ISACA Melbourne Chapter

Article July 2025 | Authors Reshma Devi and Bharat Bajaj

AI Governance and AI Security: Why You Need Both

Many organisations assume AI governance and AI security are the same. They're not; they serve distinct purposes. AI governance focuses on value creation through oversight, compliance, and accountability, ensuring AI systems are aligned with ethical and regulatory expectations. In contrast, AI security focuses on protecting AI systems from threats. Without both, AI risks increase, ranging from biased decision-making to adversarial attacks.

AI Governance: Ensuring Compliance & Trust
AI governance establishes policies and controls to ensure AI is ethical, transparent, and compliant with regulations and standards, including the Jurisdictional requirements, Privacy Act, EU AI Act, and ISO 42001 and ISO 23894.

Key Focus Areas:

- **Risk Management & Accountability:** Establishing clear oversight mechanisms for AI decision-making processes.
- **Regulatory Compliance:** Adhering to legal frameworks such as the EU AI Act and standards like ISO 42001 and ISO 23894.
- **Ethics & Bias Mitigation:** Implementing measures to ensure AI systems are fair, explainable, and free from bias.

AI Governance Standards
Standards are a structured framework to ensure ethical, secure and transparent governance throughout the AI lifecycle:

- **ISO 42001:** This standard provides a comprehensive framework for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organisations. It supports responsible AI practices by addressing ethical considerations, transparency, and continuous learning across the AI lifecycle.
- **ISO 23894:** Offers a framework for identifying, assessing, and mitigating risks associated with AI systems. It supports organisations in integrating risk management practices across all stages of the AI lifecycle from design and development to deployment, monitoring, and decommissioning.
- **EU AI Act:** This Act classifies AI systems based on their risk levels and sets out requirements for high-risk AI systems to ensure they are safe and trustworthy.
- **Australian Privacy Act:** This Act, along with the Australian Privacy Principles (APPs), governs the handling of personal information by AI systems. It ensures that AI applications comply with privacy obligations, protecting individuals' data and maintaining transparency.
- **Australian AI Ethics Principles:** These principles guide businesses and governments to design, develop, and implement AI responsibly. They focus on human, societal, and environmental well-being; human-centred values; fairness; privacy protection; reliability; transparency; contestability; and accountability.

- **Australian AI Guardrails:** The Voluntary AI Safety Standard includes ten guardrails to ensure the safe and responsible use of AI. These guardrails cover accountability processes, risk management, data governance, model testing, human oversight, and user transparency.

Key Governance Controls for AI Systems

1. **Governance Policies**
   Establish clear policies that define accountability, ethical principles, and compliance requirements for AI development and deployment.
2. **AI Risk Assessments**
   Conduct structured AI risk assessments to identify and mitigate legal, ethical, operational, and reputational risks across the AI lifecycle.
3. **Data Governance Controls**
   Ensure data quality, lineage, privacy, and security are maintained, especially for training and input data used in AI systems.
4. **Model Documentation & Traceability**
   Maintain detailed documentation of model design, training data, assumptions, and decision logic to support transparency and auditability.
5. **Bias & Explainability Audits**
   Perform regular audits to detect bias and ensure AI systems are interpretable and that the decisions can be explained to stakeholders.
6. **Human Oversight & Accountability Mechanisms**
   Define roles for human review and escalation, especially for high-impact or sensitive AI decisions.
7. **Monitoring & Performance Evaluation**
   Continuously monitor AI systems for drift, anomalies, and performance degradation, with thresholds for intervention.
8. **Incident Management & Escalation Protocols**
   Establish procedures for identifying, reporting, and responding to AI-related incidents or failures.
9. **Training & Awareness Programs**
   Educate staff and stakeholders on responsible AI use, governance policies, and ethical considerations.
10. **Independent Assurance & Audits**
    Conduct periodic internal or external audits to validate compliance with governance frameworks, standards, and policies.

Using a risk-based approach, all governance and controls should be applied in a manner proportionate to the risk level of each AI system. AI governance ensures AI systems are aligned with business goals, regulatory requirements, and societal values. By implementing robust governance frameworks, organisations can build trust and ensure the responsible use of AI technologies.

