both are conformance requirements in their own right, not optional preparation.

Stage 6 — Certification audit

An accredited certification body runs a two-stage audit — a Stage 1 review of documentation and readiness, then a Stage 2 audit of the system in operation. A successful Stage 2 results in the certificate. Periodic surveillance audits then keep it valid, which is the standard's way of ensuring the system stays alive rather than lapsing the day after the certificate is issued.

> **A note for small businesses**
> Small businesses sometimes assume ISO 42001 is out of reach. It is not. Because the standard is risk-based, a small company with a narrow AI footprint operates a correspondingly light management system. The stages are the same; the effort is proportionate to the AI you actually

> run.

6. ISO 42001 alongside SOC 2, NIST, and the EU AI Act

ISO 42001 does not replace the other things being asked of you. It is the management system that organizes them. Here is how it sits next to the frameworks a private-sector company is most often measured against.

6.1 SOC 2

A SOC 2 report attests to the security, availability, and related controls of a service organization. It is increasingly expected to address AI, but it is not AI-specific and was never designed to be. ISO 42001 is the AI-specific complement: where SOC 2 says the service is well run, ISO 42001 says the AI inside it is well governed. The two share a discipline and are best pursued together, not as rivals.

6.2 The NIST AI Risk Management Framework

The NIST AI RMF is an excellent, widely used framework for structuring how an organization thinks about AI risk. But it is voluntary guidance — there is no NIST AI RMF certificate. ISO 42001 and the RMF are complementary in the most practical way possible: the RMF helps you do the work of AI risk management well; ISO 42001 lets you prove, with a certificate, that you did. Mature programs use both.

6.3 The EU AI Act

The EU AI Act is law, not a framework — risk-tiered, with substantial obligations for high-risk AI systems phasing in through 2026. ISO 42001 is not the Act, and certification is not automatic compliance with it. But a working AI management system produces a great deal of the risk-management, quality-management, and documentation evidence the Act expects, and the standard is being adapted into a European Norm — which will draw the two closer still. For any organization in or selling into the EU, an ISO 42001 system is a substantial head start on the Act rather than a separate effort.

> The simplest way to hold these together: the NIST AI RMF helps you think, the EU AI Act tells you what the law requires, SOC 2 covers the service around the AI — and ISO 42001 is the management system that ties all of it into something you can run, evidence, and certify.

7. An ISO 42001 readiness checklist

A short diagnostic. If you cannot answer these clearly, the gaps they expose are the first work of an ISO 42001 program.

- Can you list every AI system in production — including shadow AI and AI embedded in vendor tools — not just the ones you built deliberately?
- For each AI system, do you know your role under the standard: are you developing it, providing it, or using it?
- Is there an AI policy that is actually operated — with an owner and a review cycle — rather than a published PDF nobody acts on?
- Have you assessed not only the risk AI poses to your business, but the impact each AI system has on the people it affects?
- Can you show the provenance, classification, and quality controls for the data behind each AI system?
- Is the AI you did not build — foundation models, APIs, vendor features — inside your governance, or outside it?
- If a customer's security team asked today for evidence that your AI is governed, could you produce it?
- Has anyone independent ever audited your AI governance — or would a certification audit be the first time?

> Every “no” above is a scoped, finite piece of work — not a crisis. That is the quiet advantage of a standard: it converts a vague unease about AI governance into a defined list a program can close, one item at a time, toward a certificate.

About SAI

SAI builds and secures private AI systems for enterprises and small businesses across the US, Canada, and international markets. The practice is led by Keith Patterson, a security architect with an active Top Secret clearance and thirty years spent on systems where failure was never an option. SAI is unusual in doing both halves of the job — engineering AI products and governing them — because the two were never meant to be separate disciplines.

ISO 42001 readiness is part of SAI's Assurance practice. A first engagement begins with the gap analysis described in Section 5: where your AI governance stands against the standard, and what a path to certification would involve. It can be scoped in a single 30-minute call.

Start a conversation. Book a security review at the SAI website, or write to **KP@saicloudsecure.com**. Every inquiry reaches Keith directly — not a sales desk.

Sources — market and adoption context referenced in this guide:

ISO/IEC 42001:2023, AI management system standard (International Organization for Standardization).

Industry reporting on ISO 42001 enterprise adoption and certification, 2026 (including certified organizations and EU AI Act timing).

CEN-CENELEC JTC 21 work to adapt ISO 42001 into a European Norm (prEN 18286), public enquiry 2025–2026.