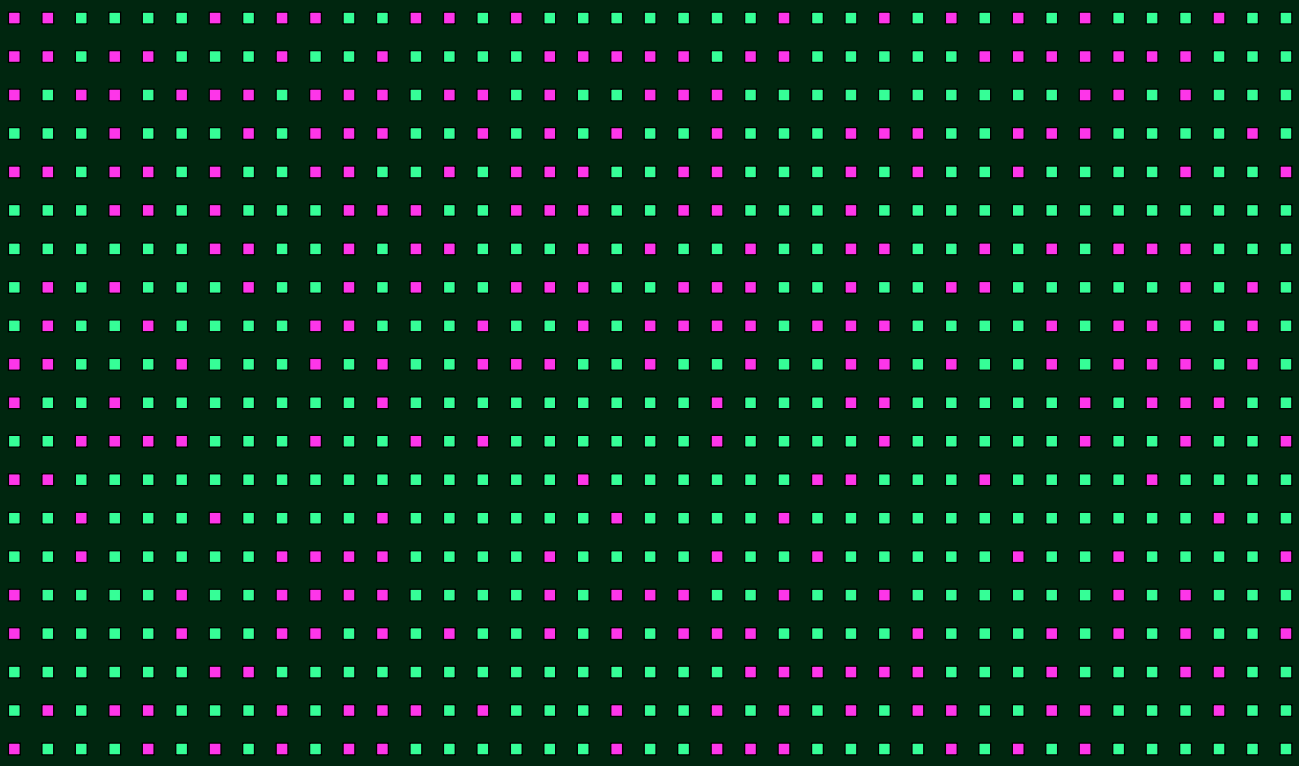
Public V1.0

Securing the AI Attack Surface

Insights from the IO State of Information Security Report on how AI is reshaping risk, governance, and resilience



The state of information security report 2025

Foreword

Chris Newton-Smith
CEO

As businesses embrace cloud, AI, and digital transformation, the risks grow just as fast.

Our State of Information Security Report 2025 reveals how organisations are adapting, where gaps remain, and what resilience looks like in the year ahead.

This report offers a focused deep dive into one of the most significant findings from our State of Information Security 2025 research: the rapid rise of artificial intelligence as both a security enabler and a new source of risk. AI promises faster decision-making, greater efficiency, and new competitive advantages. Yet its adoption also brings unseen risks: data leaks through "shadow AI", model manipulation, deepfakes, and AI-driven social engineering.

