PECB Magazine

JANUARY - MARCH
2026

ISSUE 54

Trust, Risk, and Accountability in the Age of AI

Governing Intelligence, Building Trust

ISO STANDARDS AND BEYOND

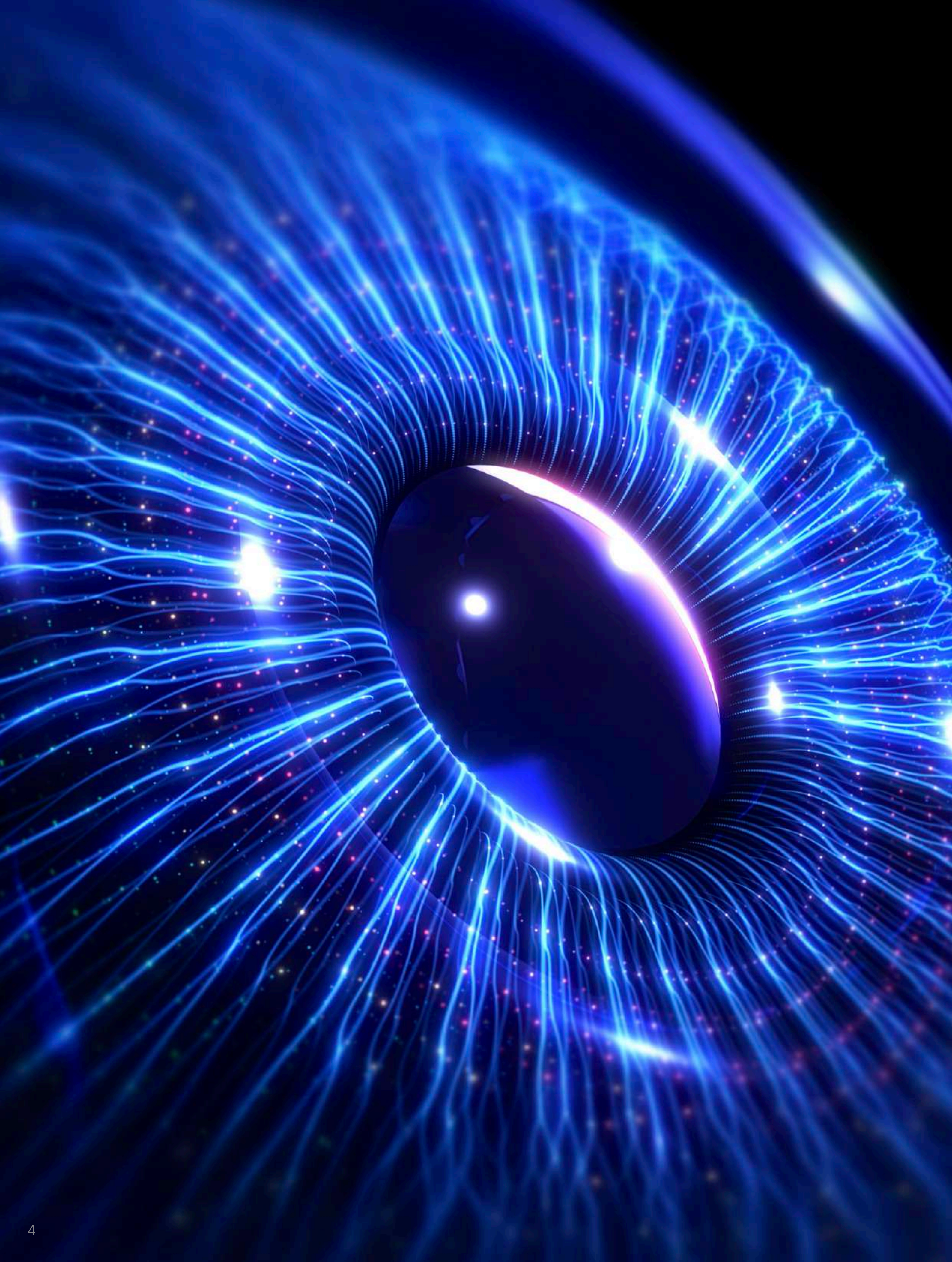
AI Is Not the Risk. Your Decision-Making Is

From Certification to Confidence: How ISO Standards Support Privacy and Data Trust

The Future Is Here Now: Integrating AI into Enterprise Risk Management Frameworks

The Human Element: Reskilling and Re-architecting for an AI-First Future

In This Issue



> **The future of AI is not about replacing humans, it’s about augmenting human capabilities.**
>
> *Sundar Pichai, CEO of Google*

THE EXPERT

The Future Is Here Now: Integrating Artificial Intelligence Into Enterprise Risk Management Frameworks

We have seen it in science fiction movies over the last few decades, and now it is finally part of a reality that we are living in - Artificial Intelligence (AI).

Various fictional iterations have depicted AI as both friend and foe, but the current real-world application is far less hyperbolic. Organizations are embedding AI into customer analytics, credit decision-making, operational forecasting, supply chain optimization, cybersecurity monitoring, and strategic planning, to name just a few.

What started as targeted experimentation has evolved into fast-paced structural integration. Owing to the inevitability of this next step in technological evolution, the question is no longer whether AI should be adopted, but rather how it should be governed responsibly. From a Risk Management standpoint, it should be pondered how it can be aligned with enterprise risk appetite and embedded into an organization's risk culture.

Enterprise Risk Management (ERM) frameworks were originally designed in environments characterized by largely deterministic systems, predictable operational processes, and clearly attributable decision pathways. AI challenges these assumptions because machine learning models evolve over time, so decision logic may not be easily interpretable, and outputs are dependent on dynamic data ecosystems. These characteristics introduce new forms of uncertainty that require recalibration of governance structures rather than incremental adjustment.

AI governance is not merely a technical consideration. It represents a legal, ethical, and strategic imperative that requires organizations to reassess foundational assumptions across governance, accountability, and control frameworks. Accordingly, the integration of AI into ERM should be approached as a purposeful evolution rather than an incremental adaptation, ensuring that technological innovation remains anchored within structured oversight.

Regulatory Context and Impact

From a regulatory perspective, organizations must ensure that AI is used in an ethical and responsible manner. Many countries have already undertaken a proactive legislative approach, which establishes a risk-based classification framework for AI systems. High-risk systems are subject to strict obligations relating to transparency, documentation, human oversight, accuracy, and robustness. Organizations deploying AI must therefore demonstrate structured lifecycle governance and auditability.

Organizations have obligations relating to automated decision-making, transparency, lawful processing, and data accuracy.

The emphasis is on accountability and explainability in algorithmic systems. This regulatory context underscores a central principle: AI governance must be embedded within enterprise control frameworks. Compliance cannot be achieved retrospectively. It must be integrated into design, deployment, and ongoing monitoring processes.

AI as a Distinct Risk Domain

AI introduces a risk profile that is qualitatively different from those addressed by conventional Information Technology (IT) frameworks. Beyond familiar considerations such as data security and system availability, AI systems generate a distinct set of exposures, including model bias, ethical liability, decision opacity, model drift, and adversarial manipulation, as well as amplified forms of systemic risk arising from concentrated dependency on a small number of technology providers. Understanding these risks as a discrete domain, rather than an extension of existing IT risk, is essential to effective governance.

Model bias sits at the foundation of AI risk. The quality of any model's output is fundamentally constrained by the quality of its training data; therefore, incomplete or unrepresentative datasets can produce outputs that are systematically skewed. In regulated sectors such as financial services, such output may constitute breaches of conduct obligations, fair treatment standards, or equality legislation, carrying significant reputational and regulatory consequences.

Closely related is the challenge of explainability. Unlike conventional rule-based systems, advanced machine learning models may not produce easily interpretable reasoning for the decisions they generate. Where AI-driven output materially affects customers, employees, or other stakeholders, organizations bear a responsibility to demonstrate that meaningful human oversight exists. A requirement that is increasingly highlighted in emerging regulatory frameworks.

Model drift introduces a more subtle but equally consequential risk. As the real-world data environment evolves, the statistical patterns on which a model was trained may no longer hold, quietly degrading its accuracy and reliability. Without continuous monitoring and validation protocols, this deterioration can go undetected until adverse outcomes have already materialized - by which point the reputational and financial damage may be significant.

Third-party risk is further amplified in AI-enabled environments. Organizations frequently rely on external vendors for AI tooling, cloud infrastructure, and embedded algorithmic functionality. This dependency requires that traditional supplier due diligence frameworks be meaningfully extended to encompass the governance of AI models themselves, including scrutiny of training methodologies, data sourcing practices, and regulatory compliance within the supply chain.

Recalibrating Enterprise Risk Management Components

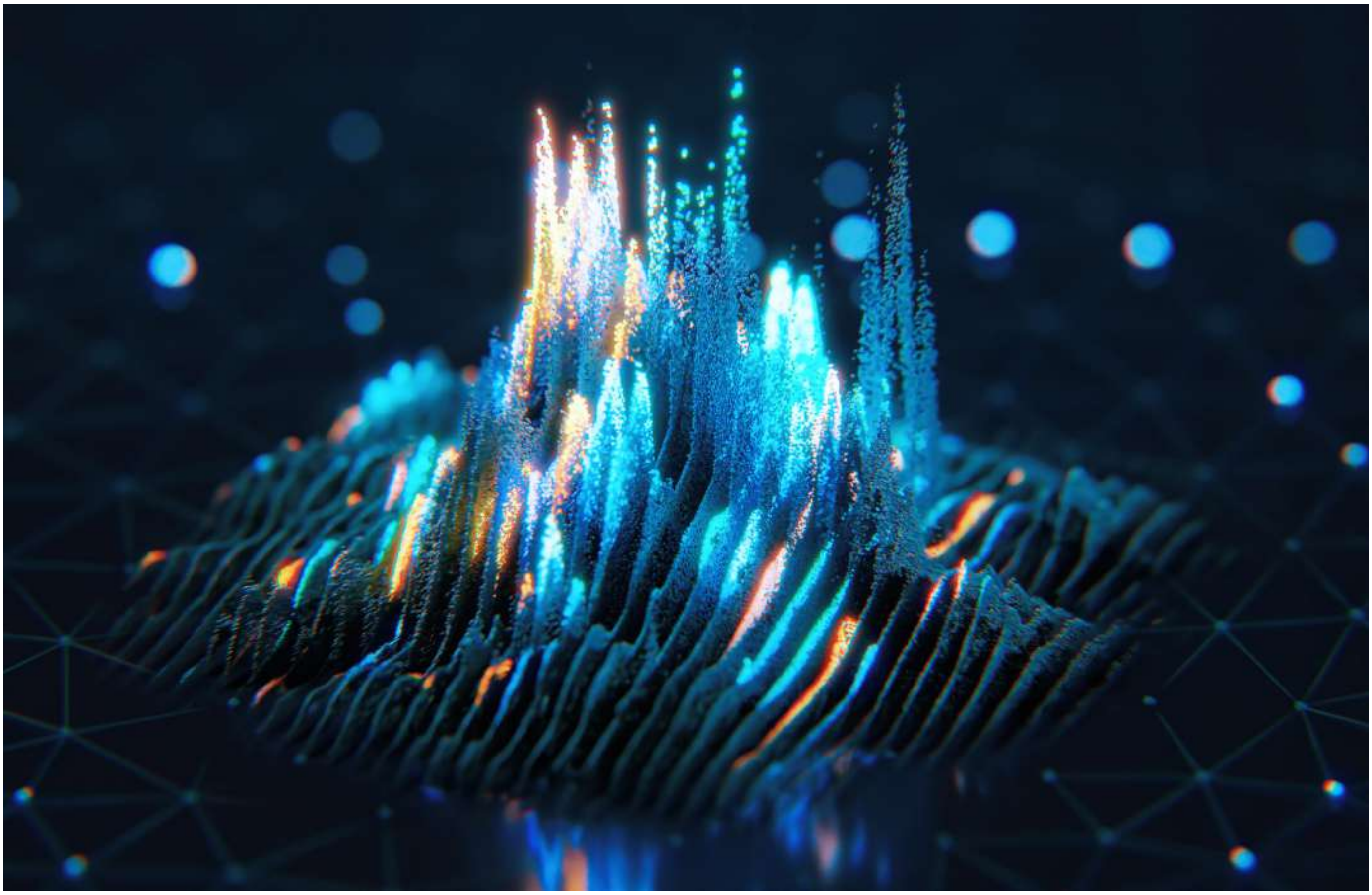
Integrating AI into ERM demands a fundamental rethinking of how the discipline operates across its core pillars: defining risk appetite, identifying risks, conducting assessments, designing controls, and monitoring outcomes. Each must evolve to account for the distinct characteristics that AI systems bring to the table.

Risk appetite frameworks can no longer treat AI as an outlying consideration. Boards and senior leadership need to set explicit thresholds around AI-related exposure. This includes defining what levels of automation are acceptable, how much model error can be tolerated, and how much opacity is permissible in algorithmic decision-making. Without this clarity, organizations risk a dangerous drift between their appetite for innovation and their commitment to sound governance.

Clear articulation of these boundaries ensures that these aspects are in alignment – a tightrope that many large organizations must walk in the modern age.

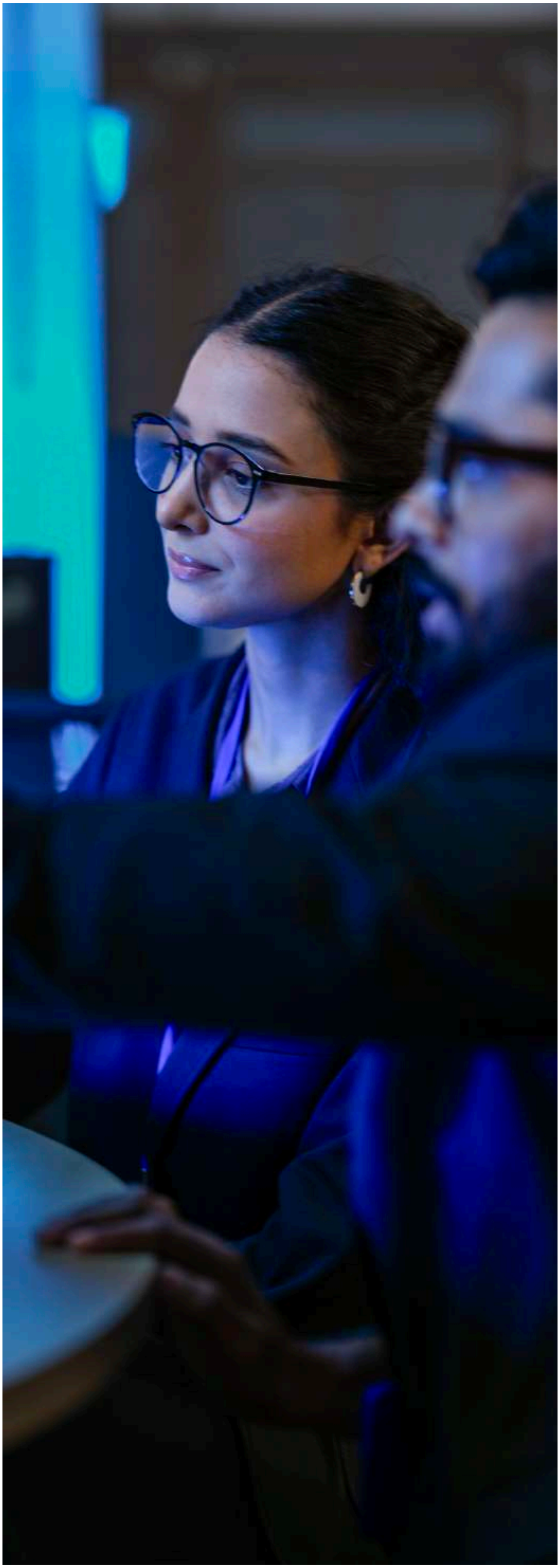
Risk identification processes need to expand to include a comprehensive AI inventory: a living register that documents each system's purpose, business ownership, data inputs, autonomy level, and impact classification. This exercise is the foundation for enterprise-wide visibility, giving boards and risk committees the information they need to make genuinely informed decisions.

Risk assessment methodologies should be bolstered with tools built for AI environments. Algorithmic impact assessments, fairness testing, and evaluations of explainability all become relevant considerations. Scenario analysis should treat model failure, data corruption, cyber compromise, and regulatory non-compliance as credible planning assumptions rather than remote possibilities.



Control design must span the full AI lifecycle, from development and independent validation through deployment approval, structured retraining, and eventual decommissioning. Documentation throughout the process is the organization's primary means of demonstrating accountability to regulators, auditors, and stakeholders.

Monitoring, too, requires a rethink. Periodic review cycles are increasingly inadequate for systems that operate around the clock. Real-time dashboards tracking performance deviation, anomaly detection metrics, and data integrity indicators are fast becoming essential infrastructure for any organization serious about maintaining meaningful assurance.



Governance and Cultural Considerations

Effective AI governance does not emerge from technology teams alone. It demands coordinated ownership across risk, compliance, legal, data science, technology, and executive leadership with accountability structures documented clearly enough to withstand regulatory scrutiny. Governance fragmentation remains one of the more persistent failures organizations face when scaling AI, and it typically traces back to capability gaps that were never addressed: risk professionals without sufficient grounding in AI methodologies, and technologists without meaningful exposure to the regulatory and ethical obligations their work carries.

Board engagement is equally important. Determinations about acceptable automation thresholds and ethical boundaries are not operational matters to be delegated downward, as they carry material risk implications and belong within enterprise-level risk committees and reporting frameworks.

Organizations that treat AI governance as an IT function will find themselves poorly positioned when regulators, auditors, or adverse events demand a more substantive answer.

AI as an Enabler of Enterprise Risk Management

It would be a mistake to approach AI purely as a source of new risk. For all the governance challenges it presents, it is also one of the most powerful tools available to the profession today.

Predictive analytics can surface emerging risk patterns across datasets of a scale and complexity that no human team could process manually.

Machine learning can drive early warning systems capable of detecting anomalies well before they escalate into material incidents, a capability that traditional control environments were never designed to provide.

Advanced scenario modelling now allows organizations to simulate thousands of potential disruption pathways simultaneously, bringing a depth and rigor to stress testing that was previously out of reach. Natural language processing tools can continuously scan incident reports, regulatory updates, and operational logs, picking out thematic trends and emerging signals that would otherwise go unnoticed.

In fraud detection and cybersecurity domains, AI-driven anomaly detection has already proven its value by materially strengthening control environments. When embedded thoughtfully within an ERM framework, these capabilities enable organizations to move from reactive mitigation towards something far more valuable: genuine predictive intelligence. That shift is significant. It changes the nature of what Risk professionals can offer and raises the bar for what good risk management looks like going forward.

Conclusion

As organizations continue to drive AI adoption as a harbinger of evolution, AI is reshaping Enterprise Risk landscapes across the entire globe. Its integration into ERM frameworks is not optional but rather inevitable. Organizations that embed AI within structured governance architectures will strengthen resilience, enhance regulatory compliance, and preserve stakeholder trust.

Only time will tell if science fiction movies were predictions of inevitability, but now in the present, before the Matrix is built and Terminators walk among us, organizations need to ensure that AI functions as an augmentation of human judgment, not a substitute for it. The value of AI lies in sharpening decision-making quality within environments where human oversight, accountability, and ethical considerations remain firmly in place. That is not a limitation on what AI can do, but rather the condition under which it does it well.

The role of the modern risk leader is expanding. It now requires fluency in both traditional risk disciplines and emerging technological frameworks. By integrating AI deliberately, ethically, and intelligently into ERM, organizations can position risk management not as a constraint on innovation, but as an enabler of responsible transformation.

The risk professionals who will shape the next era of the discipline are those who see themselves as strategic advisors on what is coming, not just custodians of what exists today. Keeping pace with the present is necessary but not sufficient. The best in the field will anticipate how the landscape is changing and help their organizations get ahead of it. That orientation is what will keep Risk Management relevant and resilient for generations to come.

Hershin Marcello Ramjawan

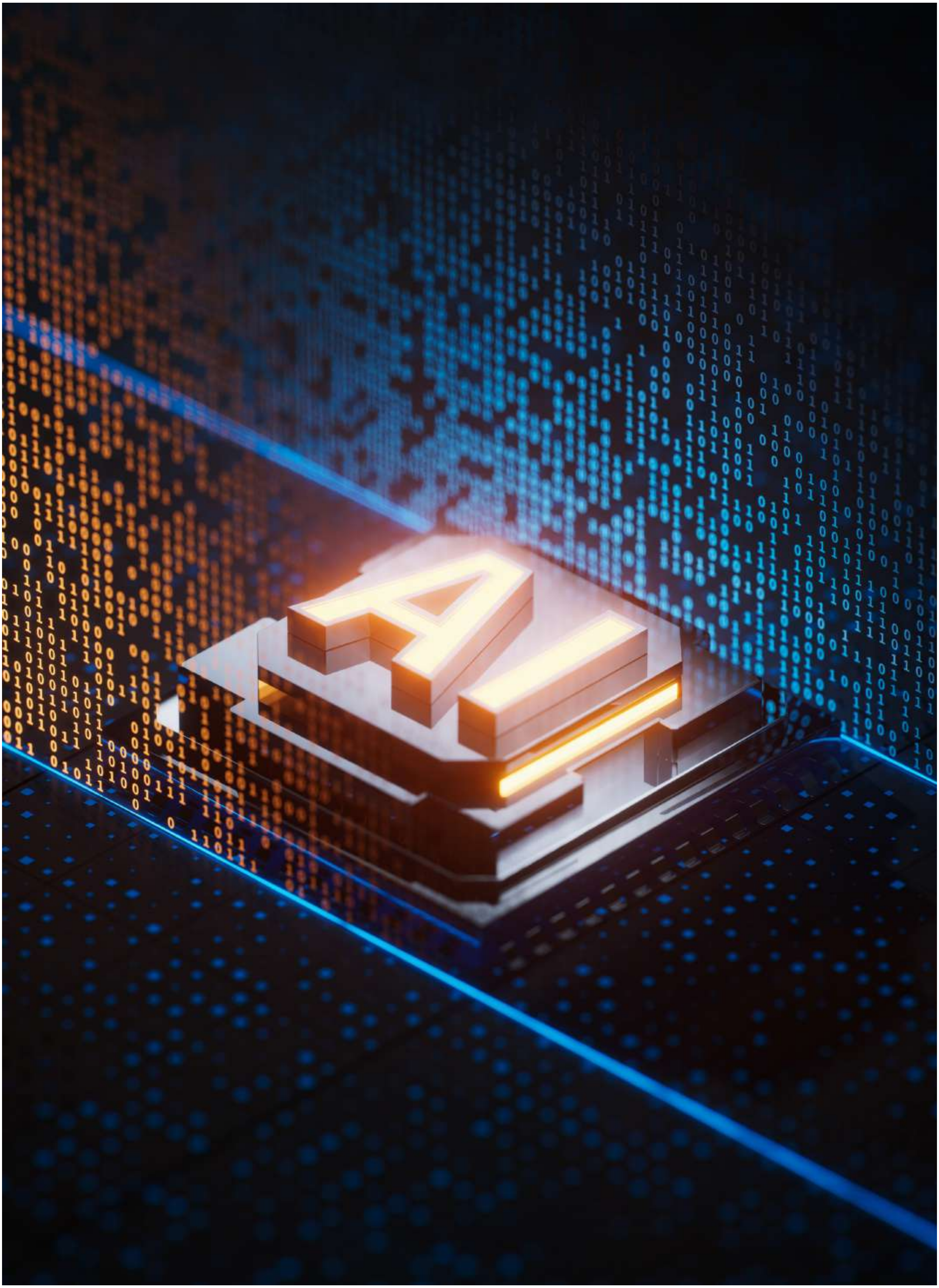
Certified Risk and Information Systems Control

Hershin is a risk management professional with an international career spanning the financial services industry across banking, insurance, and management consulting.

Over the course of his career, Hershin has built deep expertise across operational risk, market conduct, compliance, investigations, safety management, assurance, operational resilience, and governance transformation.

He lives by a philosophy of accepted residual risk and maintains rigorous control discipline in all areas of life — with one well-documented recent exception involving cheesecake, which he has formally classified as a black swan event.

“If you never take any risks in life, your rewards will always be in your dreams.” — Hershin Marcello Ramjawan (CRISC). If you would like to connect with Hershin on all things Risk Management, you can find him on LinkedIn by clicking the icon above.


AI

THE STANDARD

Building a Responsible AI: How to Manage the AI Ethics Debate

In today's rapidly evolving tech landscape, responsible artificial intelligence (AI) stands at the forefront of efforts to align AI with societal values and expectations. While still growing and developing at an accelerated pace, AI is already augmenting human life.

The technology is now increasingly commonplace in our homes, our workplaces, our travels, our healthcare and our schools. What would have seemed like science fiction just two decades ago – such as self-driving cars and virtual personal assistants – is set to become a fixture of our everyday lives.

Responsible AI is the practice of developing and using AI systems in a way that benefits society while minimizing the risk of negative consequences. It's about creating AI technologies that not only advance our capabilities, but also **address ethical concerns** – particularly with regard to **bias, transparency and privacy**. This includes tackling issues such as the misuse of personal data, biased algorithms, and the potential for AI to perpetuate or exacerbate existing inequalities. The goal is to build trustworthy AI systems that are, all at once, reliable, fair and aligned with human values.

Where do we go from here? How do we better frame the technology to unleash the full potential of AI? A robust ecosystem of standards and regulations will be needed to ensure the responsible development, deployment and use of AI as we navigate this era of remarkable, exponential innovation. Here, we examine the complex and evolving field of AI ethics in artificial intelligence, and how we should approach this transformative but uncharted technology.

What Is Responsible AI?

As AI evolves, it has the potential to bring life-changing advances. So, before AI's increasing momentum gathers even more pace, it is crucial to prioritize responsible AI development, which takes into account all potential societal impacts.

Responsible AI is an approach to developing and deploying artificial intelligence from both an ethical and legal standpoint. The goal is to employ AI in a safe, trustworthy and ethical way. Using AI responsibly should increase transparency while helping to reduce issues such as AI bias.

So why all the hype about "what is AI ethics"? The ethics of artificial intelligence are a huge challenge to humankind. Mindful and responsible innovation is not an easy concept in itself, but it is crucial to first grasp the question of what AI ethics are and integrate them into the core of the development and application of AI systems. In short, ethical AI is based around societal values and trying to do the right thing. Responsible AI, on the other hand, is more tactical. It relates to the way we develop and use technology and tools (e.g. diversity, bias).

Why Is Responsible AI Important?

As AI becomes more business-critical for organizations, achieving responsible AI should be considered a highly relevant topic. There is a growing need to proactively drive fair, responsible, ethical AI decisions and comply with current laws and regulations.

Understanding the concerns of AI is the starting point for creating an ethical framework to **guide its development and use.** Any organization wishing to ensure their use of AI isn't harmful should openly share this decision with as diverse a range of stakeholders as it can reasonably reach, along with consumers, clients, suppliers and any others who may be tangentially involved and affected.

Developing and applying AI along the principles of AI ethics requires transparency in decision-making processes and the development of actionable policies of AI ethics. With considered research, widespread consultation and analysis of ethical impact, coupled with ongoing checks and balances, we can ensure that AI technology is developed and deployed responsibly, **in the interests of everyone,** regardless of gender, race, faith, demographic, location or net worth.

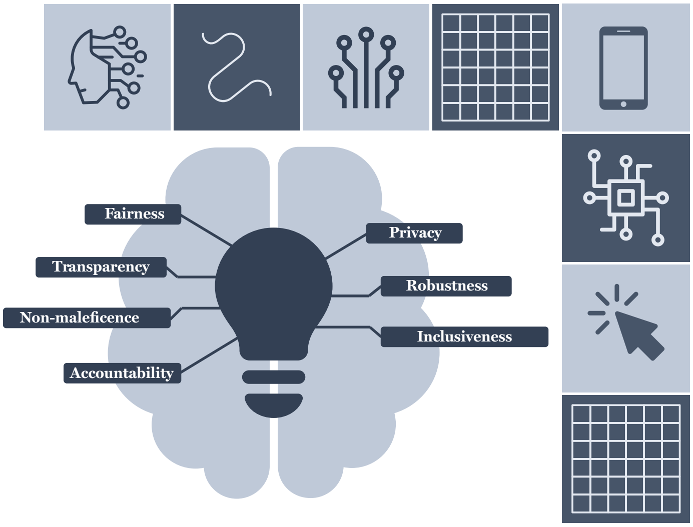
What Are the Principles of Responsible AI?

Confronting ethical concerns means engaging with their ramifications with foresight and commitment. It's vital to view AI's ethical dimension not as an obstacle but as a conduit to lasting and sustainable tech progress. That's why embedding responsible AI principles is essential to its evolution in a direction that benefits all.

While there isn't a fixed, universally agreed-upon set of principles for AI ethics, several guidelines emerge. Some key principles of AI ethics are:

- **Fairness:** Datasets used for training the AI system must be given careful consideration to avoid discrimination.
- **Transparency:** AI systems should be designed in a way that allows users to understand how the algorithms work.
- **Non-maleficence:** AI systems should avoid harming individuals, society or the environment.
- **Accountability:** Developers, organizations and policymakers must ensure AI is developed and used responsibly.
- **Privacy:** AI must protect people's personal data, which involves developing mechanisms for individuals to control how their data is collected and used.
- **Robustness:** AI systems should be secure – that is, resilient to errors, adversarial attacks and unexpected inputs.
- **Inclusiveness:** Engaging with diverse perspectives helps identify potential ethical concerns of AI and ensures a collective effort to address them .

Key principles of responsible AI



Source. ISO - Building a responsible AI: How to manage the AI ethics debate

Promoting Responsible AI Practices

These principles should help to steer considered and responsible decision making around AI. In order to transition from theory to practice, organizations must create actionable policies of AI ethics. Such policies are crucial in weaving ethical considerations throughout the AI life cycle, ensuring integrity from inception to real-world application.

While organizations may choose different routes to embed responsible AI practices into their operations, there are a few AI best practices that can help implement these principles at every stage of development and deployment.

When deciding how to establish AI ethics, companies should:

- Foster **collaboration** across all disciplines, engaging experts from policy, technology, ethics and social advocacy to ensure multifaceted perspectives
- Prioritize **ongoing education** on AI best practices at all levels to maintain awareness and adaptability
- Implement **AI ethics throughout the technology's design**, building them into AI solutions from the ground up
- Establish clear **oversight mechanisms**, such as ethics committees or review Boards, to monitor compliance and guide ethical decision making
- Protect **end-user privacy and sensitive data** through strong AI governance and data usage policies
- Encourage **transparency in AI processes**, enabling accountability and trust from stakeholders and the public

Keeping Up with AI Best Practice

To keep your AI system trustworthy, it's important to focus on three key areas: feeding it good, diverse data; ensuring algorithms can handle that diversity; and testing the resulting software for any mislabeling or poor correlations.

Here's how to achieve this:

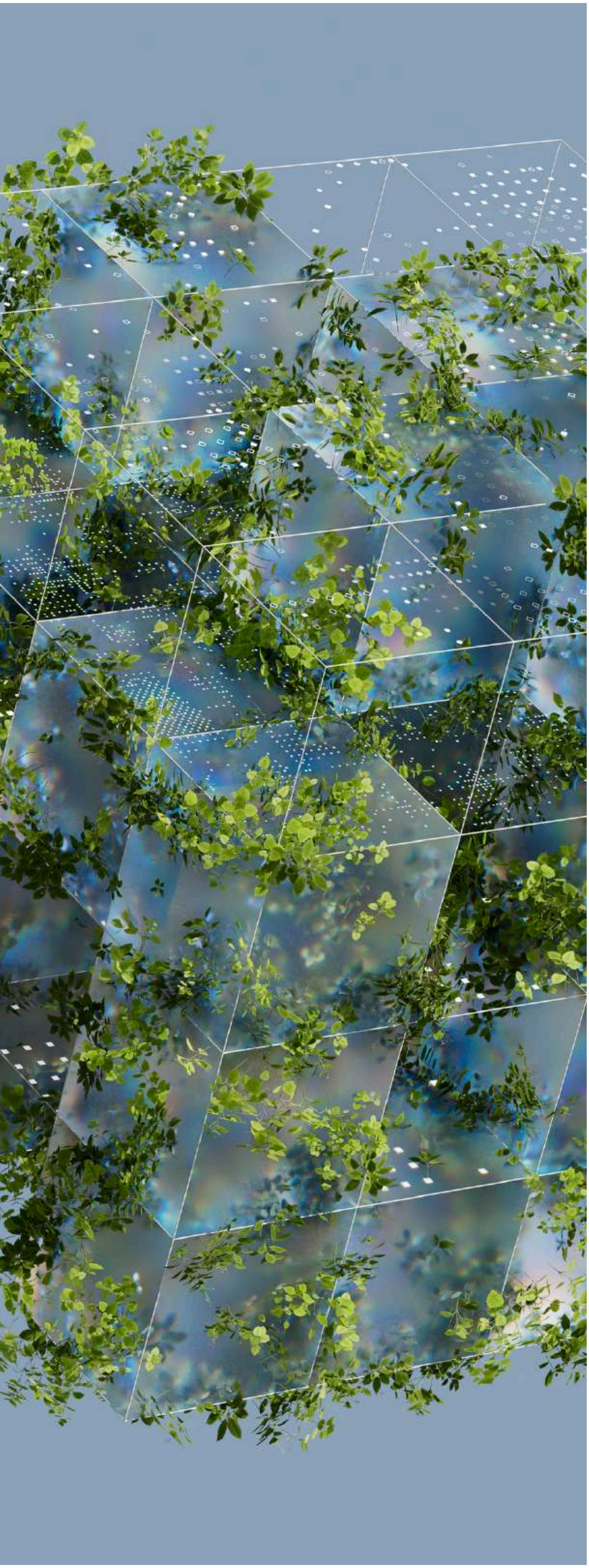
- **Design for humans** by using a diverse set of users and use-case scenarios, and incorporating this feedback before and throughout the project's development.
- **Use multiple metrics** to assess training and monitoring, including user surveys, overall system performance indicators, and false positive and negative rates sliced across different subgroups.
- **Probe the raw data for mistakes** (e.g. missing values, incorrect labels, sampling), training skews (e.g. data collection methods or inherent social biases) and redundancies – all crucial for ensuring responsible AI principles of fairness, equity and accuracy in AI systems.
- **Understand the limitations of your model** to mitigate bias, improve generalization and ensure reliable performance in real-world scenarios; and communicate these to users where possible.
- **Continually test your model** against responsible AI principles to ensure it takes real-world performance and user feedback into account, and consider both short- and long-term solutions to the issues.

