

Whitepaper

How to Mitigate Potential Costs and Risks of Unchecked AI





Table of **contents**

01	Executive Summary
02	Introduction
03	Emerging Risks from Unchecked AI
04	Cost of Failure: The Price of Unchecked AI
05	From Risk to Trust: AI Governance as Strategic Enabler
06	Conclusion & Recommendations
07	Sources / Footnotes



Executive Summary

Artificial intelligence (AI) is rapidly transforming business operations across industries. In the meantime, AI supports decision making, content generation, or customer engagement - just to name a few examples. However, recent high-profile failures highlight the risks of deploying AI without clear governance. Notable cases include AI-assisted reports with fabricated citations and flawed AI-generated U.S. judicial rulings².

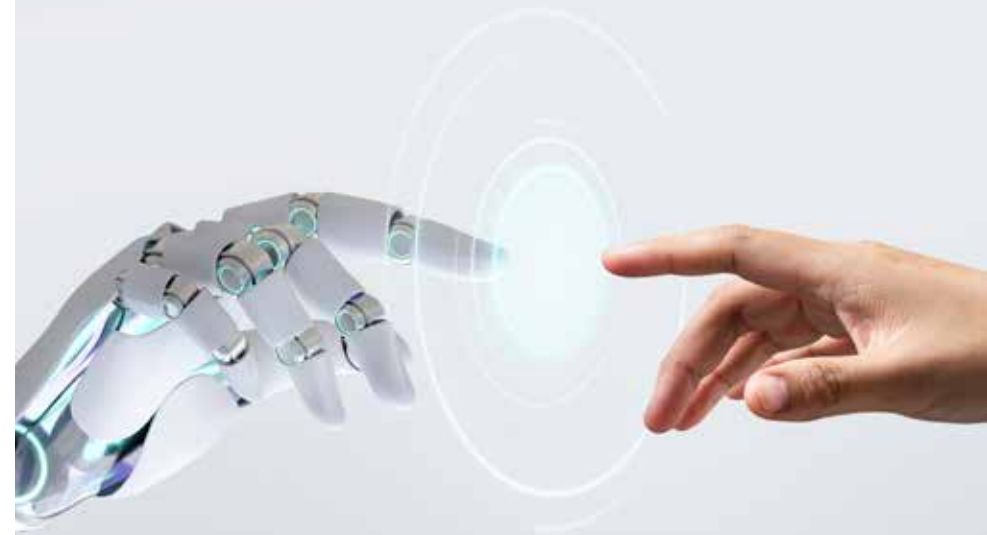
Despite widespread AI adoption, only very few organizations have embedded controls and managed AI governance⁹. Unchecked AI can create reputational, operational, legal, and privacy risks, particularly when human oversight, validation, and accountability frameworks are absent. This can lead to significant financial costs. A 2025 EY survey found that flawed AI outputs, compliance breaches, and bias resulted in US\$ 4.4 billion in losses across participating organizations^{7,8}. A key contributing factor was the cost determined by a lack of compliance^{7,8}.

This whitepaper outlines how organizations can mitigate these risks through structured AI governance, in line with standards such as ISO/IEC 42001.

Key measures include:

- ▶ embedding human-in-the-loop review,
- ▶ defining vendor governance policies,
- ▶ implementing monitoring and auditing,
- ▶ training staff on AI limitations,
- ▶ and establishing incident response protocols.

Adopting these measures ensures that organizations can safely utilize AI and enhance operational capabilities. ISO/IEC 42001 provides a practical framework for embedding AI governance into existing management systems, enabling continuous improvement and accountability. Responsible AI governance is more than a compliance requirement; it is a strategic enabler. It ensures AI accelerates value while minimizing potential costs and risks.





Introduction

In the modern world, business has wholeheartedly integrated artificial intelligence into its operations. AI tools provide support in making decisions or steering operations, generating content, and are involved in customer engagement.

