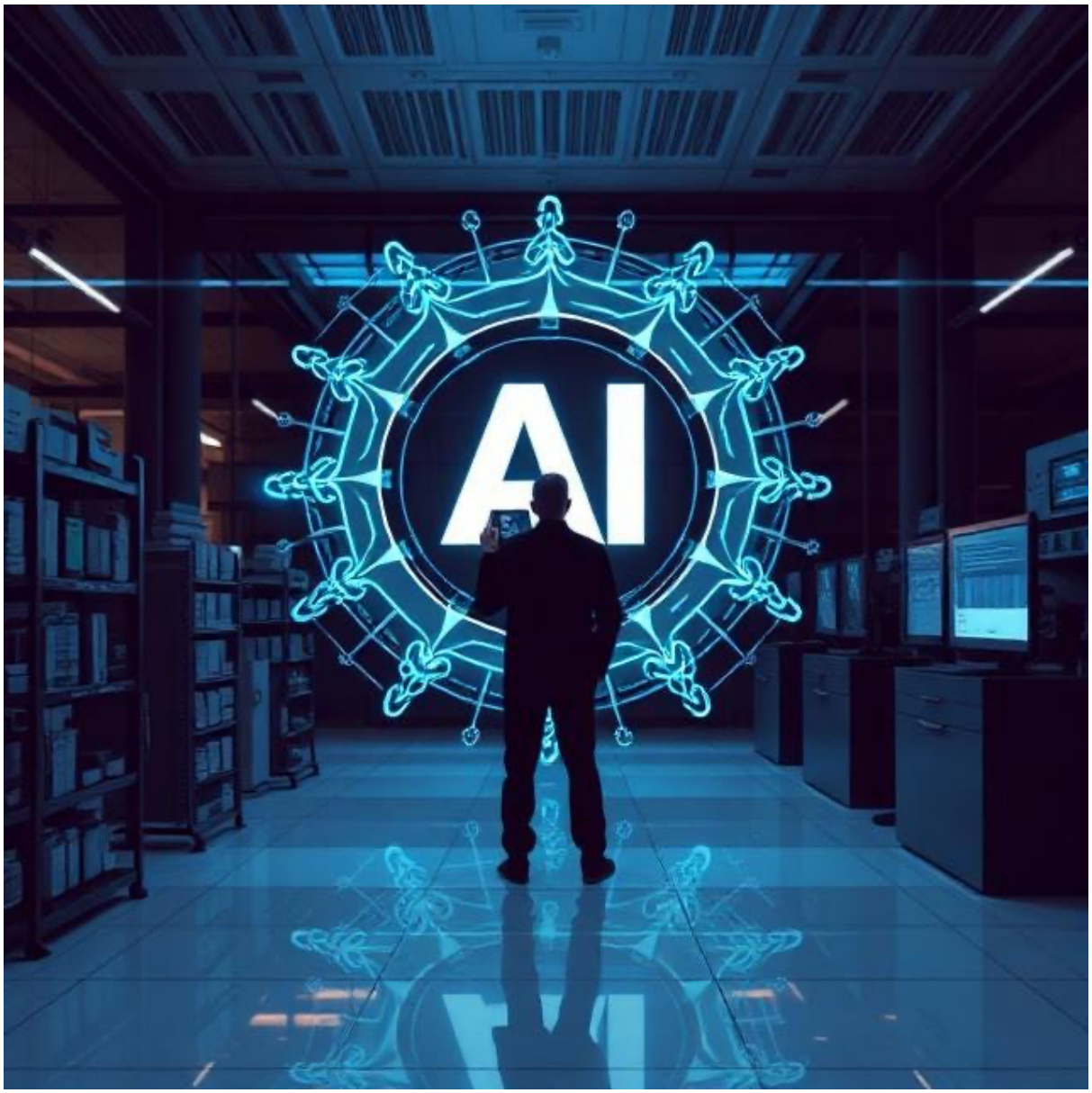
NEURAL SEVEN
ADVISORY

ISO/IEC 42001: the governance layer AI has been missing



The old fear was that AI would become too intelligent. The more immediate risk is that it's already useful enough to be connected to customer data, internal documents and business decisions before anyone has decided who is accountable for it.

That's the gap ISO/IEC 42001 is designed to close. It is not a technical security standard for an AI model, and it won't magically make an organisation compliant, private or immune from poor decisions. It is an international, certifiable management-system standard for establishing, operating and continually improving an Artificial Intelligence Management System (AIMS): the governance structure around how an organisation develops, provides or uses AI.