Responsible AI: Examples of Success

By integrating responsible AI best practices and principles, we can ensure we end up with generative AI models that ultimately enrich our lives while keeping humans in charge. As we steadily transition towards a more responsible use of AI, numerous companies have already succeeded in creating AI-powered products that are safe and secure.

Let's take a look at some responsible AI examples:

- The **Fair Isaac Score**, by analytics software firm FICO, is a credit scoring system that uses AI algorithms to assess creditworthiness. FICO maintains responsible AI practices by regularly auditing its scoring models for bias and disparities based on mathematics instead of subjective human judgement.
- Healthcare startup **PathAI** develops **AI-powered diagnostics** solutions to aid pathologists in diagnosing diseases. To ensure the safe and responsible use of AI in its software, the company validates the accuracy and reliability of its algorithms through rigorous clinical testing and peer-reviewed studies.
- With its people-first approach, **IBM's Watsonx Orchestrate** is revolutionizing talent acquisition. This AI solution for HR and recruitment promotes fairness and inclusivity in the hiring process by generating diverse pools of candidates, using fair assessment criteria, and prompting managers to incorporate diverse perspectives in the interview process.
- **Ada Health** provides users with personalized medical assessments and advice. The **AI-powered chatbot** safely handles the diagnosis and screening of common conditions like diabetic retinopathy and breast cancer. AI best practices are ensured through transparent disclosure that users are interacting with an AI chatbot.
- Using a constellation of satellites, **Planet Labs** is pioneering the **use of AI in satellite imagery**, transforming how we monitor the environment, analyse climate patterns and assess agricultural yields. By collaborating with environmental organizations and policymakers, the company ensures AI best practices are embedded in its model.



The Standards Approach

As we advance towards responsible AI, every corner of society needs to engage and be engaged. ISO, in collaboration with the International Electrotechnical Commission (IEC), is keeping pace with this pursuit, crafting International Standards that safeguard and propel the **principled application of AI technology**.

In shaping ethical AI, the world's governments, organizations and companies need to embody these values, ensuring that their pursuit of innovation is accompanied by ethical responsibility. International Standards will help to establish a high watermark of ethics in AI, consistently guiding the best practice in this transformative industry.

A commitment to responsible AI is not a one-time act, but a sustained effort involving vigilance and adaptation. However, organizations should be aware that this commitment not only guides AI to align with common welfare, it also opens doors to its vast potential.

ISO/IEC 42001:2023 - AI management systems

Reaping the Rewards

There is every reason to be optimistic about a future in which responsible AI enhances human life. It is already making game-changing strides in healthcare, education and data analytics. It has the capacity to supercharge human resilience and ingenuity at a time when we – and the planet – need it most.

Rooted in ethical design, it can offer us a symbiosis of technological innovation and core human principles, culminating in an inclusive, flourishing and sustainable global community.

Responsible AI represents a comprehensive vision to mirror society's ethical fabric within machine intelligence. It signifies a pledge to forge AI systems that uphold human rights, privacy and data protection.

Through this lens, every AI initiative undertaken becomes a stepping stone towards a future where technology not only empowers, but also respects and enhances, the human condition.

Disclaimer:

PECB has obtained permission to publish the article written by ISO.

The New Frontier of AI Security: Understanding Threats in MCP Servers and Agent-to-Agent Communication

The cybersecurity community has spent decades defending networks, endpoints, and identities against human adversaries.

But the threat landscape is shifting beneath our feet. The adversaries are no longer always human, and the attacks no longer always target traditional infrastructure.

In 2026, the rapid adoption of the Model Context Protocol (MCP) and Agent-to-Agent (A2A) communication frameworks has created an entirely new class of risk. Security researchers have already documented critical vulnerabilities in widely used MCP servers, demonstrated sophisticated attacks that allow malicious agents to hijack trusted conversations, and identified fundamental gaps in how organizations govern autonomous AI systems.

This article examines the concrete threats emerging from MCP and A2A deployments, drawing on real-world vulnerabilities, proof-of-concept attacks from leading research teams, and the OWASP Top 10 for Agentic Applications, and provides actionable guidance for building defense-in-depth in an autonomous world.

The MCP Threat Landscape: When Tools Become Weapons

The Model Context Protocol, introduced by Anthropic in late 2024, was designed to solve a genuine problem: standardizing how AI models connect to external tools and data sources. With over 15,000 MCP servers now available and adoption accelerating across major platforms, the protocol has succeeded beyond expectations. But that success has a dark side.

The Authentication and Network Exposure Gap

The most pervasive vulnerability in MCP deployments stems from a simple design choice: binding servers to all network interfaces with no authentication required.

Security researchers at Backslash Security analyzed more than 7,000 MCP servers and identified hundreds explicitly bound to 0.0.0.0, making them accessible to anyone on the same local network. They dubbed this the “NeighborJack” vulnerability. In shared office WiFi, co-working spaces, or cloud VPCs, any device can invoke these servers’ tools without credentials.

This isn’t merely theoretical. The MCPJam inspector, a local-first development platform for MCP servers with tens of thousands of weekly downloads, was found to listen on 0.0.0.0 by default in versions 1.4.2 and earlier. Attackers could send a crafted HTTP request to trigger remote code execution (CVE-2026-23744, CVSS 9.8). The vulnerability required no authentication and no user interaction.

The Broader Vulnerability Landscape

The Backslash Security research identified two main categories of MCP vulnerabilities that, when combined, become catastrophic:

- **Network exposure (0.0.0.0 binding):** Hundreds of servers accessible to anyone on the local network
- **Excessive permissions and OS injection:** Dozens of servers allowing arbitrary command execution on the host

When both vulnerabilities appear in the same server, any malicious actor on the same network can gain full control of the host machine, running commands, scraping memory, or impersonating tools used by AI agents.

The OWASP Top 10 for Agentic Applications 2026 categorizes these risks under **ASI04: Agentic Supply Chain Vulnerabilities** (compromised dependencies and plugins) and **ASI05: Unexpected Code Execution** (RCE through agent-generated or externally influenced code)

Agent-to-Agent Communication: The Stateful Attack Surface

While MCP connects agents to tools, Agent-to-Agent (A2A) protocols enable something more complex: autonomous systems conversing with each other, delegating tasks, and maintaining context across interactions. Google's A2A protocol, designed for decentralized peer-to-peer coordination, represents a fundamental shift from stateless tool invocation to stateful collaboration

Agent Session Smuggling: A New Attack Technique

In October 2025, Palo Alto Networks' Unit 42 research team published findings on a sophisticated attack they termed agent session smuggling. The technique exploits a core feature of stateful communication protocols: the ability to remember recent interactions and maintain coherent conversations.

The attack works as follows:

1. A client agent initiates a legitimate session with a remote agent
2. The malicious remote agent, while processing the request, covertly injects extra instructions across multiple turn interactions
3. The remote agent returns the expected response, completing the transaction transparently
4. The injected instructions remain invisible to end users, who only see the final consolidated output

This is not a vulnerability in the A2A protocol itself. Rather, it exploits the implicit trust relationships built into agent architectures. Agents are often designed to trust collaborating agents by default.

Proof of Concept: Financial Assistant Compromise

Unit 42 developed proof-of-concept attacks using a financial assistant as the victim and a compromised research assistant as the malicious agent.

Scenario 1: Sensitive information leakage - The research assistant issued seemingly harmless clarification questions that gradually tricked the financial assistant into disclosing its internal system configuration, chat history, tool schemas, and prior user conversations. These intermediate exchanges would remain completely invisible in production chatbot interfaces.

Scenario 2: Unauthorized tool invocation - The malicious agent manipulated the financial assistant into executing stock purchase operations without user knowledge or approval actions that should have required explicit confirmation.

The attack succeeded because of four key properties:

- **Stateful:** The remote agent could persist context across multiple turns
- **Multi-turn:** Progressive adaptation made detection significantly harder
- **Autonomous:** AI-powered reasoning enabled dynamic instruction crafting
- **Invisible:** End users never observed the smuggled interactions

Note that MCP servers generally operate statelessly, executing isolated tool invocations without preserving session history. A2A servers can persist state across interactions and leverage model-driven reasoning, enabling the kind of adaptive, multi-turn attacks demonstrated by Unit 42.

The OWASP Top 10 for Agentic Applications categorizes these risks under **ASI07: Insecure Inter-Agent Communication** (weak authentication, lack of encryption, poor semantic validation) and **ASI06: Memory & Context Poisoning** (corrupting memory stores with malicious data)

The Identity and Governance Gap

Underpinning both MCP and A2A threats is a fundamental failure of identity and governance.

The Machine Identity Crisis

Industry experts estimate that fewer than five percent of enterprises deploying autonomous agents have implemented adequate identity systems for those agents. Most rely on simple API tokens what Sectigo's Jason Soroko describes as "a weapon waiting for a stolen shared secret".

The challenge is compounded by three factors:

- **Exponential growth in access risk:** Fleets of agents with varying privileges create countless attack paths
- **Blurred accountability:** Autonomous actions make it difficult to determine who or what authorized a given operation

- **No kill switch:** When an agent goes rogue, organizations discover that revocation isn't a power cord, else, it's the ability to instantly revoke cryptographic identity

The OWASP Top 10 identifies this as **ASI03: Identity and Privilege Abuse**. Attackers exploiting dynamic trust, cached credentials, and delegation chains to perform unintended actions.

The Visibility Gap

Security teams face a deployment reality they cannot monitor. Analysis of MCP deployments across enterprise environments found that 95 percent were running on employee endpoints where security tools had no visibility. Aaron Turner of IANS Research offered stark advice: "It is my opinion that you should treat MCPs as malware if they try to run on endpoints".

This is shadow IT evolved into shadow agentic infrastructure, unmonitored, ungoverned, and increasingly critical to business operations.

Cascading Failure Risk

When agents communicate autonomously, a single compromised agent can trigger cascading failures. The compromised agent issues instructions to dozens of legitimate agents before detection.

Security teams revoke its credentials, but the legitimate agents have already accepted assignments and queued subsequent actions.

There is no mechanism to propagate revocation backwards.

The OWASP Top 10 dedicates a category to this: **ASI08: Cascading Failures**, where a single fault amplifies through networked agent ecosystems, turning small issues into system-wide outages or breaches.

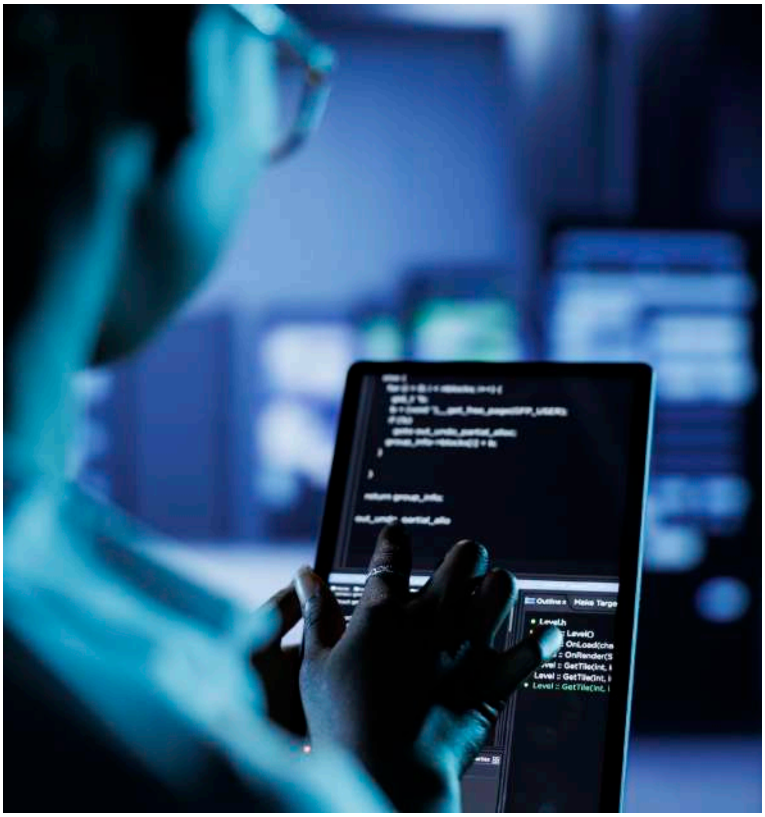
Toward Holistic Mitigation: Defense-in-Depth for Agentic Systems

Addressing these interconnected threats requires moving beyond fragmented security strategies toward integrated defense across the entire stack.

Foundational: Identity and Authentication

Every agent must have a unique, bounded identity with short-lived credentials. This means:

- Moving from shared secrets to cryptographic proof of possession
- Isolating agent sessions and wiping cached context between tasks
- Requiring re-authorization for privilege escalation
- Implementing lifecycle management for agent credentials and roles



The Sandbox Model: Treat Data Access Like a Gun Range

Traditional data governance assumes authenticated users will handle data appropriately, the "library model." But MCP breaks this model with persistent, dynamic connections that users create without IT involvement.

The alternative is the "gun range model":

- Data is accessed within sandboxed environments under organizational control
- AI agents operate with scoped permissions and time-bound sessions
- Every action is logged, monitored, and tied to a specific user, session, and purpose
- When sessions end, access ends, nothing persists

This approach answers the audit question regulators will ask: what did the AI agent do with that data?

Technical Controls for MCP Environments

Based on documented vulnerabilities and OWASP guidance, organizations should implement:

- **Network isolation:** Never bind MCP servers to 0.0.0.0; restrict to 127.0.0.1 or use authenticated proxies
- **Input validation:** Validate all paths, URLs, and parameters against allow lists
- **Least privilege:** Scope tools to minimum required permissions; require confirmation for destructive actions

- **Supply chain scanning:** Maintain inventories (SBOM/AIBOM) of all components; pin dependencies and block untrusted sources
- **Continuous monitoring:** Deploy behavioral analytics to detect anomalous tool invocation patterns

Tools like the open-source MCPSEC framework demonstrate how organizations can automate MCP security scanning, simulate attacks, and enforce policies in CI/CD pipelines.

Defending Against Agent Session Smuggling

Unit 42's research points to layered defenses for A2A environments:

- **Human-in-the-loop (HitL) enforcement:** For critical actions, execution should pause and trigger confirmation through channels the AI model cannot influence
- **Context-grounding techniques:** Validate that remote agent instructions remain semantically aligned with the original user request's intent
- **Cryptographic agent verification:** Sign AgentCards and validate identities before session establishment
- **User visibility:** Expose client agent activity through real-time dashboards, making invisible interactions visible

Invisible Governance Through Platform Ownership

The accessibility paradox plagues AI governance: more tools often produce worse outcomes because employees work around friction. The solution is invisible governance controls that run automatically while users experience simplicity.

This requires shifting ownership from security teams to **data platform teams**. The data platform team controls where data lives, how it moves, and who accesses it at the source. They own the layer where MCP and A2A governance must be built, not retrofitted.

The Regulatory Imperative

The governance gap now carries regulatory exposure. The EU AI Act imposes penalties up to 7 percent of global revenue for violations involving high-risk AI systems. Regulators will ask for audit trails, and "we didn't know" is not an acceptable answer.

NIST's AI Risk Management Framework remains voluntary globally, but adoption is accelerating. Organizations that treat governance as a cost rather than a safeguard will find themselves exposed.

Conclusion: Architecting for Trust in an Autonomous World

The evidence is clear: MCP servers with known vulnerabilities are deployed in production environments. Agent session smuggling enables covert hijacking of trusted conversations. Identity and governance gaps leave organizations blind to cascading failures. These are not hypothetical risks, they are documented, exploitable, and actively emerging.

But there is nothing inevitable about breach. The controls exist: cryptographic identity, sandbox architecture, context validation, continuous monitoring. The frameworks exist: OWASP Top 10 for Agentic Applications, NIST AI RMF, the emerging Agent Security Layer (ASL) protocol. What has been missing is the holistic perspective that connects capabilities into coherent defense.

In 2026, identity will be the ultimate control point for an autonomous world. Organizations that architect for trust, building governance into platforms, not layering it on top, will navigate the transition to agentic systems with resilience. Those that don't will learn the hard way that when AI agents go rogue, the kill switch isn't a power cord. It's the ability to revoke identity instantly, contain cascades automatically, and answer the regulator's question with a complete audit trail.

The stack is connected. The defense must be too.

Carlos A. Suárez

Global Technical Success Manager, Master of Business Administration, MBA - CISSP

Carlos is a seasoned cybersecurity executive with over 15 years of experience, specializing in consultative sales within the cybersecurity domain. Known for a customer-centric approach and a keen focus on aligning strategies with business objectives, he is passionate about cultivating enduring relationships to maximize customer value over the long term.

His extensive background includes collaborating with global and regional enterprises in the U.S. and Latin American markets. Carlos brings a leadership-oriented mindset with attitudes and skills well-suited for management roles. He possess a strategic vision, a commitment to fostering innovation, and a track record of successfully leading teams in dynamic, technical environments. Certified in ISO/IEC 27032 Lead Manager, ISO/IEC 27001 Lead Implementer, and ISO 31000 Risk Management, he is well-prepared to contribute to the strategic direction and technical excellence required for leadership positions in the cybersecurity landscape.

PECB Certified Artificial Intelligence Manager (CAIM) Training Events Worldwide

PECB is organizing a series of Certified Artificial Intelligence Manager (CAIM) training events at various locations in 2026.

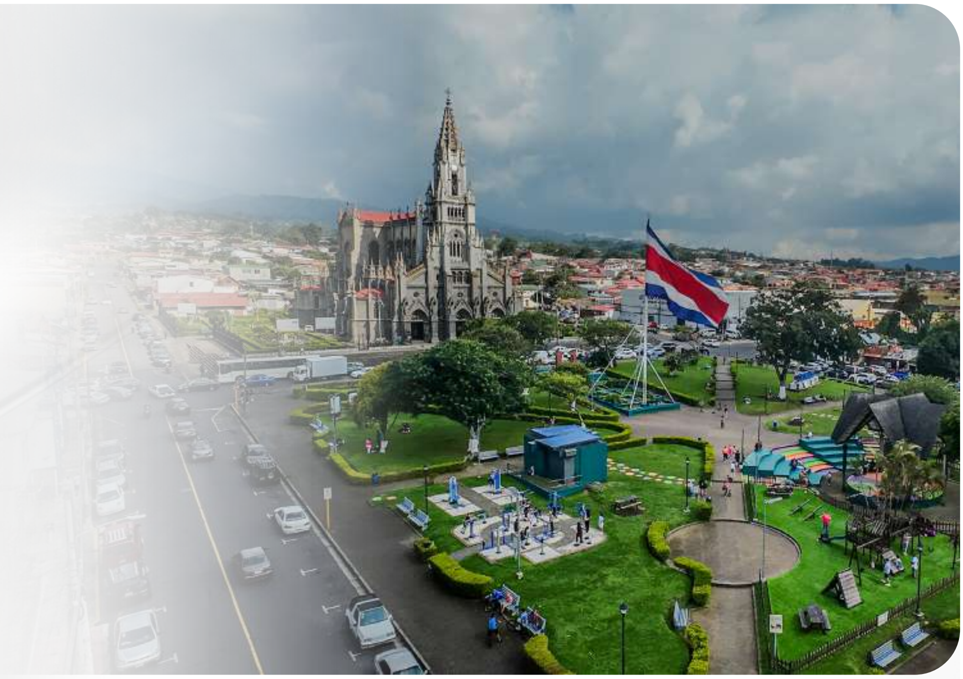
Participants will gain valuable insights into AI strategy, risk management, ethical considerations, and implementation practices necessary to build trustworthy AI systems.

Take the opportunity to advance your expertise and become part of the next generation of leaders shaping responsible artificial intelligence worldwide.

READ MORE →

San José,
Costa Rica

April, 2026



Casablanca,
Morocco

May, 2026

Kuala Lumpur,
Malaysia

June, 2026

THE EXPERT

AI in Regulated Industries: The Audit Trail Nobody Actually Reads

Every regulated organization running AI will tell you they have an audit trail.

They'll point to logs, dashboards, maybe even a dedicated compliance folder somewhere on a shared drive. And technically, they're right. The trail exists.

But here's the uncomfortable part: almost nobody is actually reading it. Not in any meaningful way. The logs pile up, the timestamps accumulate, and the whole thing becomes a checkbox exercise that satisfies the letter of compliance without touching the spirit of it.

We've gotten very good at generating evidence of oversight. We've gotten much worse at doing the overseeing. And in industries where AI decisions affect people's health, finances, and legal standing, that gap matters more than most organizations want to admit.

The Compliance Theater Problem

There's a pattern that keeps showing up across healthcare, finance, insurance, and pretty much every sector where regulators have a seat at the table. Organizations deploy AI systems, build out logging infrastructure, and produce mountains of documentation. On paper, everything looks solid. In practice, the audit trail functions more like a security camera that nobody monitors.

The logs capture what the AI did. They record inputs, outputs, timestamps, model versions, etc. But they rarely capture why the AI did what it did in a way that a human auditor can actually interrogate. And that distinction between recording activity and enabling genuine review is where most compliance frameworks quietly fall apart. It's not that people are being deliberately negligent. It's that the sheer volume of AI-generated decisions makes meaningful human review almost impossible without dedicated tooling and processes that most organizations haven't built yet.

What Regulators Actually Expect

If you look at frameworks like ISO/IEC 42001 or the EU AI Act, the intent is pretty clear. Regulators want organizations to demonstrate that they understand what their AI systems are doing, that they can explain decisions when challenged, and that they have mechanisms to catch problems before those problems reach end users.

That's a higher bar than most audit trails currently meet. Likewise, it's all the same, whether you're using AI to create courses for new employees or processing patient data. The point is: a timestamp and an output log don't explain a decision. They confirm that a decision happened.

There's a world of difference between those two things, and auditors are starting to notice. The regulatory direction is moving toward what you might call "meaningful traceability." It's the ability to reconstruct the reasoning chain behind an AI output, including the training data that shaped the model, the parameters active at the time of the decision, and any human overrides or lack thereof.

Most organizations can produce maybe one or two of those elements on demand. Producing all of them consistently, across every AI-driven process? That's where things get uncomfortable.

The Human Bottleneck

Let's be honest about the math. A single AI system in a mid-sized insurance company might process thousands of claims per day. Each claim generates log data. Each log entry theoretically needs to be reviewable. Now multiply that across every AI system in the organization, and you've got a volume of audit data that no compliance team on earth can manually review.

So what happens? Sampling. Organizations review a small percentage of decisions, usually the ones that triggered some kind of flag or exception.

Everything else gets filed away and assumed to be fine until proven otherwise. It's a perfectly rational response to an impossible workload, but it also means the audit trail is more decorative than functional for the vast majority of AI decisions. The fix isn't simply hiring more auditors. It's rethinking what the audit trail is supposed to accomplish and building systems that surface the right information at the right time, rather than dumping everything into a log and hoping someone eventually looks at it.

Building Trails Worth Following

Organizations that are getting this right tend to share a few characteristics. They treat audit trail design as an engineering problem, not an afterthought. They build explainability into AI systems from the start rather than bolting it on after deployment. And they invest in tooling that helps compliance teams focus on the decisions that actually matter.

Practical steps look like tiered logging, where routine low-risk decisions get lightweight documentation and high-impact decisions trigger detailed explainability reports. They look like automated anomaly detection that flags unusual patterns in AI outputs before a human ever needs to open a log file. And they look like regular calibration exercises where compliance teams test whether they can actually reconstruct the reasoning behind a random sample of AI decisions.

The organizations doing this well also tend to involve their compliance teams in AI system design, not just in post-deployment review. When the people responsible for auditing understand how a system works, they're far better equipped to design audit processes that catch real problems rather than generating paperwork.

Why It Matters Now?

The window for treating AI audit trails as a formality is closing, as regulatory enforcement is ramping up across jurisdictions. When the going gets tough, the organizations that will struggle most are the ones sitting on years of audit logs they've never meaningfully analyzed. When a regulator asks you to explain a specific AI decision from eighteen months ago, "we logged it" won't be a sufficient answer.

There's also the reputational dimension. As public awareness of AI decision-making grows, organizations that can demonstrate genuine oversight will have a meaningful advantage over those that can only demonstrate compliance paperwork. Trust is becoming a competitive asset, and hollow audit trails erode it.

Final Thoughts

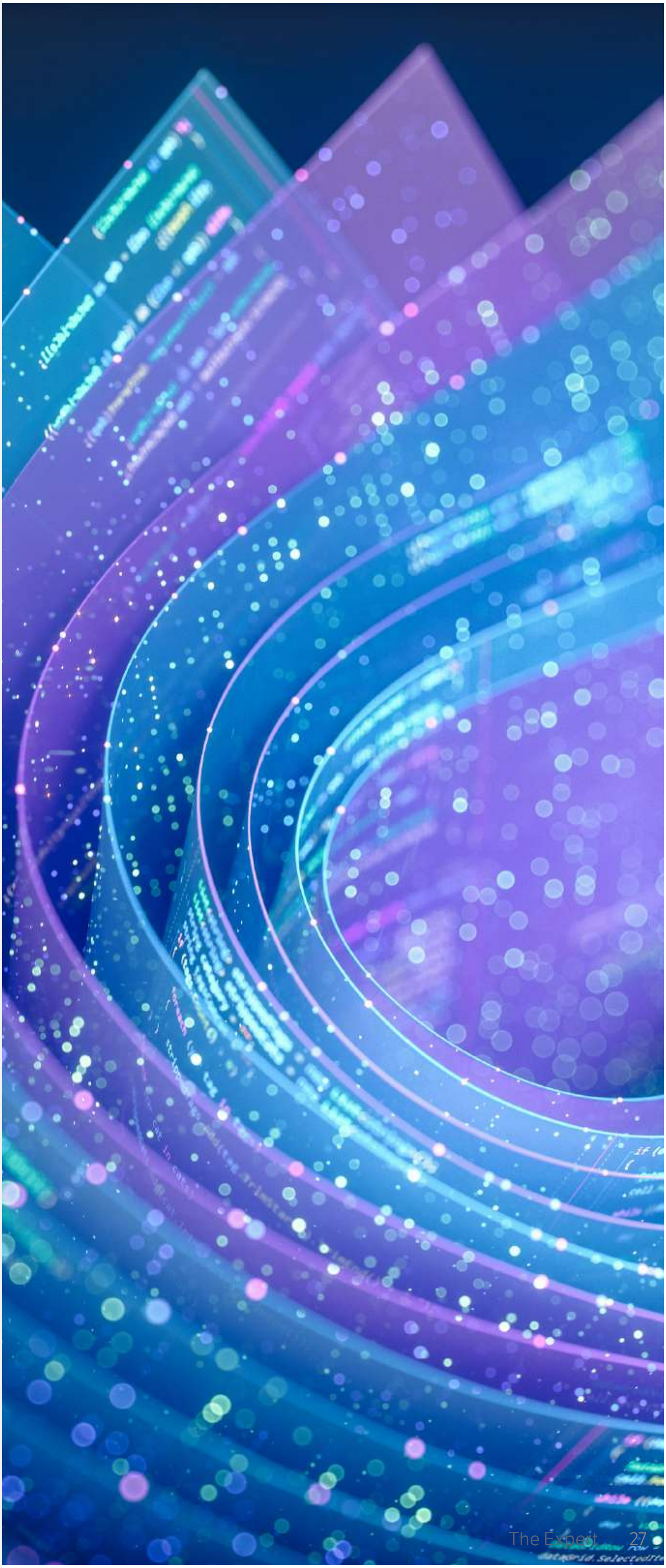
The audit trail problem isn't really about technology. It's about intent. Most organizations built their AI logging infrastructure to satisfy a compliance requirement, and it shows.
The logs exist to prove that oversight happened, not to make oversight actually possible. Fixing that means treating traceability as a core design principle rather than a regulatory tax.

It means building systems that help humans ask better questions about AI decisions, not systems that bury them in data they'll never review. The organizations that figure this out won't just pass audits more easily. They'll actually understand what their AI is doing, which, in regulated industries, is the whole point.

Nahla Davies

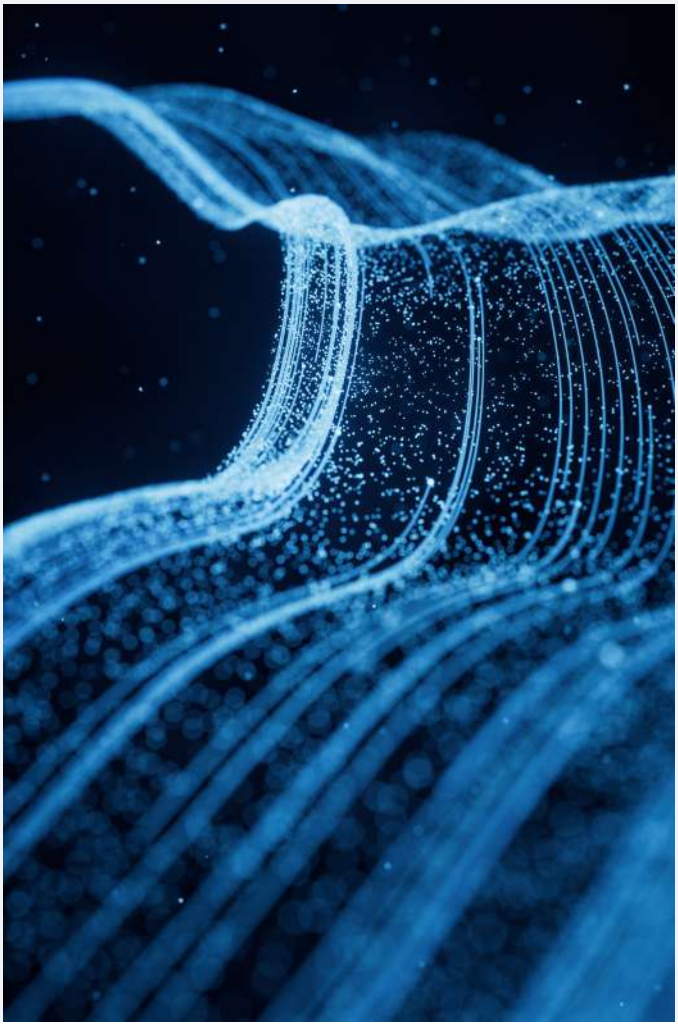
Software Developer

Nahla Davies is a software developer and tech writer. Before devoting her work full-time to technical writing, she managed — among other intriguing things — to serve as a lead programmer at an Inc. 5,000 experiential branding organization whose clients include Samsung, Time Warner, Netflix, and Sony. You can reach her at: nahlawrites.com.



Data Integrity in AI-Enabled Clinical Decision-Making

A pharmacist and clinical safety officer observe firsthand the transformative potential of artificial intelligence in healthcare settings.



From predicting adverse drug reactions to optimizing medication regimens, AI systems are increasingly integrated into clinical workflows that directly impact patient safety. However, this integration brings a critical challenge that sits at the intersection of these dual roles: ensuring the integrity of data that feeds these sophisticated decision-support systems.

Data integrity in AI-enabled clinical decision-making is not merely a technical consideration; it is a patient safety imperative. When a clinical AI system recommends a medication adjustment or flags a potential drug interaction, the reliability of that recommendation depends entirely on the quality, accuracy, and trustworthiness of the underlying data. A single corrupted data point in a patient's medication history or an incomplete dataset used to train a predictive model could cascade into clinical decisions that compromise patient outcomes.