However, the growing use of AI also introduces new risks. Recently, high-profile failures have underscored the danger of using AI without robust validation. In order to make those risks more tangible, here are a few examples:

First,

a company produced a government report which included fabricated citations, mis-referenced academic papers, and nonexistent court quotations. The firm later admitted the use of generative AI (OpenAI GPT4) without adequately validating the results. This led to a partial refund of the contract volume¹.

Second,

months later, the same corporation found itself in a similar situation again. This time, it created a health workforce report for a local governmental office. Journalists found fabricated or unverifiable academic citations and raised concerns about possible AI-generated errors. According to the company, AI was only used in limited research support and agreed to correct the citations¹⁰.

Third,

in the US federal judges unknowingly incorporated AI-generated text containing fictitious legal claims and citations into rulings. These errors were detected by the parties involved, forcing the courts to retract portions of the rulings².

This list of unfortunate events can easily be extended. These cases are not just outliers or isolated incidents; they highlight a systemic risk: as AI adoption continues, many organizations lack the structures to govern and verify its output. The risk is even more crucial in the light of recent research. According to Dialectica, more than 90% of companies are using AI whereas only 7% have implemented AI governance⁹. The gap between AI enthusiasm and governance maturity leaves many organizations exposed to operational, reputational, and compliance risk.

Emerging Risks from Unchecked AI

Unchecked AI introduces a broad landscape of risks. One major dimension is reputational risk: when AI-generated mistakes become public, stakeholders' trust suffers and media exposure can inflict long-term damage. Unfortunately, the threat goes even deeper than reputational loss. AI's ability to generate text that appears credible but is factually wrong - so-called "hallucinations" - creates accuracy and integrity risks. These risks can undermine business decisions and external communications.

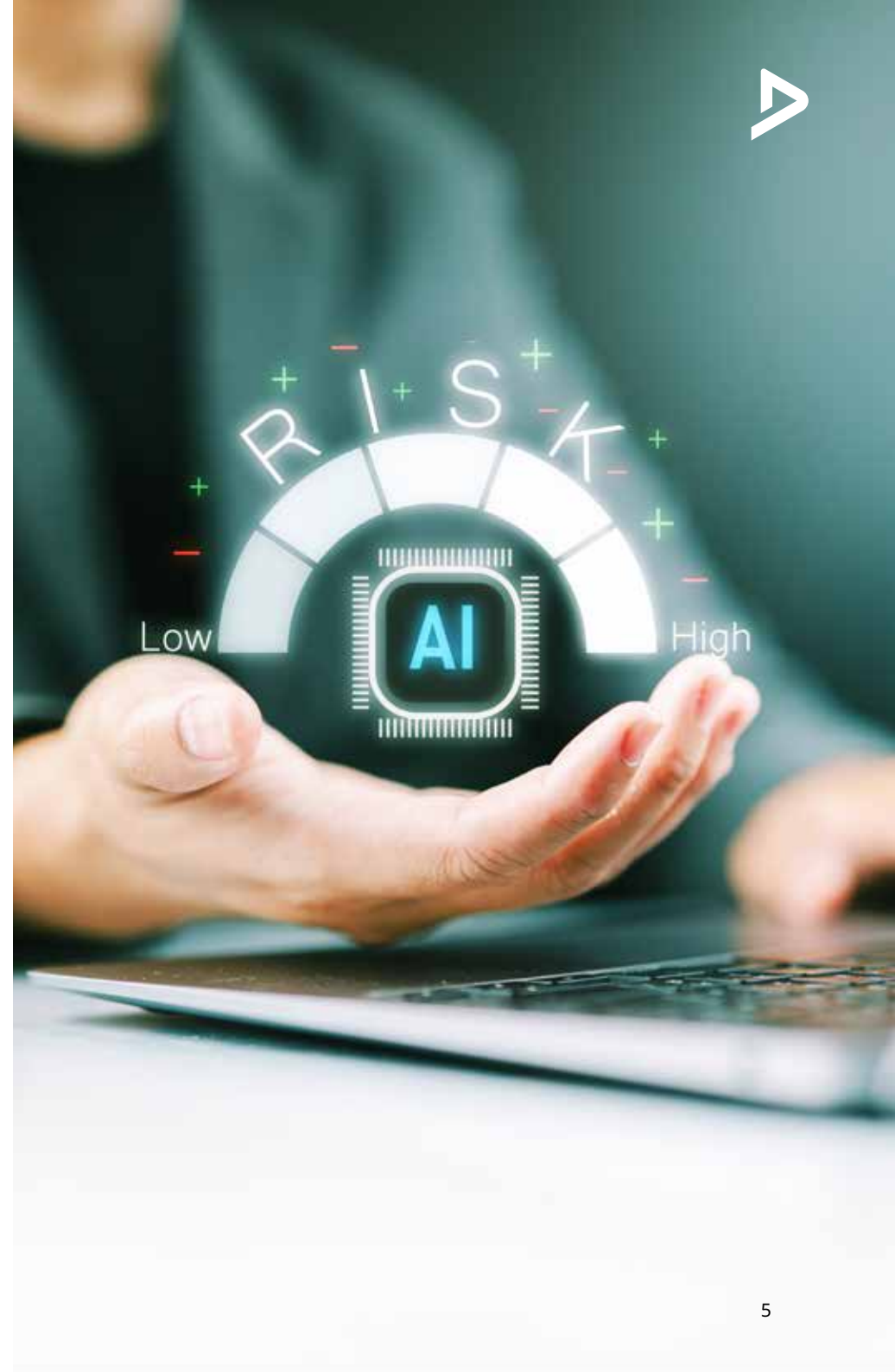
Legal and regulatory risks compound the problem. Without proper audit trails or governance, organizations relying on AI-generated outputs may face liability for non-compliance or inaccurate advice. In highly regulated sectors such as finance, healthcare, or Testing Inspection Certification (TIC), accuracy is even more crucial.

