AI legislation puts a spotlight on ISO/IEC 42001

A FUTURE OUTLOOK WHITE PAPER

sgs.com

SGS

Table of contents

Overview

This paper offers a look into the changing landscape of artificial intelligence (AI) readiness requirements for organizations that are developing or deploying AI solutions, with a focus on the first-ever legislation to pass in North America that regulates the use of AI.

You will gain an understanding of:

- The increasing relevance of 'trustworthy AI'.
- The Colorado Artificial Intelligence Act and its focus on mitigating consumer risk.
- ISO/IEC 42001: an internationally recognized management system for AI.
- How ISO/IEC 42001 prepares organizations for current or future AI-focused legislation.

THE CONTENT OF THIS PAPER IS INFORMED BY SGS EXPERTS WHO ACTIVELY CONTRIBUTED TO THE DEVELOPMENT OF THE ISO/IEC 42001 ARTIFICIAL INTELLIGENCE MANAGEMENT SYSTEM STANDARD. SGS CONSTANTLY SUPPORTS THE DEVELOPMENT OF INTERNATIONAL STANDARDS, FRAMEWORKS, SCHEMES, AND REGULATIONS. INDUSTRY SOURCES HAVE ALSO BEEN REFERENCED.

Growing reliance on AI

AI is revolutionizing modern life and business at an astonishing pace by enabling computers to learn and solve problems like humans. The attractiveness of AI as a learning or decision-making tool is owed to its ability to leverage extensive amounts of data to identify patterns and trends that might otherwise be unidentifiable to humans.

Algorithms developed by machine learning can analyze historical data and predict future outcomes, enabling quick and accurate decision-making for businesses and consumers alike. AI-based systems are being used more than ever before in both the private and public sectors, with their use only set to grow exponentially in the coming years.

However, the types of tasks becoming more reliant on AI are often highly sensitive in terms of personal well-being, for example, in hiring procedures, loan approvals or decisions that impact health. This opens up the possibility of violations to individuals or social groups, due to risks inherent in AI solutions that are powered by algorithms rather than human-based decision-making.

A PUSH TOWARDS TRUST-WORTHY AI: HUMAN AGENCY AND HUMAN OVERSIGHT

Trustworthy AI is the end goal for organizations leveraging AI solutions. It is an approach to AI development and deployment that prioritizes safety and transparency for the people who interact with it.[1] Human agency and human oversight are key to ensure that human-centric software and hardware systems adhere to ethical standards and fundamental user rights.[2] The two terms are very close but are distinct from each other.

Human agency in the context of AI refers to maintaining the autonomy of humans who either use AI systems, or who are exposed to the results of AI systems. While this can be understood as a philosophical concept, in practice, it also means supporting humans in conscious and informed interaction with machines; lacking manipulations, misinformation or the reductions of personal choice and freedom (e.g. caused by addiction). Moreover, humans should retain the right to intervene in automated decisions that have a consequential impact on their well-being. Human oversight, on the other hand, is directly connected with AI operations. It describes different levels of human-computer collaboration, where the human acts as a teacher or supervisor of an AI system and, thus, actively influences either the learning or the acting of the system. This encompasses activities such as monitoring, interpreting and intervening in AI operations, and requires humans to be capable and competent in the context of the AI application. In this sense, the human actor in place is supposed to minimize the risk of AI systems adversely affecting the health, security, safety or fundamental rights of the affected population.