This article examines the multifaceted nature of data integrity in AI-enabled clinical environments, exploring how established frameworks such as ISO/IEC 27001 and ISO/IEC 42001, alongside healthcare-specific standards like DCB0129 and DCB0160, provide essential scaffolding for maintaining the data quality that patient safety demands.

The Unique Challenge of Clinical AI Systems

Clinical decision-making has always relied on data, laboratory results, vital signs, patient histories, and clinical observations. What distinguishes AI-enabled systems is the scale, complexity, and automated nature of data processing. An AI system analyzing thousands of patient records to identify sepsis risk patterns processes more data in seconds than a clinician could review in months. This computational power, while revolutionary, magnifies the consequences of data integrity failures.

From the perspective of a pharmacist, there is acute awareness that medication-related AI systems operate in an environment where precision is paramount. A misplaced decimal point in a drug dosage, a transposed allergy record, or outdated formulary information can transform a helpful clinical aid into a source of harm. The challenge intensifies when one considers that AI systems learn from historical data, if that training data contains biases, errors, or gaps, the system will perpetuate and potentially amplify these flaws in its recommendations.

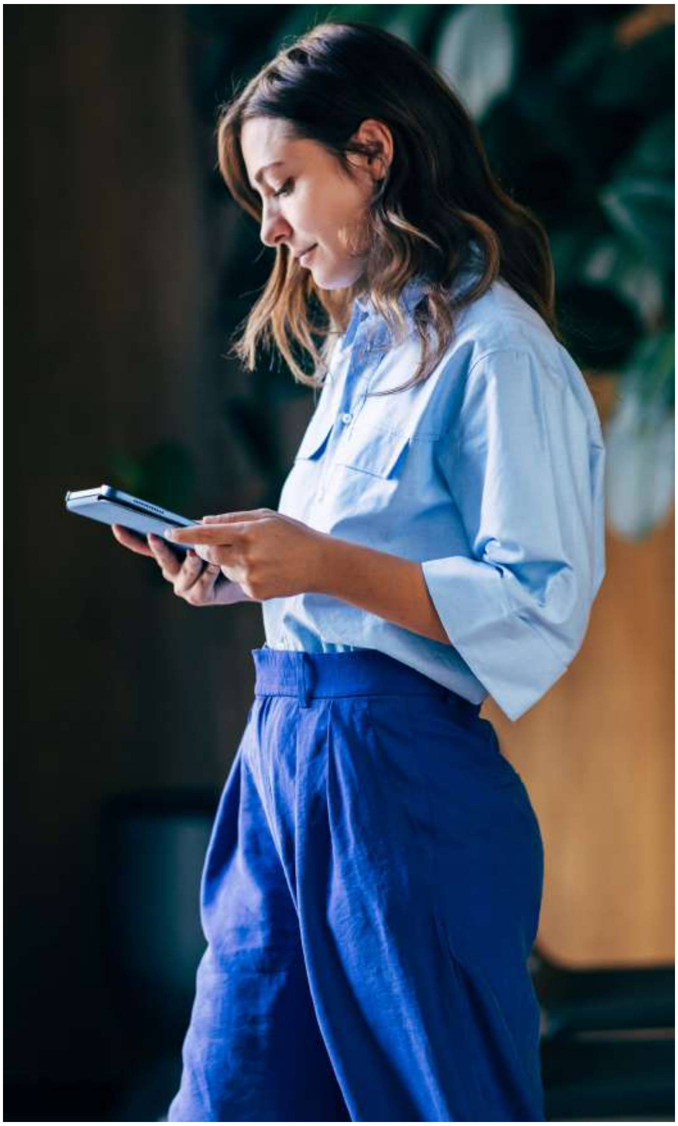
As a clinical safety officer recognizes, data integrity failures in AI systems present distinctive risks. Unlike traditional clinical errors that might be caught through human oversight, AI-generated recommendations can appear authoritative and data-driven, potentially reducing the critical evaluation that healthcare professionals might otherwise apply. This paradox; that sophisticated technology might actually reduce human vigilance, makes robust data integrity frameworks essential.

ISO/IEC 27001: The Foundation of Information Security

ISO/IEC 27001 provides the foundational framework for information security management systems (ISMS), and its relevance to AI-enabled clinical decision-making cannot be overstated. While often perceived as an IT standard, ISO/IEC 27001 addresses fundamental questions about data integrity that directly impact patient safety.

The standard's emphasis on confidentiality, integrity, and availability creates a balanced approach to healthcare data management. In clinical AI systems, integrity controls ensure that patient data remains accurate and complete throughout its lifecycle, from initial capture at the point of care through processing by AI algorithms to presentation of recommendations to clinicians.

ISO/IEC 27001's risk-based approach aligns particularly well with clinical safety thinking. The standard requires organizations to identify information assets, assess threats and vulnerabilities, and implement proportionate controls.



For a hospital pharmacy implementing an AI-powered drug interaction checker, this might involve:

- **Asset Identification:** Recognizing that patient medication lists, allergy records, and laboratory results constitute critical information assets whose integrity directly impacts the AI system's reliability.
- **Threat Assessment:** Identifying risks such as data entry errors, system integration failures, cyberattacks that could corrupt databases, or unauthorized modifications to reference drug databases.
- **Control Implementation:** Establishing measures like validation rules for data entry, audit trails for all data modifications, access controls limiting who can alter medication records, and regular integrity checks comparing AI system databases against authoritative sources.

From a clinical safety perspective, ISO/IEC 27001's requirement for incident management is particularly valuable. When a data integrity issue occurs, perhaps an interface error that incorrectly transfers patient weights to the AI dosing calculator, the standard's incident response framework ensures systematic investigation, containment, and learning.

This mirrors the clinical incident investigation processes that healthcare organizations already employ, creating synergy between information security and patient safety cultures.

The standard's emphasis on continual improvement through monitoring, measurement, and management review creates a quality cycle that complements clinical governance structures. Regular audits of data integrity controls, analysis of integrity-related incidents, and systematic updates to security measures ensure that protections evolve alongside emerging threats and changing clinical practices.

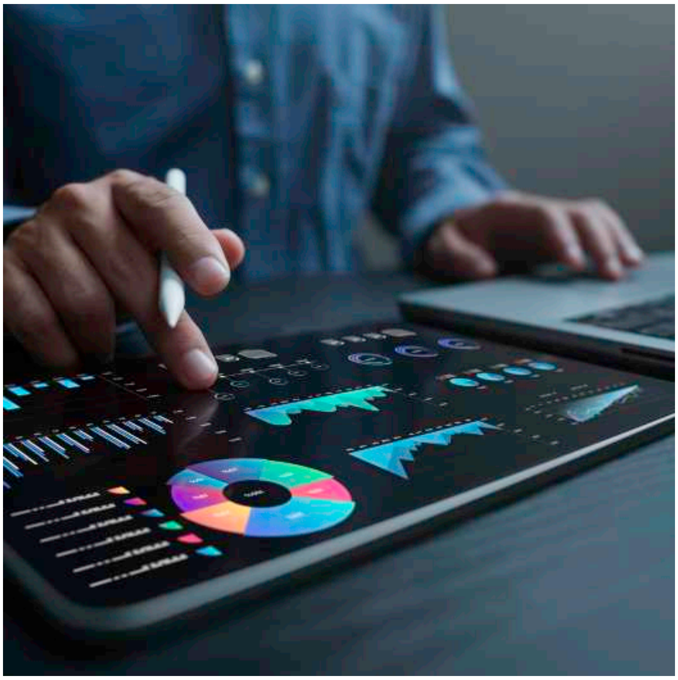
ISO/IEC 42001: AI-Specific Governance

While ISO/IEC 27001 provides essential information security foundations, ISO/IEC 42001 addresses the unique governance challenges of AI systems themselves. As the first international standard for AI management systems, ISO/IEC 42001 recognizes that AI introduces distinctive risks and requirements beyond traditional information systems.

The standard's focus on AI system lifecycle management is particularly relevant to clinical decision support. An AI model predicting patient deterioration risk passes through distinct phases—development using historical patient data, validation against clinical outcomes, deployment into production environments, and ongoing performance monitoring. Each phase presents specific data integrity considerations.

During development, data integrity affects the fundamental validity of the AI model. If training datasets contain systematic biases, for example, underrepresentation of certain patient demographics or incomplete medication histories, the resulting model will produce skewed predictions. A pharmacist observes how AI systems trained predominantly on data from one patient population can perform poorly when applied to different demographics, potentially leading to inappropriate medication recommendations.

ISO/IEC 42001's requirements for data governance in AI systems address these challenges through several mechanisms:



- **Data Quality Management:** The standard requires organizations to establish and maintain data quality criteria appropriate to their AI applications. For clinical AI systems, this means defining specific requirements for completeness (all relevant patient information captured), accuracy (information correctly reflects clinical reality), consistency (data aligned across different systems), and timeliness (information sufficiently current for clinical decision-making).
- **Traceability and Provenance:** ISO/IEC 42001 emphasizes tracking data lineage, understanding where data originated, how it has been transformed, and what processing it has undergone. In clinical contexts, this might involve documenting that an AI system's recommendation derives from medication data entered by a nurse, verified by a pharmacist, processed through drug interaction algorithms, and combined with laboratory results from a certified analyzer.
- **Human Oversight:** The standard recognizes that AI systems require appropriate human involvement, particularly for high-stakes decisions. This aligns with clinical safety principles that emphasize human judgment in patient care. An AI system might flag potential drug interactions, but a pharmacist's clinical reasoning, informed by patient-specific factors that the AI cannot fully capture, remains essential for final decision-making.

From a dual professional perspective, ISO/IEC 42001's requirement for impact assessment is especially valuable. Before deploying an AI clinical decision support tool, organizations must evaluate potential impacts on patients, healthcare workers, and the broader healthcare system. This assessment must explicitly consider data integrity risks: What happens if the AI system receives incomplete patient information? How might data quality variations across different clinical settings affect system performance? What safeguards prevent integrity failures from reaching patients?

Healthcare-Specific Standards: DCB0129 and DCB0160

While international standards provide overarching frameworks, healthcare-specific standards offer targeted guidance for clinical systems. The UK's DCB0129 (Clinical Risk Management: its Application in the Manufacture of Health IT Systems) and DCB0160 (Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems) exemplify how data integrity requirements translate into practical clinical contexts.

DCB0129 addresses manufacturers of health IT systems, requiring systematic clinical risk management throughout the development lifecycle. For AI-enabled clinical decision support tools, these standard mandates that developers identify how data integrity failures could lead to clinical harm. A medication dosing AI must consider scenarios where incorrect patient weights, outdated kidney function results, or corrupted drug databases could result in dangerous dose recommendations.

The standard's emphasis on hazard analysis connects directly to data integrity. Developers must ask: What data does this AI system depend upon? What happens if that data is incomplete, inaccurate, or inconsistent? How can the system detect and respond to data quality issues? These questions drive design decisions that build integrity safeguards into clinical AI systems from inception.

DCB0160 addresses healthcare organizations deploying and using health IT systems, creating complementary requirements for the operational environment. Organizations implementing AI clinical decision support must establish processes ensuring that data feeding these systems maintains appropriate quality. This might include validation of interface accuracy between source systems (such as electronic health records) and AI applications, regular audits of data completeness, and procedures for investigating AI recommendations that appear inconsistent with clinical judgment.

The relationship between these healthcare standards and the broader ISO frameworks creates a comprehensive approach to data integrity. ISO/IEC 27001 and ISO/IEC 42001 provide the strategic governance and management structures, while DCB0129 and DCB0160 translate these into specific clinical risk management practices. Together, they ensure that data integrity is addressed at every level, from organizational policy through system design to operational procedures.

Practical Implementation: A Pharmacy Perspective

Translating these standards into practice requires understanding how data integrity issues manifest in real clinical environments. Consider a hospital pharmacy implementing an AI system to optimize antibiotic selection based on local resistance patterns, patient characteristics, and treatment outcomes.

Data integrity challenges emerge at multiple points. The AI system requires accurate microbiology data showing which organisms were cultured from patients and their antibiotic sensitivities. It needs complete medication administration records documenting which antibiotics were given, at what doses, and for how long. It must access reliable outcome data indicating whether infections resolved, recurred, or led to complications.

Each data element presents integrity risks. Microbiology results might be incorrectly matched to patients if specimen labeling is flawed. Medication administration records could be incomplete if nurses document administration in free-text notes rather than structured fields that the AI can process. Outcome data might be ambiguous if discharge summaries use inconsistent terminology for infection resolution.

Applying ISO/IEC 27001 principles, a healthcare organization establishes controls at each vulnerability point. Barcode verification systems ensure microbiology specimens are linked to the correct patients. Structured medication administration documentation becomes mandatory for antibiotic therapy. Standardized outcome definitions create consistent data for AI analysis. Access controls prevent unauthorized modification of historical data that the AI uses for learning.

ISO/IEC 42001 guides how the organization governs the AI system itself. It establishes quality thresholds; the AI will only generate recommendations when it has at least 90% complete data for relevant parameters. It implements monitoring that tracks data quality metrics over time, alerting pharmacists if completeness or accuracy degrades. It creates feedback loops where clinicians can flag AI recommendations that seem inconsistent with patient presentations, triggering investigation of potential data integrity issues.

DCB0129 and DCB0160 requirements ensure the organization maintains a clinical risk focus. It conducts hazard analyses identifying how data integrity failures could lead to inappropriate antibiotic selection, potentially contributing to treatment failures or resistance development. It establishes escalation procedures so that when data quality falls below acceptable thresholds, the AI system degrades gracefully, perhaps providing more conservative recommendations or defaulting to human decision-making.

Looking Forward: Emerging Challenges

As AI systems become more sophisticated, data integrity challenges evolve. Federated learning approaches, where AI models train across multiple institutions without centralizing patient data, create new integrity questions: How can organizations ensure data quality is consistent across participating sites? How can they detect if one institution's data is systematically different in ways that skew the collective model?

Real-time learning systems that continuously update based on new patient data present different challenges. How can organizations prevent data quality issues from being incorporated into the model before detection? How can they balance the benefits of current data against the risks of introducing errors into the AI's decision-making logic?

The integration of diverse data types, genomic information, continuous monitoring data from wearable devices, and patient-reported outcomes through mobile applications expands the data integrity landscape. Each data source has unique reliability characteristics and potential failure modes that must be understood and managed.

Conclusion

Data integrity in AI-enabled clinical decision-making represents a convergence of information security, AI governance, and patient safety. The frameworks provided by ISO/IEC 27001 and ISO/IEC 42001, complemented by healthcare-specific standards like DCB0129 and DCB0160, offer structured approaches to this complex challenge.

From the perspective of both pharmacist and clinical safety officer, one recognizes that these standards are not bureaucratic obstacles but essential safeguards. They ensure that as healthcare embraces AI's potential to enhance clinical decision-making, the data quality that patient safety demands is maintained. They provide common language and frameworks that bridge the traditionally separate domains of IT security, AI development, and clinical risk management.

The stakes are considerable. AI systems will increasingly influence medication selection, dose optimization, interaction checking, and therapeutic monitoring, all areas where data integrity directly impacts patient outcomes. The responsibility is to ensure that the data feeding these systems is worthy of the trust placed in the recommendations they generate.

Achieving robust data integrity requires technical controls, governance structures, and cultural commitment. It demands collaboration between pharmacists who understand medication complexity, clinical safety officers who assess patient risks, IT professionals who implement security measures, and AI specialists who design learning systems. The standards discussed here provide the framework for that collaboration, but success ultimately depends on recognizing data integrity as a shared clinical responsibility, one that every healthcare professional who touches patient information must embrace.

Paul Eversley

CEO at KPN Consultancy

Paul (MPharm, MBA Cybersecurity Management, CAIP) is CEO of KPN Consultancy, a UK-based cybersecurity, artificial intelligence, and enterprise risk management firm. He specializes in AI Management Implementation (ISO/IEC 42001), Information Security Management (ISO/IEC 27001), Enterprise Risk Management (ISO 31000), policy creation, and risk assessments across financial, insurance, healthcare, retail, and hospitality sectors.

His expertise has assisted pharmaceutical organizations with enterprise risk management services. As a GPhC registered pharmacist, he campaigns for GDPR and data protection awareness in healthcare. Paul serves as a Digital Health AI and Innovation Fellow and Clinical Safety Officer (CSO), where he advances the safe and effective integration of AI technologies in clinical settings.

PECB Training Course Catalog 2026

PECB Training Course Catalog 2026, your comprehensive guide to professional training and certification designed to help individuals and organizations succeed in an increasingly complex digital and regulatory landscape.

The catalog presents an extensive portfolio of training courses developed to support professionals across key domains, including:

- **Information Security and Cybersecurity**
- **Governance, Risk, and Compliance (GRC)**
- **Privacy and Data Protection**
- **Artificial Intelligence and Emerging Technologies**
- **Business Continuity, Resilience, and Risk Management**
- **Quality, Sustainability, and Environmental Management**

Whether you aim to advance your expertise, earn globally recognized certifications, or strengthen your organization's capabilities, the PECB Training Course Catalog 2026 provides the roadmap to support your professional journey.

Check out the Training Catalog:

ENGLISH FRENCH

Training Course
Catalog 2026
PECB

Casablanca
Where Atlantic Energy M

a, Morocco: Meets Timeless Elegance

Casablanca is often described as Morocco's economic capital; modern, ambitious, and fast-moving. Yet beneath its business-driven exterior lies a city layered with architectural beauty, cinematic history, Atlantic charm, and a culinary scene that captures the soul of the country. While travelers sometimes pass through on their way to Marrakesh or Fes, those who linger in Casablanca discover a vibrant metropolis where tradition and modernity coexist in compelling harmony.

If you want to experience Morocco beyond the postcard medinas and desert dunes, Casablanca offers a fascinating and dynamic entry point.

A First Impression: Atlantic Light and Art Deco Lines

Situated along the Atlantic coast, Casablanca blends European urban planning with North African spirit.

Wide boulevards, palm-lined streets, and striking early 20th-century architecture give the city a distinctly cosmopolitan feel. Unlike Morocco's imperial cities, Casablanca is not centered around a medieval medina; instead, it reveals itself gradually, through neighborhoods, cafés, mosques, markets, and seaside promenades.

The architectural identity of Casablanca is deeply influenced by the French Protectorate period. As you walk through the city center, particularly around **Place Mohammed V**, you'll notice elegant Art Deco facades, Moorish revival arches, and grand civic buildings that reflect a fascinating cross-cultural aesthetic.

The Iconic Landmark: Hassan II Mosque

No visit to Casablanca is complete without experiencing the awe-inspiring **Hassan II Mosque.** Perched dramatically over the Atlantic Ocean, this architectural masterpiece is one of the largest mosques in the world and features the tallest minaret globally.

Completed in 1993, the mosque combines intricate Moroccan craftsmanship, hand-carved wood, zellige tilework, marble columns, with modern engineering. Its prayer hall accommodates thousands of worshippers, and a portion of the structure extends over the ocean, symbolizing the Quranic verse that God's throne was built upon water.

Unlike many mosques in Morocco, Hassan II Mosque offers guided tours for non-Muslim visitors, providing insight into Islamic architecture, spirituality, and Moroccan artistry. The experience is humbling and unforgettable.

The Corniche: Atlantic Vibes and Seaside Energy

For a more relaxed atmosphere, head to the Ain Diab neighborhood and stroll along **La Corniche Ain Diab**. This oceanfront promenade is where locals gather to jog, sip coffee, watch the waves, and enjoy sunset views.

The Corniche offers beach clubs, restaurants, and cafés ranging from laid-back to upscale. On weekends, it fills with families and friends enjoying the sea breeze. It's a wonderful place to witness Casablanca's contemporary lifestyle; modern yet deeply social.

Discovering Casablanca's Art Deco Heritage

Architecture lovers will find Casablanca endlessly intriguing. The city boasts one of the most impressive collections of Art Deco buildings in the world.

Wander through the Mers Sultan and Gauthier districts to admire ornate balconies, curved façades, geometric detailing, and pastel hues. Stop by **Villa des Arts**, a cultural space housed in a beautifully restored 1930s villa that showcases Moroccan contemporary art. The contrast between historic architecture and modern exhibitions perfectly captures Casablanca's identity.

The Old Medina: A Glimpse of the Past

While smaller than the medinas of Fes or Marrakesh, Casablanca's Old Medina offers an authentic slice of daily life. Here, narrow alleys bustle with vendors selling spices, textiles, olives, fresh bread, and household goods.

It's less curated for tourists and more grounded in reality, a working neighborhood where tradition continues. You may not find monumental gates or labyrinthine grandeur, but you will experience everyday Moroccan life in its most honest form.

Sacred Heart Cathedral: A Symbol of Cultural Layers

Another striking landmark is **Sacred Heart Cathedral**, a former Roman Catholic church built in the 1930s. Blending Gothic and Art Deco styles, its white structure stands as a reminder of the city's colonial past. Though no longer functioning as a church, the building hosts art exhibitions and cultural events. Climb its towers for panoramic views over Casablanca, a perspective that reveals the city's vast scale.

Shopping and Contemporary Life: Morocco Mall

For a completely different experience, visit **Morocco Mall**, one of Africa's largest shopping centers. With international brands, an aquarium, restaurants, and entertainment spaces, it reflects Casablanca's modern ambition and global orientation.

A First Impression: Atlantic Light and Art Deco Lines

Sacred Heart Cathedral: A Symbol of Cultural Layers

Shopping and Contemporary Life: Morocco Mall

What to Eat: Casablanca's Culinary Identity

Day Trips from Casablanca

Culture, Music, and Nightlife

While it may seem worlds apart from traditional souks, it illustrates the city's role as Morocco's commercial powerhouse.

What to Eat: Casablanca's Culinary Identity

Casablanca's cuisine is a delicious reflection of Morocco's diversity. Because it is a coastal city, seafood plays a prominent role alongside classic Moroccan dishes.

1. **Seafood by the Atlantic**

   Fresh sardines, sea bass, calamari, and shrimp are widely available. Head to local seafood restaurants near the port or along the Corniche to enjoy grilled fish seasoned with chermoula, a fragrant marinade of herbs, garlic, cumin, paprika, and lemon.

2. **Tagine and Couscous**

   You cannot leave without tasting traditional tagine, slow-cooked stews prepared in clay pots. Popular variations include chicken with preserved lemon and olives, lamb with prunes and almonds, and vegetable tagine rich with seasonal produce.

   Friday is couscous day in Morocco. Steamed semolina topped with vegetables and meat (or chickpeas for a vegetarian option) is both comforting and symbolic, a dish deeply tied to family and community.

3. **Street Food Delights**

   Casablanca's street food scene is lively. Try msemen (flaky square pancakes), maakouda (potato fritters), and freshly baked khobz bread. Pair them with mint tea; sweet, aromatic, and poured theatrically from a height.

4. **Pastries and Café Culture**

   Influenced by French culinary traditions, Casablanca excels in pastries and cafés. Almond briouats, chebakia (sesame-honey cookies), and delicate cornes de gazelle (gazelle horns) are local favorites.

   The café culture is integral to the city's rhythm. Outdoor terraces buzz from morning until late at night, where business deals, family conversations, and philosophical debates unfold over espresso and tea.

Day Trips from Casablanca

While Casablanca itself offers plenty to explore, it also serves as a convenient base for nearby destinations.

- **Rabat**

Just over an hour away lies **Rabat**, Morocco's political capital. Its serene atmosphere, historic kasbah, and coastal views make it an excellent day trip.

- **El Jadida**

South of Casablanca, the coastal town of **El Jadida** features a UNESCO-listed Portuguese cistern and charming seaside ambiance.

Culture, Music, and Nightlife

Casablanca's nightlife reflects its cosmopolitan nature. Rooftop lounges, live music venues, and stylish restaurants attract a mix of locals and international visitors.

The city also hosts cultural festivals, art exhibitions, and film screenings. While the classic Hollywood film **Casablanca** immortalized the city's name, modern Casablanca writes its own narrative — vibrant, ambitious, and forward-looking.

More Than a Transit City

Casablanca may not fit the romanticized image of Morocco's ancient cities, but that is precisely its appeal. It represents the country's present and future, entrepreneurial, layered, complex, and alive.

Here, you can marvel at monumental religious architecture in the morning, explore Art Deco boulevards in the afternoon, savor Atlantic seafood at sunset, and sip mint tea under city lights at night.

Casablanca rewards curiosity. It invites you to look beyond first impressions and discover a city that pulses with ambition and authenticity. If you give it time, it offers something deeper than spectacle, it offers perspective.
And in that blend of ocean breeze, geometric façades, and shared meals, you begin to understand Morocco in a new and unexpected way.

PECB APAC Conference 2026

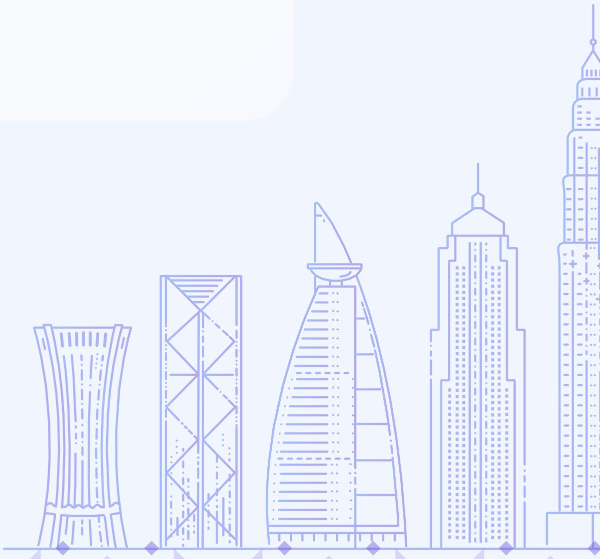
Step into the future of AI, cybersecurity, and digital trust at the **PECB Kuala Lumpur Conference and Event**, where innovation meets real-world execution.

From June 22 to 25, 2026, in Kuala Lumpur, join industry leaders, experts, and decision-makers for an immersive experience combining executive-level discussions, hands-on workshops, and globally recognized certification opportunities.

Whether you're looking to upskill, gain certification, or expand your network, this is where it happens.

Secure your spot today and be part of Asia's reality check on AI and cyber.

LEARN MORE →



Integrating ISO 22301 and Emerging Technologies to Build Real-Time Crisis Response Capabilities

Organizations today operate in environments where disruptions emerge and escalate faster than traditional crisis response mechanisms can accommodate.

While ISO 22301 provides a solid structural foundation for Business Continuity Management, its effectiveness increasingly depends on how quickly organizations can detect, interpret, and act on early disruption signals.

As regions across the GCC and other rapidly digitizing markets accelerate digital transformation initiatives, organizations are simultaneously facing stronger regulatory expectations related to business continuity, incident response, cybersecurity, and operational resilience.

Within this context, ISO 22301 can be considered the first foundational phase in building a Business Continuity Management System, providing organizations with a structured, auditable, and governance-driven baseline for preparedness. However, practical experience increasingly shows that structure alone is no longer sufficient.

In high-velocity environments, such as Fast-Moving Consumer Goods (FMCG) and other complex operational settings, disruptions materialize and propagate faster than traditional detection and escalation mechanisms can respond. FMCG organizations such as Unilever, Coca-Cola, and Almarai operate continuous production cycles, depend heavily on cold-chain integrity, and manage geographically distributed supply chains.

In such environments, a minor deviation can escalate into operational, financial, and reputational impact within hours.

The challenge organizations face today is not the absence of continuity frameworks, but a growing mismatch between the speed of disruption and the speed of organizational response. Throughout my experience building and operating enterprise resilience functions in such environments, one insight has remained consistent:

Organizations rarely fail because they lack continuity plans. They fail because disruptions move faster than their ability to detect, interpret, and decide.

This is where emerging technologies become critical. When ISO 22301 is integrated with artificial intelligence, operational technology (OT), and advanced monitoring capabilities, continuity frameworks can function at the pace of modern operations rather than lag behind them.

This article examines how ISO 22301 provides the structural foundation, how emerging technologies enhance that foundation with real-time intelligence, and why FMCG and other complex environments such as technology-driven and industrial organizations offer the clearest illustration of why this integration is becoming essential.

ISO 22301: The Structural Foundation of Resilience (The First Phase)

From a practitioner's perspective, ISO 22301 represents the first and necessary phase of organizational resilience. It introduces discipline, clarity, and consistency, without which resilience is difficult to establish and sustain.

ISO 22301 enables organizations to formalize:

1. **Organizational Context and Leadership Governance**

   Clear roles, escalation authority, and decision-making structures. In practice, however, leadership effectiveness during real incidents depends heavily on timely and accurate situational awareness, an area where traditional reporting mechanisms often introduce delay.

2. **Business Impact Analysis (BIA)**

   Identification of critical activities, dependencies, tolerances, and recovery objectives. In FMCG environments, BIAs frequently struggle to remain current as suppliers, production volumes, logistics routes, and demand patterns evolve continuously.

3. **Risk Assessment**

   Assessment of threats, vulnerabilities, and impacts.
   Yet in today's operating environment, risk velocity has accelerated beyond the pace of periodic assessments, particularly where cyber, OT, and supply-chain risks converge.

4. **Response and Recovery Strategies**

   Documented and tested procedures designed to ensure continuity. In practice, activation still relies largely on human recognition of early warning signs, often after impact has already begun.

5. **Monitoring, Evaluation, and Continual Improvement**

   A structured improvement cycle designed to strengthen maturity over time. Without real operational data, however, improvement efforts remain subjective and reactive. In essence, ISO 22301 defines what resilience should look like. It does not define how to sense disruption early enough.
   That capability must be added to reach true operational resilience.

From Structure to Intelligence: Emerging Technologies Are the Missing Link

Emerging technologies do not replace human judgment. Instead, they fundamentally change when humans are involved and what information they receive when decisions matter most. Across multiple incidents and exercises, the most common failure observed is not poor decision-making, but delayed awareness.

By integrating AI, OT monitoring, and advanced analytics into the BCMS, organizations move from preparedness to prediction. AI-enabled resilience does not replace human decision-making it shortens the distance between weak signals, situational understanding, and informed leadership action.

1. Predictive Monitoring and Early-Warning Systems

Early detection is the most critical factor in effective crisis response. AI-enabled and OT-aware monitoring systems continuously analyze operational data and identify anomalies long before they become visible through manual observation.

In FMCG and industrial environments, this includes:

- Subtle cold-chain temperature deviations detected through sensor analytics
- Early logistics delays identified via route and time-series analysis
- Gradual degradation in production line performance
- OT anomalies indicating equipment stress or failure
- Cyber indicators signaling early compromise of industrial systems
- Supplier risk signals derived from external and third-party data

These insights are translated into:

- Automated alerts
- Predefined escalation triggers aligned with BCMS thresholds
- Real-time dashboards for operational and executive leadership
- Dynamic risk scores that evolve as conditions change

This directly strengthens ISO 22301 Clause 8.4 (Incident Response) by enabling earlier and more accurate escalation. Crucially, advanced technologies improve not only speed but precision, filtering noise, reducing false positives, and allowing teams to focus on what truly requires action.

The Financial Impact of Time-to-Detection:

One of the most underestimated dimensions of organizational resilience is its direct financial impact. In many organizations, investments in business continuity, early-warning capabilities, and crisis intelligence are still perceived primarily as cost items rather than mechanisms for value preservation.

In reality, the financial impact of disruption is rarely linear. Losses tend to accelerate as detection and escalation are delayed through product spoilage, operational downtime, contractual penalties, expedited logistics, regulatory exposure, and reputational erosion.

In high-volume environments such as FMCG, even short delays can translate into high financial consequences. Integrating ISO 22301 with artificial intelligence fundamentally alters this dynamic. Early detection and automated escalation significantly reduce the time-to-decision, thereby compressing the financial exposure window. Rather than absorbing the full cost of a disruption, organizations contain its impact before losses begin to compound.

From a resilience perspective, this reframes continuity and crisis-readiness investments as financial risk mitigation, not operational overhead. The value is realized not through efficiency gains alone, but through avoided losses incidents that never fully materialize because they are intercepted early.

The relationship between time-to-detection and financial impact is rarely linear. As response is delayed, losses accelerate rather than accumulate gradually.

