BS ISO/IEC 42001:2023

INTERNATIONAL STANDARD

ISO/IEC 42001

First edition
2023-12

Information technology — Artificial intelligence — Management system

Technologies de l'information — Intelligence artificielle — Système de management

Reference number
ISO/IEC 42001:2023(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword v
Introduction vi
1 Scope 1
2 Normative references 1
3 Terms and definitions 1
4 Context of the organization 5
4.1 Understanding the organization and its context 5
4.2 Understanding the needs and expectations of interested parties 6
4.3 Determining the scope of the AI management system 6
4.4 AI management system 6
5 Leadership 7
5.1 Leadership and commitment 7
5.2 AI policy 7
5.3 Roles, responsibilities and authorities 8
6 Planning 8
6.1 Actions to address risks and opportunities 8
6.1.1 General 8
6.1.2 AI risk assessment 9
6.1.3 AI risk treatment 9
6.1.4 AI system impact assessment 10
6.2 AI objectives and planning to achieve them 10
6.3 Planning of changes 11
7 Support 11
7.1 Resources 11
7.2 Competence 11
7.3 Awareness 12
7.4 Communication 12
7.5 Documented information 12
7.5.1 General 12
7.5.2 Creating and updating documented information 12
7.5.3 Control of documented information 13
8 Operation 13
8.1 Operational planning and control 13
8.2 AI risk assessment 13
8.3 AI risk treatment 14
8.4 AI system impact assessment 14
9 Performance evaluation 14
9.1 Monitoring, measurement, analysis and evaluation 14
9.2 Internal audit 14
9.2.1 General 14
9.2.2 Internal audit programme 14
9.3 Management review 15
9.3.1 General 15
9.3.2 Management review inputs 15
9.3.3 Management review results 15
10 Improvement 15
10.1 Continual improvement 15
10.2 Nonconformity and corrective action 16
Annex A (normative) **Reference control objectives and controls** 17

Annex B (normative) **Implementation guidance for AI controls** **21**
Annex C (informative) **Potential AI-related organizational objectives and risk sources** **46**
Annex D (informative) **Use of the AI management system across domains or sectors** **49**
Bibliography **51**

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

Artificial intelligence (AI) is increasingly applied across all sectors utilizing information technology and is expected to be one of the main economic drivers. A consequence of this trend is that certain applications can give rise to societal challenges over the coming years.

This document intends to help organizations responsibly perform their role with respect to AI systems (e.g. to use, develop, monitor or provide products or services that utilize AI). AI potentially raises specific considerations such as:

— The use of AI for automatic decision-making, sometimes in a non-transparent and non-explainable way, can require specific management beyond the management of classical IT systems.

— The use of data analysis, insight and machine learning, rather than human-coded logic to design systems, both increases the application opportunities for AI systems and changes the way that such systems are developed, justified and deployed.

— AI systems that perform continuous learning change their behaviour during use. They require special consideration to ensure their responsible use continues with changing behaviour.

This document provides requirements for establishing, implementing, maintaining and continually improving an AI management system within the context of an organization. Organizations are expected to focus their application of requirements on features that are unique to AI. Certain features of AI, such as the ability to continuously learn and improve or a lack of transparency or explainability, can warrant different safeguards if they raise additional concerns compared to how the task would traditionally be performed. The adoption of an AI management system to extend the existing management structures is a strategic decision for an organization.

The organization's needs and objectives, processes, size and structure as well as the expectations of various interested parties influence the establishment and implementation of the AI management system. Another set of factors that influence the establishment and implementation of the AI management system are the many use cases for AI and the need to strike the appropriate balance between governance mechanisms and innovation. Organizations can elect to apply these requirements using a risk-based approach to ensure that the appropriate level of control is applied for the particular AI use cases, services or products within the organization's scope. All these influencing factors are expected to change and be reviewed from time to time.

The AI management system should be integrated with the organization's processes and overall management structure. Specific issues related to AI should be considered in the design of processes, information systems and controls. Crucial examples of such management processes are:

— determination of organizational objectives, involvement of interested parties and organizational policy;

— management of risks and opportunities;

— processes for the management of concerns related to the trustworthiness of AI systems such as security, safety, fairness, transparency, data quality and quality of AI systems throughout their life cycle;

— processes for the management of suppliers, partners and third parties that provide or develop AI systems for the organization.

