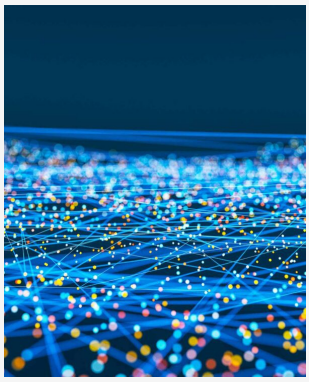
OSLER

The role of ISO/IEC 42001 in AI governance

JULY 10, 2024 7 MIN READ



Related Expertise

- Artificial Intelligence
- Commercial Technology Transactions
- Corporate Governance
- Privacy and Data Management
- Risk Management and Crisis Response
- Technology

Authors: Sam Ip, Joseph Ierullo, Iman Jaffari

Global overview of emerging AI standards

The rapid advancement of artificial intelligence (AI) has brought about a complex landscape of regulatory challenges and ethical considerations. To navigate them, several key regulatory frameworks and standards have been developed.

In Canada, the proposed *Artificial Intelligence and Data Act* (Part 3 of Bill C-27) aims to regulate AI systems comprehensively, focusing on safeguarding individuals and regulating responsible AI development and adoption. In the European Union (EU), the *Artificial Intelligence Act* was finalized and received approval from the Council of the EU on May 21, 2024. This Act takes a risk-based approach to regulation, categorizing AI systems by risk level and imposing corresponding regulatory requirements. Meanwhile, the U.S. National Institute of Standards and Technology's AI Risk Management Framework offers guidelines for organizations to assess and mitigate AI-related risks. In addition to these frameworks, the AI landscape is replete with various proposed laws, principles, guidelines, and codes of conduct. This patchwork of instruments underscores the need and role for international standards to ensure consistency and coherence with respect to the use, adoption, and governance of AI.

Overview of ISO/IEC 42001: AI management system

Amidst these various developments, ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) have created ISO/IEC 42001 through a collaborative effort of a diverse group of AI stakeholders, the first global and unified standard for AI management systems. Published recently in December 2023, this international standard provides a certifiable and comprehensive framework for organizations to establish, implement, and continuously improve their AI governance systems. ISO/IEC 42001 focuses on ethics, transparency, accountability, bias mitigation, safety, and privacy, covering essential elements of AI development and deployment.

ISO/IEC 42001 does not exist in isolation. It is part of a broader series of AI-related standards developed by ISO and IEC, including the following:

- ISO/IEC 22989: establishes common-language definitions of AI-related terminology and

outlines emerging concepts within the field

- ISO/IEC 23053: establishes a framework for describing generic AI systems that utilize machine learning technology, promoting interoperability among AI systems and their components
- ISO/IEC 23894: establishes guidance for managing AI-related risks in organizations developing deploying AI products and services, outlining processes for integrating AI risk management strategies into organizational activities, helping to identify, assess and mitigate such risks effectively

Scope and application

While these standards address specific aspects and applications of AI, ISO/IEC 42001 stands out as a comprehensive management system standard that offers a practical approach for managing AI-related risks and opportunities across the entirety of an organization. By adopting this standard and its holistic approach across various AI applications, organizations can promote responsible AI use, enhance trust in their AI applications, and support compliance with legal and regulatory standards. This framework not only promotes accountability, but also encourages innovation within a well-defined structure.

Moreover, ISO/IEC 42001 is designed for organizations of all sizes that develop, provide and utilize AI-based products or services across various industries. This standard applies to both private and public sector entities, including companies, non-profits, and government agencies, covering the entire AI system lifecycle from development to deployment.

Key concepts of ISO/IEC 42001

The key concepts that form the foundation of an effective AI management system under ISO/IEC 42001 include:

- Organizational Context: Organizations must thoroughly understand their internal and external environments when identifying the needs and expectations of stakeholders such as customers, regulatory bodies, and employees. By identifying these contexts, organizations can align their AI systems to be relevant and effective, addressing both internal capabilities and external demands.
- Leadership Commitment: Senior management needs to establish clear AI policies, define roles and responsibilities, and integrate AI governance into its overall strategic objectives. This top-down approach ensures support and prioritization for AI initiatives and projects.
- AI management system planning: Organizations must be proactive in their approach to managing AI systems by identifying risks and opportunities early on. Examples of this include collaborating with various stakeholders to create comprehensive plans for AI objectives, risk management strategies, and methods to leverage AI-related opportunities in the organization.
- Resource allocation: Adequate support for AI systems is essential for its effective management and implementation. This requires organizations to allocate sufficient financial, technological, and human resources in maintaining and improving competence in AI operations. Examples of this include budgeting for AI governance tools, hiring AI

ethics specialists, and ongoing training on internal AI systems.