AI Security: Safeguarding AI Systems against Threats

AI security protects the models, data, and infrastructure from threats such as adversarial manipulation, data poisoning, and model theft. Frameworks such as MITRE ATLAS, Cloud Security Alliance LLM Threats Taxonomy, and OWASP Foundation Top 10 for LLMs highlight key risks and provide guidelines for mitigating them.

Key Focus Areas for AI Security

1. **Security Governance & Policy**
   Define AI-specific security policies aligned with broader cybersecurity frameworks. Include roles, responsibilities, and escalation protocols for AI-related incidents.
2. **Secure Design & Development**
   Embed security principles into AI system design, including threat modelling, secure coding practices, and privacy-by-design approaches.
3. **Model Integrity & Resilience**
   Ensure models can withstand adversarial attacks and manipulation. Techniques like adversarial training and robustness testing are essential.
4. **Data Protection & Privacy**
   Secure training and operational data using encryption, access controls, and anonymisation. Ensure compliance with data protection regulations.
5. **Access & Identity Management**
   Control access to AI models, data, and infrastructure using role-based access, multi-factor authentication, and audit trails.
6. **Threat Detection & Monitoring**
   Continuously monitor AI systems for anomalies, adversarial behaviour, and unauthorised access using AI-aware security tools.
7. **Vulnerability Management & Threat Mitigation**
   Conduct regular vulnerability assessments, penetration testing, and apply patches to mitigate risks from evolving threats.
8. **Incident Response & Recovery**
   Establish AI-specific incident response plans, including rollback procedures, forensic analysis, and stakeholder communication protocols.
9. **Third-Party Risk Management**
   Assess and manage risks from external AI models, APIs, and data sources, including supply chain vulnerabilities.
10. **Security Audits & Assurance**
    Perform independent audits using frameworks like MITRE ATLAS, OWASP Top 10 for LLMs, and CSA's AI threat taxonomies to validate controls and identify gaps.

AI Security Standards:

- **ISO 27001:** This standard outlines requirements for establishing, implementing, maintaining, and continually improving an information security management system (ISMS). It includes controls for AI data protection and access management.
- **MITRE ATLAS:** This framework details adversarial attack techniques for AI, helping organisations understand and defend against potential threats. It includes a comprehensive taxonomy of attack vectors and mitigation strategies.
- **OWASP AI Security and Privacy Guide:** This guide provides actionable guidance on designing, developing, testing, and procuring secure, privacy-preserving AI systems. It addresses common AI-specific vulnerabilities, such as data leakage, model inversion attacks, and unintended memorisation.
- **NIST AI Risk Management Framework:** Developed by the National Institute of Standards and Technology, this framework helps organisations manage AI-related risks,

including security risks. It provides guidelines for identifying, assessing, and mitigating AI-related risks.

- **ISM (Information Security Manual):** Provides principles and controls for securing government and enterprise systems. It includes guidance on securing AI workloads, protecting sensitive data, and managing system vulnerabilities.
- **Essential Eight:** A set of baseline mitigation strategies developed by the Australian Signals Directorate to help organisations protect against cyber threats. These strategies are increasingly applied to AI systems to ensure secure configuration, patching, access control, and incident response.

Key AI Security Threats:

- **Model Manipulation:** Adversarial attacks that modify AI outputs by manipulating input data. These attacks can cause AI systems to make incorrect decisions, potentially leading to harmful outcomes.
- **Data Poisoning:** Corrupting training data to alter AI behaviour. This can result in biased or incorrect AI models that produce unreliable outputs.
- **Model Theft:** Stealing or extracting AI models, which can lead to intellectual property theft and unauthorised use of AI technologies.
- **Prompt Injection:** Manipulating AI-generated responses by injecting malicious inputs. This can compromise the integrity of AI outputs and lead to unintended consequences.
- **Insecure Supply Chain:** Introducing vulnerabilities through third-party AI dependencies. Ensuring the security of the entire AI supply chain is crucial to prevent exploitation.