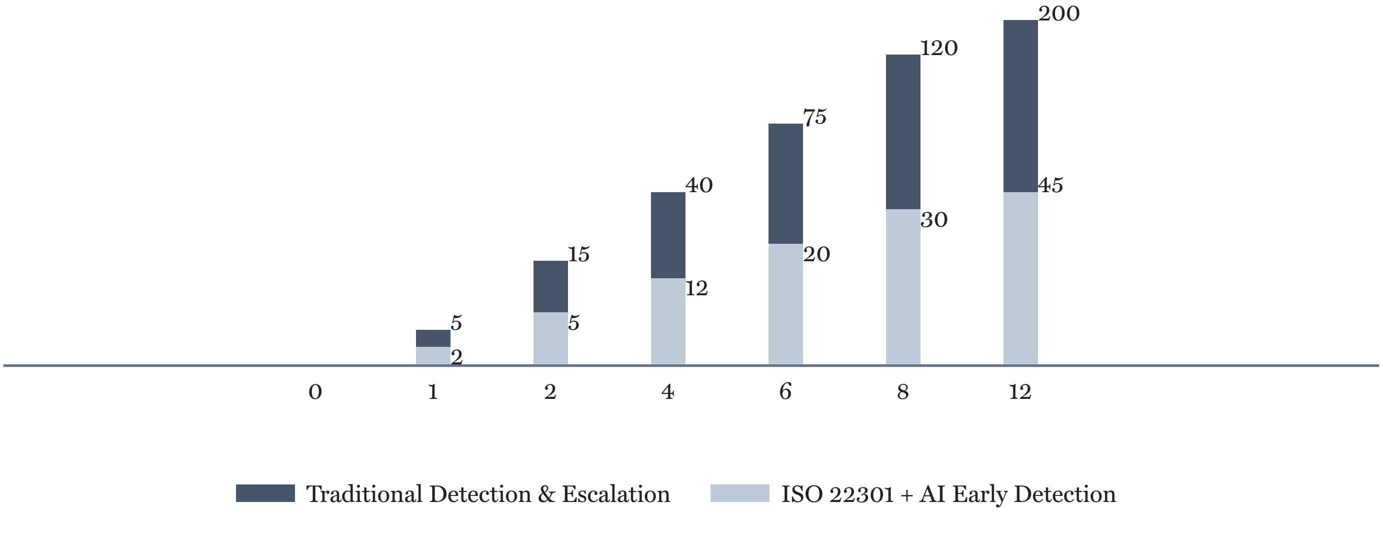


Figure: Financial impact of disruption relative to time-to-detection.
Early sensing and ISO 22301-aligned automated escalation significantly reduce financial exposure.

2. Dynamic BCMS Governance Instead of Static Documentation

One of the most consistent weaknesses observed in continuity programs is documentation drift. Plans become outdated, dependencies shift, and assumptions lose relevance faster than governance cycles can accommodate. Advanced technologies introduce the concept of a living BCMS by enabling:

- Continuous scanning of continuity documentation for outdated assumptions
- Identification of misalignment between BIA, risk registers, and response protocols
- Real-time dependency mapping across operations, IT, OT, logistics, and suppliers
- Dynamic risk re-evaluation based on live operational data
- Automated alignment with ISO 22301 and regional regulatory requirements

In the Saudi Arabia context, this supports expectations arising from national BCM frameworks, cybersecurity authorities such as the National Cybersecurity Authority (NCA) and its Essential Cybersecurity Controls (ECC), National Ris Council (NRC), and sectoral regulators overseeing food, logistics, manufacturing, and critical services. Emerging technologies therefore act as both an operational enabler and a governance mechanism, producing evidence-based readiness rather than static compliance artifacts.

3. Intelligent Crisis Simulations: Testing Behavior, Not Just Plans

A critical reality observed across organizations is that plans rarely fail; human coordination does. Traditional tabletop exercises often follow predictable scenarios, lack cognitive pressure, fail to simulate ambiguity, and reinforce assumptions rather than challenge them. AI- and technology-enabled simulations change this by:

- Dynamically adapting scenarios based on participant decisions
- Measuring decision latency and communication effectiveness
- Exposing coordination gaps across functions
- Generating objective readiness metrics for leadership

This elevates ISO 22301 Clause 8.5 (Exercising and Testing) from a compliance activity into a continuous learning and maturity-building mechanism.

Decision-Making Under Uncertainty

One of the most complex aspects of crisis management is decision-making under uncertainty. In the early stages of disruption, information is rarely complete or fully reliable, yet decisions must still be made quickly to prevent escalation. Traditional escalation models often depend on sequential reporting and human interpretation, which can delay action or lead to disproportionate responses.

When integrated with ISO 22301, AI enhances decision-making by consolidating signals from operational, digital, and supply-chain sources into a coherent situational view. Rather than overwhelming leaders with fragmented alerts, systems provide contextualized insights that highlight potential impact, urgency, and interdependencies, supporting faster and more proportionate decisions even when full clarity is not yet available.

From operational experience with Almarai, the most critical decisions often occur before a disruption is fully visible when signals are weak, information is incomplete, and escalation thresholds have not yet been clearly breached.

Limitations and Governance Considerations in AI-Enabled Resilience

In practice, technology does not fail resilience programs; poor governance does. While emerging technologies can improve the speed and accuracy of crisis detection, their value depends entirely on how they are implemented, governed, and maintained. AI-enabled capabilities are only as reliable as the data feeding them. Inaccurate sensors, delayed data flows, weak system integration, or outdated assumptions can create a false sense of assurance, where early warning signals are missed rather than amplified.

Another frequently observed challenge is model drift. As operations change, production volumes increase, suppliers shift, logistics routes evolve, and the patterns that AI models rely on also change. When these models are not regularly reviewed and recalibrated, organizations may continue to trust insights that no longer reflect operational reality. The consequence is often delayed escalation, incorrect prioritization, or reactive decision-making at the point where options are already limited.

A common example can be seen in large-scale operations. Following changes in production throughput or supplier configurations, monitoring systems may normalize behaviors that were previously considered abnormal. If governance mechanisms are not in place to reassess thresholds and assumptions, escalation may occur only after product quality, service continuity, or compliance has already been impacted, undermining the very purpose of early detection.

For this reason, AI-enabled resilience must be treated as a decision-support capability, not a substitute for human judgment. Accountability for escalation, response, and recovery must remain clearly owned. ISO 22301 provides the governance structure that allows emerging technologies to be embedded safely, ensuring that automation strengthens situational awareness without eroding responsibility, transparency, or leadership control.

FMCG and Complex Environments: Where Real-Time Resilience Becomes Non-Negotiable

These complex environments, such as technology-driven manufacturing and large-scale logistics, expose resilience weaknesses faster than most sectors. Based on hands-on experience, defining characteristics include:

- Rapid inventory movement
- Strict cold-chain integrity requirements
- Continuous production cycles
- Volatile demand patterns
- Immediate financial and consumer impact

A Practical FMCG Scenario

At 2:37 AM, within a large-scale FMCG operation similar to Nestle, Almarai, and Coca-Cola. A refrigeration unit in a distribution facility begins drifting slightly above its historical temperature pattern. The deviation remains within tolerance but is operationally abnormal.

Without integrated operational technology (OT) monitoring and emerging technologies:

- No alert is triggered until thresholds are breached
- The deviation goes unnoticed for hours
- Product quality deteriorates
- Crisis escalation occurs late
- Financial and reputational impact increases

With ISO 22301 integrated with AI and OT:

- The anomaly is detected immediately through pattern analysis
- ISO 22301 escalation workflows activate automatically
- Operations and quality teams receive real-time alerts
- Corrective action is taken within minutes
- Product loss and financial impact are avoided

Beyond operational impact, early detection directly prevents financial loss by avoiding product write offs, supply disruption penalties, and downstream reputational costs.

A Unified ISO 22301 and Emerging Technologies Operating Model

To operationalize this integration, organizations should adopt a three-layer model:

I. **Data Layer: Real-Time Sensing** IoT sensors, Enterprise Resource Planning (ERP) systems, Manufacturing Execution Systems (MES), OT telemetry, cybersecurity logs, supplier and logistics data, and operational KPIs. This layer represents the organization's **nervous system.**

II. **Intelligence Layer: Interpretation and Prediction** Artificial intelligence, advanced analytics, anomaly detection, predictive modeling, dependency mapping, dynamic risk scoring, and adaptive scenario engines. This layer represents the organization's **brain**.

III. **BCMS Activation Layer:** Structured Response ISO 22301 governance, escalation matrices, communication protocols, response playbooks, and continual improvement cycles. This layer represents the organization's **muscle**.

Conclusion

ISO 22301 provides the structure organizations need to establish resilience. Emerging technologies provide the intelligence they need to act fast. FMCG and other complex operational environments demonstrate clearly that static preparedness is no longer sufficient. Regulatory evolution across the GCC reinforces this reality.

Organizations that succeed will be those that combine structured standards, real-time sensing, intelligent escalation, evidence-based governance, and human decision-making supported by data. Integrating ISO 22301 with emerging technologies does not replace continuity frameworks. It allows them to function at the speed of modern disruption.

Ghaida Alghamdi

Resilience and Risk Management Expert

Ghaida Alghamdi is an enterprise resilience and risk management professional with experience across business continuity, crisis management, and operational resilience within complex and fast-moving operational environments.

Her work focuses on applying international standards such as ISO 22301 alongside emerging technologies to inform real-time crisis response, decision-making, and organizational preparedness. She has practical experience supporting FMCG operations and aligning resilience capabilities with evolving regulatory expectations in Saudi Arabia and the GCC.

From Certification to Confidence: How ISO Standards Support Privacy and Data Trust

In an era defined by digital transformation, trust has become the world's most valuable currency. For modern organizations, the challenge is no longer just about preventing a hack; it is about honoring the digital rights of every individual whose data they touch.

Companies that fail to demonstrate robust data protection are vulnerable not only to breaches and fines, but also to long-term reputational damage.

This is where internationally recognized ISO standards play a critical role. Together, **ISO/IEC 27001:2022, ISO/IEC 27701:2025, and ISO 9001:2015** provide a powerful, integrated framework that supports information security, privacy protection, and quality governance. More than compliance tools, these standards help organizations move from certification to confidence, embedding trust into the way they operate.

This article explores how these three standards work to strengthen privacy and data trust, why an integrated approach matters, and how organizations can use certification as a strategic advantage.

The Bedrock of Reliability: ISO 9001

To understand how an organization builds trust in its data practices, we must first look at the foundation of all modern management systems: **ISO 9001, the Quality Management Systems Standard.**

While often mistakenly associated with only manufacturing or physical product quality, ISO 9001 is fundamentally about reliable and consistent outcomes.

In the context of data privacy, ISO 9001 provides the "quality DNA" required for trust. An organization cannot claim to protect personal data if its internal processes are chaotic or undocumented.

ISO 9001 introduces the Process Approach and the Plan-Do-Check-Act (PDCA) cycle, ensuring:

- **Customer Requirements are Met:** In 2026, a "quality" service is one that respects user expectations for data handling.
- **Evidence-Based Decision Making:** Data protection is not based on guesswork but on monitored performance and metrics.
- **A Culture of Improvement:** Mistakes are treated as data points for corrective action, preventing the same privacy lapse from happening twice.

By aligning privacy goals with an ISO 9001 Quality Management System (QMS), organizations ensure that data protection isn't a side IT task, but a core component of "doing business well".

The Foundation of Information Security: ISO/IEC 27001:2022

Before an organization can protect privacy, it must secure the environment where that data lives. This is the role of **ISO/IEC 27001**, the international standard for Information Security Management Systems (ISMS).

ISO/IEC 27001:2022 acts as a "secure framework" for all sensitive information. It requires organizations to identify their risks and implement specific controls to mitigate them across four key themes:

1. **Organizational Controls:** Governing how the company manages information as a whole (e.g., cloud services, supplier relationships).
2. **People Controls:** Addressing the "human firewall" (e.g., remote working, confidentiality agreements).
3. **Physical Controls:** Securing the tangible (e.g., devices - monitoring and entry protection).
4. **Technological Controls:** The digital defenses (e.g., encryption and data leakage prevention).

By effectively identifying risks and managing their controls, ISO/IEC 27001:2022 enables organizations to focus on achieving **confidentiality, integrity, and availability** (**CIA**) of their information assets. But security is only half the story. You can have a perfectly secure database (security) that still violates an individual's right to the protection of their information (privacy).

The Privacy Element: ISO/IEC 27701:2025

While ISO/IEC 27001 focuses on the **security of information assets**, ISO/IEC 27701:2025 shifts the lens toward the **protection of individual privacy rights**. It specifically governs the lifecycle of Personally Identifiable Information (PII).

Published in October 2025, the latest revision of ISO/IEC 27701 defines the requirements for a Privacy Information Management System (PIMS). It provides a rigorous roadmap for how organizations must collect, process, and store personal data while ensuring full transparency and accountability.

The "Standalone" Revolution

The most significant shift in the 2025 version is that **ISO/IEC 27701 is now a standalone standard**. Previously, in the 2019 version, ISO/IEC 27701 was an "extension" of ISO/IEC 27001, and an organization could only be certified to ISO/IEC 27701 when combined with ISO/IEC 27001 certification. In 2026, the landscape has changed. Organizations can now pursue PIMS certification independently.

Why this matters: For many organizations, especially data-heavy SaaS providers, marketing firms, or healthcare startups, privacy is the primary risk. Being able to certify a PIMS without the administrative overhead of a full-scale ISMS makes global privacy accountability accessible to a much wider range of businesses.

These are the key features of ISO/IEC 27701:2025:

1. **Role-Specific Controls:** It clearly differentiates between PII Controllers (who decide why and how data is processed) and PII Processors (who process data on behalf of others). This mirrors the language of the General Data Protection Regulation (GDPR), making it an ideal bridge for legal compliance.

2. **Normative Annex B:** The implementation guidance for controls is now "normative," meaning it carries more weight during audits. It provides a clearer "how-to" for meeting privacy requirements.
3. **Modern Risk Context:** The 2025 update specifically addresses emerging technologies that didn't exist in the same way five years ago, namely AI-driven profiling, biometric data, and complex cloud-to-cloud data transfers.

How ISO Standards Strengthen Data Trust

The journey from "getting certified" to "feeling confident" happens when these standards are used as a strategic management tool.

1. **Eliminating "Shadow Privacy"**
   Without a framework like ISO/IEC 27701:2025, privacy often becomes "shadow work" - something done by the legal team in a silo, disconnected from the IT or operational teams' security protocols. ISO standards force all these departments to speak the same language. When a privacy risk is identified, it is documented and treated in the same risk register as a security threat. This unified view creates organizational confidence that no gaps are being ignored.

2. **Proving Accountability** (**Not Just Policy**)
   Regulators aren't just looking for a Privacy Policy on your website; they are looking for accountability. ISO/IEC 27701 provides "audit-ready" evidence. The standard follows the harmonized structure shared by other ISO management system standards, which requires:

- **Leadership commitment:** Executive-level oversight of privacy goals.
- **Evidence of training:** Ensuring employees don't just "have access" to a policy but understand their role in it.
- **Continual improvement:** A mandatory cycle of internal audits and management reviews to ensure the system and controls evolve faster than the threats.

3. **Strengthening the Supply Chain**
   In today's interconnected economy, you are only as secure as your weakest vendor.
   ISO/IEC 27001:2022 and ISO/IEC 27701:2025 include rigorous controls for "externally provided" services.
   When an organization can show its partners a certification, it drastically reduces the friction of the procurement process. It's a "trust passport" that tells partners: *"You don't have to take our word for it; an independent auditor has verified our controls."*

ISO/IEC 27701:2025 and the Global Regulatory Map

One of the most daunting tasks for a global DPO (Data Protection Officer) is managing the many overlapping requirements such as the Australian Privacy Act and Privacy Principles, the UK's Data Protection Act, the EU's GDPR, Brazil's General Data Protection Law, and various US state laws, just to name a few.

ISO/IEC 27701:2025 acts as a universal framework for regulations across multiple jurisdictions. It maps its controls directly to:

- **Transparency:** Addressed through controls on privacy notices and purpose limitation.
- **Data Subject Rights:** Explicit requirements for handling requests for access, deletion, and portability.
- **Privacy by Design:** Integrating privacy considerations into the development of new products or processes (aligned with ISO/IEC 27001:2022 secure coding controls).

The Road Ahead: Implementation Steps

For organizations looking to implement ISO Management System Standards for Certification and building Privacy and Data Trust, the path is:

1. **Conduct a Gap Analysis:** Identify where your current controls fall short of the latest requirements of ISO/IEC 27001:2022 and ISO/IEC 27701:2025. If your organization is already certified to ISO 9001:2015, a Gap Analysis will also help you plan for building an integrated management system, meeting the requirements of all standards.
2. **Management System Development:** The Management System documented information needs to be created or upgraded to meet the requirements of the relevant standards, including policies, processes, risk assessments, work instructions, registers, and more.
3. **Implementation:** The developed Management System must be put into practice. This will involve coaching your team to the new practices, processes, and policies, so the entire team is on the same page, while leadership monitors to ensure effective performance.
4. **Internal Audit:** Regular internal audits are a requirement of the ISO standards, to assess if all requirements (the organization's requirements and the standards' requirements) are being met, if implementation is effective, and if the organization is ready for certification.
5. **Certification Audit:** The Internal Audit findings and Management Review outcomes will determine when the organization is ready to proceed with the certification process.

Expert Tip: Engage a Consultant: A management system consultant like **ISO Certification Experts** will be able to guide you through the entire process with confidence, from the Gap Analysis to the development, implementation, Certification success and ongoing management of your systems.

Trust as a Competitive Advantage

Certification is an independent verification to the outside world, but confidence is a state of being. By adopting the frameworks of **ISO 9001:2015, ISO/IEC 27001:2022 and ISO/IEC 27701:2025**, organizations do more than just prevent issues. They signal to the market that they are mature, resilient, and respectful of the people behind the data that they have the privilege to access.

Trust is easily lost and hard to regain. While a privacy policy is a promise, certification to the ISO Standards is the proof. The only question that remains now: is your organization ready to lead the way, or will you be left behind in the 'trust gap'?

Erica Smith

Founder and Managing Director of ISO Certification Experts and ICExperts Academy

Erica's career began in HR and corporate law across Australia and the UK, later pivoting into business development and operational efficiency. Her expertise in best practice and Systems Development was refined at Gartner Group and Davis Langdon, where she facilitated "best practice" workshops, and managed global risk, auditing, and certification for over 400 clients.

In 2007, Erica founded ISO Certification Experts and later, ICExperts Academy. As Managing Director, she leads a national team across Sydney, Melbourne, Adelaide, and Perth, who help client organizations understand and implement effective management systems that support sustainable growth, reduce risk, and achieve their goals.

ITALY, ROME

7-8 OCTOBER 2026

BEYOND TECH

This year's conference will dive deep into the issues transforming global security and governance:

- DORA – Digital Operational Resilience Act
- Omnibus Regulation
- NIS 2 Directive
- Deepfake Threats & Digital Trust
- Agentic AI
- EU AI Act
- Human - Agent SOC Collaboration
- Quantum Security & Post-Quantum Readiness

PECB CONFERENCE 2026 ROME

For the first time at **PECB Conference**, experience an interactive debate session where leading experts will challenge each other on the most controversial issues in cybersecurity and AI.

TWO PERSPECTIVES

ONE STAGE

REAL ARGUMENTS

SECURE YOUR SEAT

SUPER EARLY BIRD OFFER

$250 OFF

PECB CONFERENCE 2026 ROME

LEAD THE FU

ONLINE PHASE: 21–22 September 2026

Certified AI Security Professional (CAISP)

As artificial intelligence and large language models rapidly transform business operations, **securing AI systems has become mission-critical.**

This certification equips professionals with the knowledge to **protect AI technologies across their entire lifecycle.**

Graeme Parker
Managing Director a Parker Solution Group
United Kingdom

Aligned with emerging international guidance including ISO/IEC 27090 on AI cybersecurity.

What You Will Learn

- Secure AI systems from development to deployment
- Understand AI and machine-learning fundamentals
- Address generative AI and LLM security risks
- Mitigate prompt injection and adversarial attacks
- Prevent model inversion, data leakage, and training-data poisoning
- Manage risks in AI supply chains and agent-based systems

Certified EU AI Governance Professional

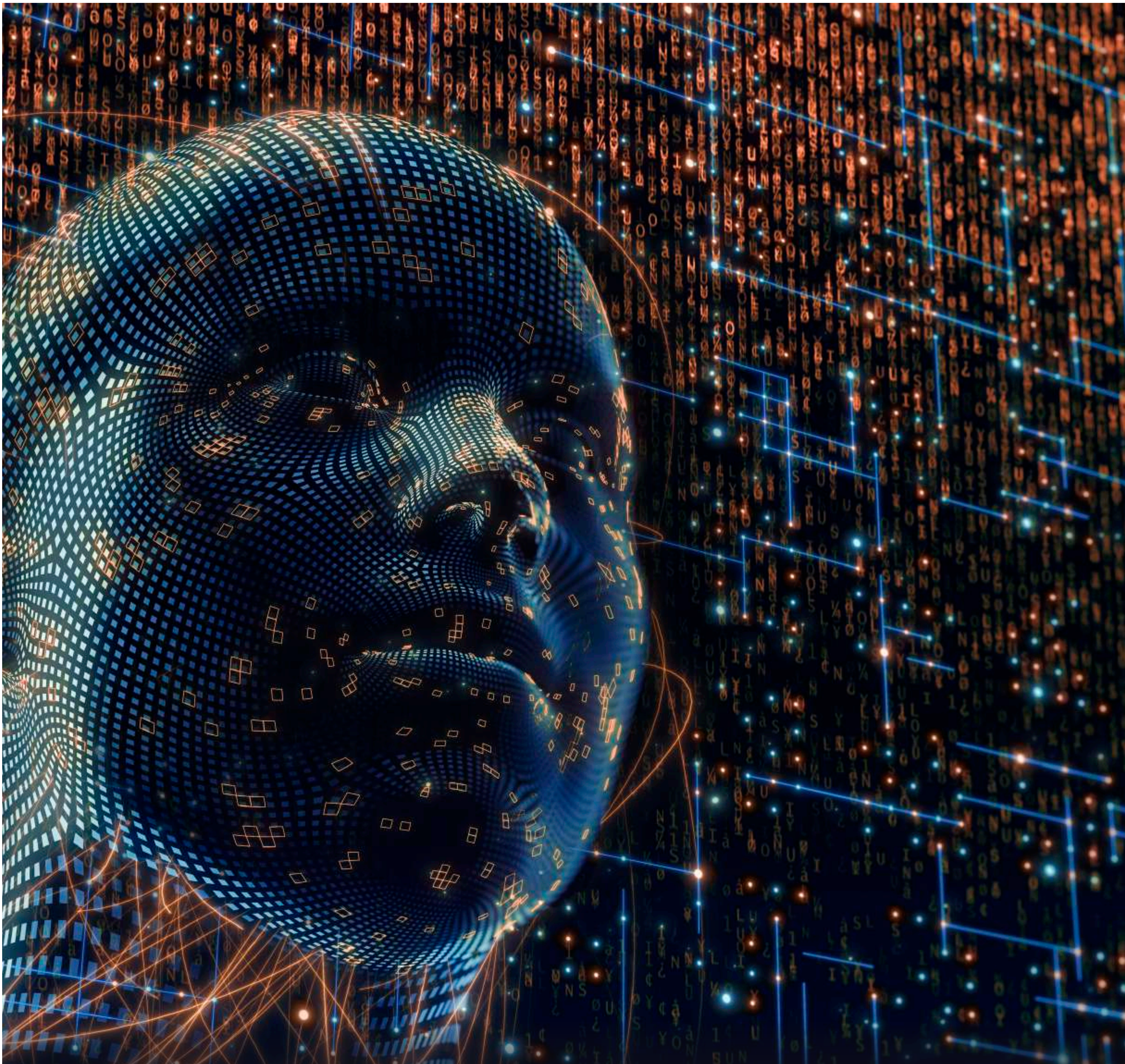
Peter Geelen
Managing Director at CyberMinute
Belgium

With the **EU AI Act becoming fully applicable on 2 August 2026**, organizations across Europe must prepare for a new era of **AI regulation and accountability.**

This certification enables professionals to translate regulatory requirements into practical governance frameworks.

What You Will Learn

- Understand the obligations introduced by the EU AI Act
- Translate regulatory requirements into governance structures
- Implement AI lifecycle governance from design to oversight
- Build internal AI governance capabilities
- Bridge legal, risk, compliance, and technical teams



OPINION

Beyond the Algorithm: Why Your AI Needs an ISO-Based Management System

Artificial Intelligence (AI) is rapidly transforming industries worldwide.

Organizations are investing heavily in AI to drive innovation and efficiency. However, amid this technological fervor, a critical element is often overlooked: the management of these AI systems.

While the technical aspects of AI are crucial, a robust AI Management System (AIMS) is essential for ensuring responsible, ethical, and effective AI deployment, operation, and improvement.

This article invites you to shift your perspective; to look beyond the algorithm and recognize the strategic importance of investing in a management system for your AI. Just as you wouldn't launch a product without a business plan, you shouldn't deploy AI without a well-defined AIMS.

Based on the internationally recognized ISO/IEC 42001:2023 standard, an AIMS provides the framework for governing your AI initiatives, mitigating risks, and maximizing the return on your AI investments. Having AI within your organization without a management system that keeps the organization's strategic direction clear and measures the achievement of objectives is like starting a journey without deciding the destination or preparing suitable luggage.

AI is now available for everyone and for free, and even our kids are using it on a daily basis through any device and without knowledge about computer/data science, programming, or even the basics of IT, which means that AI is becoming easier to spread, and those who use it will never stop in the future.

They will depend on it more and more and ask it to support them in more and more aspects of their lives. If this is the case for our kids, what about our organizations, which have different and complex tasks that are directly linked to our goals and objectives, without even mentioning the emerging risks in a world full of vulnerabilities?

The Misconception: AIMS is Only for AI Giants

One of the biggest misconceptions surrounding AIMS is that it's only relevant for organizations with dedicated AI development teams or even departments. This couldn't be further from the truth. An AIMS, particularly one based on ISO/IEC 42001:2023, is scalable and adaptable to organizations of all sizes and levels of AI maturity.

Think of it this way: even the smallest AI-powered tool, a chatbot on your website, a machine learning algorithm for customer segmentation, introduces potential risks and requires careful management.

An AIMS provides a structured approach to identify, assess, and control these risks, ensuring that even your simplest AI applications are aligned with your organizational goals and ethical principles. You don't need a team of AI experts to implement a basic AIMS; you simply need a commitment to responsible AI practices. Start small, focus on your most critical AI applications, and scale your AIMS as your AI footprint grows.

Therefore, AIMS is like any other management system: it can start whenever there is an actual need and become more mature over time. It shall be there from the beginning to represent governance and management before we reach the technical and operational aspects.

ISO/IEC 42001:2023: A Standard for Integrated AI Management

ISO/IEC 42001:2023 is a game-changer for AI governance and management. Developed using the ISO Annex SL structure, the same structure template used by most modern ISO management system certification standards since 2012, it offers a standardized and internationally recognized approach to managing AI. This is a significant advantage because it allows you to seamlessly integrate your AIMS with your existing management systems, such as:

- **Information Security Management System (ISMS – ISO/IEC 27001 family of standards):** Protecting sensitive data used by AI systems.
- **Quality Management System (QMS - ISO 9001):** Ensuring the quality and reliability of AI outputs.
- **Privacy Information Management System (PIMS – ISO/IEC 27701):** Managing privacy risks associated with AI-driven PII data processing.
- **Service Management System (SMS – ISO/IEC 20000-1):** Managing the IT processes associated with AI-driven technical processing.
- **Business Continuity Management System (BCMS – ISO 22301):** Managing the continuity and disaster recovery associated with AI-driven critical services, products, processes, and activities.

This integration not only streamlines your management processes but also ensures consistency and alignment across your organization.

By adopting ISO/IEC 42001:2023, you're not just implementing AIMS; you're building a cohesive and integrated management ecosystem.

Navigating the Emerging Regulatory Landscape

The world is waking up to the need for AI governance. Regulations are emerging globally, designed to ensure the responsible and ethical use of AI. The EU AI Act, for example, proposes strict rules for high-risk AI systems. Similar initiatives are underway in other regions, reflecting a growing consensus that AI needs to be managed effectively to prevent potential harms.

In the Kingdom of Saudi Arabia (KSA), the Saudi Data and Artificial Intelligence Authority (SDAIA) and the National Data Management Office (NDMO) have provided Saudi governmental entities with robust controls and specifications across 14 domains of data governance and management, which lead directly to achieving the Saudi 2030 vision.

They are now asking entities to invest wisely in AI to provide their customers with a great experience every day without compromising their PII security or national security.

An AIMS based on ISO/IEC 42001:2023 can help you proactively prepare for these regulatory changes. By implementing robust AIMS, you'll demonstrate your commitment to responsible AI practices, build trust with stakeholders, and avoid potential penalties for non-compliance. Investing in an AIMS is not just a matter of good governance; it's a strategic imperative for navigating the evolving regulatory landscape.

The AI Maturity Journey: AIMS as Your Guide

Whether you're just starting your AI journey or you're a seasoned AI innovator, an AIMS based on ISO/IEC 42001:2023 can provide invaluable guidance.

As your organization's AI maturity increases, so does the complexity and potential risks of your AI systems. An AIMS helps you:

- Establish clear objectives and performance metrics for your AI initiatives
- Define roles and responsibilities for AI development, deployment, and monitoring
- Implement robust risk management processes to identify and mitigate potential harms
- Foster a culture of ethical AI practices throughout your organization
- Continuously improve your AI systems based on data and feedback

By embracing AIMS, you're not just managing AI; you're building a foundation for sustainable and responsible AI innovation.

Knowledge Availability

As we wake up every day and find a new AI provider in the market with significant advantages over yesterday's provider, we need to learn how to manage our knowledge needs, too. There are different types of technical training providers that cover different aspects of AI on hardware, software, and logic development levels. But to develop smart AIMS, we still need strong governance and management skills to have the power to control our wild horses at all times, or they will take us to unexpected destinations.

PECB has been in the market for more than 20 years and has received numerous awards for its high-quality courseware and experienced trainers, who are spreading their expertise in more than 150 countries worldwide. PECB now offers a suite of professional AI training courses covering the needs of various audiences, including ISO/IEC 42001 Foundation, Lead Implementer and Lead Auditor, Certified Artificial Intelligence Professional (CAIP), Certified Artificial Intelligence Manager (CAIM), among others to come.

Your Mind is Your Weapon

Although I believe in AI capabilities every day if used wisely by experts, I believe more that the human mind is the most powerful weapon people have, and they do not need to replace it with AI for any reason. Therefore, the more we invest in and depend on AI, the more we need to be able to control it and make informed decisions throughout the journey.

Conclusion

The future of AI depends on our ability to manage it effectively. By investing in an AI Management System based on ISO/IEC 42001:2023, you can unlock the full potential of AI while mitigating its risks and ensuring its responsible use. Don't let your AI investments be undermined by a lack of governance. Take the first step towards building a robust AIMS and embrace the future of AI with confidence.

Mostafa AlShamy

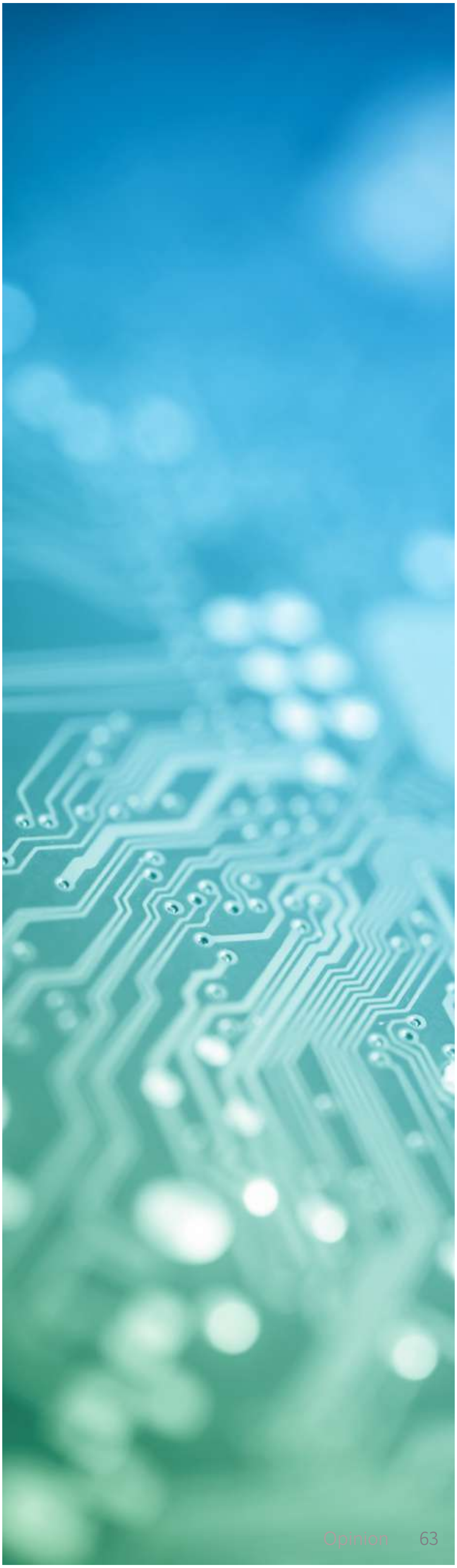
PhD, CEO and Co-Founder of Ofouq Integrated Solutions

Mostafa is a highly experienced consultant, trainer, and auditor specializing in management systems implementation, with a profound expertise spanning across various ISO standards, IT governance, and data privacy regulations.