For Australian organisations, that matters because AI risk is increasingly ordinary business risk: data leakage, unreliable decisions, unapproved suppliers, poorly controlled access, misleading customer outcomes and a lack of evidence when the board asks the obvious question: "How do we know this is under control?"

What ISO 42001 actually means

Think of ISO 42001 as the operating model for responsible AI. It asks an organisation to define its AI objectives and accountabilities, identify and assess risks and impacts, select appropriate controls, monitor outcomes, manage changes, and improve over time. It applies to organisations that build AI products, embed AI in services, or simply use AI tools in normal operations.

That last group is often overlooked.

A mid-sized professional-services firm using Microsoft Copilot, a customer-service platform with generative AI, and a recruitment team using AI to screen candidates may not see themselves as “an AI company”. Yet each is making decisions about data, access, transparency, supplier assurance and human oversight. Those are AI governance decisions whether they are recorded as such or not.

ISO 42001 is trying to protect more than information. It is designed to support trustworthy AI practices around accountability, transparency, privacy, security, fairness, reliability and human oversight. Its practical value is in forcing an organisation to connect those concepts to actual systems, actual owners and actual evidence.

Why this has become urgent

The AI conversation has moved from experimentation to operational dependency faster than most governance models have kept pace. McKinsey’s 2026 research found that 74% of respondents considered inaccuracy a highly relevant AI risk, while 72% identified cybersecurity as highly relevant.

Those results should not surprise any CIO or CISO. An AI system that invents a policy interpretation, discloses information through an overly broad connector, or makes a recommendation nobody can explain is not merely an “AI issue”. It is a risk-management and accountability issue.

Identity has become central to the equation as well. CyberArk reported that 69% of APAC organisations lacked identity-security controls for AI, and 46% said they could not secure shadow-AI use in their organisation. As AI agents increasingly act on behalf of people and systems, the question shifts from “Who has access?” to “What can this identity, service account or agent do, with whose authority, and under what constraints?”

That is a board-level question.

The shadow AI problem

Shadow AI is the use of unapproved AI tools, plugins, browser extensions, agents or services outside an organisation’s agreed security, privacy and procurement processes. It is not necessarily malicious. Usually, it starts with someone trying to finish work faster.

A sales manager pastes a proposal into a public AI tool to improve its tone. A finance analyst uploads a spreadsheet to explain a variance. A developer uses an unapproved coding assistant connected to a source repository. An executive enables an AI meeting tool that records, transcribes and stores commercially sensitive discussions in a separate cloud environment.

Individually, these choices can feel harmless. Collectively, they create an unmanaged data estate.

Australian evidence points to a serious visibility problem. Josys' 2025 survey of 500 Australian technology decision makers found that 70% of organisations had limited or no visibility of the AI tools being used, while 36% of employees reported uploading sensitive company information to AI tools and 24% said they had uploaded customer personally identifiable information.

The global picture is no more reassuring. IBM's 2025 Cost of a Data Breach report found that 63% of organisations lacked governance policies to manage AI or prevent shadow-AI risk, and 97% of AI-related breaches involved missing proper access controls.

ISO 42001 does not solve shadow AI by banning it. Blanket bans rarely survive contact with reality. What it does is create the disciplines needed to manage it: an AI inventory, acceptable-use boundaries, risk assessment, approval paths, supplier assessment, identity controls, monitoring, incident response and a clear owner who is expected to care.

What risk looks like on a Tuesday

The most damaging AI incidents may not look dramatic at first.

Picture a customer-service assistant deployed to reduce call-centre demand. It is connected to a knowledge base and a customer-record system. A configuration change gives it access to notes intended for internal staff, and a prompt sequence causes it to expose account information that should never have appeared in a customer

conversation. The failure is not simply “the model said something wrong”. It is a failure of data classification, access design, testing, change control, monitoring and escalation.

Or consider a regional business using a generative AI platform to analyse customer feedback. Someone exports a large dataset containing contact details, transaction history and complaint information into a workspace hosted outside Australia. The vendor’s terms, data-retention settings and cross-border processing arrangements were never properly assessed. The business may now face privacy, contractual and reputational questions long before it can determine precisely where the data went.

Neither example requires science fiction. Both are plausible outcomes of normal operational shortcuts.

IBM found that organisations with high levels of shadow AI incurred average breach costs USD 670,000 higher than organisations with low or no shadow AI. The exact financial impact will differ by industry and incident, but the direction is hard to dismiss: unmanaged AI use has a measurable business cost.