AI security protects models, data, and infrastructure from threats like adversarial manipulation, data poisoning, and model theft, ensuring system integrity and resilience. By implementing robust security measures and adhering to established standards and frameworks, organisations can protect AI assets and build trust in its AI technologies.

Implementation, where to start:

Implementing AI governance and security can seem daunting yet breaking it down into manageable steps can help. Here's a detailed guide on where to start:

1. If You Lack AI Governance:

- Implement ISO 42001: This standard provides a structured framework for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS). It covers ethical considerations, transparency, continuous learning, and security measures to protect AI systems.
- Establish Clear Policies and Controls: Develop comprehensive policies that outline the ethical use of AI, data governance, and accountability structures. Ensure these policies align with regulations such as the EU AI Act and the Australian Privacy Act.
- Conduct Regular Audits and Risk Assessments: Regularly audit AI systems to ensure compliance with ethical standards and regulatory requirements. Conduct risk assessments to identify and mitigate potential legal and ethical risks.

- Engage Stakeholders: Involve a wide range of stakeholders, including AI developers, users, policymakers, and ethicists, to ensure that AI systems are developed and used in alignment with societal values.

2. If You Lack AI Security:

- Apply ISO 27034 & MITRE ATLAS Techniques: ISO 27034 focuses on secure AI application development, providing guidelines for integrating security measures throughout the AI software development lifecycle. MITRE ATLAS provides detailed descriptions of adversarial attack techniques for AI, helping organisations understand and defend against potential threats.
- Implement Robust Data Protection Measures: Ensure AI data is secure and tamper-resistant by using robust encryption, secure communication protocols, and regular security audits.
- Conduct Vulnerability Assessments and Penetration Testing: Regularly test AI systems for vulnerabilities and apply security patches promptly to mitigate risks.
- Develop a Comprehensive Security Framework: Establish a security framework that includes guidelines for data protection, model integrity, and threat mitigation. This framework should align with standards such as ISO 27001 and the OWASP AI Security and Privacy Guide.

3. If You Have Neither:

- Build a Cross-Functional AI Governance & Security Task Force: Form a task force that includes members from various departments such as IT, legal, compliance, and ethics. This team will be responsible for overseeing the implementation of AI governance and security measures.
- Define Roles and Responsibilities: Clearly outline the roles and responsibilities of each team member. Ensure that there is a Chief AI Officer or a similar role to provide leadership and accountability.
- Develop a Roadmap for Implementation: Create a detailed roadmap that outlines the steps for implementing AI governance and security. This roadmap should include timelines, milestones, and key performance indicators (KPIs) to track progress.
- Provide Training and Resources: Ensure that all team members are adequately trained in AI governance and security best practices. Provide resources such as guidelines, toolkits, and frameworks to support their efforts.

By following these steps, organisations can establish robust AI governance and security frameworks that ensure the ethical and secure use of AI technologies. This approach not only mitigates risks but also builds trust and fosters innovation.

Conclusion

AI governance and AI security are both crucial for effectively managing AI technologies. While AI governance ensures ethical, transparent, and compliant AI systems through oversight and accountability, AI security focuses on protecting these systems from threats like adversarial attacks and data breaches. Implementing standards such as ISO 42001 for governance and ISO 27034 for secure AI development, along with frameworks like MITRE ATLAS, helps organisations mitigate risks and build trust. By integrating both governance and security, organisations can mitigate risks, build trust, and ensure AI aligns with both business objectives and societal expectations.

Other Useful References:

1. https://www.digital.gov.au/policy/ai/AI-technical-standard
2. https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/guidance-on-privacy-and-the-use-of-commercially-available-ai-products
3. https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/guidance-on-privacy-and-developing-and-training-generative-ai-models
4. https://www.cyber.gov.au/business-government/secure-design/artificial-intelligence/deploying-ai-systems-securely
5. https://www.cyber.gov.au/business-government/secure-design/artificial-intelligence/guidelines-for-secure-ai-system-development
6. https://www.cyber.gov.au/business-government/secure-design/artificial-intelligence/engaging-with-artificial-intelligence