This document provides guidelines for the deployment of applicable controls to support such processes.

This document avoids specific guidance on management processes. The organization can combine generally accepted frameworks, other International Standards and its own experience to implement crucial processes such as risk management, life cycle management and data quality management which are appropriate for the specific AI use cases, products or services within the scope.

An organization conforming with the requirements in this document can generate evidence of its responsibility and accountability regarding its role with respect to AI systems.

The order in which requirements are presented in this document does not reflect their importance or imply the order in which they are implemented. The list items are enumerated for reference purposes only.

Compatibility with other management system standards

This document applies the harmonized structure (identical clause numbers, clause titles, text and common terms and core definitions) developed to enhance alignment among management system standards (MSS). The AI management system provides requirements specific to managing the issues and risks arising from using AI in an organization. This common approach facilitates implementation and consistency with other management system standards, e.g. related to quality, safety, security and privacy.

Information technology — Artificial intelligence — Management system

1 Scope

This document specifies the requirements and provides guidance for establishing, implementing, maintaining and continually improving an AI (artificial intelligence) management system within the context of an organization.

This document is intended for use by an organization providing or using products or services that utilize AI systems. This document is intended to help the organization develop, provide or use AI systems responsibly in pursuing its objectives and meet applicable requirements, obligations related to interested parties and expectations from them.

This document is applicable to any organization, regardless of size, type and nature, that provides or uses products or services that utilize AI systems.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
organization
person or group of people that has its own functions with responsibilities, authorities and relationships to achieve its *objectives* (3.6)

Note 1 to entry: The concept of organization includes, but is not limited to, sole-trader, company, corporation, firm, enterprise, authority, partnership, charity or institution or part or combination thereof, whether incorporated or not, public or private.

Note 2 to entry: If the organization is part of a larger entity, the term "organization" refers only to the part of the larger entity that is within the scope of the AI *management system* (3.4).

3.2
interested party
person or *organization* (3.1) that can affect, be affected by, or perceive itself to be affected by a decision or activity

Note 1 to entry: An overview of interested parties in AI is provided in ISO/IEC 22989:2022, 5.19.

3.3
top management
person or group of people who directs and controls an *organization* (3.1) at the highest level

Note 1 to entry: Top management has the power to delegate authority and provide resources within the organization.

Note 2 to entry: If the scope of the *management system* (3.4) covers only part of an organization, then top management refers to those who direct and control that part of the organization.

3.4
management system
set of interrelated or interacting elements of an *organization* (3.1) to establish *policies* (3.5) and *objectives* (3.6), as well as *processes* (3.8) to achieve those objectives

Note 1 to entry: A management system can address a single discipline or several disciplines.

Note 2 to entry: The management system elements include the organization's structure, roles and responsibilities, planning and operation.

3.5
policy
intentions and direction of an *organization* (3.1) as formally expressed by its *top management* (3.3)

3.6
objective
result to be achieved

Note 1 to entry: An objective can be strategic, tactical, or operational.

Note 2 to entry: Objectives can relate to different disciplines (such as finance, health and safety, and environment). They can be, for example, organization-wide or specific to a project, product or *process* (3.8).

Note 3 to entry: An objective can be expressed in other ways, e.g. as an intended result, as a purpose, as an operational criterion, as an AI objective or by the use of other words with similar meaning (e.g. aim, goal, or target).

Note 4 to entry: In the context of AI *management systems* (3.4), AI objectives are set by the *organization* (3.1), consistent with the AI *policy* (3.5), to achieve specific results.

3.7
risk
effect of uncertainty

Note 1 to entry: An effect is a deviation from the expected — positive or negative.

Note 2 to entry: Uncertainty is the state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood.

Note 3 to entry: Risk is often characterized by reference to potential events (as defined in ISO Guide 73) and consequences (as defined in ISO Guide 73), or a combination of these.

Note 4 to entry: Risk is often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated likelihood (as defined in ISO Guide 73) of occurrence.

3.8
process
set of interrelated or interacting activities that uses or transforms inputs to deliver a result

Note 1 to entry: Whether the result of a process is called an output, a product or a service depends on the context of the reference.

3.9
competence
ability to apply knowledge and skills to achieve intended results

3.10
documented information
information required to be controlled and maintained by an *organization* (3.1) and the medium on which it is contained

Note 1 to entry: Documented information can be in any format and media and from any source.