A recent global survey from the Schwartz Reisman Institute for Technology and Society identified that a significant number of respondents were concerned about the impact AI has on violating the privacy of citizens. Privacy concerns ranked third, right after concerns about misuse and impact on jobs, specifically the replacement of human jobs due to AI. Interestingly, the survey revealed that although respondents are relatively open to using AI to assist with various tasks, their level of trust in the AI performing the task effectively tends to be lower than their level of openness. Further, willingness and trust are lower for applications linked with personal identity, expression or emotions. For example, respondents are unsure whether they will use AI to select their clothes or potential romantic partners, but are more amenable to AI helping to plan their vacations or choose their groceries.[3]

North America's first-ever AI legislation

THE COLORADO ARTIFICIAL INTELLIGENCE ACT

On May 17, 2024, the state of Colorado enacted the first comprehensive law regulating AI in the United States. Known as The Colorado Artificial Intelligence Act - or Consumer Protections for Interactions with Artificial Intelligence - this marks the first-ever North American legislation that regulates several aspects related to the development and deployment of AI solutions. The law applies to 'high-risk artificial intelligence systems' – specifically, to all developers who create or modify such systems or to deployers who leverage these systems for operation in Colorado, with potential impact on organizations across their supply chains.

Violations of the Act will be treated as violations of Colorado's General Consumer Protection law, which includes a maximum civil penalty of USD $20,000 per violation.[4] A violation constitutes a breach of the law per consumer, implying that organizations could be subject to hefty fines if a system impacts multiple users simultaneously. Since this legislation will not carry any repercussions until February 1, 2026, organizations have some time to ramp up efforts to ensure that they are compliant.

A future outlook anticipates more legislation of this type being announced in other states and in Canadian provinces, with federal legislations potentially also on the horizon. The provisions and recommendations of this law therefore offer insight into what future legislation might entail, in order to develop and deploy robust AI systems.

Designed to Mitigate Consumer Risk

Although there are several ways in which risk needs to be mitigated when developing or deploying AI solutions, the key area that this legislation focuses on is risk to the consumer. Particularly, the legislation seeks to ensure that consumers are protected when interacting with artificial intelligence systems.[5]

> *"The act requires a developer or a deployer of a high-risk artificial intelligence system (high-risk system) to use reasonable care to protect consumers from any known or reasonably foreseeable risks of algorithmic discrimination in the high-risk system."* [6]

WHAT IS A 'HIGH-RISK AI SYSTEM'?

The law constitutes a 'high-risk AI system' as one that assists in making a consequential decision, that is capable of changing the outcome of a substantial decision, or that generates a consequential decision.

AI SYSTEMS OPERATING FOR THE FOLLOWING PURPOSES ARE CONSIDERED TO BE OF HIGH-RISK:

- Educational enrollment/opportunities
- Employment/employment opportunities
- Financial or lending services
- Essential government services
- Healthcare services
- Housing
- Insurance
- Legal services

IN THE ABOVE AREAS, THE LAW EXCLUDES AI SYSTEMS THAT EITHER:

1. Perform narrow procedural tasks; or
2. Detect decision-making patterns or deviations from prior decision-making patterns and are not intended to replace or influence human assessment or review.

Other technologies are also excluded from the law, such as cybersecurity software or apps that filter spam – since these do not serve as a substantial factor in making consequential decisions.[7]

COMPLEX RESPONSIBILITIES OF DEVELOPERS & DEPLOYERS

The core tenet of the Colorado law is to ensure that developers and deployers of high-risk AI systems use reasonable care to avoid algorithmic discrimination. The law outlines the responsibilities of both developers and deployers to ensure that they account for and prevent algorithmic discrimination in both the development of AI solutions and the delivery of systems to consumers.

Algorithmic discrimination results in differential treatment or impact. It occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people, for example, based on their race, color, ethnicity, sex, religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. Depending on the specific circumstances, such algorithmic discrimination may violate legal protections.[8]

The responsibilities of accountable stakeholders are complex since consideration needs to be given to all members of a given AI system's ecosystem, while also considering technology development and deployment from infancy to the retirement of a given system. The law however excludes any discrimination that may result from the use of a high-risk AI system for the sole purposes of:

1. Self-testing their own systems to identify and rectify incidents or risks of discriminatory behavior/outputs.

2. Expanding an applicant, customer, or participant pool to increase diversity or redress historical discrimination.

The European Union AI Act came into effect in March 2024. It is the first ever AI legislation and has a wider scope than the Colorado legislation. Although it also regulates high-risk AI systems, it categorizes a broader set of technology systems under the umbrella of 'high-risk'. These systems include any that pose significant risk of harm to people's health, safety or fundamental rights.[9]

View the bill summary.

