Implementation of 42001

SCSD Lecture

Controls on Lifecycle

A.6 AI system life cycle		
A.6.1 Management guidance for AI system development		
Objective: To ensure that the organization identifies and documents objectives and implements processes for the responsible design and development of AI systems.		
	Topic	Control
A.6.1.2	Objectives for responsible development of AI system	The organization shall identify and document objectives to guide the responsible development AI systems, and take those objectives into account and integrate measures to achieve them in the development life cycle.
A.6.1.3	Processes for responsible AI system design and development	The organization shall define and document the specific processes for the responsible design and development of the AI system.

Controls

- The controls in the annexe correspond to the section in the main standard.
- The Annex is normative (not that the controls contain “shall”)
- The controls are listed in Annex A but implementation is provided in Annex B which is also normative.
- Here we will explore the “AI lifecycle” section.

AI System Lifecycle

A.6.2 AI system life cycle		
Objective: To define the criteria and requirements for each stage of the AI system life cycle.		
	Topic	Control
A.6.2.2	AI system requirements and specification	The organization shall specify and document requirements for new AI systems or material enhancements to existing systems.
A.6.2.3	Documentation of AI system design and development	The organization shall document the AI system design and development based on organizational objectives, documented requirements and specification criteria.
A.6.2.4	AI system verification and validation	The organization shall define and document verification and validation measures for the AI system and specify criteria for their use.
A.6.2.5	AI system deployment	The organization shall document a deployment plan and ensure that appropriate requirements are met prior to deployment.

A.6.2.6	AI system operation and monitoring	The organization shall define and document the necessary elements for the ongoing operation of the AI system. At the minimum, this should include system and performance monitoring, repairs, updates and support.
A.6.2.7	AI system technical documentation	The organization shall determine what AI system technical documentation is needed for each relevant category of interested parties, such as users, partners, supervisory authorities, and provide the technical documentation to them in the appropriate form.
A.6.2.8	AI system recording of event logs	The organization shall determine at which phases of the AI system life cycle, record keeping of event logs should be enabled, but at the minimum when the AI system is in use.

AI Lifecycle controls

- Documents the entire lifecycle:
  - Requirements
  - Design
  - V&V
  - Deployment
  - Operation and Monitoring
  - Documentation
  - Event logging
- The stages are defined and the controls place requirements on the process but leave many details undefined.

Implementation of Controls

Annex B
(normative)

Implementation guidance for AI controls

B.1 General

The implementation guidance documented in this annex relates to the controls listed in Table A.1. It provides information to support the implementation of the controls listed in Table A.1 and to meet the control objective, but organizations do not have to document or justify inclusion or exclusion of implementation guidance in the statement of applicability (see 6.1.3).

The implementation guidance is not always suitable or sufficient in all situations and does not always fulfil the organization's specific control requirements. The organization can extend or modify the implementation guidance or define their own implementation of a control according to their specific requirements and risk treatment needs.

This annex is to be used as guidance for determining and implementing controls for AI risk treatment in the AI management system defined in this document. Additional organizational and technical controls other than those included in this annex can be determined (see AI system management risk treatment in 6.1.3). This annex can be regarded as a starting point for developing organization-specific implementation of controls.

Implementing the lifecycle

B.6.2 AI system life cycle

B.6.2.1 Objective

To define the criteria and requirements for each stage of the AI system life cycle.

B.6.2.2 AI system requirements and specification

Control

The organization should specify and document requirements for new AI systems or material enhancements to existing systems.

Guidance and Cross Reference

Implementation guidance

The organization should document the rationale for developing an AI system and its goals. Some of the factors that should be considered, documented and understood can include:

a) why the AI system is to be developed, for example, is this driven by a business case, customer request or by government policy;

b) how the model can be trained and how data requirements can be achieved.

AI system requirements should be specified and should span the entire AI system life cycle. Such requirements should be revisited in cases where the developed AI system is unable to operate as intended or new information arises that can be used to change and to improve the requirements. For instance, it can become unfeasible from a financial perspective to develop the AI system.

Other information

The processes for describing the AI system life cycle are provided by ISO/IEC 5338. For more information about human-centred design for interactive systems, see ISO 9241-210.

Guidance

- Note that this is guidance
- Notice the use of “should”
- Notice the references to other standards that cover different aspects
  - AI System Lifecycle
  - Human-centered design

In more detail – consider V&V

B.6.2.4 AI system verification and validation

Control

The organization should define and document verification and validation measures for the AI system and specify criteria for their use.

Implementation guidance

The verification and validation measures can include, but are not limited to:

— testing methodologies and tools;

— selection of test data and their representation of the intended domain of use;

— release criteria requirements.