Note 2 to entry: Documented information can refer to:

— the *management system* (3.4), including related *processes* (3.8);

— information created in order for the organization to operate (documentation);

— evidence of results achieved (records).

3.11
performance
measurable result

Note 1 to entry: Performance can relate either to quantitative or qualitative findings.

Note 2 to entry: Performance can relate to managing activities, *processes* (3.8), products, services, systems or *organizations* (3.1).

Note 3 to entry: In the context of this document, performance refers both to results achieved by using AI systems and results related to the AI *management system* (3.4). The correct interpretation of the term is clear from the context of its use.

3.12
continual improvement
recurring activity to enhance *performance* (3.11)

3.13
effectiveness
extent to which planned activities are realized and planned results are achieved

3.14
requirement
need or expectation that is stated, generally implied or obligatory

Note 1 to entry: "Generally implied" means that it is custom or common practice for the *organization* (3.1) and *interested parties* (3.2) that the need or expectation under consideration is implied.

Note 2 to entry: A specified requirement is one that is stated, e.g. in *documented information* (3.10).

3.15
conformity
fulfilment of a *requirement* (3.14)

3.16
nonconformity
non-fulfilment of a *requirement* (3.14)

3.17
corrective action
action to eliminate the cause(s) of a *nonconformity* (3.16) and to prevent recurrence

3.18
audit
systematic and independent *process* (3.8) for obtaining evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled

Note 1 to entry: An audit can be an internal audit (first party) or an external audit (second party or third party), and it can be a combined audit (combining two or more disciplines).

Note 2 to entry: An internal audit is conducted by the *organization* (3.1) itself, or by an external party on its behalf.

Note 3 to entry: "Audit evidence" and "audit criteria" are defined in ISO 19011.

3.19
measurement
process (3.8) to determine a value

3.20
monitoring
determining the status of a system, a *process* (3.8) or an activity

Note 1 to entry: To determine the status, there can be a need to check, supervise or critically observe.

3.21
control
<risk> measure that maintains and/or modifies *risk* (3.7)

Note 1 to entry: Controls include, but are not limited to, any process, policy, device, practice or other conditions and/or actions which maintain and/or modify risk.

Note 2 to entry: Controls may not always exert the intended or assumed modifying effect.

[SOURCE: ISO 31000:2018, 3.8, modified — Added <risk> as application domain]

3.22
governing body
person or group of people who are accountable for the performance and conformance of the organization

Note 1 to entry: Not all organizations, particularly small organizations, will have a governing body separate from top management.

Note 2 to entry: A governing body can include, but is not limited to, board of directors, committees of the board, supervisory board, trustees or overseers.

[SOURCE: ISO/IEC 38500:2015 2.9, modified — Added Notes to entry.]

3.23
information security
preservation of confidentiality, integrity and availability of information

Note 1 to entry: Other properties such as authenticity, accountability, non-repudiation and reliability can also be involved.

[SOURCE: ISO/IEC 27000:2018, 3.28]

3.24
AI system impact assessment
formal, documented process by which the impacts on individuals, groups of individuals, or both, and societies are identified, evaluated and addressed by an organization developing, providing or using products or services utilizing artificial intelligence

3.25
data quality
characteristic of data that the data meet the organization's data requirements for a specific context

[SOURCE: ISO/IEC 5259-1:—[1)], 3.4]

3.26
statement of applicability
documentation of all necessary *controls* (3.23) and justification for inclusion or exclusion of controls

Note 1 to entry: Organizations may not require all controls listed in Annex A or may even exceed the list in Annex A with additional controls established by the organization itself.

Note 2 to entry: All identified risks shall be documented by the organization according to the requirements of this document. All identified risks and the risk management measures (controls) established to address them shall be reflected in the statement of applicability.

4 Context of the organization

4.1 Understanding the organization and its context

The organization shall determine external and internal issues that are relevant to its purpose and that affect its ability to achieve the intended result(s) of its AI management system.

The organization shall determine whether climate change is a relevant issue.

The organization shall consider the intended purpose of the AI systems that are developed, provided or used by the organization. The organization shall determine its roles with respect to these AI systems.