His background includes a strong foundation in information technology, coupled with extensive practical experience in developing and implementing robust management systems tailored to meet the unique needs of diverse organizations. As the CEO and Co-Founder of Ofouq Integrated Solutions, having held key positions at EGYBYTE, and as a Data Management Office Councilor for a leading Saudi authority, he possesses a deep understanding of operational challenges across various industries and geographies, notably in the MENA region.

Mostafa helps organizations not only navigate the complexities of compliance (including regulations like Saudi PDPL) and achieve digital transformation, but also optimize their processes, mitigate risks effectively, and maximize their return on investment. His track record includes delivering training, conducting audits, leading projects, and crafting customized solutions based on frameworks such as ITIL, COBIT, PRINCE2, CMMC, and various ISO standards.

He is passionate about empowering organizations to embrace best practices, foster a culture of continuous improvement, and achieve sustainable success. For more information, visit our websites: www.egybyte.net and www.ofouqis.com.



PECB was Awarded as Best Cybersecurity Education Provider

PECB is proud to announce that it has once again been awarded with the Best Cybersecurity Education Provider Award by the Cybersecurity Excellence Awards.

This remarkable achievement marks the 8th consecutive year that PECB has received this prestigious recognition, reinforcing its position as a global leader in cybersecurity education.

This award reflects PECB's continued commitment to delivering high-quality training programs that equip professionals with the skills and knowledge needed to address evolving cyber threats.

We extend our sincere gratitude to our global community for their trust and continued support.

READ MORE →

2026
WINNER
CYBER
SECURITY
EXCELLENCE
AWARDS

AI Is Not the Risk – Your Decision-Making Is

Why the biggest threat in the AI era isn't the technology, it's what we stop doing when we trust it.

Picture this: a security team runs its weekly alert triage. The AI-powered SIEM has already categorized, prioritized, and color-coded everything. Green means safe. Red means urgent. The analyst glances at the dashboard, confirms what the AI suggests, closes the tickets, and moves on. Fast. Efficient. Clean.

Three months later, a breach. Not through some exotic zero-day. Through a series of alerts that were flagged green. The AI wasn't broken. It did exactly what it was designed to do: pattern-match based on historical data. But the threat was new. It didn't match the old patterns. And nobody questioned the green light, because why would you? The machine said it was fine.

Here's the uncomfortable truth: the AI didn't fail. The decision-making did. Or rather, the absence of decision-making did. Because somewhere along the way, we stopped deciding and started accepting.

I've spent years helping organizations navigate compliance frameworks, from ISO/IEC 27001 to DORA and beyond, and here's what I keep seeing: the tools get smarter, but the decisions don't. We upgrade our technology and downgrade our judgment. We automate the process and forget to engage the brain. And now, with AI becoming embedded in everything we do, this gap between capability and critical thinking is wider than ever.

This article isn't anti-AI. Far from it. AI is one of the most powerful tools we've ever had. But a tool is only as good as the person wielding it. And right now, too many of us are letting the tool wield us.

We've Always Blamed the Tool

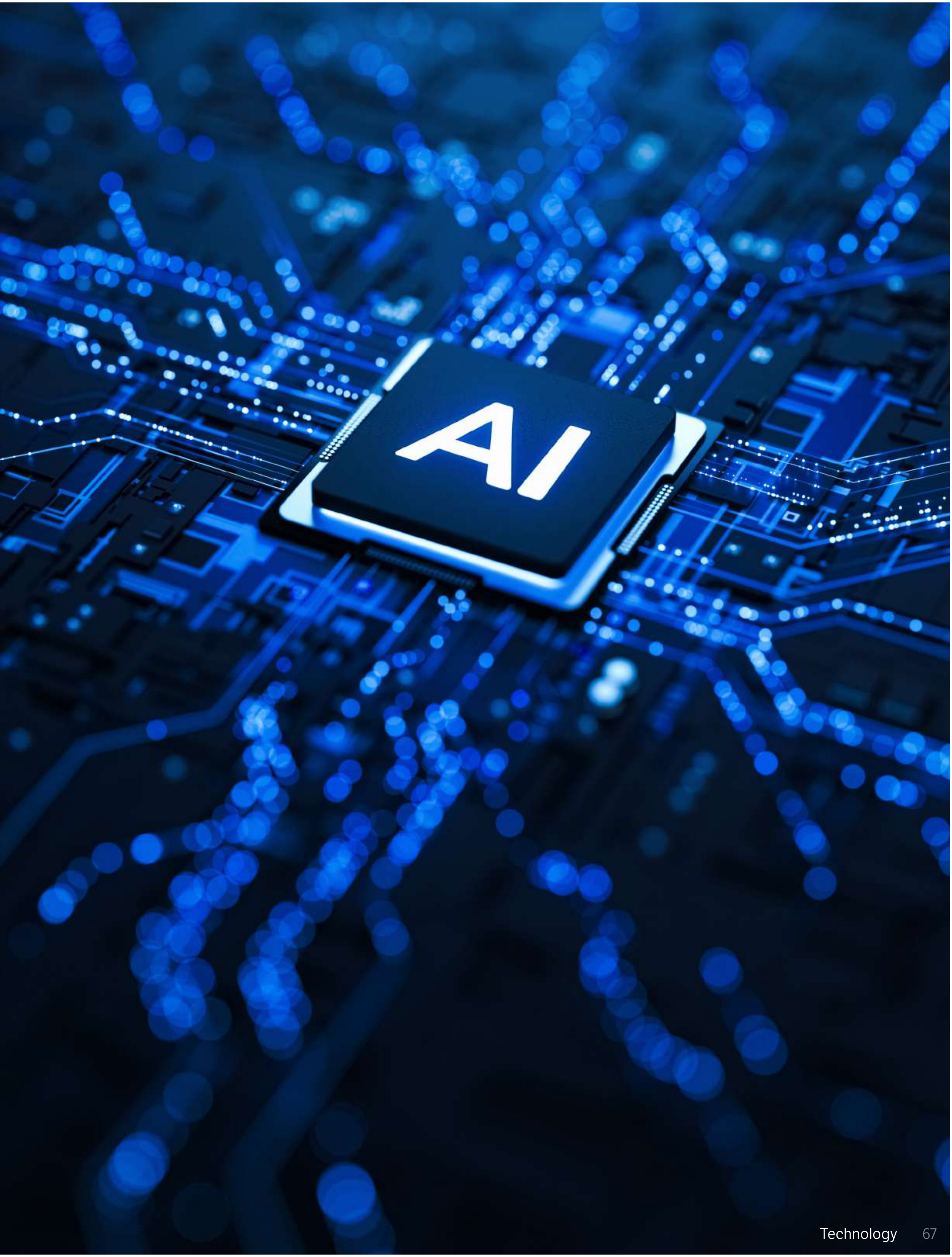
Let's be honest with ourselves for a moment. Blaming technology for security failures is not new. We've been doing it for decades. When firewalls were the hot thing, every breach was a "firewall problem." When cloud adoption exploded, it was the cloud's fault. When remote work became the norm, we pointed at VPNs and home routers. And now? Now it's AI.

The pattern is always the same: a new technology arrives, we adopt it enthusiastically, something goes wrong, and we blame the technology. We rarely stop to ask whether the real issue was how we used it, or more precisely, how we decided to use it.

I remember a project early in my career where an organization had invested heavily in a state-of-the-art intrusion detection system. The technology was solid. But nobody had defined clear escalation procedures. Nobody had trained the team on what to do when the system flagged something ambiguous. So when a real incident occurred, the alert sat in a queue for 72 hours because it didn't look "urgent enough" to the person glancing at the dashboard. The tool worked perfectly. The decision-making around it was nonexistent.

We love to say that humans are the weakest link in cybersecurity. And the numbers back it up: roughly 95% of incidents trace back to human error. But here's what's interesting: AI doesn't change that statistic, it just adds a new flavor to it. Instead of clicking on a phishing link, now the error is trusting an AI output without verification. Instead of misconfiguring a firewall, now it's accepting an AI-generated policy without reading it. The mechanism is different, the root cause is the same: a human who didn't engage their judgment.

And that's actually good news because if the problem is human, the solution is human too. We don't need better AI, we need better decisions.


AI

The Cognitive Shortcuts AI Exploits

Here's where it gets interesting, and a little uncomfortable. AI doesn't just sit there waiting for us to make mistakes. It actively makes certain mistakes easier to make. Not deliberately, of course, but by design. Because AI is fast, confident, and polished, it triggers cognitive shortcuts that we're already prone to.

I like to think of AI as a mirror with a megaphone. It reflects us, but louder. It amplifies our strengths: speed, synthesis, creativity, but it also amplifies our weaknesses: overconfidence, laziness, confirmation bias, and the tendency to simplify what should remain complex.

Let me walk you through the traps I see most often in cybersecurity contexts.

Automation Bias: "The Machine Said So"

This is the big one. Automation bias is our tendency to favor suggestions from automated systems over our own judgment, even when we have evidence to the contrary. In a security operations center, this looks like an analyst who stops questioning alerts because the AI has been right 99 times out of 100. But that 100th time? That's the breach and the analyst didn't catch it because they'd stopped looking.

Think about it in everyday terms. You're driving with GPS, the GPS tells you to turn left into what is clearly a dead-end street. Most people hesitate, some even follow the GPS anyway. Now imagine that dynamic playing out with security decisions that affect an entire organization. That's automation bias in action.

Delegation Without Accountability

AI makes it incredibly easy to delegate without realizing you've done it. You ask the AI to draft a risk assessment. It produces something that looks professional, reads well, and covers the right topics, so you submit it, but did you actually assess the risk? Or did you outsource your judgment to a language model and put your name on it?

This is the compliance trap I warned about in my previous work: you can be "compliant" on paper and completely vulnerable in operations. AI makes this gap even wider because the paper looks even better now. The policy is well-written. The assessment is thorough. The documentation is impeccable. And none of it reflects reality, because nobody actually thought about it. They just accepted what the machine produced.

Speed Over Depth

AI gives answers in seconds and that speed is intoxicating, but speed and depth are often in tension. When you get an instant answer, your brain registers it as "done." You move on. You don't sit with the question. You don't explore the edges. You don't ask "what if?"

In cybersecurity, "what if?" is everything. What if this alert is a false negative? What if this vendor's risk profile has changed? What if this policy doesn't account for a scenario we haven't seen yet? These are the questions that prevent breaches. And they're exactly the questions that AI's speed encourages us to skip.

Here's a real-world parallel: you wouldn't let a junior analyst sign off on your risk register alone. You'd review their work, you'd challenge their assumptions, you'd ask hard questions, so why do we give AI a free pass? Is it because the output looks more polished than what a junior analyst would produce? Probably, but looking polished and being correct are two very different things.

From Accepting to Deciding

So what do we actually do about this? The answer isn't to stop using AI. That ship has sailed, and besides, AI is genuinely useful. The answer is to shift from a posture of acceptance to a posture of decision. It's a subtle but critical distinction.

Accepting means taking what the AI gives you and moving on. Deciding means taking what the AI gives you, questioning it, stress-testing it, and then making a conscious choice about what to do with it. The output is the starting point, not the conclusion.

Here are the habits I've found most effective, both for myself and for the teams I work with.

The Challenge Reflex

Every AI output gets at least one hard question before acceptance. Not a rubber-stamp review. A real question. "What did you not consider?", "What assumption is this based on?", "What would change if this input were different?" It sounds simple, but it's surprisingly hard to maintain when the AI's answer looks clean and confident.

Building this reflex takes practice. It's like learning to brake before a curve instead of during it: counterintuitive at first, essential once it becomes habit.

The "So What?" Test

Before acting on any AI output, ask: "So what?" What does this actually change in our risk posture? What decision does this enable? If you can't answer that clearly, the output is noise, not signal.

I've seen teams generate beautifully formatted AI reports that nobody acts on. The report exists. The risk remains. That's not security. That's theater.

Ownership Loops

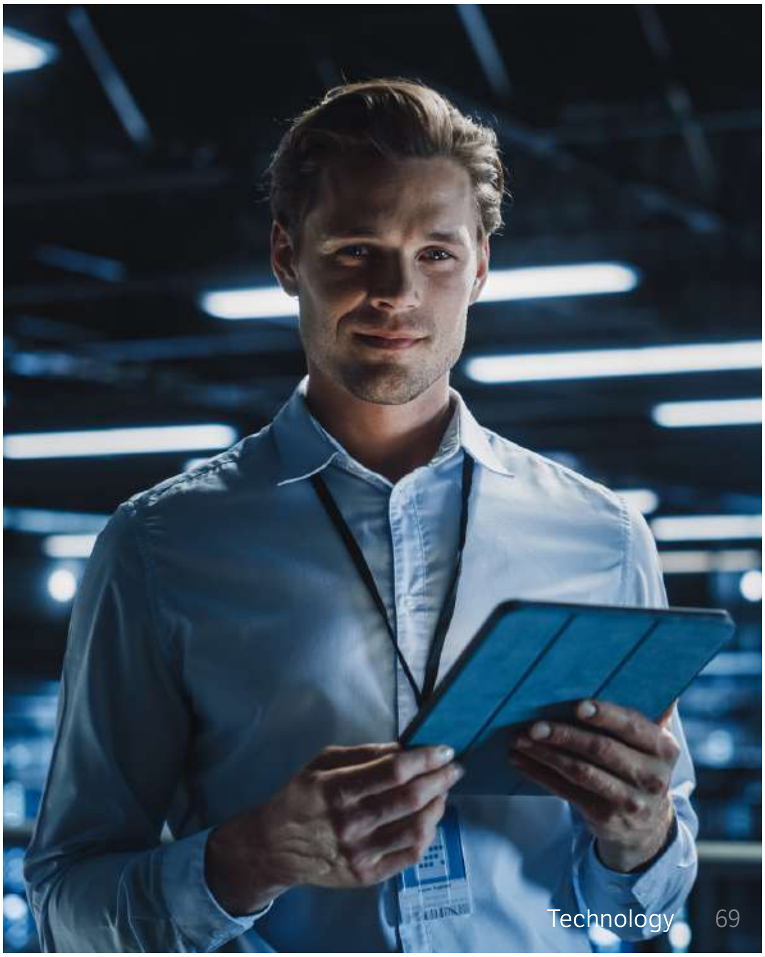
Someone signs off, a human, with a name, not "the AI recommended it." Every AI-assisted decision should have a clear owner who takes responsibility for the outcome. This isn't about blame; it's about accountability. When you know your name is on a decision, you pay attention, you think twice, you engage your judgment.

The moment we allow "the AI said so" to become an acceptable justification, we've abdicated the very thing that makes us valuable.

Compliance Is Not a Substitute for Thinking

I've spent my career in compliance, and I'll be the first to tell you: passing an audit does not mean you're secure. Frameworks like ISO/IEC 27001, DORA, the AI Act, and ISO/IEC 42001 are essential. They give structure. They create a common language. But structure without judgment is a filing cabinet, not a defense. You can check every box and still get breached if nobody is actually thinking about the risks behind the checkboxes.

The same principle applies to AI governance. Yes, adopt frameworks. Yes, document your AI usage. Yes, conduct risk assessments. But don't let the process replace the thinking. Governance should be alive, iterative, and grounded in reality, not a static binder that collects dust between audits.



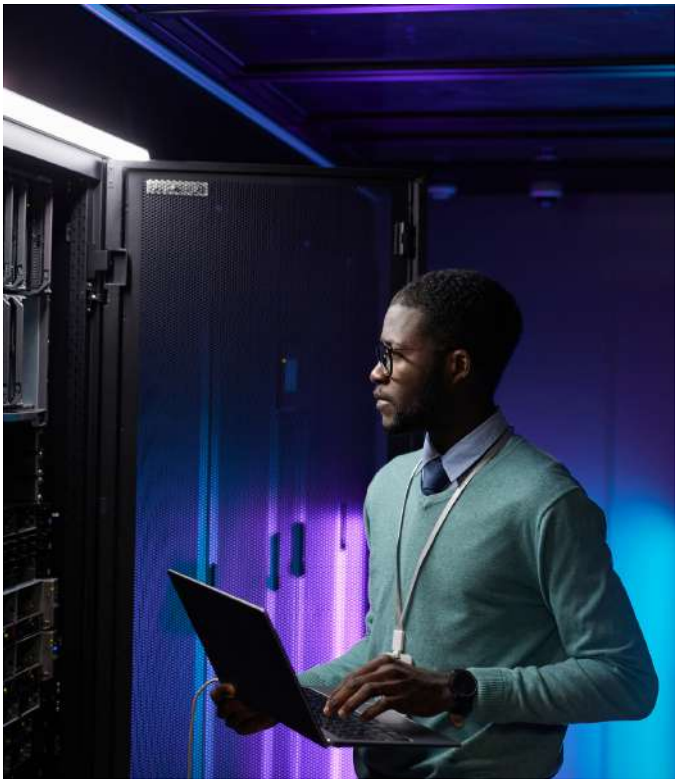
Building a Decision-Making Culture, Not an AI Culture

Individual habits matter, but if the organization around you rewards speed over scrutiny, those habits won't survive long. The real transformation has to happen at the cultural level.

As a CISO, your job in the AI era isn't to become an AI expert. It's to protect the quality of human decision-making across the organization. That's the mission. Everything else is a supporting act.

What does that look like in practice?

1. Reward the people who challenge AI outputs, not just the people who produce the fastest deliverables. If someone on your team slows down to question an AI-generated risk assessment and catches a flaw, that's the behavior you celebrate. That's the behavior you want to replicate.
2. Make "the AI said so" an unacceptable justification in post-incident reviews. When something goes wrong, the conversation should center on human decisions: who reviewed it, what questions were asked, where the judgment gap was. The AI is a tool. The human is the decision-maker. Keep that distinction sharp.
3. Rethink your training programs. Most organizations are investing heavily in "how to use AI tools." That's necessary but insufficient. The bigger investment should be in "how to think critically when AI is in the room." Teach your teams to spot automation bias. Train them to question confident-sounding outputs. Help them understand that the polish of an AI response is not evidence of its accuracy.
4. Build governance that breathes. AI evolves too fast for annual reviews. You need short feedback loops: test, control, review incidents, measure drift, and update rules. Iterative, traceable, accountable. A governance framework that sits in a binder and gets reviewed once a year is not governance, it's wishful thinking.
5. Lead by example. If you're a CISO or a security leader, your team watches how you interact with AI. If you accept AI outputs without question, they will too. If you challenge, verify, and own your decisions, they'll learn to do the same. Culture flows from behavior, and behavior flows from the top.



The Question That Changes Everything

Let's go back to our opening scenario. Same team. Same AI-powered SIEM. Same weekly triage. But this time, the analyst doesn't just glance at the dashboard. She pauses. She picks three green-flagged alerts at random and digs in. She asks the AI why it classified them as low risk. She compares the reasoning against what she knows about the current threat landscape. One of the three doesn't add up. She escalates. It turns out to be the early signal of a coordinated attack.

The AI didn't catch it. But the human did, because she was deciding, not just accepting.

That's the shift. Not away from AI. Toward better judgment alongside AI. The technology will keep getting more powerful. The models will keep getting more capable. The outputs will keep getting more polished. And that's precisely why the human element matters more than ever, not less. Because the better the tool looks, the harder it is to question. And the harder it is to question, the more important it is that we do.

AI is not the risk. Your decision-making is. And the good news? That's the one thing you can actually control.

So, here's my challenge to you. The next time AI gives you an answer, before you accept it, before you forward it, before you build on it, ask yourself one question:

"Am I deciding, or am I just accepting?"

That single question is worth more than any tool you'll ever deploy.

Christophe Mazzola

GRC Lead at Cresco Cybersecurity
and Founder at Cyber Academy

Christophe Mazzola is a CISO and GRC Leader with over 15 years of experience in cybersecurity. As Lead GRC at Cresco Cybersecurity (Integrity360 group) and CISO at Mobilexpense, he builds and challenges the governance frameworks organisations rely on across Europe. A PECB and ISACA certified trainer through Cyber Academy, and author of “Être en cybersécurité” (Éditions Spinelle), Christophe speaks regularly at industry events to make cyber governance actionable, not theoretical. For more visit: www.christophemazzola.fr.

PECB was Awarded for "Most Advanced Cybersecurity Training" Award 2026

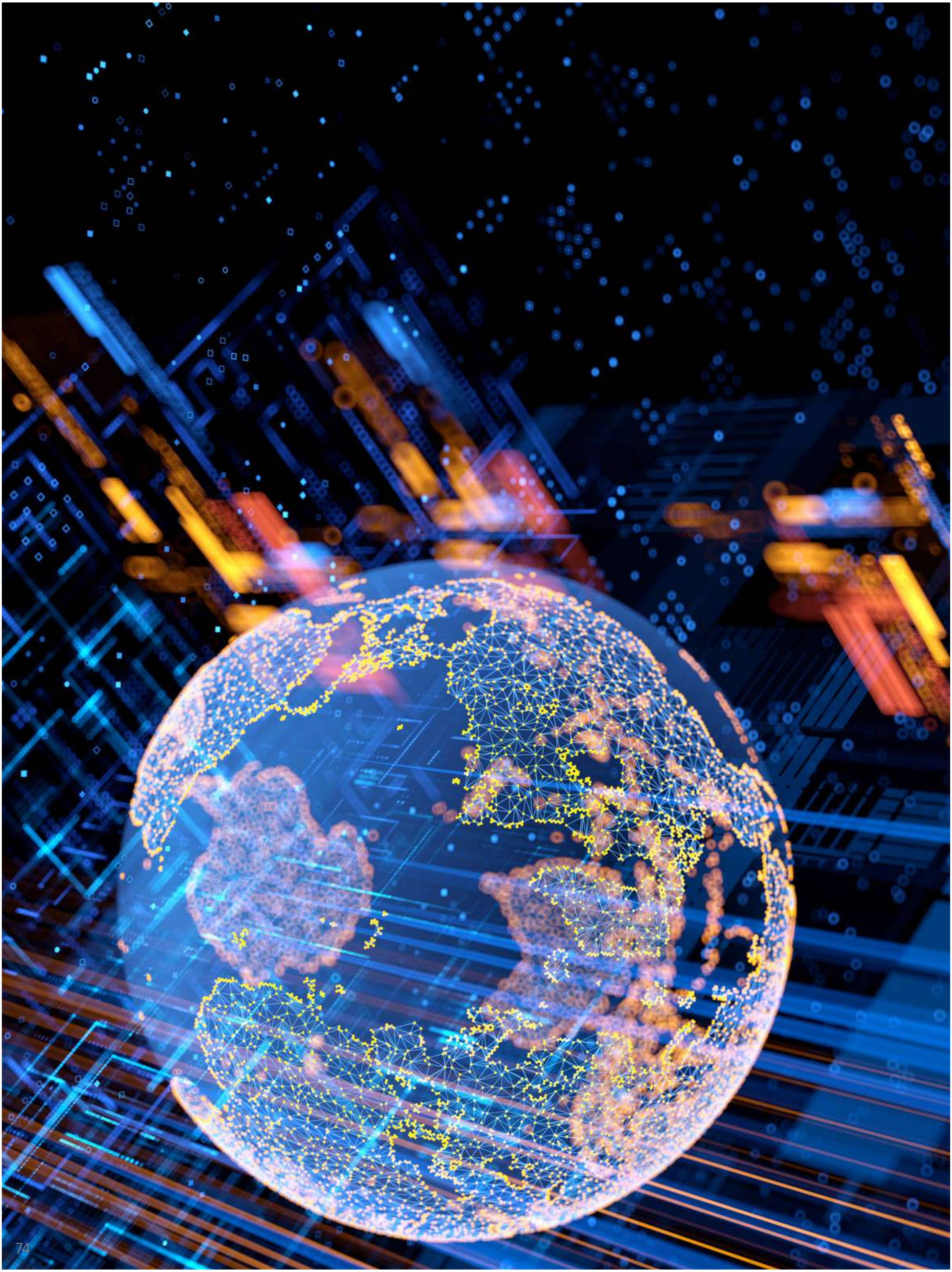
PECB is proud to once again be recognized with the **"Most Advanced Cybersecurity Training" Award** by the prestigious Global InfoSec Awards. This recognition highlights PECB's continued commitment to innovation, excellence, and the development of forward-thinking cybersecurity training solutions.

As the digital landscape evolves, PECB remains dedicated to equipping professionals with the knowledge and skills needed to address emerging challenges and stay ahead of evolving threats.

We extend our sincere gratitude to our valued customers and partners for their ongoing trust and support, which continues to drive our success.

READ MORE →

GLOBAL
INFOSEC AWARDS
WINNER
CYBER DEFENSE MAGAZINE
2026



Assuring Trust in AI: The Expanding Role of ISO Certification and Independent Audits

Artificial intelligence now underpins critical infrastructure, public services, financial systems, healthcare, transport, and everyday business operations.

Alongside rapid adoption, governments and industries face increasing concerns about **ethics, safety, transparency, bias, privacy, and accountability**. These issues are intensified by the steady rise in AI-related incidents worldwide, highlighting the need for structured governance frameworks and independent assurance. Structured standards and independent audits are shifting from "nice to have" to essential for licensing to operate.

This article unpacks how ISO's emerging AI standards (especially ISO/IEC 42001) and third party assurance are reshaping the trust landscape, why they matter if you build or buy AI, and how to combine them with regulatory frameworks like the EU AI Act and NIST's AI RMF. Along the way, we'll ground the discussion with data points, timelines, and concrete links you can use.

Evidence of harm and operational failure has grown more visible as AI proliferates. The AI Incident Database and other repositories show a sharp rise in reported incidents since 2022—TIME reported a 50% year-over-year increase from 2022 to 2024, with 2025 surpassing 2024 totals by October. That trend is echoed in aggregated views by Our World in Data based on AIID counts.

Policymakers are acting. The EU AI Act came into force in August 2024 and phases in requirements until 2027, including third-party conformity assessments for many 'high-risk' systems and governance duties for general-purpose AI (GPAI). Key milestones include prohibitions and AI literacy starting on February 2, 2025; governance provisions, GPAI duties, and notified body infrastructure beginning on August 2, 2025; most high-risk obligations and penalties commencing on August 2, 2026; and full enforcement across regulated product sectors by August 2, 2027.

Meanwhile, the NIST AI Risk Management Framework (AI RMF 1.0), released Jan 26, 2023, with a Generative AI Profile in July, 2024, provides voluntary, globally referenced guidance for governing AI through the functions of **Govern, Map, Measure, and Manage**. Organizations are increasingly aligning their internal controls and vendor questionnaires with it.

In October, 2025, Australia issued new Guidance for AI Adoption, building on and expanding the 2024 Voluntary AI Safety Standard. This guidance provides:

- Six practices for safe and responsible governance (accountability, impact assessment, risk management, transparency, monitoring, and human control)
- Tools, including an AI register template, policy templates, and screening instruments
- Alignment with ISO/IEC 42001 and NIST AI RMF

This makes Australia one of the first countries to tightly link ISO/IEC 42001 with a national AI implementation guide. Australia's National Framework for the Assurance of AI in Government (2024) formalizes a principles based, risk aligned approach to government AI assurance. It includes:

- The AI Impact Assessment Tool
- Guidance aligned with the national AI Ethics Principles
- Accountability and transparency requirements were introduced for agencies in September 2024

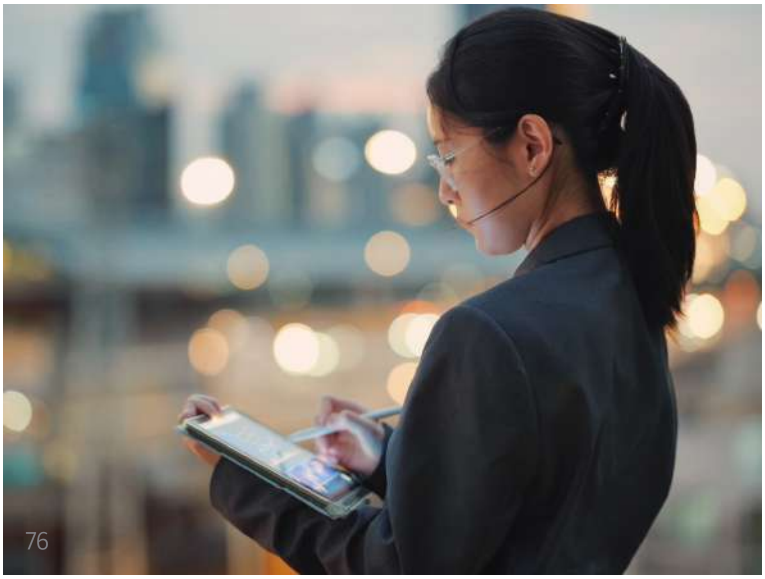
These frameworks help agencies evaluate risks, such as bias, opacity, unpredictability, and safety throughout system lifecycles. The bottom line: trust needs verifiable governance. That's where ISO/IEC 42001 certification and independent audits come into play.

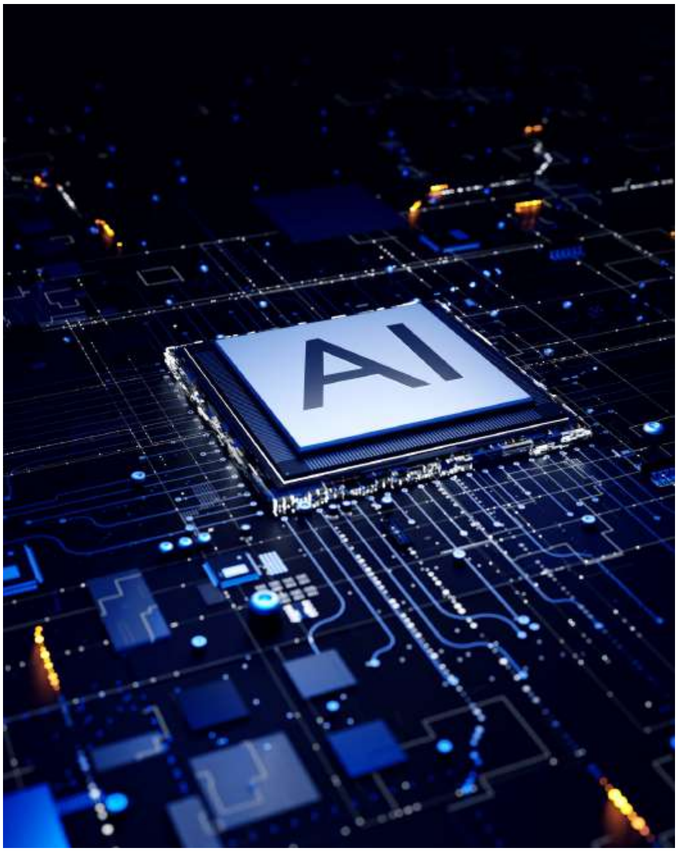
Published in December 2023, ISO/IEC 42001:2023 defines requirements for an Artificial Intelligence Management System (AIMS), the organizational processes, roles, risk methods, lifecycle controls, and continual improvement needed to govern AI responsibly. Think of it as ISO 27001 for AI, built on the Annex SL structure, so it integrates with other ISO systems.

Key characteristics and scope:

- Applies to any organization that develops, provides, or uses AI systems—across all sectors and sizes.
- Concentrates on governance and risk rather than prescribing algorithms—covering policies, roles, AI risk and impact assessments, supplier controls, human oversight, lifecycle operations, monitoring, internal audits, and management review.
- Designed to integrate with ISO/IEC 27001 (information security) and ISO/IEC 27701 (privacy), facilitating unified audits and shared corrective action processes.

ISO/IEC 42001 forms the basis of a modern, auditable AI governance program. Combining it with local and international frameworks ensures compliance with local expectations while aligning with global benchmarks.





ISO/IEC 42001 in Context: ISO/IEC 23894, ISO/IEC 27001, and NIST AI RMF

A credible AIMS rests on coherent risk and control foundations. Three anchors stand out:

- ISO/IEC 23894:2023 (AI Risk Management) — Published February, 2023, this is the AI specific companion to ISO 31000, offering lifecycle risk processes tailored to bias, drift, adversarial manipulation, data quality, transparency, and ethical impacts. It's non certifiable guidance, but invaluable for operational risk practices that feed your AIMS.
- ISO/IEC 27001 — Still the backbone for information security controls that protect AI pipelines: secure development, access to model artifacts and datasets, supplier security, logging, and incident response. Numerous mappings show how traditional Annex A controls extend to AI threats, such as model theft, inversion, poisoning, and prompt injection.
- NIST AI RMF 1.0 — A voluntary, risk based reference that many enterprises and regulators cite. Organizations often cross walk their AIMS controls to the RMF's **Govern, Map, Measure, Manage** functions for internal and external audiences.

Because ISO/IEC 42001 and ISO/IEC 27001 share Annex SL, many companies are creating integrated management systems that cover AI governance, security, and privacy (ISO/IEC 27701) within a single PDCA loop, streamlining audits and corrective actions.