Our State of Information Security 2025 data shows that 79% of organisations adopted AI or machine learning in the past year, with 37% already concerned about unsanctioned use. The pace of change is reminiscent of the early cloud era, innovation outpacing governance.

At the same time, interest in ISO 42001, the new global standard for AI management systems, has soared from 2% to 28% year-on-year, signalling that organisations are seeking structure and assurance amid the uncertainty.

Drawing on exclusive data from the State of Information Security 2025 report, this deep dive examines how AI is reshaping the attack surface, what it means for governance and compliance, and why resilience in the AI era depends as much on how we manage and oversee AI as on how we secure data.

The reality is that threats will keep changing. What matters is that we are better prepared, treating information security not as a back-office function, but as part of how we build resilience, earn trust and grow.

The AI surge and the expanding attack surface

Our State of Information Security 2025 findings reveal a striking acceleration in AI adoption from experimentation to enterprise integration.

Across sectors, organisations are embedding AI into workflows, analytics, customer interactions, and product design. In our survey, nearly four in five respondents reported adopting AI or ML, while one in five more plan to do so within a year.

The benefits are undeniable: faster insight, reduced manual workload, and more adaptive systems. But each new dataset, model, and API also expands the potential for exposure.

A growing concern is “shadow AI”, the unsanctioned use of generative AI tools by employees. Just as “shadow IT” once proliferated when cloud apps first appeared, staff are now turning to AI platforms to boost productivity without considering data privacy or compliance risks. Sensitive information can easily find its way into external systems beyond an organisation’s control.

The speed of adoption often outpaces governance. Over half of organisations (54%) admit they deployed AI faster than they could assess the associated risk. The result: innovation has outstripped oversight.

Yet the rise in ISO 42001 adoption interest shows a turning tide. Businesses are beginning to recognise that the next phase of AI maturity isn’t technical, it’s structural. As one industry leader noted:

“The AI revolution has moved faster than the security and compliance frameworks designed to govern it. Organisations are now racing to close that gap.”

To what extent do you agree or disagree with the following statements about the current state of the information security landscape?

The AI attack surface:

Where risk lives now

The modern attack surface is no longer limited to networks or endpoints. Respondents to the State of Information Security 2025 survey identified a new set of risks emerging from AI integration, which now includes algorithms, data pipelines, and decision-making models, each a potential vector for compromise.

In the State of Information Security 2025, respondents identified a range of emerging AI-related threats. These risks stretch across every layer of the digital enterprise, from employees and data assets to suppliers and customers.

- **AI data/model poisoning (26%)** attackers manipulating training data or inserting hidden triggers to bias model outcomes or introduce backdoors.
- **AI-generated social engineering (38%)** generative models crafting hyper-personalised phishing or vishing messages that evade traditional detection.
- **Deepfakes and impersonation (27%)** synthetic audio and video eroding trust in verification processes.
- **Shadow AI** data leakage through unsanctioned or poorly configured AI tools, undermining GDPR and contractual compliance.
- **Supply chain exposure** AI features embedded into third-party software without adequate oversight.

A simple way to visualise the modern AI attack surface is through five domains:

Data	Models	People	Systems	Supply chain
compromised or unverified training inputs	vulnerabilities within deployed AI systems	misuse, error, or social engineering	integration points and APIs	third-party AI dependencies

AI risk is no longer hypothetical. It is operational, measurable, and already reshaping security priorities. As AI becomes embedded across business processes, resilience will depend on knowing where and how AI is being used, and who is accountable for its outcomes.

From innovation to exposure:

The governance and compliance gap

The report's data also highlights that whilst adoption accelerates, governance has lagged behind. Most organisations acknowledge that their frameworks for managing AI risk are incomplete or fragmented, pointing to a growing "governance gap".

In the report, 95% of respondents said they are planning on investing in AI governance, yet many admit they are unclear on what "good" looks like. Only a minority have formalised AI policies, audit mechanisms, or ethics committees.