NOTE 1 To understand the organization and its context, it can be helpful for the organization to determine its role relative to the AI system. These roles can include, but are not limited to, one or more of the following:

— AI providers, including AI platform providers, AI product or service providers;

— AI producers, including AI developers, AI designers, AI operators, AI testers and evaluators, AI deployers, AI human factor professionals, domain experts, AI impact assessors, procurers, AI governance and oversight professionals;

— AI customers, including AI users;

— AI partners, including AI system integrators and data providers;

— AI subjects, including data subjects and other subjects;

— relevant authorities, including policymakers and regulators.

A detailed description of these roles is provided by ISO/IEC 22989. Furthermore, the types of roles and their relationship to the AI system life cycle are also described in the NIST AI risk management framework.[29] The organization's roles can determine the applicability and extent of applicability of the requirements and controls in this document.

NOTE 2 External and internal issues to be addressed under this clause can vary according to the organization's roles and jurisdiction and their impact on its ability to achieve the intended outcome(s) of its AI management system. These can include, but are not limited to:

a) external context related considerations such as:

   1) applicable legal requirements, including prohibited uses of AI;

   2) policies, guidelines and decisions from regulators that have an impact on the interpretation or enforcement of legal requirements in the development and use of AI systems;

1) Under preparation. Stage at the time of publication ISO/IEC DIS 5259-1:2023.

3) incentives or consequences associated with the intended purpose and the use of AI systems;

4) culture, traditions, values, norms and ethics with respect to development and use of AI;

5) competitive landscape and trends for new products and services using AI systems;

b) internal context related considerations such as:

1) organizational context, governance, objectives (see 6.2), policies and procedures;

2) contractual obligations;

3) intended purpose of the AI system to be developed or used.

NOTE 3 Role determination can be formed by obligations related to categories of data the organization processes (e.g. personally identifiable information (PII) processor or PII controller when processing PII). See ISO/IEC 29100 for PII and related roles. Roles can also be informed by legal requirements specific to AI systems.

4.2 Understanding the needs and expectations of interested parties

The organization shall determine:

— the interested parties that are relevant to the AI management system;

— the relevant requirements of these interested parties;

— which of these requirements will be addressed through the AI management system.

NOTE Relevant interested parties can have requirements related to climate change.

4.3 Determining the scope of the AI management system

The organization shall determine the boundaries and applicability of the AI management system to establish its scope.

When determining this scope, the organization shall consider:

— the external and internal issues referred to in 4.1;

— the requirements referred to in 4.2.

The scope shall be available as documented information.

The scope of the AI management system shall determine the organization's activities with respect to this document's requirements on the AI management system, leadership, planning, support, operation, performance, evaluation, improvement, controls and objectives.

4.4 AI management system

The organization shall establish, implement, maintain, continually improve and document an AI management system, including the processes needed and their interactions, in accordance with the requirements of this document.

5 Leadership

5.1 Leadership and commitment

Top management shall demonstrate leadership and commitment with respect to the AI management system by:

— ensuring that the AI policy (see 5.2) and AI objectives (see 6.2) are established and are compatible with the strategic direction of the organization;

— ensuring the integration of the AI management system requirements into the organization's business processes;

— ensuring that the resources needed for the AI management system are available;

— communicating the importance of effective AI management and of conforming to the AI management system requirements;

— ensuring that the AI management system achieves its intended result(s);

— directing and supporting persons to contribute to the effectiveness of the AI management system;

— promoting continual improvement;

— supporting other relevant roles to demonstrate their leadership as it applies to their areas of responsibility.

NOTE 1 Reference to "business" in this document can be interpreted broadly to mean those activities that are core to the purposes of the organization's existence.

NOTE 2 Establishing, encouraging and modelling a culture within the organization, to take a responsible approach to using, development and governing AI systems can be an important demonstration of commitment and leadership by top management. Ensuring awareness of and compliance with such a responsible approach and in support of the AI management system through leadership can aid the success of the AI management system.

5.2 AI policy

Top management shall establish an AI policy that:

a) is appropriate to the purpose of the organization;

b) provides a framework for setting AI objectives (see 6.2);

c) includes a commitment to meet applicable requirements;

d) includes a commitment to continual improvement of the AI management system.

The AI policy shall:

— be available as documented information;

— refer as relevant to other organizational policies;

— be communicated within the organization;

— be available to interested parties, as appropriate.

Control objectives and controls for establishing an AI policy are provided in A.2 in Table A.1. Implementation guidance for these controls is provided in B.2.

NOTE Considerations for organizations when developing AI policies are provided in ISO/IEC 38507.