From an operational perspective, over-reliance on AI without human oversight can lead to flawed or misaligned outputs, potentially disrupting core business processes or decision workflows. On the data protection and privacy front, AI systems may bypass existing controls, creating vulnerabilities for sensitive data exposure or regulatory breaches.

Another key risk is the governance and accountability gap. Many organizations lack formal structures around AI ownership, incident response, and oversight. Those are key components of tracing responsibility when things go wrong. This is particularly relevant in the TIC industry, where documentation, traceability, and auditability are central to trust.

The above-mentioned risks are not speculative. They are being observed and studied in academic settings. For example, research on verifiable generative AI explores how model outputs can be linked to their training data to improve trust³.

Meanwhile, other studies argue for "safe harbors" to encourage independent red-teaming and evaluation of AI systems, stressing the importance of open, accountable governance⁴.

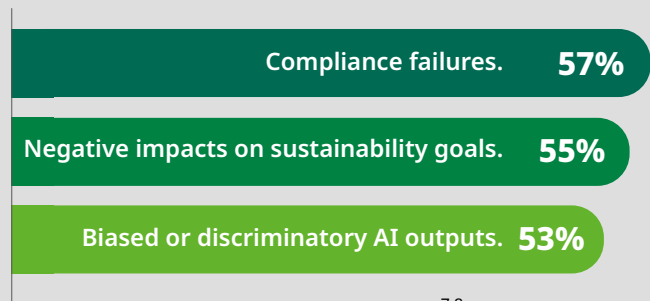


Cost of Failure: The Price of Unchecked AI

Unchecked AI deployments are not simply about reputational or compliance concerns, but they also have a financial impact. There are several examples out there, where significant financial losses occurred. According to a survey of EY in 2025, organizations had to bear cost of US\$ 4.4 billion in total due to flawed AI outputs, compliance failures, and bias⁷. The survey addressed worldwide operating corporations with annual sales of more than 1 billion USD^{7,8}. In the report, there are the following risk-categories cited as common contributors to AI-related financial loss^{7,8}.