This governance gap leaves businesses exposed on several fronts:

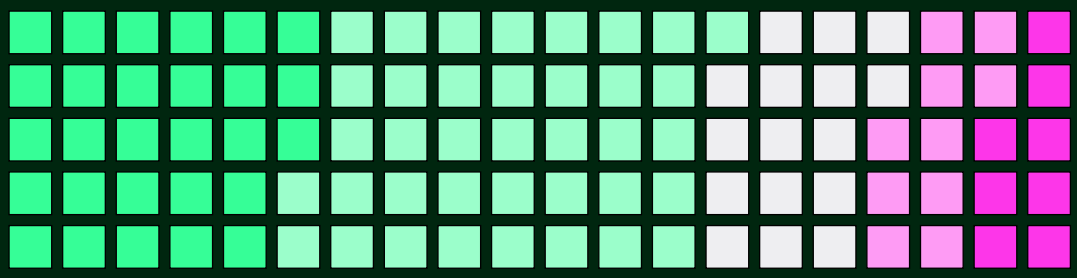
Regulatory readiness
emerging laws such as the EU AI Act will soon require demonstrable controls over model development, training data, and decision transparency.

Accountability
unclear ownership of AI decisions or errors complicates incident response and liability.

Compliance overlap
AI interacts with existing frameworks like ISO 27001, GDPR, NIS 2, and DORA, each with its own data-handling requirements.

Audit fatigue
compliance teams struggle to align evolving standards across different departments and vendors.

Two-thirds of respondents (66%) say the pace of regulatory change itself is becoming a risk factor. Many are still grappling with foundational information security obligations and now face the additional challenge of governing AI at scale.



To what extent do you agree that the speed and volume of regulatory change make it increasingly difficult to stay compliant with information security standards

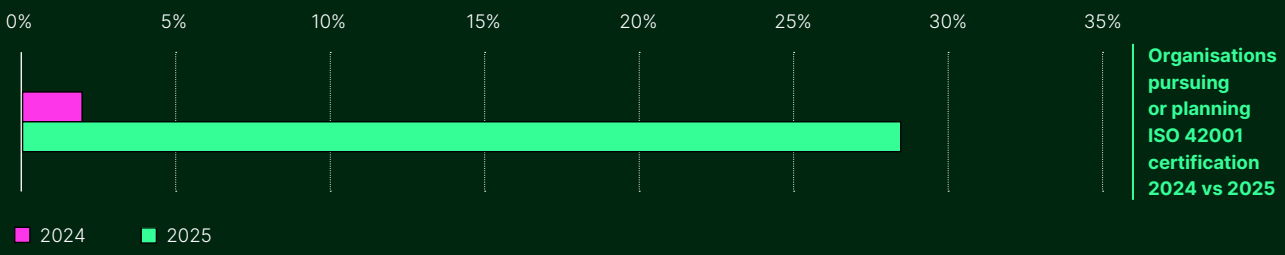
AI governance isn't a blocker to innovation; it's the only way to scale it safely. As the governance gap narrows, attention is shifting to frameworks capable of providing assurance and structure, starting with ISO 42001.

Building trust through standards:

Why ISO 42001 matters

Findings from the 2025 reports show that interest in ISO 42001, the new global AI management system standard, has surged.

ISO 42001 provides a management system framework, similar to ISO 27001 for information security, focused explicitly on the lifecycle of AI systems. Interest in the standard has grown dramatically. Our data shows a jump from 2% to 28% year-on-year among organisations pursuing or planning ISO 42001 certification. This signals a broader move toward structured governance and transparent accountability.



At its core, ISO 42001 helps organisations operationalise four principles of trustworthy AI:

Ethical and explainable AI ensuring transparency in automated decisions.

Data quality and integrity validating the inputs that shape AI behaviour.

Risk management and incident response addressing errors, bias, or misuse systematically.

Supplier governance requiring third-party AI components to meet defined assurance standards.