- Operational controls: Implementing processes for responsible AI development, deployment, and use throughout the lifecycle is critical. Examples of this includes collaborating with relevant stakeholders in establishing specific and detailed protocols to ensure security, privacy and fairness in AI operations across the entirety of the organization.
- Performance monitoring: Creating continuous evaluation mechanisms to assess AI performance and identify areas for improvement is crucial for ensuring accuracy and compliance. This can include conducting periodic audits and spot checks to ensure alignment with clearly articulated metrics, organizational objectives, and regulatory standards.
- Continuous improvement: To complement the overall management process, it is essential to establish steps for updating and enhancing AI systems based on performance evaluations, emerging technologies, and evolving regulatory requirements. This involves fostering continuous improvement by incorporating specific feedback, refining AI governance practices, and ensuring the AI management system remains aligned with organizational objectives and compliant with applicable regulations.

ISO/IEC 42001 also provides four annexes in offering additional and complementing guidance:

- Annex A [Reference Control Objectives]: Provides detailed guidelines for building AI systems, including specific steps for design, development and testing. This Annex provides reference objectives and controls, covering areas such as data governance, system quality management, model selection, and performance evaluation.
- Annex B [Guidance for Implementing Controls]: Offers specific guidance on implementing controls, detailing measures for risk mitigation, data governance, and ethical AI practices. The annex includes practical examples and checklists for each enumerated control area, helping organizations ensure comprehensive coverage of AI management requirements.
- Annex C [AI Objectives and Risk Sources]: Provides a framework for organizations to navigate the complexities of the implementation of AI management systems. This annex emphasizes key objectives such as fairness, security, safety and privacy while detailing potential risks, such as risks of bias, data poisoning, system reliability concerns, and performance drifts.
- Annex D [Standards for Specific Domains and Industries]: Provides general guidance for applying the standard in specific industries such as healthcare, finance, and defence. This annex also highlights the importance of integrating AI management practices with other sector-specific standards to enhance compliance and effectiveness across diverse operational contexts.

Implications

The adoption of ISO/IEC 42001 not only assists organizations in ensuring their AI systems are developed and managed responsibly, but also enhances their trustworthiness. Many organizations are already familiar with ISO/IEC 27001, the international standard for

information security management systems. ISO/IEC 27001 emphasizes risk management, continuous improvement, and organizational commitment to security and ethics. ISO/IEC 42001 complements ISO/IEC 27001 by providing a holistic approach to mitigating information security risks. It addresses AI-specific challenges such as confidentiality in model training and privacy, integrity concerns like bias and tampering, and ensures availability for critical AI processes, as needed, for authorized use. This integration enables organizations to establish cohesive policies that maintain consistency and effectively address AI-related risks within their existing information management security systems. Certification and audits against a globally recognized standard, like ISO/IEC 27001 and ISO/IEC 42001, can further enhance stakeholder trust in an organization's AI capabilities and provide competitive differentiation in the marketplace.

For organizations interested, initial steps for adoption involve becoming familiar with the requirements and guidelines of ISO/IEC 42001, assessing current AI practices to identify gaps, and developing a comprehensive implementation plan. Engaging consultants with expertise in both ISO/IEC 42001 and ISO/IEC 27001 can provide valuable guidance during the process. This joint expertise can streamline the risk assessment and policy creation process by integrating AI ethical usage, transparency, and data privacy elements. A gap assessment may reveal necessary changes in organizational roles specified within the standard, mandating responsibilities for AI governance. It also spotlights necessary updates to security awareness programs and investments in technologies supporting AI and information security management systems.

Looking ahead

While the *Artificial Intelligence and Data Act* is currently before parliamentary committee for a clause-by-clause review, its potential passage would mandate organizations to establish an accountability framework for the deployment and adoption of AI. It is noteworthy that certifiable standards such as ISO/IEC 42001 could play a crucial role, especially for organizations operating or doing business across multiple jurisdictions, by providing a framework for establishing and maintaining an AI system with international relevance.