- ▶ **Compliance failures: The most frequently cited risk.**
57% of the companies report this issue as a key factor⁸.
- ▶ **Negative impacts on sustainability goals.**
55% of respondents reported it as a contributor to losses⁸.
- ▶ **Biased or discriminatory AI outputs.**
Reported by 53% of companies as a loss factor⁸.

Most common contributors to AI-related financial loss



Source: Survey by EY in 2025^{7,8}

Collectively, these costs highlight the financial stakes of deploying AI without robust governance, emphasizing the need for structured AI management systems.



From Risk to Trust:

AI Governance as Strategic Enabler

There is good news: risks linked to AI are not any different from other types of risks which have already successfully been managed in the field of Information and Cyber Security such as data breaches or insider threats.

Consequently, the right tools are available and, therewith, risks posed by AI misuse are manageable. A pre-requisite though, is that organizations must adopt structured, intentional AI governance.

The first step on this road is the implementation of an artificial intelligence management system (AIMS) based on ISO/IEC 42001. The norm was published in December 2023 and is known as one of the world's first AI Management System (AIMS) standards⁵. It is structured into core clauses as described below:

Clause	Explanation
Context of the organization	Understanding the context of the organization is key for implementing an effective AIMS. The context sets the stage of how AI systems are managed, governed, and aligned with the organization's overall goals.
Leadership	Leadership is more than oversight; it's about actively shaping and guiding the direction of AI management within the organization.
Planning	Planning ensures AI management is proactive, structured, and aligned with the organization's broader objectives. Risk assessment, risk treatment, and impact assessment play a pivotal role in this context.
Support	Adequate support structures ensuring effective implementation and maintenance of an AIMS. Support underpins all other elements of AI management.
Operations	Ensuring the organization has a systematic and ongoing process for managing its AI risks and controls.
Performance evaluation	Performance evaluation is crucial for ensuring AI systems operate as intended, meet organizational goals, and comply with relevant standards.
Improvement	Continuous improvement is key to maintaining the AIMS' effectiveness, adaptability, and alignment with the organization's evolving goals and external challenges.

Source: ISO/IEC 42001:2023⁵





From Risk to Trust: AI Governance as Strategic Enabler

The above-mentioned clauses urge organizations to define policies, roles, processes, and continuous improvement practices to govern AI responsibly. Furthermore, ISO/IEC 42001 complements existing management systems such as ISO/IEC 27001, making integration smoother.

Key mitigation practices include embedding human-in-the-loop (HITL) review. If HITL is implemented well, subject matter experts validate AI-generated outputs before being used externally. This ensures AI output remains under human scrutiny and is treated more as an assisted human worker rather than an autonomous intelligence.

Training and awareness are equally important. Under ISO/IEC 42001, organizations are encouraged to build competence, provide targeted training, and communicate clearly about AI's fallibility⁶. Everyone using AI should clearly understand that "AI-assisted" does not mean "automatically correct."

Vendor governance is another pillar. Contracts should require transparency about AI use, mandate validation steps, and include audit rights so organizations can verify how third-party tools are used and managed.

Additionally, continuous monitoring and auditing of AI systems are crucial. Organizations should measure error rates, bias incidents, near-misses, and operational performance. These metrics feed into a governance feedback loop and consequently enable continuous improvement.

Incident response planning must not be an afterthought. Effective governance means having predefined playbooks containing issues, communicating transparently, and remediating quickly. ISO/IEC 42001 explicitly supports establishing these processes⁵.

From a practical perspective, ISO/IEC 42001 operationalizes governance through planning, building, verifying, deploying, monitoring, and improving. An organization, for example, might assess its risk exposure (Plan), define policies and controls (Build), perform internal or third-party audits (Verify), launch AI systems under guardrails (Deploy), track performance (Monitor), and then loop back into refining the system (Improve)^{5,6}.