Crucially, ISO 42001 is designed to integrate with existing frameworks. Organisations already certified to ISO 27001 (information security) or ISO 27701 (privacy) can extend their management systems to include AI-specific controls with relative ease. This convergence reduces duplication, simplifies audits, and strengthens confidence among customers, regulators, and partners.

Forward-thinking organisations view ISO 42001 not merely as a compliance checkbox but as a strategic enabler, a way to build trust, demonstrate accountability, and differentiate through responsible innovation.

The human element:
Awareness, training & trust

Every security evolution eventually comes back to people. AI is no exception.

While 38% of respondents cite lack of employee awareness as a key challenge, the risks extend beyond phishing or password hygiene. Staff now interact directly with AI, prompting models, uploading data, and relying on outputs to make decisions.

Without clear guidance, well-intentioned employees may inadvertently expose sensitive information or delegate critical decisions to unverified systems. That's why AI literacy is emerging as a cornerstone of cybersecurity culture.

Organisations should:

- Define acceptable-use policies for generative AI tools.
- Provide training on responsible AI use and data handling.
- Encourage human oversight for AI-assisted decisions, particularly in regulated contexts.
- Integrate AI awareness into existing security education programmes.

"AI literacy is the new security awareness. Every employee using AI is now part of the attack surface, and part of the solution."

What are the common types of information security/ cybersecurity mistakes made by your employees in the last 12 months? (top responses)

Building trust in AI means ensuring people understand both its power and its limits. A well-trained workforce remains the strongest defence against the risks of shadow AI, data leakage, and algorithmic bias.

Turning risk into resilience:

The role of governance platforms

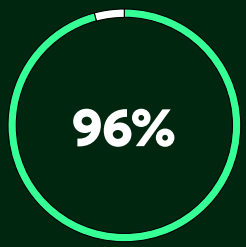
AI may amplify threats, but it can also strengthen defences. When deployed responsibly, AI supports faster detection, adaptive response, and smarter compliance.

Almost 96% of organisations plan to invest in AI-enabled threat detection, and 30% specifically aim to enhance defences against AI-generated attacks. These technologies are increasingly capable of analysing anomalies, correlating incidents, and identifying patterns beyond human capacity.

However, technology alone is insufficient. Governance remains the foundation that ensures AI systems are used securely, not merely built securely. Governance platforms play a vital role by connecting policies, risks, and controls across AI, privacy, and information security domains.

A unified approach provides:

- **Visibility**
  an inventory of all AI systems, owners, and associated risks.
- **Accountability**
  clear mapping of controls to ISO 42001 or AI Act requirements.
- **Readiness**
  audit-ready evidence of compliance across multiple frameworks.
- **Resilience**
  integrated incident response and continuous improvement processes.



Plan to invest in GenAI threat detection & defence

Are enhancing defences against AI-generated threats

When governance and technology evolve together, resilience becomes proactive rather than reactive. Businesses can anticipate regulatory shifts, reduce duplication of effort, and demonstrate to stakeholders that AI innovation is under control.

Governance is the new AI advantage

AI is now central to how organisations compete, communicate, and create value. It is also redefining how they must think about risk. The attack surface is expanding, but so too are the frameworks to manage it.

The message from the State of Information Security 2025 is unambiguous: AI governance has moved from a niche consideration to a board-level priority. With ISO 42001 adoption rising from 2% to 28% in just one year, businesses are recognising that structure and accountability are the true enablers of innovation.

Resilience in the AI era will depend on visibility, control, and continuous improvement, the same principles that underpin ISO 27001. Organisations that embrace integrated governance will not only comply with emerging regulations but also earn the trust that sustains long-term growth.

Governance isn't bureaucracy; it's resilience by design.

Standard Focus Role in AI Governance

ISO 27001	ISO 27701	ISO 42001
Information Security	**Data Privacy**	**Responsible AI**
Protects data and systems	Manages personal information	Governs AI lifecycle and accountability

Together, these frameworks form the foundation of trustworthy, compliant, and resilient digital operations, an approach that unites technology, process, and people under one structured management system.

Get the full story in

The state of information security report 2025

Read the full report →

Explore more at **isms.online**