What good looks like

A mature ISO 42001 programme should not become a bureaucratic obstacle course for every employee who wants to use an approved writing assistant. Done well, it gives people a safe route to use AI rather than encouraging them to work around security.

It starts with knowing where AI is being used: customer-facing products, internal automation, embedded features in enterprise software, analytics tools, development assistants and emerging AI agents. That inventory should identify the business owner, intended purpose, data involved, supplier, decision criticality, integration points and the identities or permissions the system uses.

From there, the organisation can apply proportionate controls. A tool used to summarise public meeting notes does not warrant the same assessment as an AI system supporting employment decisions, handling health information, making credit-related recommendations or accessing privileged systems.

Good governance also means treating AI as a lifecycle issue. Before deployment, teams assess purpose, data provenance, privacy, security, bias, explainability, human oversight and supplier terms. In production, they monitor performance, access, unexpected behaviour, complaints, changes in the model or vendor service, and whether the original risk assumptions still hold.

And crucially, it means clear decision rights. The business owns the outcome. Technology owns the architecture and operational controls. Privacy, legal, risk and security provide challenge and assurance. The board sets the risk appetite and receives meaningful reporting, rather than a slide declaring that “AI governance is in place”.

The Australian privacy connection

ISO 42001 is not a substitute for Australian privacy law. Certification does not relieve an organisation of its responsibilities under the *Privacy Act 1988* or the Australian Privacy Principles (APPs).

The Office of the Australian Information Commissioner is explicit: the Privacy Act and APPs apply to AI uses involving personal information, including where personal information is used to train, test or operate an AI system. That includes information entered into a commercial AI product and AI-generated output where it relates to an identified or reasonably identifiable person.

This is where ISO 42001 is useful. It provides a management-system structure through which privacy obligations can be operationalised: data-use decisions, impact assessment, records of processing and purpose, transparency, access controls, supplier oversight, testing, review and corrective action.

It also fits naturally alongside ISO/IEC 27001. ISO 27001 provides the information-security management foundation: confidentiality, integrity, availability, security risks and controls. ISO 42001 adds the AI-specific governance layer: model behaviour, data and system impacts, human oversight, AI lifecycle controls and stakeholder considerations. They overlap, but neither makes the other redundant.

For organisations operating across jurisdictions, ISO 42001 can also create useful common ground with emerging AI regulation and customer assurance expectations. It does not confer legal compliance by itself, but a well-run AIMS produces something often more valuable than a policy statement: defensible evidence.

Should you seek certification?

Not every organisation needs an ISO 42001 certificate immediately.

For many SMEs, the first sensible move is to adopt the standard's disciplines without rushing into an external audit. If AI use is limited, low-risk and internal, a right-sized governance model, documented risk assessment, approved-tool register, privacy review and strong identity controls may deliver most of the early value.

Certification becomes more compelling when AI is central to what the organisation sells, when customers demand assurance, when sensitive or regulated data is involved, or when the business wants a credible way to distinguish itself in procurement and partner conversations. It can also be valuable for organisations already operating ISO 27001, because the management-system muscles already exist: scope, leadership commitment, internal audit, management review, corrective action and continual improvement.

There is a cost, though. Certification takes executive sponsorship, defined scope, evidence, internal effort, staff awareness, risk

assessments and recurring audit discipline. It should not be purchased as a badge for a website footer.

The better question is not, “Will certification make us look responsible?” It is, “Would the discipline of certification materially improve how we make, use and assure AI-driven decisions?”

If the answer is yes, certification may be a sound investment.

What are the next steps

- Do we have a reliable inventory of every AI capability in use, including embedded features, plugins, copilots, agents and supplier-provided tools?
- Which AI use cases process personal information, commercially sensitive data, regulated information or decisions that materially affect customers, employees or partners?
- Are we looking to implement ISO 42001 as a practical governance model first, or is an accredited certification genuinely required by customers, tenders, investors or our risk profile?
- How much effort will certification take once we define a realistic scope, and which ISO 27001, privacy, risk and supplier-assurance artefacts can we reuse?

- What is our plan for shadow AI when visibility is poor and staff are already using tools faster than formal governance can approve them?
- Can we demonstrate that AI agents, service accounts and integrations have only the access they need, with monitoring, review and a clear accountable owner?

These are not questions to leave solely with the IT team. They determine how confidently an organisation can pursue AI-enabled growth without discovering, after an incident or audit, that its controls existed only in people's assumptions.

ISO 42001 will not remove the uncertainty that comes with AI. But it can replace ad hoc adoption with a repeatable, auditable way to make better decisions. For organisations that intend to use AI seriously, that is far more valuable than a certificate on the wall.