5.3 Roles, responsibilities and authorities

Top management shall ensure that the responsibilities and authorities for relevant roles are assigned and communicated within the organization.

Top management shall assign the responsibility and authority for:

a) ensuring that the AI management system conforms to the requirements of this document;

b) reporting on the performance of the AI management system to top management.

NOTE A control for defining and allocating roles and responsibilities is provided in A.3.2 in Table A.1. Implementation guidance for this control is provided in B.3.2.

6 Planning

6.1 Actions to address risks and opportunities

6.1.1 General

When planning for the AI management system, the organization shall consider the issues referred to in 4.1 and the requirements referred to in 4.2 and determine the risks and opportunities that need to be addressed to:

— give assurance that the AI management system can achieve its intended result(s);

— prevent or reduce undesired effects;

— achieve continual improvement.

The organization shall establish and maintain AI risk criteria that support:

— distinguishing acceptable from non-acceptable risks;

— performing AI risk assessments;

— conducting AI risk treatment;

— assessing AI risk impacts.

NOTE 1 Considerations to determine the amount and type of risk that an organization is willing to pursue or retain are provided in ISO/IEC 38507 and ISO/IEC 23894.

The organization shall determine the risks and opportunities according to:

— the domain and application context of an AI system;

— the intended use;

— the external and internal context described in 4.1.

NOTE 2 More than one AI system can be considered in the scope of the AI management system. In this case the determination of opportunities and uses is performed for each AI system or groupings of AI systems.

The organization shall plan:

a) actions to address these risks and opportunities;

b) how to:

1) integrate and implement the actions into its AI management system processes;

2) evaluate the effectiveness of these actions.

The organization shall retain documented information on actions taken to identify and address AI risks and AI opportunities.

NOTE 3 Guidance on how to implement risk management for organizations developing, providing or using AI products, systems and services is provided in ISO/IEC 23894.

NOTE 4 The context of the organization and its activities can have an impact on the organization's risk management activities.

NOTE 5 The way of defining risk and therefore of envisioning risk management can vary across sectors and industries. The definition of risk in 3.7 allows a broad vision of risk adaptable to any sector, such as the sectors mentioned in Annex D. In any case, it is the role of the organization, as part of risk assessment, to first adopt a vision of risk adapted to its context. This can include approaching risk through definitions used in sectors where the AI system is developed for and used, such as the definition from ISO/IEC Guide 51.

6.1.2 AI risk assessment

The organization shall define and establish an AI risk assessment process that:

a) is informed by and aligned with the AI policy (see 5.2) and AI objectives (see 6.2);

NOTE When assessing the consequences as part of 6.1.2 d) 1), the organization can utilize an AI system impact assessment as indicated in 6.1.4.

b) is designed such that repeated AI risk assessments can produce consistent, valid and comparable results;

c) identifies risks that aid or prevent achieving its AI objectives;

d) analyses the AI risks to:

1) assess the potential consequences to the organization, individuals and societies that would result if the identified risks were to materialize;

2) assess, where applicable, the realistic likelihood of the identified risks;

3) determine the levels of risk;

e) evaluates the AI risks to:

1) compare the results of the risk analysis with the risk criteria (see 6.1.1);

2) prioritize the assessed risks for risk treatment.

The organization shall retain documented information about the AI risk assessment process.

6.1.3 AI risk treatment

Taking the risk assessment results into account, the organization shall define an AI risk treatment process to:

a) select appropriate AI risk treatment options;

b) determine all controls that are necessary to implement the AI risk treatment options chosen and compare the controls with those in Annex A to verify that no necessary controls have been omitted;

NOTE 1 Annex A provides reference controls for meeting organizational objectives and addressing risks related to the design and use of AI systems.

c) consider the controls from Annex A that are relevant for the implementation of the AI risk treatment options;

d) identify if additional controls are necessary beyond those in Annex A in order to implement all risk treatment options;

e) consider the guidance in Annex B for the implementation of controls determined in b) and c);

NOTE 2 Control objectives are implicitly included in the controls chosen. The organization can select an appropriate set of control objectives and controls from Annex A. The Annex A controls are not exhaustive and additional control objectives and controls can be needed. If different or additional controls are necessary beyond those in Annex A, the organization can design such controls or take them from existing sources. AI risk management can be integrated in other management systems, if applicable.

f) produce a statement of applicability that contains the necessary controls [see b), c) and d)] and provide justification for inclusion and exclusion of controls. Justification for exclusion can include where the controls are not deemed necessary by the risk assessment and where they are not required by (or are subject to exceptions under) applicable external requirements.