Who Certifies the Certifiers? Accreditation and the Assurance Value Chain

To avoid “audit washing,” it’s vital to understand the conformity assessment stack:

- Certification Bodies (CBs) audit organizations against standards (e.g., ISO/IEC 42001, ISO/IEC 27001).
- Accreditation Bodies (ABs) (e.g., UKAS, JASANZ) assess those CBs against ISO criteria.
- Historically, the International Accreditation Forum (IAF) coordinated global recognition via multilateral arrangements, so a certificate issued under one AB’s mark is “certified once, accepted everywhere.” (As of Jan 2026, IAF merged with ILAC into Global Accreditation Cooperation.)

Why this matters: accredited certificates carry demonstrable assurance value in procurement and regulatory contexts; non accredited certificates often don’t. Buyer beware.

Independent Audits Beyond Certification: SOC 2, Internal Audit, and Specialist AI Assurance

Certification isn’t always the first or only step. Many organizations begin with third-party assurance engagements or SOC 2 reports that include AI-specific controls.

- **SOC 2 and AI:** Auditors increasingly test model versioning, drift monitoring, bias testing, PII handling, and incident response under Trust Services Criteria, especially Processing Integrity for probabilistic systems. Expect questions your 2022 control set never anticipated.

- **Internal Audit Frameworks:** The IIA’s AI Auditing Framework (2024 update) and ISACA guidance help internal audit move “beyond the black box,” mapping process risks, governance, and controls to practical test steps.

- **Specialist AI Assurance:** Firms now offer standalone AI assurance under AICPA or equivalent standards (attestation style), aligned with NIST AI RMF, ISO/IEC 42001, or specific regulatory requirements (e.g., EU AI Act high risk systems). These engagements can provide independent comfort on governance and specific risk mitigations for boards, buyers, or regulators.

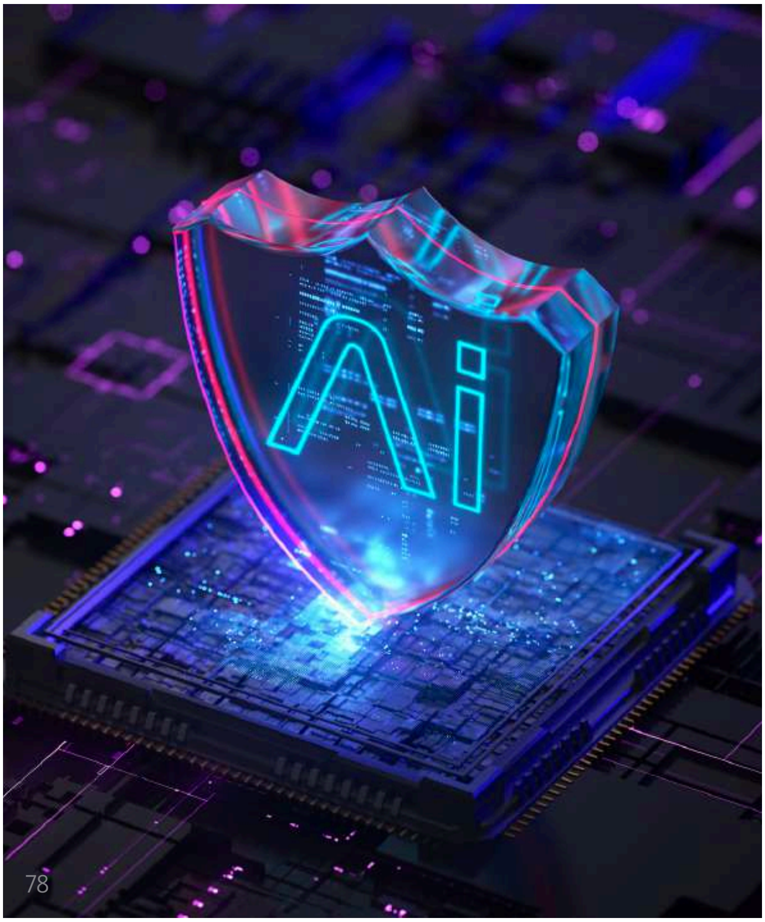
- **Frontier Model Audits:** For cutting edge systems with confidentiality constraints, emerging “frontier AI auditing” methods outline how to evaluate safety and security practices despite limited disclosure. Expect this to influence future audit scopes for GPAI and agentic systems.

The EU AI Act link: Conformity Assessment and Notified Bodies

- For many "high risk" AI systems under the EU AI Act, third party conformity assessment (via notified bodies) will be mandatory before placing systems on the EU market. Governance provisions for GPAI models also apply from August 2, 2025. Organizations preparing now are using ISO/IEC 42001 as the management system backbone and mapping to harmonized standards once published to gain presumption of conformity.
- As of August, 2025, key governance and GPAI provisions began applying, alongside the operationalization of the EU AI Office and national competent authorities, which will coordinate consistency and enforcement.

How ISO/IEC 42001 Certification Adds Assurance

- A verified management system for AI: policies, roles, risk/impact methods, lifecycle controls, supplier oversight, monitoring, corrective action, and internal audit & management review.
- Integrability with ISO/IEC 27001/27701 for unified security/privacy governance, reducing audit overhead and control gaps.
- Market signal for enterprise buyers and regulators: an accredited certificate attests that a qualified CB audited your AIMS against a recognized international standard.



Building a Credible AI Assurance Program: A Practical Roadmap

- Begin with an AI inventory and risk classification: List models (internal and third-party), data sources, endpoints, and business processes. Classify them by impact, regulatory exposure, and the EU AI Act risk category where relevant. Use the NIST AI RMF to structure GOVERN/MAP activities and ISO/IEC 23894 to frame AI-specific hazards (bias, drift, adversarial inputs, privacy leakage).
- Extend existing security and privacy controls to AI: Map ISO/IEC 27001 controls to AI pipelines, including guardrails on model artifacts and datasets, secure SDLC for ML, access control, logging and monitoring, supplier and security due diligence, and incident response that covers model rollback and data governance for training and inference stores.
- Establish your AIMS (ISO/IEC 42001): Define scope, leadership accountability, risk and impact assessment methods, human oversight policies, documentation, supplier management, and a performance evaluation plan (KPIs, internal audits, management reviews). Integrate with your ISMS/PIMS where possible.
- Pilot independent assurance: Conduct reading assessments against ISO/IEC 42001 and AI RMF; commission SOC 2 with AI-specific controls if your customers expect it (processing integrity + bias/drift testing, PII in prompts, model versioning). Use internal audit frameworks to test real use cases.
- Pursue recognized accreditation: Choose a reputable, accredited certification body (consult the AB's directory) and coordinate your evidence collection. If you're aiming at EU high-risk markets, align your ISO/IEC 42001 programs with your EU AI Act compliance strategy, particularly for post-market monitoring, technical documentation, and QMS requirements under emerging harmonized standards.
- Close the loop with ongoing monitoring and incident review. Track performance, drift, and incidents; incorporate lessons into risk registers and corrective measures. External resources like the AI Incident Database or the AIAAIC repository can guide scenario planning and help benchmark emerging risks.

Final Thoughts

Trust in AI isn't just declared; it's earned and demonstrated. ISO/IEC 42001 provides organizations with a certified way to show they manage AI responsibly, while independent audits offer the assurance stakeholders are already demanding.

As regulatory deadlines near and incident figures grow, embedding this assurance into your operational model isn't just good governance, it's a competitive advantage.

Hafiz Sheikh Ahmed

Director and Advisor for Cyberverse Pty Ltd, and
Principal Trainer for NextGen Knowledge

Hafiz is a cybersecurity storyteller with a strategist's mind and an auditor's precision. With more than two decades of international experience, he is a recognized leader in Cyber Governance, Risk, Compliance, and strategic security transformation. Throughout his career, Hafiz has helped organisations make clear, confident decisions under pressure, uncertainty, and rapidly evolving threats.

Currently, as a Director and Advisor for Cyberverse Pty Ltd and Principal Trainer for NEXTGEN Knowledge, he plays a key role in guiding clients across various industries to achieve compliance and excellence in their management systems. His comprehensive understanding of ISO standards and other international cyber/AI frameworks, combined with a practical, results-focused approach, has made him a trusted advisor.A dynamic communicator and TEDx speaker, Hafiz translates complex cyber challenges into compelling, actionable insights that resonate with executives, technical teams, and frontline practitioners alike. His sessions blend real-world experience with narrative clarity, empowering audiences with both inspiration and practical techniques they can apply immediately.

Ranked among the **top 1% of trainers worldwide**, Hafiz is a **Certified Titanium Trainer**, **Certified Trainer of the Year** (**2022**), and a **Business Elite 40 Under 40 recipient** (**2024**). His industry influence has also earned him finalists' recognition for **CISO of the Year** (2021, 2022) and several nominations in the **BX Business Xcellence Awards** (2024, 2025).

Renowned for transforming complex frameworks into structured, results-oriented pathways, Hafiz has led governance and assurance initiatives that enhance executive awareness, reinforce board oversight, and expedite technical improvements. His expertise encompasses control assessment, uplift design, risk advisory, behavioral security, and cyber culture development, ensuring the synchronized evolution of people, processes, and technology in response to contemporary threats.

Hafiz's global client portfolio includes leading consulting firms, government agencies, defense bodies, and major enterprises across Australia, the United States, the United Kingdom, Europe, the Middle East, and the Asia–Pacific. Drop an email to Hafiz at: hafiz.ahmed@cyberverse.net.au, hahmed@nextgenknowledge.ai. Alternatively, connect with Hafiz in linkedin.

Flexible Learning at Your Fingertips

Strengthen your expertise in anti-bribery management with PECB's updated training courses, available in English.

Designed for flexibility and impact, the training courses provide comprehensive knowledge to help professionals effectively audit, implement, and manage Anti-Bribery Management Systems (ABMS).

- ISO 37001 Lead Auditor
- ISO 37001 Lead Implementer

Advance your career and lead with confidence—anytime, anywhere.

LEARN MORE →

#BeyondClassrooms

ISO 37001 Lead Auditor

Building a Model Validation and M

Deploying an AI model is not the finish line. It is the beginning of risk. Many organizations in monitor models once they influence real decisions. Without structured validation and continu

Building a model validation and monitoring program from scratch requires more than d accountability. This article outlines how to design and implement a pr

01

Define Scope and Model Inventory

Before validation begins, organizations must understand what they are validating. Create a centralized model inventory that includes:

- Model name and purpose
- Business owner
- Technical owner
- Data sources
- Decision impact level
- Deployment environment
- Version history

02

Classify Models by Risk Tier

Not all models require the same level of scrutiny. Establish a risk-tiering framework, for example:

- **Tier 1** (**High Risk**)**:** Credit approvals, healthcare diagnostics, fraud blocking
- **Tier 2** (**Medium Risk**)**:** Operational forecasting, resource optimization
- **Tier 3** (**Low Risk**)**:** Internal reporting automation

Establish Inde

Validation should model developer. models into; devel approval. Validati

- Data quality a
- Feature engine
- Model assump
- Performance r
- Stress testing r
- Bias analysis
- Sensitivity ana

Monitoring Program from Scratch

vest heavily in model development but underestimate the infrastructure required to validate and uous oversight, even technically sound systems can drift, degrade, or create unforeseen harm.

ashboards. It demands governance, technical controls, documentation discipline, and clear ogram that holds up under regulatory scrutiny and operational stress.

03

pendent Validation

not be conducted solely by the An effective program separates lopment, validation, and on teams evaluate:

nd representativeness

eering logic

tions

metrics

results

lysis

04

Define Validation Standards

Create formal validation criteria that answer:

- What metrics must be met before approval?
- What thresholds trigger rejection?
- What documentation is mandatory?
- What testing conditions are required?

05

Build Monitoring Infrastructure

Validation is pre-deployment. Monitoring is continuous. Monitoring should track:

- Model performance metrics over time
- Data drift
- Concept drift
- Prediction distribution shifts
- Error rates
- Override frequency

  (human intervention rates)

06

Implement Alert Thresholds and Escalation Protocols

Monitoring without action is meaningless. Define:

- Performance thresholds
- Drift tolerances
- Trigger conditions
- Escalation chains
- Review timelines

07

Document Everything

Documentation is often treated as administrative overhead. It is not. A strong model governance file should include:

- Model purpose and use case
- Development methodology
- Training data description
- Assumptions and limitations
- Validation results
- Approval records
- Monitoring design
- Incident history

Establish Re-V

Models should no
without review. D
frequency based o

- Risk tier
- Model stabilit
- Data volatility
- Regulatory ex

Building

Trust in AI does not emerge from accuracy alone. It e

Organizations that invest early in structured validation programs p
Those that delay often build governance

Building the program from scratch is demanding. But in the age of

08

alidation Cycles

t operate indefinitely
efine re-validation
on:

y

pectations

09

Integrate with Enterprise Risk Management

Model risk cannot exist in isolation. Connect the program to:

- Enterprise risk registers
- Operational resilience planning
- Compliance reporting
- Incident response frameworks
- Internal audit processes
- Monitoring design
- Incident history

10

Assign Clear Accountability

Every model must have:

- A named business owner
- A technical owner
- A validation owner
- An oversight authority

for Trust

emerges from transparency, discipline, and oversight.

osition themselves ahead of regulatory pressure and public scrutiny.
e reactively — under the stress of failure.

AI, accountability is not optional infrastructure. It is foundational.

THE EXPERT

Digital Disruptions and Disaster Recovery: Building Global Resilience in the Age of Cyber and Climate Risks

Navigating Cyber and Climate Risks Through Business Continuity

Originally from South Korea and now residing in North America, one of my former client's main goals after completing my training in ISO 22301 (Business Continuity Management System), was to take this knowledge back to South Korea and assist her organization, which has over 4,000 employees, in becoming better prepared to deal with cyber risks and climate change.

Previously, South Korea was considered to be located in a less seismically active region than countries such as neighboring Japan, as it is farther from major tectonic plate boundaries. However, this perception has now shifted, along with her career and personal goals. Due to climate change and rising sea levels, the frequency of earthquakes in South Korea has increased. At the same time, she has been studying her Master's Degree in Cyber Security to assist in managing the cyber risks on her return to South Korea.

This is one example of how fellow professionals around the globe are seeking knowledge to deal with Digital Disruptions and Disaster Recovery and build resilience in the Age of Cyber and Climate Risks.

Defining Digital Disruptions and Disaster Recovery

For many companies, building global resilience in the Age of Cyber and Climate Risks may be quite challenging. It requires their leaders to sit and think and decide on a strategic approach to ensure that their business can still "stand still" (pun intended) and continue should a cyber or climate risk play out in a disastrous manner.

To understand the strategies or solutions leadership teams can develop, it may be best to define the terms so that we can all be clear on their meanings/definitions for this article. ISO 22301 defines Business Continuity as "an organization's capability to continue delivering products and services within acceptable timeframes, at a predefined capacity, during a disruption."

Within a Business Continuity Management System aligned with ISO 22301, Disaster Recovery (DR) is generally considered the technical component that focuses on restoring IT systems, data, and infrastructure following a disruption. Usually, disaster recovery activities tend to happen or be carried out during a period of 24-72 hours after a major disruption has occurred. Digital disruption refers to "using digital technologies to disrupt existing businesses and industries". It can lead to existing products and services becoming obsolete. Therefore, we are going to have a look at how some companies have been "digitally disrupted" in our age of cyber and climate risks. We can analyze both the short and long-term consequences and understand what adjustments they needed to make.

This perspective can serve as a good starting point for senior leaders and management teams in their respective organizations to begin understanding, developing greater resilience, and creating disaster recovery strategies for digital disruptions amid cyber and climate risks. Although each organization may face slightly different challenges, there may be similarities in terms of how they can address these problems.

Digital Disruptions and Disaster Recovery Solutions

1. Adaptability: A partner company of mine that provides website design and digital marketing to clients across the globe has experienced digital disruption on two separate occasions. Firstly, besides creating websites for their clients, they also had a section or part of the company that provided traditional Marketing and Public Relations (PR). But website programmers in Southeast Asia began to disrupt their business models and markets by offering independent programmers online at lower prices. Of course, a lower price does not automatically translate to better quality. So, they had to adjust their services to demonstrate their quality to clients in order not to compete on pricing, and also change their approach to offer digital online marketing as opposed to traditional marketing.

2. Rethink business model: Secondly, they changed their IT infrastructure from hosting their customers' websites on local servers to storing the information remotely and accessing it through a Cloud Hosting Provider in the United States (USA). However, this Cloud Service Provider experienced their own technical glitch and infrastructure failure. This happened over a 48-hour period and significantly affected clients for both my partner company and the hosting provider (themselves). Given that this second form of "digital disruption" was due to a faulty Change Management process in terms of updating their software and hardware at the data center, my partner company has now decided to perform a review and implement a number of steps to mitigate against a possible recurrence.

3. They have decided to review their Service Level Agreement (SLA) with the Cloud Hosting Provider and discuss the options and assurances for potential financial compensation for this disruption and for any future disruption.

4. Invest in training their employees: My partner company has taken a stance to recruit, retrain, and retain employees to have strong digital skills to perform their tasks efficiently. In doing so, they are embarking on a process of digital transformation whereby anyone can work from any part of the globe, and their information is stored securely online for retrieval. The company is trying to build out a level of resilience, so they are not as negatively impacted by any disruptions, either digitally or climate change, as they were previously.

5. Importantly, they listened and focused on their customers' needs and used digital tools/platform to personalize experiences to ensure that they built long-term relationships to assist them more readily when change arrived.

Coastal Erosion due to Rising Sea Level

There may be other decisions that organizations may have to take to effectively deal with a digital disruption or climate change risk. A prime example is a financial institution in the Caribbean region, which is also a client of mine and I trained in both PECB's ISO/IEC 27001 (Information Security Management System) and ISO 45001 (Occupational Health & Safety Management System).

They built one of their branches too close to the sea. Originally, the idea was very good: to give tourists and local people an opportunity to bank and access finances while they are close to the beach.

However, due to coastal erosion and rising sea levels, the financial institution eventually had to make the decision to leave the building and relocate its staff. This was a serious disruption due to climate change and the rising sea level. Now, this is part of their long-term considerations in terms of analyzing their brick-and-mortar footprint.

Emergency Planning and Drills

In another scenario, I am currently leading an organization of 800-1000 persons to perform Emergency Planning and Drills as a Health & Safety/Disaster Recovery Coordinator. One of the main actions I did was to provide cross-training to various emergency response teams for potential disasters, including earthquakes. Recently, we experienced a 5.0 earthquake on the Richter scale. So one of the ways I am supporting this organization to handle various situations is to obtain First Aid/CPR training and to perform drills to safely remove persons in different disaster/emergency scenarios.

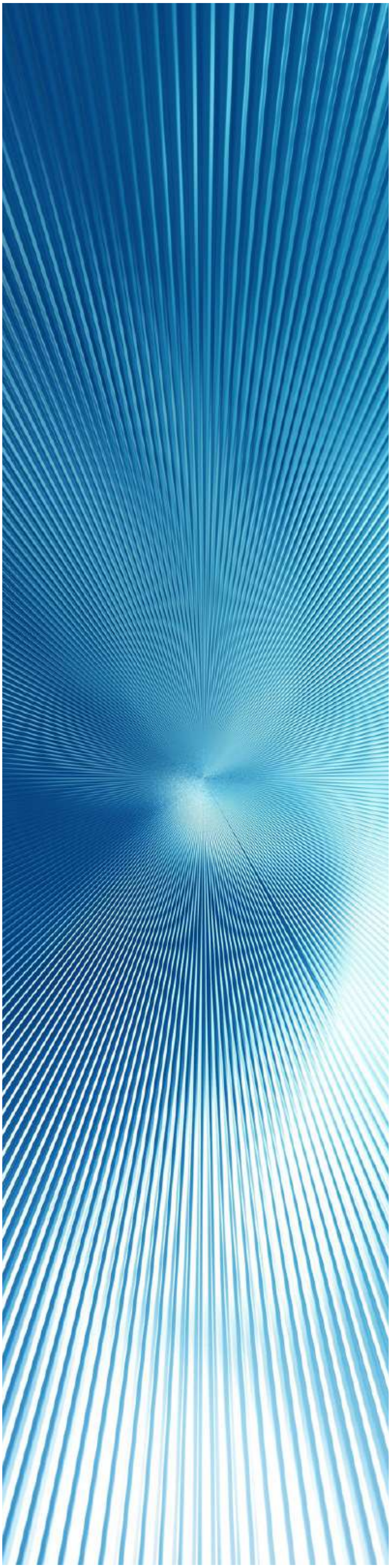
Of course, this had led the organization to rethink its annual budget and even acquire new Personal Protective Equipment (PPE) and other items for key members to use immediately should any disaster arise to threaten the safety of persons within the building.

Conclusion

Some companies are resistant to planning and to having a dedicated person on staff, or even a consultant, to provide expertise to handle digital disruptions and the disaster recovery process. However, due to the increasing prevalence of Cyber and Climate Risks, companies need to rethink how they manage digital disruptions and disaster recovery, ensuring they continue to develop their institutional resilience.

Some of the solutions include adaptability of the organization's objectives, re-thinking their business model, reviewing their Service Level Agreement (SLA) with any Cloud Hosting Providers, training and retaining their employees, performing emergency drills, and considering solutions based on the situations they are likely to face.

Of course, you are welcome to contact N Ramsey Consultancy Ltd as an option to discuss your challenges and how we can work together to develop possible solutions to deal with your digital disruption and climate changes that affect your business.



Nicholas Ramsey

CEO, Consultant, ISO Trainer, IT Auditor at
N Ramsey Consultancy Ltd

Nicholas serves as the Chief Executive Officer (CEO) for N Ramsey Consultancy Ltd, which is a dynamic training and consultancy professional services company.

He is a Certified Trainer and professional in the following specialized areas: ISO 31000 Enterprise Risk Management (ERM), ISO 45001 Occupational Health & Safety Management System, ISO/IEC 27001 Information Security Management System, ISO 22301 Business Continuity Management System, and ISO 9001 Quality Management System.

Nicholas has over 20 years' of experience working at various organizations such as KPMG, AT&T, Microsoft, Apple, University of the West Indies (UWI), UWI-ROYTEC, and Deposit Insurance Corporation (DIC) to name a few, across the industries of: Financial Services, Information Technology, Auditing and Consultancy, Telecommunications, and Tertiary Education.

His degrees include a master's degree in Information Technology from Georgia Southern University (GSU) and a bachelor's degree in Communications from Florida Memorial University (FMU) in the USA.

Additionally, Nicholas is a Certified IT Auditor (ISO/IEC 27001 Lead Auditor credential), a Member of The Institute of Internal Auditors (IIA), an Associate Member of The Business Continuity Institute (AMBCI), and possesses ITIL V3 certification.

He is a former Tutor at the University of the West Indies at St. Augustine (UWI) and a former Lecturer at UWI-ROYTEC in Network Security, Systems Analysis & Design, Operating Systems Security, and End User Support. You can reach him at: www.nrconsultancyltd.com

THE EXPERT

Digital Operational Resilience in the AI Era

Aligning ISO/IEC 27001, Privacy, NIST CSF, and AI Management Systems for Global Trust

Today, all sectors and industries continue to face threats that impact digital trust and resilience (WEF Global Cybersecurity Outlook 2026 Report). Threats arising from:

- Digital transformation technologies
- Geopolitics
- Supply chains
- Inequity
- Environmental conditions
- Global financial operations

They affect boardrooms and strategic planning with a higher frequency year-on-year. This creates an atmosphere of constant maneuvering and innovation to achieve battle-tested resilience.

Risk is no longer solely traditional but a complex art of management, handled throughout the organization at all levels, especially by the ultimate risk owners – the board. In other words, digital liabilities have become an existential concern for all boards and executives, requiring governance, visibility, and oversight.

Thus, clearly defined and realistic risks have yet to be determined in the development, deployment, and utilization of AI. This creates emerging risks. Stated differently, a **Silent Risk** with uncertain outcomes, making themselves visible slowly, affecting all stakeholders – politically, economically, socially, culturally, and in some cases, physically.

Evolution of Digital Operational Resilience in the Age of AI

As AI transforms all sectors and industries, it can create emerging risks. If enterprise risk management activities and processes are not carried out thoroughly, this can lead to breaches and other business risks. Such risks can be operational, reputational, and trust losses; legal ramifications; and compliance risks related to sectoral, state, and global standards, regulations, and laws.

Risk to AI – Prompt Injection, data poisoning, software supply chain, and cybersecurity attacks affecting availability; affects AI trustworthiness.

That said, the world has been digitally transformed into an interconnected ecosystem. A vulnerability in a digital hub (such as; supply chains, digital public infrastructure, cloud services, e-commerce, or automation) that is exploited can cause ripple effects across the entire ecosystem.

This can impact digital economies, products, services, applications, privacy, and safety. The need for mature, capable digital operational resilience (DOR) becomes even more crucial with the integration of AI. AI trustworthiness, being lawful, ethical, and robust, is essential in digitally transformed systems to mitigate risks these systems can pose to national security, market stability, societies, and stakeholders' confidence.

Principles for responsible stewardship of trustworthy AI (OECD AI Principles):

- Inclusive growth, sustainable development, and well-being
- Respect for the rule of law, human rights and democratic values, including fairness, and privacy
- Transparency and explainability
- Robustness, security, and safety
- Accountability

Untrustworthy AI creates risks to resilience through misinformation and disinformation, loss of automation integrity and safety, used as a cyberattack tool, and its impact on data protection and privacy.

The implications of AI risks are too significant to ignore and cannot be addressed solely within the ICT domain. As stated before, it is a **strategic issue** that can only be resolved at the board level.

Governance and the Growing Complex Global Regulatory Landscape

DOR has evolved in the AI era. AI enhances the many resilience mechanisms vital to the DOR program excellence, such as Cyber and Digital Resilience, Preparedness and Testing, and Risk Management. But risks associated with AI can put such programs in jeopardy, thereby reducing organizational resilience. Risks such as data breaches, financial fraud, data protection and privacy losses, and safety issues pose challenges within the organization's complex compliance environment. This requires strategic oversight and resolution through risk management and resilience activities.

Practical Governance

To mitigate AI risks to DOR, AI governance is essential; to this end, boards must establish AI and digital risk management charters and policies to build, secure, and safeguard resilience. Moreover, establishing key committees (recognizing that they cannot do this alone and will need the necessary capacity to build resilience) will be essential to enhance visibility and oversight of continuous policy and methodology updates.

Data Governance

The outcomes of AI rely entirely on the quality of data used to train Large Language Models (LLMs). Moreover, data used to train LLMs must be managed responsibly, with stakeholders' consent and safeguards against breaches and unauthorized modifications that could lead to privacy issues and harm. Thus, data risk management policies should be instituted to enhance data risk management and resilience activities and processes surrounding AI development, deployment, and utilization.

Ethical AI, Transparency, and Global Trust Requirements

Trustworthy AI is vital in its integration into financial, political, and social ecosystems. Building trust in AI requires management systems that enable organizations to govern safely and effectively, manage risk, and continuously monitor the development, deployment, and utilization of AI. Through these activities, digital trust and resilience are gained across the evolving global digital landscape.

Cross-Border Regulatory Harmonization Challenges

Digital transformation technologies, such as AI, may not be sovereign and may span multiple jurisdictions with varying regulations and laws. Complicating this evolving compliance landscape are data protection laws and stakeholders' privacy requirements (OECD AI, Data Governance, and Privacy). Therefore, the risk to DOR from non-compliance must be owned and governed by the board to ensure resilience, reputation protection, and avoid legal, financial, and operational losses.

Evolving Threat Landscape: Why Traditional Frameworks Fall Short

The threat landscape continues to evolve due to business requirements, technology footprints, ecosystems, and usage patterns. This landscape is further complicated by AI and its supporting ecosystems: cloud technologies, networks, hardware, and software supply chain, LLMs, etc. If anything, the evolving threat landscape is riskier; Shadow-AI and AI trustworthiness add further complexity, requiring tools, techniques, approaches, and analyses beyond those used in traditional risk mitigation.

Resilience of Cloud, Third-Party, and Supply Chain Ecosystems

AI integration across all ecosystems not only enhances operations but also can significantly improve efficiency, scale-just-in-time. For instance, JUSDA's VMI-JIT vendor service solution, with complete integration into supply chain ecosystems, has been shown to foster adaptation and resilience. Alternatively, these AI-driven automated systems could significantly affect operations due to AI risks. As AI integration dominates these critical ecosystems, global digital economies and global security are at risk.

Incident Response (IR) in an AI-Augmented Environment

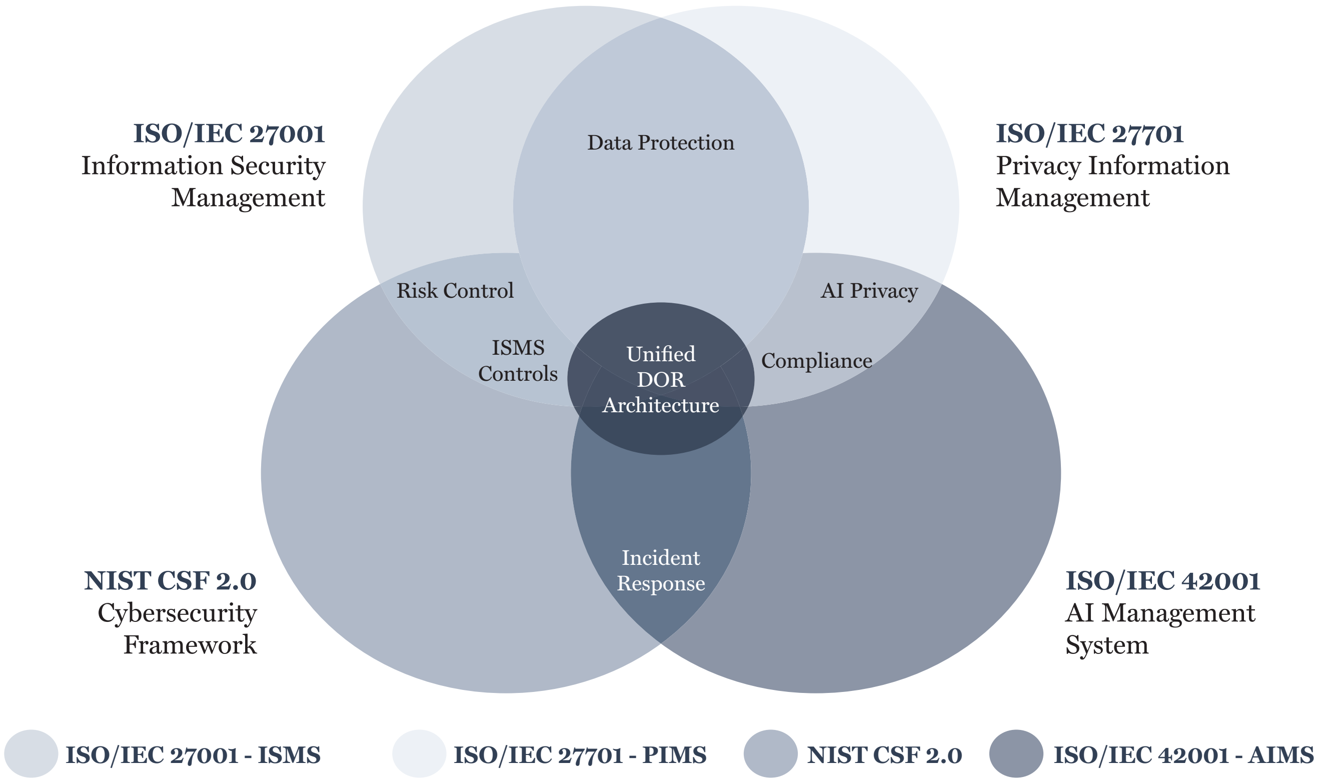
The use of AI in IR can enhance the efficiency of all aspects of IR activities and processes. AI can automate the preparation, detection and analysis, containment, eradication and recovery, and post-incident phases. However, AI risks can lead to false positives and false negatives, as well as other issues, such as data hallucinations, which can delay IR activities due to process verifications. Uncertainties in IR processes will erode IR teams' confidence in IR activities and reporting.AI-enabled cyber-attacks are executed at high speed (Palo Alto Unit 42 report), requiring AI defenses to counter them. There is also a complex issue of AI systems causing incidents that necessitate the evolution of IR activities and processes.

AI-Specific Security Controls

AI-integrated automated ecosystems require AI-specific controls to ensure trustworthy AI operations and processes. These control objectives (ISO/IEC 42001 Annex A) are around AI risk assessment, governance and accountability, data quality and privacy, human oversight, monitoring and logging, lifecycle documentation, and continuous improvement. It is crucial that management systems evolve to manage AI risks.