Conclusion & Recommendations

AI is now standard in many business processes and is highly likely here to stay. A world without AI has become inconceivable. Undoubtedly, AI has enormous potential to influence our lives for the better, however, there are also some challenges which come along with it. In this setting, proper governance is key to avoiding AI becoming a liability rather than a competitive asset. By adopting a structured governance system, embedding human oversight, training users, monitoring performance, and aligning with standards such as ISO/IEC 42001, organizations can harness AI responsibly while minimizing risks.

As employees are already using AI, it is generally recommended for organizations to start with the following concrete steps:

- 1.** Implement training and coaching for all AI users to foster responsibility and competence.
- 2.** Use controlled AI tools (e.g., enterprise-grade ChatGPT, Copilot, etc.) under clear governance protocols. This requires defining and maintaining an accurate and updated AI inventory, avoiding the spread of “shadow AI”.
- 3.** Conduct a risk assessment from both security and governance perspectives, and classify AI use cases by risk and impact, be it on customers, employees, health & safety, and the environment.
- 4.** Test, document, and monitor. Before and after deployment, there must be a reliable tracking of tests directed at performance, security, reliability, explainability, and legal compliance.
- 5.** Manage third parties and assess the AI supply chain. The organization remains accountable for how it uses a model and cannot simply treat associated risks as the vendor’s responsibility.
- 6.** Leverage TIC services (audit, certification) to get an independent, third-party evaluation of your AI governance.

Adopting these measures not only reduces risks but also strengthens trust with customers, regulators, and stakeholders.

AI Use Disclosure: AI tools were used to assist with the drafting of this paper; all content was reviewed, validated, and finalized by the author.

Sources / Footnotes

1. Computerworld. (2025). Deloitte's AI governance failure exposes critical gap in enterprise quality controls. last consolidated 2026-05-02; [Deloitte's AI governance failure exposes critical gap in enterprise quality controls – Computerworld](#)
2. Heise Online. (2025). AI garbage exposed in US judgments. last consolidated 2026-05-02; [AI Garbage Exposed in US Judgments | heise online](#)
3. Tang, N., Yang, C., Fan, J., Cao, L., Luo, Y., & Halevy, A. (2023). VerifAI: Verified generative AI (arXiv:2307.02796). arXiv. last consolidated 2026-05-02; [\[2307.02796\] VerifAI: Verified Generative AI](#)
4. Longpre, S., Kapoor, S., Klyman, K., et al. (2024). A safe harbor for AI evaluation and red teaming (arXiv:2403.04893). arXiv. last consolidated 2026-05-02; [\[2403.04893\] A Safe Harbor for AI Evaluation and Red Teaming](#)
5. International Organization for Standardization. (2023). ISO/IEC 42001:2023 – Information technology — Artificial intelligence — Management systems. last consolidated 2026-05-02; [ISO/IEC 42001:2023 - AI management systems](#)
6. ISMS.online. (2025). Why AI governance demands fairness, transparency & accountability – ISO 42001 explained. last consolidated 2026-05-02; [Why AI Governance Demands Fairness, Transparency & Accountability: ISO 42001 Explained](#)
7. EY. (2025). Global AI risk survey. Reuters. last consolidated 2026-05-02; [Most companies suffer some risk-related financial loss deploying AI, EY survey shows | Reuters](#)
8. EYGS LLP. (2025). EY survey: Companies advancing responsible AI governance linked to better business outcomes. FinanzNachrichten. last consolidated 2026-05-02; [EYGS LLP: EY survey: companies advancing responsible AI governance linked to better business outcomes](#)
9. Dialectica. (2025). The market for AI governance and ISO 42001 adoption. last consolidated 2026-05-02; [The Market for AI Governance and ISO 42001 Adoption](#)
10. Financial Express. (2025). After Australia, Deloitte hit by AI-linked citation row in Canada. last consolidated 2026-05-02; [After Australia, Deloitte hit by AI-linked citation row in Canada - Business News | The Financial Express](#)

Would you like more information?

Contact Us!