What to include

The organization should define and document evaluation criteria such as, but not limited to:

- a plan to evaluate the AI system components and the whole AI system for risks related to impacts on individuals or groups of individuals, or both, and societies;
- the evaluation plan can be based on, for example:
  - reliability and safety requirements of the AI system, including acceptable error rates for the AI system performance;
  - responsible AI system development and use objectives such as those in B.6.1.2 and B.9.3;
  - operational factors such as quality of data, intended use, including acceptable ranges of each operational factor;
  - any intended uses which can require more rigorous operational factors to be defined, including different acceptable ranges for operational factors or lower error rates;

Mitigating issues

— the methods, guidance or metrics to be used to evaluate whether relevant interested parties who make decisions or are subject to decisions based on the AI system outputs can adequately interpret the AI system outputs. The frequency of evaluation should be determined and can be based upon results from an AI system impact assessment;

— any acceptable factors that can account for an inability to meet a target minimum performance level, especially when the AI system is evaluated for impacts on individuals or groups of individuals, or both, and societies (e.g. poor image resolution for computer vision systems or background noise affecting speech recognition systems). Mechanisms to deal with poor AI system performance as a result of these factors should also be documented.

Further Issues

The AI system should be evaluated against the documented criteria for evaluation.

Where the AI system cannot meet the documented criteria for evaluation, especially against responsible AI system development and use objectives (see B.6.1.2 and B.9.3), the organization should reconsider or manage the deficiencies of the intended use of the AI system, its performance requirements and how the organization can effectively address the impacts to individuals or groups of individuals, or both, and societies.

NOTE Further information on how to deal with robustness of neural networks can be found in ISO/IEC TR 24029-1.

Annex B

- This includes implementation guidance for each of the controls.
- These are generally “should” statements
- The organization should follow these where possible
- The standards are not prescriptive in many areas
- Standards become more prescriptive the more specialized the domain becomes

References

- ISO 42001 refers out to many other standards:
  - Domain specific standards e.g. Food, Medical
  - Safety standards
  - Ergonomics/Usability/Accessibility
  - Performance
  - Risk Management
  - Engineering
  - Security
  - Governance
  - Lifecycles

Bibliography

[1] ISO 8000-2, *Data quality — Part 2: Vocabulary*

[2] ISO 9001, *Quality management systems — Requirements*

[3] ISO 9241-210, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*

[4] ISO 13485, *Medical devices — Quality management systems — Requirements for regulatory purposes*

[5] ISO 22000, *Food safety management systems — Requirements for any organization in the food chain*

[6] IEC 62304, *Medical device software — Software life cycle processes*

[7] ISO/IEC Guide 51, *Safety aspects — Guidelines for their inclusion in standards*

[8] ISO/IEC TS 4213, *Information technology — Artificial intelligence — Assessment of machine learning classification performance*

[9] ISO/IEC 5259 (all parts[2]), *Data quality for analytics and machine learning (ML)*

[10] ISO/IEC 5338, *Information technology — Artificial intelligence — AI system life cycle process*

[11] ISO/IEC 17065, *Conformity assessment — Requirements for bodies certifying products, processes and services*

[12] ISO/IEC 19944-1, *Cloud computing and distributed platforms — Data flow, data categories and data use — Part 1: Fundamentals*

[13] ISO/IEC 23053, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

[14] ISO/IEC 23894, *Information technology — Artificial intelligence — Guidance on risk management*

[15] ISO/IEC TR 24027, *Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*

[16] ISO/IEC TR 24029-1, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*

[17] ISO/IEC TR 24368, *Information technology — Artificial intelligence — Overview of ethical and societal concerns*

[18] ISO/IEC 25024, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality*

[19] ISO/IEC 25059, *Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems*

[20] ISO/IEC 27000:2018, *Information technology — Security techniques — Information security management systems — Overview and vocabulary*

[21] ISO/IEC 27701, *Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines*

[22] ISO/IEC 27001, *Information security, cybersecurity and privacy protection — Information security management systems — Requirements*

[23] ISO/IEC 29100, *Information technology — Security techniques — Privacy framework*

[24] ISO 31000:2018, *Risk management — Guidelines*

[25] ISO 37002, *Whistleblowing management systems — Guidelines*

[26] ISO/IEC 38500:2015, *Information technology — Governance of IT for the organization*

[27] ISO/IEC 38507, *Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations*

[28] Lifecycle D.D.I. 3.3, 2020-04-15. Data Documentation Initiative (DDI) Alliance. [viewed on 2022-02-19]. Available at: https://ddialliance.org/Specification/DDI-Lifecycle/3.3/

[29] Risk Framework N.I.S.T.-A.I. 1.0, 2023-01-26. National Institute of Technology (NIST) [viewed on 2023-04-17] https://www.nist.gov/itl/ai-risk-management-framework