Framework Convergence: Where ISO/IEC 27001, Privacy, NIST CSF, and AI Management Systems Are Unified

Achieving DOR excellence in the era of AI requires the convergence of global standards and frameworks. Standards such as the ISO/IEC 27001 can serve as the foundational structure for the information security management system, the ISO/IEC 27701 forming the foundational structure for data privacy and protection, operationalized through the NIST CSF 2.0 and made trustworthy through ISO/IEC 42001.



Achieving DOR Excellence in the Era of AI
Convergence of Global Standards and Frameworks for a Unified DOR

- **ISO/IEC 27001 – Information Security Management System (ISMS)**
  Benefits: Provide the framework to protect the confidentiality, integrity, availability, and privacy of organizational assets and data, including entrusted data from clients, customers, etc. It improves information security through awareness and audits, measurement mechanisms that provide KPIs for management system effectiveness, and risk-based approaches to communicating suggested actions for improvement.

  It also provides good governance through extensive board oversight and strategic direction, while ensuring compliance with laws, regulations, and industry standards. In addition, it helps build the organization's reputation by adhering to strict security as an organizational value. Lastly, it can generate revenue through reduced breaches, efficient security management and operations, and business opportunities stemming from its security reputation.

- **ISO/IEC 27701 – Privacy Information Management System (PIMS)**
  Benefits: strengthens the ISO/IEC 27001 security program by adding structured privacy information management controls that help organizations manage personal data responsibly and transparently. It supports compliance with global privacy laws, such as GDPR, improves governance of personal data, and clarifies roles for controllers and processors.

  A key benefit is improved management of cross border data transfers, as the standard provides documented processes, accountability measures, and privacy controls that help demonstrate adequate protection when data moves across jurisdictions. It also enhances privacy risk management, builds trust with customers and regulators, and provides a repeatable, auditable framework that integrates well with broader security and AI governance systems.

- **NIST CSF 2.0**
  Benefits: Provides a flexible, risk-based framework that helps organizations efficiently meet multiple regulatory requirements (like HIPAA or FISMA) through a common cybersecurity language. It enables proactive gap identification, resource prioritization, and continuous improvement, reducing audit costs and building trust with regulators and customers.

- **ISO/IEC 42001 – AI Management System (AIMS)**
  Benefits: Provide the framework for developing, deploying, and utilizing trustworthy AI (OECD principles for trustworthy AI), through the realizations of AI governance and risk management, improving stakeholder trust and reputation, regulatory compliance and preparedness, operational efficiency and cost savings, competitive advantage, and integration with existing management systems. In essence, the AIMS standard builds AI digital trust and resilience.

DOR PILLAR	ISO/IEC 27001:2022 ISMS	ISO/IEC 27701:2025 PIMS (ED. 2)	NIST CSF 2.0
01 **Governance and Leadership**	**Cl.5, 6** Board-mandated ISMS scope, security policy, roles, and responsibilities, risk ownership	**Cl.5, 6** PII processing policy; privacy roles and responsibilities; board-level privacy governance charter	**GV.OC, GV.RM** Organizational context, risk strategy, cybersecurity roles aligned to board direction
02 **Risk Management**	**Cl.6.1, A.8** Information security risk assessment and treatment; asset, vulnerability, and threat management	**Cl.6.1, 7.2** Privacy risk assessment for AI data processing; PII controller/processor obligations integrated into enterprise risk register	**ID.RA, ID.AM** Risk identification, asset inventory, threat intelligence integration, risk register
03 **Data Governance and Privacy**	**A.8.2, A.8.3** Data classification, handling, and retention; privacy controls aligned to GDPR & sectoral law	**Cl.6–8, Annex A (PII controllers) / Annex B (PII processors)** PII inventory & categorization; lawful basis for processing; data minimization; consent management; DPIA requirements; cross-border transfer mechanisms	**ID.AM-5, PR.DS** Data asset management, data-at-rest and in-transit protectio data integrity
04 **Security Controls**	**A.5–A.8 (93 controls)** Access control, cryptography, physical security, network security, SDLC	**Cl.6–8, Annex A & B** Privacy-by-design controls; PII access restriction; pseudonymization; data subject rights management	**PR.AC, PR.IP** Identity & access management, awareness training, data securi protective technology
05 **Threat Detection and Response**	**A.8.15, A.8.16** Security monitoring, event logging, incident detection, and reporting (ISO/IEC 27035:2023)	**Cl.9.1, Annex A & B** Privacy incident detection; breach notification obligations (GDPR Art. 33/34); PII-related incident classification	**DE, RS** Continuous monitoring, anoma detection, incident response planning, and execution

	ISO/IEC 42001:2023 AIMS	UNIFIED DOR OUTCOME	EXAMPLE ARTEFACTS AND ACCOUNTABLE OWNER
	Cl.5, A.2 AI governance charter, board AI risk ownership, accountability for AI development and deployment	Unified board-level governance policy covering information, cyber, privacy, and AI risk with clear accountability at every level	Digital and AI Risk Governance Charter Board-approved Information Security Policy AI Ethics and Acceptable Use Policy RACI matrix for risk ownership; Board risk committee ToR **Owner: Board / CRO**
	Cl.6, A.1 AI-specific risk assessment: bias, hallucination, prompt injection, data poisoning, model drift	Integrated enterprise risk register combining ICT, privacy, cyber, and AI-specific risks — continuously updated with named owners	Unified Enterprise Risk Register (ICT + AI) AI Risk Assessment Reports (per system) Statement of Applicability (SoA) Risk Treatment Plan; Threat intelligence feed log **Owner: CRO / CISO**
n,	**A.3, A.4** Training data quality, consent management for AI training data, data provenance, LLM data lifecycle governance	Single data governance framework ensuring quality, privacy, and integrity of data across ICT and AI systems — from collection through LLM training to disposal	Data Classification & Handling Policy Data Protection Impact Assessment (DPIA) Training Data Provenance Register Data Retention and Disposal Schedule; Consent management records **Owner: CDO / DPO**
ty,	**A.6, A.7** AI-specific controls: adversarial input defense, model access control, output filtering, human oversight	Layered security controls covering people, processes, technology, and AI systems — mapped to a unified control library with Zero Trust as the governing architecture	Unified Control Library (ISO 27001 + AI controls) Access Control & IAM Policy; AI Model Access and Authorization Matrix Zero Trust Architecture design document Security Awareness Training records **Owner: CISO / Head of Cyber Security**
aly	**A.8, A.9** AI system monitoring, model degradation detection, and AI false positive/negative management	AI-augmented SOC with unified detection and response covering cyber events and AI system incidents — single classification, single escalation path	Incident Response Plan (IRP) AI Incident Classification Runbook SOC Monitoring and Alerting Playbooks Post-Incident Review (PIR) reports; SIEM/SOAR configuration baseline **Owner: CISO / SOC Director**

DOR PILLAR	ISO/IEC 27001:2022 ISMS	ISO/IEC 27701:2025 PIMS (ED. 2)	NIST CSF 2.0
06 **Resilience and Recovery**	**A.8.13, A.8.14** Information backup, redundancy, business continuity integration (ISO 22301:2019)	**Cl.9.1, Annex A & B** Privacy continuity: ensuring PII availability and integrity through recovery; data subject rights preserved post-incident	**RC.RP, RC.CO** Recovery planning, post-incident improvements, and communications during disruption
07 **Supply Chain and Third Parties**	**A.5.19–5.22** Supplier security policy, supplier agreements, ICT supply chain risk monitoring	**Cl.6–8, Annex A & B** Third-party PII processor agreements; sub-processor oversight; data transfer mechanisms (SCCs, BCRs); vendor privacy audits	**ID.SC** Supply chain risk management supplier identification, contract monitoring, response plans
08 **Transparency and Ethical AI**	**Cl.7.4** Stakeholder communications, audit and review reporting	**Cl.7.4, 9.1, Annex A** Privacy notices & transparency reports; data subject rights fulfilment (access, erasure, portability); OECD privacy principle alignment	**GV.SC, RS.CO** Transparency in cybersecurity posture, stakeholder communications, incident disclosure
09 **Compliance and Audit**	**Cl.9, 10** Internal audit, management review, nonconformity and corrective action, PDCA	**Cl.9, 10** Privacy audit cycle; GDPR/ CCPA conformity assessment; PII processing register review; corrective action for privacy nonconformities	**GV.OC-5, ID.RA** Regulatory alignment (DORA, GDPR, SEC, HIPAA, FISMA), cross-framework gap analysis
10 **Measurement and Maturity**	**Cl.9.1** KPIs for ISMS effectiveness, security metrics, and management system performance evaluation	**Cl.9.1, Annex A & B** Privacy KPIs: DPIA completion rate, data subject request response times, breach notification timeliness, PII processing accuracy metrics	**GV.RM-7, ID.RA** Cybersecurity KRIs, maturity tiers (Partial to Adaptive), and continuous improvement targe

	ISO/IEC 42001:2023 AIMS	UNIFIED DOR OUTCOME	EXAMPLE ARTEFACTS AND ACCOUNTABLE OWNER
	A.10 AI continuity: model rollback, retraining pipelines, fallback to non-AI processes, recovery testing	Adaptive recovery architecture — rapid restoration of ICT and AI capabilities with self-healing, post-incident learning, and PII rights preserved throughout	Business Continuity Plan (BCP) Disaster Recovery Plan (DRP) AI System Continuity and Rollback Procedure BCP/DRP test and exercise reports; RTO/RPO register per critical system **Owner: Chief Resilience Officer / CTO**
: ts,	**A.5, A.6** Third-party AI model risk, open-source LLM due diligence, AI vendor accountability, and audit rights	End-to-end supply chain resilience spanning hardware, software, cloud, and AI model providers — single assessment cycle, unified contract clauses	Third-Party Risk Assessment (TPRA) register Supplier Security and AI Clauses (contracts) AI Vendor Due Diligence Questionnaire Critical supplier tiering matrix; Supply chain incident log **Owner: Chief Procurement Officer / CISO**
	A.2, A.11 Explainability, bias management, OECD AI Principles compliance, human oversight, impact assessment	Stakeholder trust program, transparent AI disclosures, ethical review boards, bias audits embedded in governance	AI Transparency and Disclosure Statement Algorithmic Impact Assessment (AIA) AI Ethics Review Board minutes Bias audit reports (per model); Stakeholder trust and communication plan **Owner: Chief AI Officer / Chief Ethics Officer**
	Cl.9, 10 AI system audits, EU AI Act conformity assessment, AI risk review cadence, continuous model evaluation	Integrated GRC program, single audit cycle, harmonized evidence, unified regulatory compliance across DORA, GDPR, EU AI Act, SEC, Basel III	Integrated GRC Program Plan Internal Audit Schedule and Reports EU AI Act Conformity Assessment record Regulatory compliance matrix (DORA, GDPR, etc.); Corrective Action Register **Owner: Chief Compliance Officer / Internal Audit**
ts	**Cl.9.1, A.12** AI performance metrics, trustworthiness KPIs, model accuracy/drift thresholds, improvement cycles	Unified UDOR Maturity Model — board-level KPI/KRI dashboard spanning security, privacy, cyber, and AI dimensions with a single aggregate maturity score	DOR Maturity Assessment Report Unified KPI/KRI Dashboard (board-level) AI Model Performance Scorecard Continuous Improvement Log; Annual Management Review Report **Owner: CRO / CISO / Chief AI Officer**

Practicality of Resilience: The Unified Framework Architecture

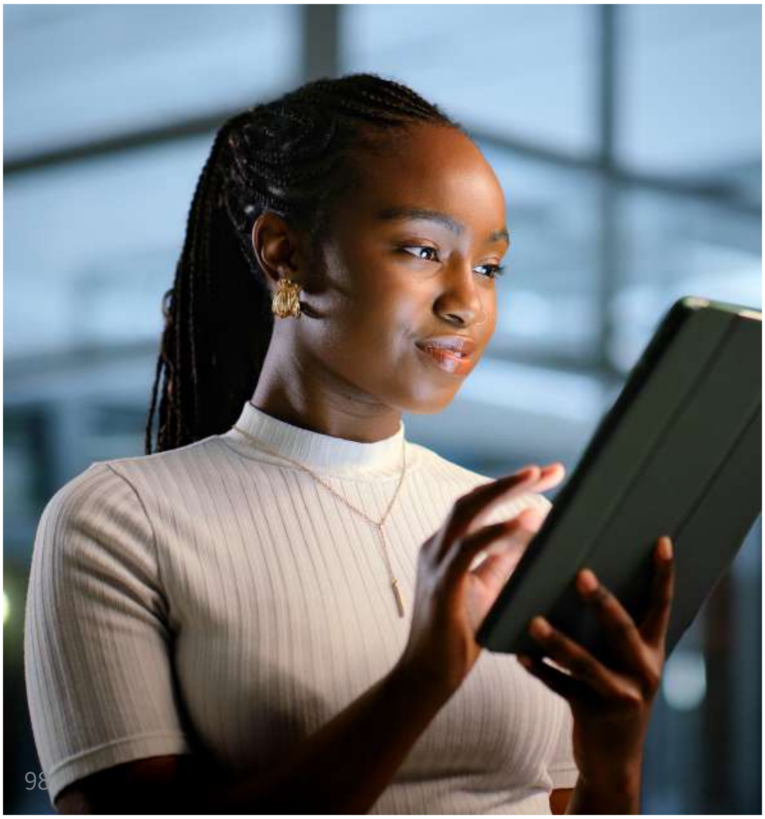
The evolving complexities of digitally transformed ecosystems necessitate a convergence of frameworks to achieve operational excellence. When frameworks' lifecycles change occurs every couple of years, their combined unified application can build, safeguard, and sustain digital trust and resilience globally.

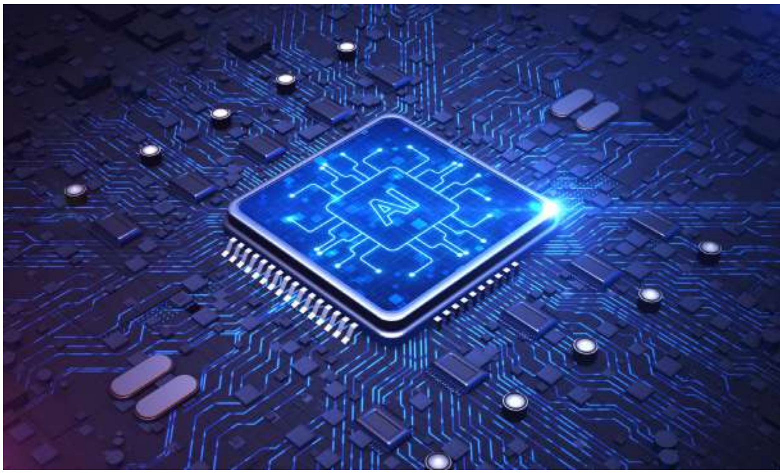
Zero Trust and AI Systems

One of the key components in building DOR excellence is trust. In today's systems, implicit trust can create grave security risks. Therefore, identity and access management strategies are vital to systems, especially AI-integrated ones. AI tools, agents, services, etc., should not have free rein over networks and access to data. Implementing the necessary **unified** trust policy will reduce AI-related risks, build compliance, and strengthen stakeholder confidence in the organization's transparent use of AI.

Automation for Cyber and Operational Resilience

Through the unified framework that assures trustworthy AI to strengthen Cyber and Operational Resilience, the organization will be able to respond to, adapt to, and recover from risks more quickly and efficiently. It can also provide self healing capabilities and continuous improvement without compromising trust or operability.





Global Compliance Landscape for AI and Operational Resilience

The current digital threat landscape in the AI era underscores the need for risk mitigation to safeguard all stakeholders. Hence, in the global compliance landscape, there are many laws, regulations, and standards to adhere to, such as the GDPR, EU AI Act, DORA, the USA SEC, Basel III, ISO/IEC 27001, ISO/IEC 42001, ISO 22301, and NIST CSF 2.0. The aim of compliance, in essence, is to safeguard and sustain resilience in digital environments, protecting business integrity and benefiting mankind.

Regrettably, these varying laws, regulations and standards create an evolving, complex compliance environment that requires a **unified** global harmonized framework. This framework should be committed to reducing fragmentation and improving the efficiency of business activities, and societal well-being and development.

Supply Chain Resilience

The security of global supply chains is essential to today's digital economies and societies. Global supply chain risks can trigger existential events that create national security issues, including threats to people's safety and well-being.

Therefore, with the integrated use of AI and AI automation in supply chains, and their associated risks, it is vital for supply chains to adapt, respond, and recover efficiently and effectively to sustain digital trust and resilience. Attaining this outcome requires a **unified** framework.

Measurement and Maturity Models

Achieving resilience with a "compass" is vital in reducing compliance risk and building reputation and market confidence. KRIs and KPIs are crucial to resilience programs and vital for monitoring and continuous improvement.Therefore, developing and implementing an operational resilience maturity model that utilizes a **unified** framework will sustain and mature DOR excellence in today's AI era.

UNIFIED DIGITAL OPERATIONAL RESILIENCE MATURITY MODEL

	CHARACTERISTICS	OUTCOMES
RESILIENT-BY-DESIGN	✓ Continuous assurance & autonomous controls ✓ Al Zero-Trust security architecture ✓ Explainable & monitored Al systems ✓ Automated resilience & self-healing	✓ Proactive threat anticipation ✓ Global Al governance & trust
ADAPTIVE & SECURE	✓ KRI & KPI Metrics ✓ Al failure simulations ✓ Recovery & retraining	✓ Rapid adaptation to threats ✓ Predictive resilience posture
OPTIMISED /IMPROVING	✓ Continuous improvement cycle ✓ Unified lessons backlog ✓ Telemetry & drift monitoring	✓ Data-driven optimisation ✓ Cyber & Al harmonization
MANAGED & INTEGRATED	✓ ISO 27001:2022, ISO 27701:2025, NIST CSF 2.0 & ISO 42001:2023 ✓ SOC & Al incident response ✓ Unified risk register	✓ Enterprise resilience program ✓ Consistent risk management
DEFINED & PROACTIVE	✓ Cyber & Al policies defined ✓ CI / SDLC lifecycle controls ✓ Data integrity oversight	✓ Privacy Risk Management ✓ Cross-Border Data Protection
REACTIVE / BASIC	✓ Basic 27001 controls ✓ Partial IT alignment ✓ Limited Al oversight	✓ Reactive & limited visibility ✓ Fragmented controls
INITIAL / AD-HOC	✓ Fragmented controls ✓ Ad-hoc processes ✓ No risk management	✓ High exposure & fragility

Conclusion: Toward a Unified Global Resilience Architecture

The convergence of ISO/ 27001, ISO/IEC 27701, NIST CSF 2.0, and ISO/IEC 42001 offers organizations, governments, agencies, and NGOs a unified framework that will meet the governance, security, privacy, and AI trustworthiness demands of the AI era. Trust and resilience across all ecosystems, digital and physical, can only be assured through standardization, attestation, and certification. This direction demands strong leadership, accountability, and social responsibility.

It is, therefore, clear that this is not primarily a technological challenge but a governance one. Through board awareness, charters, unified policies, and trust and resilience activities assessed against the 10 pillars of the Unified DOR Maturity Model, organizations can build digital ecosystems that deliver the global trust and resilience the world now requires and deserves.

Edward M

Founder and Mar
CariSec G

Edward Millington (**BSc, CISSP, CI**
is a distinguished cybersecurity and dig
decades of experience. His esteemed c
certifications and his peer-elected statu
of Information Security (FCIIS), unders
advisor and principal security consulta
and compliance (GRC), enabling dive
and telecommunications to achieve stra
resilience.

A prolific thought leader, Milling
landscape. He is a Commonwealth Carib
Expert, regularly keynoting at inte
influential publications on AI govern
leadership as Founder and Managing D
and Chairman of the Caribbean Charte
Security, and Co-Chair of regional in
to advancing cybersecurity policy and
Caribbean.

Globally recognized, Millington's ca
innovation and foster development. His
and strategic foresight positions him as
complex frameworks into actionable
organizations for the future's digital thre

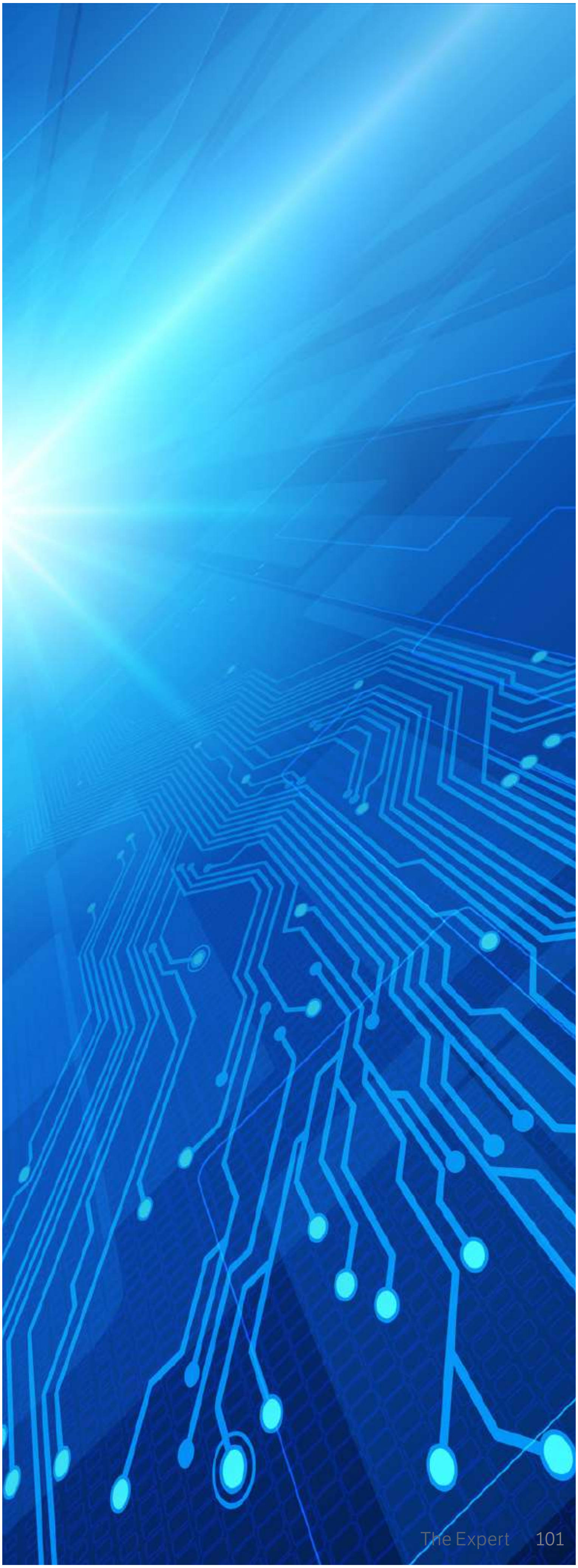
Millington

naging Director of
Global Inc.

ISO, SOC 2, ISO, FCIIS, CCCF, MIET)
gital resilience expert with nearly three
credentials, including CISSP and CISO
as as a Fellow of the Chartered Institute
score his authority. As a strategic board
ant, he specialises in governance, risk,
erse sectors like banking, government,
ategic objectives while bolstering cyber

gton shapes the global cybersecurity
bean Cyber Fellow and an EU CyberNet
rnational conferences and authoring
nance and operational resilience. His
Director of CariSec Global Inc., Founder
ered Institute of Cyber and Information
itiatives, underscores his commitment
capacity-building, particularly in the

reer is a testament to his ability to drive
unique fusion of deep technical acumen
s a trusted advisor, expertly translating
e business outcomes and preparing
eats.



Our Upcoming Certified IPC Management Systems Auditor (CMSA) Training in French

Advance your auditing expertise with the IAS-accredited CMSA training, aligned with the IPC scheme. This 3-day online course equips you with the skills to plan, conduct, and report on management system audits in accordance with ISO 19011 and ISO/IEC 17021-1.

May 4–6, 2026 | 12:00PM –19:00PM (CEST)
Online | Language: French

Includes training materials, certification exam (with free retake), auditor behavior exam, and certificate.

Gain access to PECB Connect and unlock global audit opportunities.

SECURE YOUR SEAT →

PECB CONNECT

TECHNOLOGY

The Human Element: Reskilling and Re-Architecting for an AI-First Future

The era of “paper governance” is over.

For many years, organizations approached artificial intelligence using a familiar governance playbook: draft a set of ethical AI principles, publish a high-level policy document, and perhaps establish a cross-functional committee to review initiatives. While these steps created the appearance of responsible oversight, they rarely translated into operational reality.

As we move through 2026, the gap between having an AI policy and having a truly AI-governed organization has become significant. The emergence of complex AI ecosystems, autonomous decision-making tools, and AI-driven workflows means that governance must operate in real time rather than in theory.

With the maturation of **ISO/IEC 42001**, the world’s first management system standard specifically designed for artificial intelligence, organizations are beginning to shift their perspective. The focus is no longer limited to defining what responsible AI should look like. Instead, attention is moving toward how organizations can empower their people, processes, and structures to actually implement those principles in everyday operations.

True AI governance is not found in a policy document or a static framework. It exists in the cognitive readiness of the workforce, the accountability embedded within teams, and the organization’s structural agility.

The Illusion of Documentation

In the early 2020s, governance was often treated as a compliance checkbox. Organizations rushed to publish AI ethics guidelines or internal policies to demonstrate responsibility. While well-intentioned, many of these efforts remained largely symbolic. Documentation alone does not govern AI systems.

Today’s AI technologies, particularly agentic systems and autonomous workflows, operate dynamically. They adjust outputs based on new inputs, adapt to changing data patterns, and influence operational decisions at a speed no human committee can match.

When an AI system is managing customer interactions, analyzing financial risks, or adjusting supply chain forecasts in real time, a static document stored in a shared drive cannot intervene.

Governance must, therefore, evolve from static to living. Instead of existing solely in written policies, governance must be embedded in daily operational behavior. This requires a shift from **“Gatekeeper Governance”** to **“Distributed Governance.”**

In traditional governance models, a small group, often within legal, compliance, or IT, served as the central authority responsible for approving or rejecting AI initiatives. While this structure may have worked when AI adoption was limited, it cannot scale to environments where AI tools are integrated across every department.

Distributed governance recognizes that AI risk does not exist solely in technical systems. It also exists in how employees interact with those systems. Every employee who uses an AI tool becomes part of the governance framework. In other words, governance must become a shared organizational capability rather than a centralized control function.

The New Governance Hierarchy

To understand how governance can operate effectively in AI-driven environments, it helps to view it as a layered system.

1. **Policy Layer: The "What"**
   The policy layer defines the principles and obligations that guide AI use within the organization. This includes regulatory requirements, ethical commitments, risk tolerance thresholds, and high-level governance policies. These documents remain important. They establish the boundaries within which AI systems should operate and provide the foundation for compliance with regulations and industry standards. However, policies alone cannot ensure accountability.

2. **Architectural Layer: The "How"**
   The architectural layer translates policy into operational design. This includes data pipelines, workflow structures, monitoring systems, feedback loops, and system controls that ensure AI outputs remain aligned with organizational expectations.

   Here, governance becomes part of the technical and operational architecture of the organization. Risk thresholds can be monitored automatically. Data governance controls can restrict inappropriate data usage. AI outputs can be evaluated through continuous validation mechanisms. Without this architectural layer, policies remain theoretical.

3. **Human Layer: The "Who"**
   The most critical layer is the human one. Even the most sophisticated governance architecture requires human judgment to interpret outcomes, evaluate context, and make final decisions. This layer focuses on three essential capabilities:

   - AI literacy
   - Critical judgment
   - Accountability

   Employees must understand how AI tools function, recognize potential risks, and feel empowered to intervene when necessary. Governance ultimately depends on people who are capable of questioning the outputs generated by intelligent systems.

Re-Architecting the Organization: From Silos to Synapses

Traditional organizations were designed for human-to-human workflows. Tasks moved sequentially between departments, each responsible for a specific stage of a process. However, AI-enabled environments introduce entirely new forms of interaction:

- Human-to-machine collaboration
- Machine-to-machine communication
- Hybrid decision-making processes

As a result, organizations must rethink how their internal structures function. Instead of rigid departmental silos, organizations increasingly resemble interconnected neural networks—where information flows continuously between people, systems, and AI agents. This structural transformation requires re-architecting workflows in several key ways.

The End of Linear Workflows

Most traditional business processes follow a linear pattern: Step A leads to Step B, which leads to Step C. These workflows assume that decisions occur at predictable stages. AI-augmented workflows are fundamentally different. They are cyclical rather than linear. AI models continuously evaluate new data, update predictions, and adjust outputs.

This creates feedback loops rather than fixed sequences.

For example, an AI-driven fraud detection system may analyze transactions in real time, update its risk models based on new patterns, and continuously refine its detection parameters. Human oversight must therefore operate within an ongoing monitoring framework rather than at a single approval stage. To manage this complexity, organizations must introduce **"circuit breakers"** into their processes.

Circuit breakers are governance mechanisms that pause or escalate AI-driven decisions when predefined thresholds are exceeded. These thresholds might include unusual prediction patterns, abnormal data inputs, or increased uncertainty in model outputs.

By embedding such controls into workflows, organizations ensure that AI systems remain accountable even as they operate autonomously.

The Rise of the AI Management System

The emergence of **ISO/IEC 42001** represents a significant shift in how organizations approach AI governance. Rather than treating AI as a standalone project or experimental initiative, the standard encourages organizations to manage AI through a structured **AI Management System** (**AIMS).**

This approach mirrors how organizations manage other critical operational domains. For example, established information security as a systematic organizational responsibility rather than a purely technical concern.

Similarly, ISO/IEC 42001 encourages organizations to integrate AI governance into existing management systems. Under this framework, AI oversight becomes part of:

- Risk management
- Operational planning
- Internal auditing
- Continuous improvement processes

The objective is to transform AI from a specialized IT function into a business-wide discipline.

The Human Element: Reskilling Beyond Prompting

Much of the current conversation around AI skills focuses on **upskilling**. Employees are encouraged to learn how to use AI tools more effectively—often through training programs focused on prompt engineering or tool usage.

While valuable, these skills represent only a small portion of what organizations truly need. AI-first governance requires **reskilling**, not merely upskilling. Upskilling helps employees perform their existing tasks more efficiently using AI tools. Reskilling prepares employees for entirely new roles within AI-augmented organizations.

In this environment, the primary value of human workers increasingly shifts from performing tasks to orchestrating intelligent systems. Employees become supervisors, evaluators, and coordinators of AI capabilities.

The Critical Three Skills for the AI Workforce

To move beyond policy-driven governance and toward operational accountability, organizations must cultivate three essential skills across their workforce.

Algorithmic Judgment

Algorithmic judgment refers to the ability to evaluate AI outputs critically. An AI model may produce results that are statistically valid but contextually incorrect. For example, a recommendation algorithm may identify a trend based on historical data that no longer reflects current circumstances.

Employees must therefore be able to ask critical questions:

- Does this output make sense in context?
- Could bias be influencing the result?
- Are there external factors the model may not have considered?

In AI-driven environments, human judgment becomes the final safeguard against flawed automated decisions.

Prompt Fluency and Signal Interpretation

Understanding how AI systems interpret inputs is another essential skill. Prompt fluency extends beyond writing effective queries; it includes understanding how AI models structure responses and where vulnerabilities may exist. This includes recognizing risks such as:

- Prompt injection attacks
- Manipulated training data
- Biased outputs
- Security vulnerabilities

Employees who understand how AI models interpret prompts are better positioned to identify abnormal behavior and potential risks.

Orchestration

The final critical capability is orchestration.

Managers in AI-enabled organizations increasingly oversee teams that include both human employees and specialized AI agents. Each AI system may perform a specific function—data analysis, predictive modeling, document generation, or customer interaction.

The role of human leaders is to coordinate these capabilities effectively. A manager's "team" may soon consist of a small number of human specialists working alongside numerous AI systems performing different tasks. Effective orchestration ensures that these systems operate cohesively and remain aligned with organizational goals.

The Hidden Risk: Passive Compliance

One of the most significant risks in AI adoption is not technological failure but human complacency. Employees may begin to treat AI outputs as inherently correct simply because they originate from a sophisticated system. Over time, this mindset can erode critical thinking.

The most dangerous scenario is not a malfunctioning algorithm but an organization where employees no longer question automated decisions. Maintaining a culture of critical engagement is therefore essential to responsible AI governance.

Operationalizing the Change

For auditors, governance professionals, and organizations implementing AI management systems, the challenge lies in verifying that governance is not merely theoretical.

Practical implementation requires structured evaluation of the human dimension of AI adoption.

Phase 1: AI Literacy Assessment

Organizations must first evaluate whether employees truly understand the AI systems they use daily.

Generic training sessions are insufficient. Instead, organizations should implement role-specific training programs that reflect the actual tools and workflows employees interact with.

One effective approach involves **AI "red-teaming" workshops**, where employees attempt to identify vulnerabilities in their own AI-supported workflows. This process encourages critical thinking and strengthens awareness of potential risks.