IMPORTANTLY, THE COLORADO LAW ALSO OFFERS RECOMMENDATIONS FOR THE ENFORCEMENT OF RESPONSIBILITIES, TO PREVENT ALGORITHMIC DISCRIMINATION AND TO ENSURE THAT CONSUMERS ARE AWARE OF WHAT THEY ARE BEING EXPOSED TO WHEN INTERACTING WITH A HIGH-RISK AI SYSTEM.

ISO/IEC 42001 for AI legislation compliance

GOVERNANCE FRAME-WORKS NAMED IN THE LEGISLATION

Under the new law, organizations are expected to adopt and implement a 'risk management policy and program' to govern the use of applicable high-risk AI systems. This is explicitly required to be an 'iterative process - planned, implemented and regularly and systematically reviewed and updated over the life cycle of a high-risk artificial intelligence system, requiring a regular, systematic review and updates.'

To meet this standard, the policy and program must also 'specify and incorporate the principles, processes, and personnel that the deployer uses to identify, document, and mitigate known or reasonably foreseeable risks of algorithmic discrimination.'

RECOMMENDED GOVERNANCE PATHS:

- The International Organization for Standardization's (ISO) ISO/IEC 42001 Standard
  or
- The National Institute of Standards and Technology's (NIST) AI Risk Management Framework

ISO/IEC 42001 MITIGATES RISK FOR ORGANIZATIONS DEPLOYING OR DEVELOPING HIGH-RISK AI SYSTEMS.

WHY AN INTERNATIONAL STANDARD?

The International Organization for Standardization (ISO) is a global, non-governmental organization that represents 160 countries through a single standards body for each country. ISO delivers credibility through the implementation of internationally recognized best practices. In December 2023, ISO introduced the ISO/IEC 42001 International Standard - the world's first AI management system (AIMS) standard. It offers a framework to develop trustworthy AI systems. This standard ensures responsible AI development, deployment, and operation - critical for successful adoption and broader digital transformation.

ISO/IEC 42001 was developed, as are all international standards, after extensive discussion and work by a technical committee comprised of experts drawn from across the globe, representing numerous industries. The committee on Artificial Intelligence, ISO/IEC JTC 1/SC 42, repeatedly met over a two-year period to work out the standard's requirements and nuances. The committee continues to meet and includes experts from SGS.

Technical committees ensure standards apply to all types of organizations. They also help to define what needs to be done to manage obligations from a legal and regulatory perspective based on various industries and geographic jurisdictions of operation.

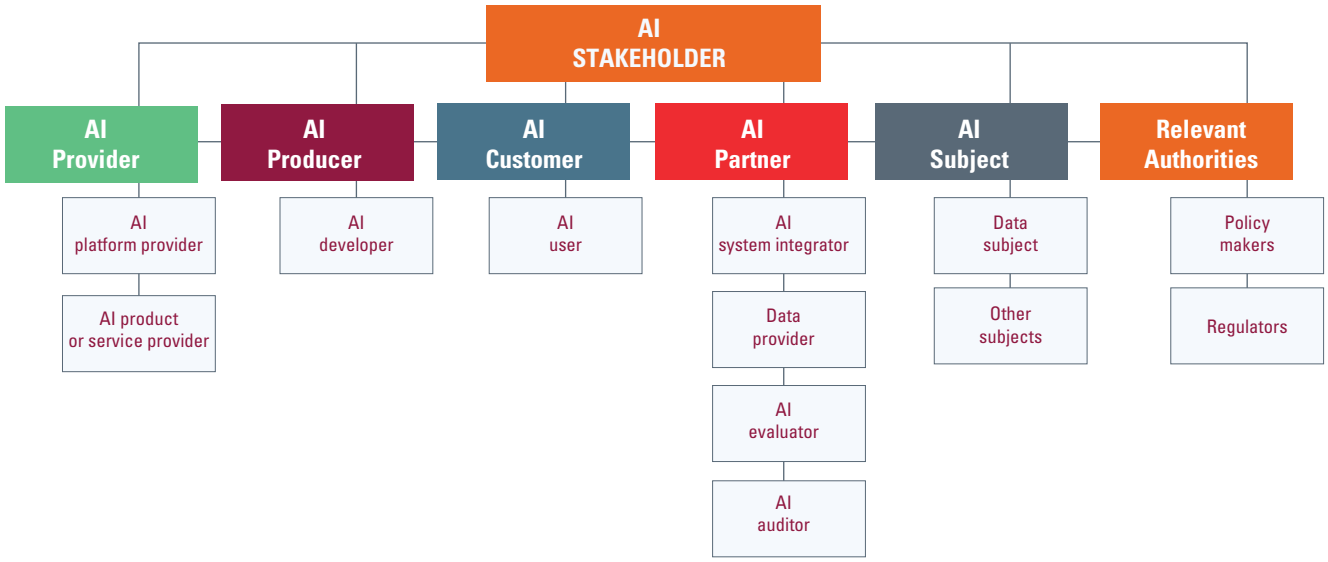
WILLY FABRITIUS
Global Head of Strategy
& Business Development, SGS