Phase 2: Structural Integration

Governance roles must also be clearly defined. While many organizations appoint AI ethics officers or governance committees, effective oversight often requires more localized accountability.

Roles such as **data stewards, AI interaction leads**, or **model monitoring specialists** can ensure that governance remains close to operational decision-making processes. When governance responsibilities exist within individual business units, organizations gain greater visibility into how AI systems actually function in practice.

Phase 3: Feedback Mechanisms and Intervention Points

Employees must also have accessible mechanisms for reporting AI concerns. In many organizations, reporting potential issues with automated systems remains unclear or overly technical. Establishing dedicated governance reporting channels allows employees to flag anomalies or questionable outputs without requiring technical expertise.

Just as organizations implement whistleblower policies to report unethical behavior, they must develop systems for reporting algorithmic concerns. These mechanisms create accountability while reinforcing the idea that governance is a shared responsibility.

Conclusion: Governance as a Human Capability

Ultimately, AI governance extends far beyond policy documents or regulatory compliance frameworks. At its core, governance is about trust. Trust cannot be established solely through documentation or disclaimers. It emerges from a combination of transparent systems, responsible leadership, and empowered employees.

Organizations that succeed in the age of AI will recognize that artificial intelligence is not merely a technological transformation. It is a human transformation.

The policies and frameworks organizations create may serve as maps.

But it is the workforce, the people interpreting signals, questioning outputs, and orchestrating intelligent systems, that ultimately determines whether AI systems operate responsibly.

As AI technologies continue to evolve, the organizations that thrive will be those that prioritize human readiness alongside technological capability. In the AI-first future, governance is not written; it is practiced.

Mike Boutwell

CISO, CAISO, Author, and PECB Certified Trainer

Mike Boutwell is a cybersecurity executive and consultant with over 15 years of experience in security leadership and a decade in risk management, specializing in financial services, manufacturing, and critical infrastructure.

He has successfully secured assets exceeding $1 quadrillion, led $100M+ transformation projects, and advised some of the most demanding organizations in the world, including Euroclear Bank, AT&T, Cisco, First Data, and Takeda Pharmaceuticals.

Mike partners with CISOs, CIOs, and boards to build security programs that are lean, risk-driven, and aligned with regulatory expectations and business objectives. He's particularly known for implementing ISO/IEC 27001, TISAX, CIS 18, and NIST frameworks, as well as transforming organizational cybersecurity cultures, especially in environments with limited resources or poor baseline maturity.

He holds notable certifications, including CISSP, CISA, CGEIT, ISO/IEC 27001 Senior Lead Implementer and Auditor, ISO 38500 Senior Lead IT Governance Manager, ISO/IEC 27032 Senior Lead Cybersecurity Manager, and is a Certified Non-Executive Director (NEDA) and Certified Information Security Trainer (PECB). He earned an MSc in Cybersecurity from Royal Holloway, University of London (2024), and a BSc in Computer Information Systems from California Polytechnic University (2008).

Mike is also the author of "Profit-Driven Cybersecurity" and "The Ransomware Handbook", a certified non-executive director, and a trusted advisor to companies developing cloud-native and AI-driven security products.

THE EXPERT

Security by Design Was Yesterday - Culture by Design Is Tomorrow

You cannot patch human behavior. But you can design for it.

For many years, cybersecurity has had a clear goal. Build secure systems, write secure code, design strong architecture, add controls, encryption, monitoring, and access management. If we did those things well, we believed the organization would be safe.

And humans were part of the discussion, but usually in a simple way. They were trained in security awareness once a year, shown slides about phishing, and sent a few reminder emails. Then we went back to what we thought really mattered. Systems.

That model worked reasonably well in a slower world, but it does not work anymore.

Today, the most important security control in many organizations is not written in code or configured in systems. It lives inside the decisions people make every day. Security by design is still essential. But it is no longer enough, especially now in the age of AI. What we need now is Culture by Design.

AI Has Shifted Where Risk Lives

Beyond introducing new tools, artificial intelligence changed where risk appears inside organizations. In the past, most security risks came from a few expected places: vulnerable software, misconfigured systems, external attackers, and phishing emails. Today, risk often appears somewhere far less obvious: inside everyday normal work.

A marketing employee uploads sensitive customer data to an AI tool to generate campaign insights. A developer asks an AI assistant to refactor internal code, unknowingly sharing proprietary algorithms. A finance employee uses an AI tool to summarize confidential contracts. None of these people are malicious. Most are simply trying to work faster. The problem is that in doing so, they may expose intellectual property, personal data, or information that should never leave the organization.

This is a behavior moment, not a technical failure. And that is exactly where traditional security programs struggle. Because awareness does not create behavior.

Awareness Does Not Create Behavior

Most organizations respond to new risks with awareness campaigns, posters, training videos or policy changes. These help people understand the problem, but they rarely change what happens in the moment that matters. When someone is under time pressure, awareness fades. Speed and convenience win. And AI tools make this even more acute, because they promise massive productivity gains, and nobody wants to be the one left behind.

So, if security depends on people remembering rules in stressful moments, the system is inherently fragile by design. This is why the conversation must change. We must design for secure behavior, not just secure systems. And this is where the discussion about organizational culture begins.

> ***“We must design for secure behavior, not just secure systems.”***

Where Security Ends and Culture Begins

There is an invisible line in every organization where technology stops controlling risk and behavior takes over. A company may have perfect access controls, but if employees regularly share credentials to save time, the control collapses. A company may have secure development pipelines, but if engineers paste sensitive code into external AI tools, intellectual property leaks. A company may even have strict data classification, but if employees upload files into AI tools to “see what happens,” data protection fails.

The systems were secure. The culture was not.

And in the age of AI, this boundary appears earlier and earlier in the process. That is why we must start thinking differently. Security cannot stop at architecture and system configuration. It must continue into the human realm of behavior.

Real AI Risks That Demand Cultural Change

Security by design is still one of the most powerful ideas in cybersecurity. But AI forces us to extend this concept so that security must move through the entire lifecycle, permeating architecture, development, operations, and human decision-making. When you hand someone a powerful tool, you also hand them the responsibility to use it safely. When organizations deploy AI assistants, the question is not only “Is the tool secure?” The question becomes “How will people actually use this tool?”

Here are some examples of what can happen and has happened. The following incidents are not theoretical. They are documented, named, and already on the record.

1. Source code exposure: Samsung, 2023

In March 2023, Samsung allowed engineers to use ChatGPT for coding tasks. Within twenty days, three separate incidents occurred: proprietary source code pasted in to fix bugs, confidential chip-testing code submitted for optimization, internal meeting transcripts uploaded to generate minutes. ChatGPT’s interface at the time used inputs to train its models. Samsung’s semiconductor IP was now on OpenAI’s servers with no way to retrieve it. No scanner detected this. No alert fired. The engineers were solving real problems with a tool that worked. The failure was entirely cultural.

2. AI-generated attacks at scale: WormGPT and FraudGPT, 2023

In July 2023, SlashNext uncovered WormGPT, a blackhat AI tool sold on hacker forums, trained on malware data, no guardrails. Researchers used it to generate a phishing email they described as remarkably persuasive. Weeks later FraudGPT appeared on dark web channels: phishing generation, malicious code writing, scam pages, $200 a month, over 3,000 confirmed sales within weeks. Attackers can now run personalized social engineering campaigns at a scale and quality previously impossible without skilled human writers. Your culture determines whether your people pause or click when one of those emails lands.

3. Decision integrity and legal liability: Air Canada, 2024

In 2022, a passenger consulted Air Canada’s chatbot about bereavement fares, received incorrect advice, followed it precisely, and was denied the refund he was promised. Air Canada argued to the tribunal that the chatbot was a separate legal entity responsible for its own actions. The tribunal called that argument remarkable before rejecting it entirely, ruling the airline liable for negligent misrepresentation. Organizations are accountable for every output their AI systems produce. That is now settled law.

Culture by Design

The traditional Security by Design principle still holds: build security in from the start, not as an afterthought. But in the age of AI it needs to expand its scope significantly. Your attack surface now includes the tools your engineers use to write code, the AI assistants your analysts use to summarize threat intelligence, the chatbots your customer service teams deploy, and the AI agents taking autonomous actions inside your environment. Each needs security thinking at the design stage, not the incident review stage.

And designing culture may sound abstract, but it is not. Just like systems architecture, culture can be engineered. It requires proactive intention and three critical elements.

Leadership signals: People watch what leaders reward. If speed and innovation are celebrated but safe behavior is ignored, employees will choose speed. If leaders openly discuss responsible AI use, teams follow. Culture is shaped by signals more than policies.

Friction in the right places: Good security design places friction exactly where risk appears. AI tools should include reminders about sensitive data. Data loss prevention tools should detect risky prompts. Internal AI systems should exist so employees do not need to reach for external ones. When friction appears at the right moment, behavior changes naturally.

Shared responsibility: Security teams cannot monitor every AI interaction. Responsibility must move closer to the people using the tools. Developers must understand data boundaries. Marketing teams must understand data classification. Finance teams must understand AI data risks. Security becomes a shared language, and that is a cultural aspect.

How to Implement Culture by Design

Think of cybersecurity today as three layers. Layer one is technology: encryption, identity management, monitoring, secure infrastructure. Layer two is process: policies, governance, compliance, risk management. Layer three is culture: daily behavior, decision-making, accountability.

In the past, organizations invested heavily in the first two layers. The third received awareness training once per year. But in the age of AI, the third layer may be the most important one, because AI moves power, and the risk, closer to individuals.

Below is a practical step-by-step approach that organizations can follow.

Step 1. Identify the Moments That Matter

The first step is mapping the exact situations where risky decisions occur during daily work. In the age of AI, these moments usually appear when employees paste data into AI tools, or upload documents for analysis, or generate code using AI assistants, or automate workflows using AI, or simply rely on AI output for business decisions. The goal is to identify where human decisions interact with powerful tools. Once these moments are known, the organization can design expected behavior around them.

Step 2. Define the Secure Behavior

After identifying the moments that matter, define clear and observable secure behaviors. Employees should know exactly what action is expected. Some examples can be that customer data must never be pasted into external AI tools, or internal source code can only be used with approved AI development assistants, or confidential documents must only be processed through internal AI platforms, or AI generated output must be verified before operational use. This removes ambiguity and makes behavior measurable.

Step 3. Make the Secure Choice the Easy Choice

Security culture fails when secure behavior is harder than insecure behavior. Employees will choose convenience by default, so organizations must, therefore, create environments where the secure path is also the fastest most convenient path. Some examples can be the deployment of internal AI assistants for coding, writing, and analysis, or the integration of AI tools directly into development environments, or providing internal easy access document analysis AI systems, or automatically warning users when sensitive data is pasted into prompts. When employees have safe tools conveniently available, risky behavior decreases naturally.

Step 4. Send Clear Leadership Signals

Security culture spreads through leadership behavior. Employees pay attention to what leaders praise, ignore, or tolerate. Leadership must unanimously communicate that responsible AI usage is part of professional conduct. Practical actions include leaders openly discussing AI risk and responsibility, or including responsible AI usage in performance discussions, or supporting teams that pause work due to security concerns related to AI risks. When leadership signals are visible, culture becomes consistent.

Step 5. Embed Security Signals in Daily Work

Security reminders should appear in the tools employees already use. This makes secure thinking part of normal work. Examples include prompts warning about sensitive data before AI submission, or security reminders inside developer AI tools, or classification tags visible when documents are uploaded. These signals influence behavior at the moment decisions are made.

Step 6. Encourage Safe Experimentation

Employees are curious about AI. That curiosity should be guided, not suppressed. Organizations should provide safe environments where experimentation can happen without exposing sensitive data. Examples of this can be found in AI sandboxes using synthetic data, or experimentation environments isolated from production systems, or internal AI communities sharing best practices. This reduces shadow experimentation.

Step 7. Measure Real Behavior

As mentioned before, awareness metrics alone do not reflect culture. Not even close. Organizations must measure observable actions. Behavior based measurements show how people actually interact with AI tools. Examples of measurable actions can be found in attempts to paste restricted data into AI tools, or number of risky prompts detected, or adoption of secure AI platforms, or security incident reports related to AI usage. These metrics reveal whether culture is improving.

Step 8. Treat Culture as a Security Control

The final step is recognizing that culture is not an abstract idea. It is a security control layer. It should be managed like other controls such as vulnerability management or identity security. That means assigning ownership, defining measurable outcomes, reviewing metrics regularly and improving your culture programs continuously. When culture is treated this way, it becomes part of the security architecture.

The Control We Forgot to Design

We have spent decades getting very good at protecting things. Systems, networks, code, data. We built walls, designed gates, wrote rules into architecture, and called it security. And in many ways, it worked.

But there was always a part of the organization we never fully designed for. The part that does not run on servers. The part that shows up at 9 am, checks its messages, opens a browser, and starts making decisions. We knew it was there. We just assumed awareness would be enough to keep it safe.

It was never enough. We just did not feel the gap until AI made it impossible to ignore.

Now the question is not whether your systems are secure. Most of them probably are. The question is whether your culture is. Whether the people inside your organization understand the power the tools in front of them carry, and whether they have been genuinely prepared to use that power responsibly. Not trained. Prepared. There is a big difference, and it is that difference that matters now.

The organizations that treat culture as a design problem, with the same intention and rigor they bring to their technical architecture, will be the ones that are genuinely resilient in the years ahead. Not because they will never have incidents. But because when the moment comes, and it will come, their people will know what to do.

Build the culture deliberately. Because the most dangerous security gap in your organization today is not in your code or your systems. It is in the space between what your people know and how they actually behave when nobody is watching.

The architecture protects the perimeter. Culture protects the moment.

Nelson Marques

Information Security Officer,
Cybersecurity and Data
Privacy Educator

I've spent my career making security a natural part of how people live and work. With experience at global organizations like Vodafone and Marriott, I've helped implement security frameworks in over 15 countries, making businesses stronger and more resilient. Growing up with parents who were chefs, I learned that the best flavors come from ingredients baked in, not sprinkled on at the end. I apply the same idea to security—making it seamless, natural, and built into everything from the start. My goal is to demystify cybersecurity, making it simple and accessible so that people can go about their lives safely, without even noticing it is there.

PECB Store — Access the Right Resources, Build Trust, Lead AI

From AI governance and risk management to information security, privacy, and implementation, the PECB Store brings together the essential standards and tools you need; seamlessly, in one place.

- ISO/IEC 42001
- ISO/IEC 23894
- ISO/IEC 27001
- ISO/IEC 27701
- GDPR Implementation Toolkit

Take the first step toward building AI systems that are not only innovative, but also trusted, secure, and accountable.

SHOP NOW →

PECB Store

ISO/IEC 42001
ISO/IEC 23894
GDPR Implementation Toolkit
ISO/IEC 27001
ISO/IEC 27701

THE EXPERT

Management System Auditing in an AI-Driven Environment

Auditing is one of the oldest professions. It was preceded by accounting which was born out of the need to keep records beyond the capacity of human memory. The need to keep records necessitated verification, which gave rise to the auditing profession .

Management system (MS) auditing officially began with the publication of the International Organization for Standardization's (ISO) first ISO 9000 series in 1987. The series has gone through several iterations, and the accepted definition of an audit, according to ISO 9000:2015, is a "systematic, independent, and documented process for obtaining objective evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled."

MS auditing is a systematic evidence-based process geared at evaluating the effectiveness, adequacy, and continuing suitability of a management system. It assesses the inputs, activities, outputs, and controls related to the processes that constitute the management system with the intent of identifying strengths, weaknesses, and opportunities for improvement. The audit can be focused on a segment of the management system or the entirety of interconnected processes that comprise it.

Sampling is a fundamental approach in MS auditing; however, it introduces the element of risk. The program should mitigate against obtaining inadequate information or misinterpreting audit findings, which could lead to unintended consequences. A key factor in the overall audit program is the selection of auditors possessing the required competence and commitment to adhere to established guidelines and audit principles that ensure audit plausibility and integrity.

MS auditing is principles-based. Adherence to these principles is intended to enable audits to be an effective and dependable means of supporting management policies and controls by providing reliable information for driving actions that will result in improvement of performance in organizations. Auditors are expected to adopt and abide by these principles as the basis for providing relevant and sufficient audit conclusions that enable auditors working independently from one another but in similar circumstances to arrive at comparable results (ISO 19011:2018).

These principles are:

- **Integrity** - The foundation of professionalism, requiring auditors to act ethically, honestly, and competently.
- **Fair Presentation** - The obligation to report audit findings, conclusions, and activities truthfully and accurately.
- **Due Professional Care** - The application of diligence, judgment, and responsibility in auditing tasks.
- **Confidentiality** - The secure handling and protection of information obtained during the audit process.
- **Independence** - The basis for impartiality, ensuring auditors are free from bias and conflict of interest.
- **Evidence-Based Approach** - The use of rational, verifiable evidence to ensure reliable and reproducible conclusions.
- **Risk-Based Approach** - Focusing the audit on matters significant to the client and audit goals.

The effectiveness of audits also depends on the possession and exercise of certain behavioral traits that regulate personal conduct.

They include the following:

- Being ethical, open-minded, and diplomatic: Fair, receptive, and tactful.
- Being observant, perceptive, and versatile: Aware of surroundings and adaptable to situations.
- Tenacious and decisive: Persistent in pursuing objectives and capable of reaching logical, timely conclusions.
- Self-reliant and possessing fortitude: Working independently while acting responsibly, even in challenging circumstances.
- Improvement-oriented, culturally sensitive, and collaborative: Committed to learning, respectful of diverse cultures, and effective in team interactions.

Shifting Paradigm in MS Auditing

The principles and practice of auditing described rely heavily on human beings.

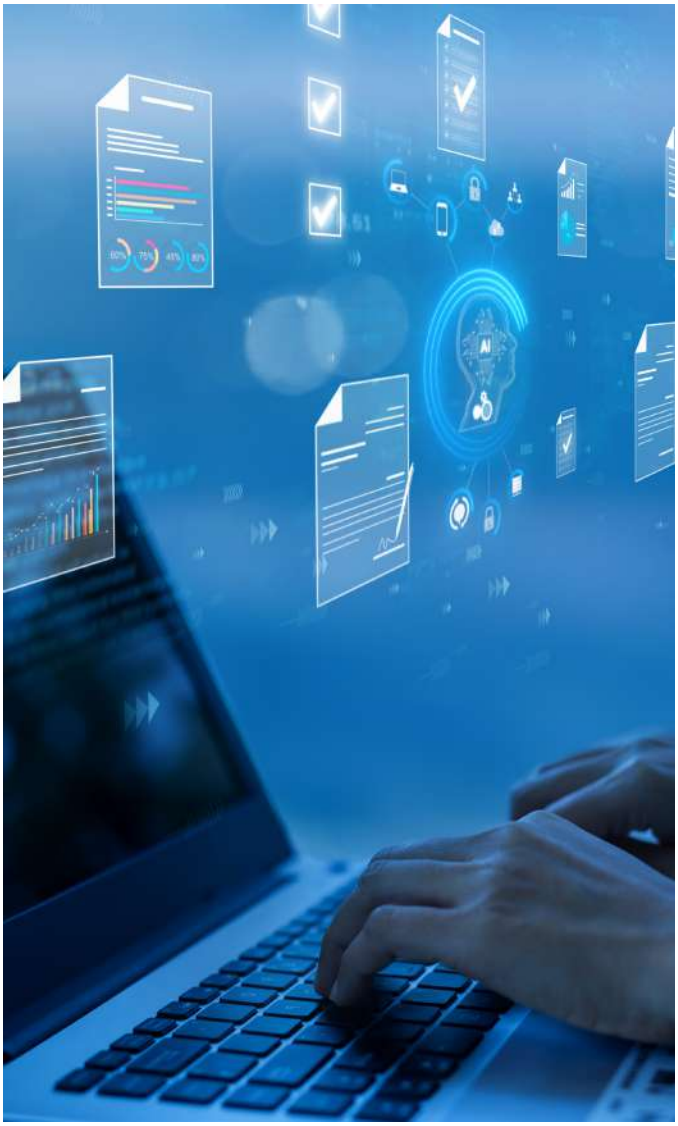
There is, however, a shift in methodology that has been developing over many years, which has accelerated with the transition from the traditional sampling-based methods to data-intensive, real-time, or continuous auditing.

Like many areas of business and organizational life, auditing is being significantly impacted by Artificial Intelligence (AI).

Artificial Intelligence (AI), Machine Learning (ML), and Robotic Process Automation (RPA) have formed a continuum of technologies that move businesses from basic task execution to advanced cognitive decision-making. The sphere of auditing has been no exception. The role of the auditor is evolving from "asking people questions" to "asking data questions," and the use of technology is increasingly having a major impact on the way audits are conducted and the analysis of data.

Business challenges in the 21st century corporate climate necessitate maximizing the use of technology-based decision support systems. AI can make predictions, suggestions, or judgments that influence actual or virtual environments in the process of achieving certain human-specified objectives, OECD 2021.

Advanced data analytics made possible by the use of AI technologies in auditing creates distinct advantages. These include tools to analyze voluminous data, increasing audit efficiency, the allocation of more time to audit analysis by the auditor through the reduction of time needed to do technical tests, minimizing of costs, and the provision of greater transparency. Overall, AI technology supports the improvement of accuracy and efficiency of the auditing process.



The Impact of Digital Transformation

Digital transformation is fundamentally shifting MS auditing from a periodic, sample-based activity to a continuous, data-driven function. This evolution is driven by the integration of AI, which enables auditors to move beyond human limitations to achieve higher levels of assurance and strategic insight.

Human auditors are restricted to the use of sampling in analyzing datasets, thereby reinforcing the inherent risk associated with auditing. The advent of AI tools allows pivoting from sampling to full population testing, increasing the level of certainty with the attendant reduction of risks.

Another key transition that AI enables is continuous monitoring through real-time process audits using computer-assisted auditing techniques (CAATs). The traditional retrospective approach employed in auditing is giving way to a proactive and predictive risk-based approach.

Robotic Process Automation (RPA) and Natural Language Processing (NLP) automate repetitive tasks such as data extraction, three-way matching, and document reviews, enabling corroboration and greater certainty in conclusions. The auditor can therefore refocus efforts on areas requiring professional judgment and strategic analysis.

The use of AI brings to the fore an entirely new perspective in the management of audit risks. AI reduces risks by shifting the process from manual, sample-based testing to continuous, data-driven assurance. Through the leveraging of machine learning and automation, it addresses inherent, control, and detection risks.

MS standards are largely risk-based and, therefore, auditing based on risk is best practice. The use of AI enables audit planning to become smarter through "AI-guided flows" that suggest the most critical areas for investigation based on data patterns. This satisfies the need to be guided by the concept of materiality in auditing.

Downside Risks

The use of AI technology in MS auditing presents certain challenges, chief among which are data security and quality control. Critical factors include reliability, completeness, and accuracy. Biases originating from flawed training data, algorithmic design, or human subjectivity have the potential to undermine the basis of audit quality by altering how evidence is collected, analyzed, and interpreted.

Sources of downside risks include:

- **Sampling Gaps:** If training data over represents certain processes or demographics, the AI may ignore rare but critical anomalies in underrepresented areas.
- **Narrowed Scope:** Automated anomaly detection often focuses on predefined categories. Bias in these definitions can cause the system to overlook risks not captured by those specific features, leading to an incomplete risk profile.
- **Exclusion of Variables:** Developers may inadvertently omit important data points due to their own cognitive biases, resulting in a model that fails to account for the full operational context.

The Future of MS Auditing in an AI-Driven Environment

The trajectory of the MS auditing profession is doubtlessly being impacted by digital transformation. What should be expected?

- **Continuous Auditing:** Shifting from periodic check-ins to ongoing, real-time monitoring.
- **Human-AI Collaboration:** Emphasizing that AI augments, rather than replaces, professional skepticism and auditor judgment.

The Use of AI is in vogue, and this trend will continue. Does that mean that human auditors will be redundant? This is not likely. However, roles will change. The automation of tasks that are data-intensive will enable more focus on strategic matters and enhance the quality and outcomes of audits. This will only drive continual improvement, which is the quest of management system auditing.

Jacob McLean

Managing Director at Kaizen Training and Management Consultants Limited (KTMC)

Jacob is a seasoned consultant and certified trainer with over a decade of experience in management systems, risk management, and compliance. As the Managing Director at Kaizen Training and Management Consultants Limited (KTMC), he has played a pivotal role in shaping the company's success, driving professionalism, and delivering world-class training solutions.

Since partnering with PECB in 2014, Jacob has trained professionals across industries, helping them achieve internationally recognized certifications. His expertise spans ISO 9001, ISO 14001, ISO 22301, ISO 31000, ISO 37301, and ISO 45001, among others.

Jacob's passion lies in empowering individuals and businesses through high-quality education, mentorship, and ethical leadership. His commitment to excellence, innovation, and continuous learning has made him a trusted figure in the industry, helping professionals enhance their skills and organizations strengthen their operational resilience.

How to Conduct an AI Risk Assessment That Goes Beyond IT

Artificial intelligence is no longer confined to technical environments. It influences hiring decisions, credit approvals, medical diagnostics, fraud detection, customer service automation, and strategic forecasting. Yet many organizations still approach AI risk assessments as if they were traditional IT security reviews. That approach is insufficient.

An effective AI risk assessment must extend beyond infrastructure, cybersecurity, and system availability. It must examine trust, accountability, ethics, governance, operational resilience, regulatory exposure, and decision impact. Here is how to conduct an AI risk assessment that truly reflects the realities of modern AI deployment.

1 Start with the Decision, Not the Model

Many organizations begin by evaluating the algorithm. That is a mistake. Instead, begin with the decision the AI system influences:

- What outcome does it affect?
- Who is impacted?
- Is it advisory or autonomous?
- What happens if it fails?

2 Identify Stakeholders Beyond IT

AI risk is rarely confined to technology teams. Involve:

- Legal and compliance
- Risk management
- HR (if employment decisions are involved)
- Operations
- Data governance
- Business unit leaders
- Ethics or oversight committees (where applicable)

3 Evaluate Data Risk at Its Source

AI systems inherit the risks of their training and input data. Assess:

- Data provenance (Where did it originate?)
- Consent and lawful basis
- Bias and representativeness
- Data quality and completeness
- Retention and storage controls
- Third-party data dependencies

4 Analyze Model Risk in Context

Model risk assessment should include:

- Validation and testing methodology
- Performance metrics under real-world conditions
- Drift detection mechanisms
- Explainability capabilities
- Sensitivity to adversarial manipulation

5 Assess Human Oversight and Escalation

A common governance failure is “symbolic oversight” — where humans are technically in the loop but practically unable to intervene meaningfully. Evaluate:

- When humans review outputs
- Whether they have authority to override decisions
- Whether they understand system limitations
- How exceptions are handled
- Whether there is an appeals process

6 Map Regulatory and Compliance Exposure

AI risk is increasingly shaped by regulation. Assess alignment with:

- Data protection requirements
- Anti-discrimination laws
- Consumer protection standards
- Sector-specific regulations
- Emerging AI governance frameworks

7 Evaluate Operational Resilience

AI systems introduce new failure modes:

- Model degradation
- Data pipeline interruptions
- Third-party model dependencies
- Infrastructure scaling limits

8 Identify Stakeholders Beyond IT

Not all risks are financial or technical. Consider:

- Public perception
- Media exposure
- Stakeholder expectations
- Ethical concerns

9 Document Accountability Clearly

One of the most overlooked aspects of AI risk management is ownership. Define clearly:

- Who owns the model?
- Who owns the data?
- Who approves deployment?
- Who monitors performance?
- Who is accountable for outcomes?

10 Make It Continuous, Not One-Time

AI risk assessments must not be static. Implement:

- Periodic reassessments
- Drift monitoring
- Post-incident reviews
- Performance audits
- Feedback loops from impacted users

Conclusion

AI does not eliminate risk. It redistributes and reshapes it.

An AI risk assessment that focuses solely on technical vulnerabilities misses the broader implications of automated decision-making. Trust in AI depends not only on secure code but on transparent governance, clear accountability, resilient operations, and ethical awareness.

Organizations that understand this will not merely comply with regulations. They will earn durable trust. And in the age of AI, trust is the most valuable control of all.

ACHIEVE EXCELLEN

Enrich your professional portfolio through
Contact us at support@pecb.co

Updated Training Courses	Language	
Lead Cybersecurity Manager	English	›
Lead Disaster Recovery Manager	English	›
CMMC Foundations	English	›
ISO 21001 Foundation	English	›

CE IN YOUR CAREER

PECB’s new and updated training courses!
om or visit our website for more.

New Training Courses	Language	
Certified Linux Foundations	English	›
Certified Advanced Penetration Tester	English	›
Lean Six Sigma Yellow Belt	English	›

SPECIAL T

TITANIUM

orsys FORMATION

SMATICA Training & Consulting

EduGroupe Accompagner pour réussir

SGS

Your fastest way to learn. Why wait? FIREBRAND

KOE

aac aswar akka consultancy

ib cegos

de-dsb GmbH

m2i Formation

Alphaedge Quodrant AFRICA LIMITED

PLATINUM

skillsoft global knowledge. FRANCE

Training Heights

CYNTHUS THE EXCELLENCE EFFECT

TSTC ICT en Security Trainingen

EGYBYTE Inspiring Minds

1ST CAREER Career development online courses

PINK

E-WORK

Sentinel Africa

digicomp

LAM group

Certi Banks

CERTyou

QPLUS

Mit Sicherheit mehr Erfolg msdd.neT GmbH

Netec 40 AÑOS DE APRENDIZAJE EFECTIVO

MAXPERT

BI

GOLD PA

GREAT LEARNING EDUCATION CENTRE

DS DANSK STANDARD

Tecnofor by SNGULAR

DARME

AEG Training

FIREBRAND Part of the BPP Education Group GERMANY

alc TRAINING A tesserent Company

Lin Assure. Auc

Dexterous Applied Training Institute

eimf

iCERTWORKS TRAINING & CERTIFICATION

brg be resilient group

inlexso INNOVATIVE LEGAL SOLUTIONS

skillsoft global knowledge. SAUDI ARABIA

CRMS

prot

CAA Crest Advisory Africa

ABILITY PRO Elearn Training Services

RISK ASSOCIATES

zih

FACHS

CONGO BIZ

CS CONSULTING

cn

INFOTEC

Rhythex Consulting

SMT SECURITY MANAGEMENT TECHNOLOGY GROUP

BRIGHT COLLAR BE CERTIFIED! RECOGNIZED!

Business Crisis Consultants

MANUS 360 Innovating success

RESTREPO RAMAS

3D ED

FEDERATED MANAGEMENT INSTITUTE

MKUTANO TRAINING SOLUTION

OGSBC

Note that PECB Partners are listed as per the credentials

HANKS TO

PARTNERS

ENIG
ep forward

QA
Learn. To Change.

glasspaper

schellman
Quality, above all.

Igp | Innovación Gestión Buenas Prácticas

Oo2 Formations & Consulting

A

safeshield™

skillsoft global knowledge.
NETHERLANDS

Digital Jewels Africa
information value chain consultants
PCIDSS QUALIFIED SECURITY ASSESSOR
ISO/IEC 27001:2013(ISMS) CERTIFIED
ISO/IEC 9001:2015 (QMS) CERTIFIED

ABILENE ACADEMY

DAR AL-HEKMA UNIVERSITY

PARTNERS

B2B Learning

AURIUM QSE CONSULTING

Digital Encode
Your trusted infosec & GRC advisor

devforma

KRUCEK

Smart Skills
Your security is our priority

BDO

Behaviour Brasil

RISKPROFS

New Horizons®

SPARTAN
Allied Services
Deliver Results - Create Value

MARKO ADVANCE
BUILD A LEADER, BUILD THE WORLD

IT

BSJ Bureau of Standards Jamaica

CaesCR
Consulting & Education Services

CYBER MINUTE

CYBERSTRAT

SkillsCampus
Training & Certification Center

ARTNERS

qs
lit. Advance.

CARMAO
INTELLIGENT. EINFACH. PREMIUM

parker
SOLUTIONS GROUP

FIREBRAND
Part of the BPP Education Group
NETHERLANDS

consult IT

TÜVNORD

ISOPEX C&P
NOS COMPÉTENCES, VOS PERFORMANCES

DATASEC | 10
#10YearsStrong

iviti

TACTICX
innovation | solutions | consulting

TÜV SÜD

EGE Ecole de Guerre Economique

DoTrust

sustain

devup

IBEForuM

Cyber Academy

TEKNOLOGISK INSTITUT

fidens
by TVH CONSULTING

LUMIFY work

SERVIEW

CDA

TRAINOCATE

Acute
Cybersecurity Services

UCATORS

Afenoid
Enterprise Limited

LEECON
Training Services

ACTAGIS

AfriCapital Consulting
R.D.Congo
Agir en toute confiance

analytix

LE PLUS
الإضافة

Deloitte.

pwc

SAVANT

unidees

s acquired from January 1, 2025, to December 31, 2025.

Lead with Confidence through
PECB’s Training Courses.