"Extensive input from technical experts representing private industry and the public sector culminated in the much anticipated ISO/IEC 42001, which now defines requirements and controls for organizations of all types and sizes to embark on their AI governance journeys."

AI STAKEHOLDER ROLES

In 2022, the ISO/IEC 22989 standard was published, which defined AI terminologies. The new ISO/IEC 42001 standard however defines requirements and controls for the implementation of an AI Management System (AIMS). For continuity, the AI stakeholder roles defined in ISO/IEC 22989 apply within the ISO/IEC 42001 standard. All stakeholders, whether liable for the outcomes of an AI system or impacted by the outcomes of an AI system, are depicted and defined below.

ISO/IEC 22989:2022(E)



AI PROVIDER

An AI provider is an organization or entity that provides products or services that use one or more AI system. AI providers encompass AI platform providers and AI product or service providers.

AI PRODUCER

An AI producer is an organization or entity that designs, develops, tests and deploys products or services that use one or more AI system.

AI DEVELOPER

An AI developer is an organization or entity that is concerned with the development of AI services and products.

AI CUSTOMER

An AI customer is an organization or entity that uses an AI product or service either directly or by its provision to AI users.

AI PARTNER

An AI partner is an organization or entity that provides services in the context of AI.

AI USERS

An AI user is an organization or entity that uses AI products or services.

AI SUBJECT

An AI subject is an organization or entity that is impacted by an AI system, service or product.

RELEVANT AUTHORITIES

Relevant authorities are organizations or entities that can have an impact on an AI system, service or product.

PROTECTING STAKEHOLDERS FROM LIABILITY AND IMPACT

With ISO/IEC 42001, AI 'providers', 'producers', 'developers', 'customers' and 'partners' become equipped with a governance framework that specifies the requirements that are needed to establish, implement, maintain, and continually improve an AI Management System (AIMS). It includes requirements for assessing and treating AI risks, to strengthen accuracy and relevancy of learning and decision-making, as well as to meet the expectations of AI 'users', 'customers', 'subjects' or 'relevant authorities'.

While users of an AI system can be negatively impacted, their actions can also cause liability for their organization. So, in this sense, an 'AI user' is an end user of a system which can pose harm – but at the same time they, the end user, can also cause liability to their organization if improperly using a given AI system. ISO/IEC 42001 outlines requirements to identify risks stemming from the usage of AI systems and offers controls to organizations to protect them from both scenarios.

A SYSTEMATIC RISK-BASED APPROACH FOR DATA AND DECISIONS GENERATED BY AI.

1

Understand, document and prioritize risks related to the AI system under development, deployments and/or usage.

2

Gain visibility into your supply chain (e.g. data sets used to train the AI system) and control the quality of services being procured to develop and maintain the system.

3

Determine and document a plan for communicating incidents to users, customers, subjects and/or relevant authorities of the AI system.

4

Determine and manage the lifecycle of the AI System – from inception to retirement.

CORE CONTROLS AND PROCESSES

RISK MANAGEMENT

Organizations are required to implement processes to identify, analyze, evaluate, and monitor the risks and opportunities during the entire AI system's lifecycle.

IMPACT ASSESSMENT

Organizations must define a process for assessing the potential consequences for individuals or groups of individuals (or both) and societies, that can result from the development, provision or use of AI systems.

ISO/IEC 42005 is another AI-focused standard which is currently being drafted. It will provide more robust guidance on AI system impact assessments including how and when one should be performed. Importantly, it will also offer guidance on how an AI system can be integrated into AI risk management and AI management systems (e.g, integration with ISO 23984 and ISO 42001).

"The AI system impact assessment shall determine the potential consequences an AI system's deployment, intended use and foreseeable misuse has on individuals or groups of individuals, or both, and societies."

"The AI system impact assessment shall take into account the specific technical and societal context where the AI system is deployed and applicable jurisdictions."

– ISO/IEC 42001

SYSTEM LIFECYCLE MANAGEMENT

Organizations must take care of all the aspects of the development of the AI system, including planning, testing and remediating the findings and even the consideration of what happens to the AI at the end of its service (or retirement).

PERFORMANCE OPTIMIZATION

Organizations must continuously improve the effectiveness of their AI system, placing a strong emphasis on not only current, but also future performance.

SUPPLIER MANAGEMENT

Organizations need to extend controls to cover suppliers, who must be aligned with the organization's principles and approach. In the context of AI, "supplier management" includes (but is not limited to) suppliers of the training data used to train the AI.

ORGANIZATIONS CERTIFIED TO THE ISO/IEC 42001 STANDARD DEMONSTRATE THE INTEGRITY OF THEIR AI DATA AND SYSTEMS, AND THEIR COMMITMENT TO INTEGRITY IN AI GENERATED DECISION-MAKING.

Establishing a governance system for trustworthy AI

The requirements and controls involved in establishing a governance system for trustworthy AI pose a daunting task that organizations need to tackle from inception to retirement of their AI systems. Moreover, what is considered a low-risk AI system today could be considered a high-risk AI system in the future; one that is subject to legislation and penalties.

Whether operating in Colorado or beyond, ISO/IEC 42001 is a great place to start your AI governance journey, while continuing to build upon the maturity of your information security posture or to complement management systems that govern other areas of your business.

Every organization has some form of a management system. A management system is simply how an organization is governed. As such, it is often referred to as a governance system. It incorporates the way the organization is run; its policies, procedures and processes. As such, a company can have multiple management systems to cover different areas of the business. For example, an organization may have a management system to cover information security that includes each area of the business that processes or has access to corporate and/or personnel data. Alternatively, a management system may be specific to a very narrow area of the business, for example, finance. For many management systems that span the organization's activities, integrating respective policies and procedures can lead to increased efficiency, expanded efficacy and reduced waste and costs.

ISO/IEC 42001 is now one of several complementary, harmonized standards that exist to govern processes, quality assurance and continuous improvement in areas such as technology and sustainability. With regard to the latter, ISO/IEC 42001 also contributes to UN Sustainable Development Goals 5, 7, 8, 9, 10, 12 and 14.[10]

ISO/IEC 42001 CERTIFICATION: PROCESS & BENEFITS

A cornerstone of trustworthy AI is compliance with standards and regulations, demonstrated through conformity assessments, carried out by accredited and independent third parties. As the world's leading testing, inspection, and certification company, SGS is ready to help you on your AI knowledge-building and certification journey.

Our active contribution to the ongoing development of the family of ISO/IEC 4200x standards firmly positions us to offer the verification and assurance required to develop and/or deploy any type of AI system.

PROCESS

APPLICATION & QUOTE

Obtain a quote for your certification project

COMPETENCE

Identify any skills and competence gaps your staff may have

GAP ASSESSMENT

Identification of any weaknesses

STAGE 1

Cofirmation that the implementation of the management system is on the track

STAGE 2

Confirmation that the management system is fully implemented

CERTIFICATION

Following final technical review, share your success with the world

ONGOING IMPROVEMENT

Regular surveillance visits will ensure your management system remains effective

After the successful completion of Stage 1 and Stage 2 audits and after final review by our technical team, we issue a certificate, confirming your organization's effective implementation of the requirements of ISO/IEC 42001.

BENEFITS

Certifying your organization to ISO/IEC 42001 will provide your organization with the knowledge and verification it needs to be ready for current or upcoming AI legislation. It enables you to move forward with a strong foundation to:

Implement AI safely, with evidence of responsibility and accountability.

Consider security, safety, fairness, transparency and data and AI system quality throughout the AI system lifecycle.

Demonstrate that introducing AI is a strategic decision with clear objectives.

Create strong governance concerning AI.

Strike a balance between governance and innovation.

Ensure that AI is used responsibly, especially concerning its continuous learning.

Combine key frameworks with experience to implement crucial processes like risk, lifecycle and data quality management.

Ensure that all relevant safeguards are in place.

Contribute to multiple UN Sustainable Development Goals (SDGs).

SGS AI resources

WEBINARS

UNLOCK THE SAFE USE OF AI THROUGH STANDARDS, LEGISLATION AND BEST PRACTICE

WATCH NOW

Hear from our experts on the latest developments and updates to standards and regulations on this fast-paced subject. We also discuss the trustworthiness of AI and give direction on how to integrate AI safely and effectively into a business.

ROBUSTNESS AND PERFORMANCE OF AI APPLICATIONS

WATCH NOW

Our experts delve into the vital aspects of AI system performance and robustness, providing an overview of current practices in evaluating AI accuracy and ensuring reliability, as well as a discussion on the unresolved issues and emerging challenges.

TRUSTWORTHY AI: PRIVACY AND SECURITY

WATCH NOW

We explore how to maintain trust and uphold ethical AI use and how advancements in cryptography and privacy-enhancing technologies can safeguard data.

TRUSTWORTHY AI: TRANSPARENCY AND EXPLAINABILITY

WATCH NOW

We address the significance of explainability, reviewing existing approaches and discussing challenges and future prospects.

DIGITAL TRUST CERTIFICATIONS: INFORMATION TECHNOLOGY & AI

Digital Trust Label Certification	ISO/IEC 27001 Certification - Information Security, Cybersecurity & Privacy Protection	ISO/IEC 27701 Certification - Privacy Information Management System	EuroPrivacy – GDPR Certification Information Management System	ISO/IEC 42001 Certification - Artificial Intelligence (AI) Management System

AI TRAINING

Click below to access training.

ISO/IEC 42001:2023 AIMS Requirements Training (Exemplar Global RTP)	ISO/IEC 42001 Foundation Self-Study Training Course	ISO/IEC 42001 Lead Auditor Self-Study Training Course	ISO/IEC 42001 Lead Implementer Self-Study Training Course

REACH OUT TO US

certification@sgs.com

www.sgs.com

References

1
blogs.nvidia.com/blog/what-is-trustworthy-ai/

2
ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.html

3
srinstitute.utoronto.ca/public-opinion-ai

4
iapp.org/news/a/the-colorado-ai-act-what-you-need-to-know

5
leg.colorado.gov/bills/sb24-205

6
leg.colorado.gov/bills/sb24-205

7
www.foley.com/insights/publications/2024/05/colorado-passes-new-ai-law-protect-consumer-interactions/

8
www.whitehouse.gov/ostp/ai-bill-of-rights/

9
www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

10
www.iso.org/standard/81230.html

When you need to be sure

certification@sgs.com

SGS North America Inc.
201 Route 17 North
7th Floor
Rutherford, New Jersey 07070
United States

sgs.com