NOTE 3 The organization can provide documented justifications for excluding any control objectives in general or for specific AI systems, whether those listed in Annex A or established by the organization itself.

g) formulate an AI risk treatment plan.

The organization shall obtain approval from the designated management for the AI risk treatment plan and for acceptance of the residual AI risks. The necessary controls shall be:

— aligned to the objectives in 6.2;

— available as documented information;

— communicated within the organization;

— available to interested parties, as appropriate.

The organization shall retain documented information about the AI risk treatment process.

6.1.4 AI system impact assessment

The organization shall 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.

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.

The result of the AI system impact assessment shall be documented. Where appropriate, the result of the system impact assessment can be made available to relevant interested parties as defined by the organization.

The organization shall consider the results of the AI system impact assessment in the risk assessment (see 6.1.2). A.5 in Table A.1 provides controls for assessing impacts of AI systems.

NOTE In some contexts (such as safety or privacy critical AI systems), the organization can require that discipline-specific AI system impact assessments (e.g. safety, privacy or security impact) be performed as part of the overall risk management activities of an organization.

6.2 AI objectives and planning to achieve them

The organization shall establish AI objectives at relevant functions and levels.

The AI objectives shall:

a) be consistent with the AI policy (see 5.2);

b) be measurable (if practicable);

c) take into account applicable requirements;

d) be monitored;

e) be communicated;

f) be updated as appropriate;

g) be available as documented information.

When planning how to achieve its AI objectives, the organization shall determine:

— what will be done;

— what resources will be required;

— who will be responsible;

— when it will be completed;

— how the results will be evaluated.

NOTE A non-exclusive list of AI objectives relating to risk management is provided in Annex C. Control objectives and controls for identifying objectives for responsible development and use of AI systems and measures to achieve them are provided in A.6.1 and A.9.3 in Table A.1. Implementation guidance for these controls is provided in B.6.1 and B.9.3.

6.3 Planning of changes

When the organization determines the need for changes to the AI management system, the changes shall be carried out in a planned manner.

7 Support

7.1 Resources

The organization shall determine and provide the resources needed for the establishment, implementation, maintenance and continual improvement of the AI management system.

NOTE Control objectives and controls for AI resources are provided in A.4 in Table A.1. Implementation guidance for these controls is provided in Clause B.4.

7.2 Competence

The organization shall:

— determine the necessary competence of person(s) doing work under its control that affects its AI performance;

— ensure that these persons are competent on the basis of appropriate education, training or experience;

— where applicable, take actions to acquire the necessary competence, and evaluate the effectiveness of the actions taken.

Appropriate documented information shall be available as evidence of competence.

NOTE 1 Implementation guidance for human resources including consideration of necessary expertise is provided in B.4.6.

NOTE 2 Applicable actions can include, for example: the provision of training to, the mentoring of, or the re-assignment of currently employed persons; or the hiring or contracting of competent persons.

7.3 Awareness

Persons doing work under the organization's control shall be aware of:

— the AI policy (see 5.2);

— their contribution to the effectiveness of the AI management system, including the benefits of improved AI performance;

— the implications of not conforming with the AI management system requirements.

7.4 Communication

The organization shall determine the internal and external communications relevant to the AI management system including:

— what it will communicate;

— when to communicate;

— with whom to communicate;

— how to communicate.

7.5 Documented information

7.5.1 General

The organization's AI management system shall include:

a) documented information required by this document;

b) documented information determined by the organization as being necessary for the effectiveness of the AI management system.

NOTE The extent of documented information for an AI management system can differ from one organization to another due to:

— the size of organization and its type of activities, processes, products and services;

— the complexity of processes and their interactions;

— the competence of persons.

7.5.2 Creating and updating documented information

When creating and updating documented information, the organization shall ensure appropriate:

— identification and description (e.g. a title, date, author or reference number);

— format (e.g. language, software version, graphics) and media (e.g. paper, electronic);

— review and approval for suitability and adequacy.

7.5.3 Control of documented information

Documented information required by the AI management system and by this document shall be controlled to ensure:

a) it is available and suitable for use, where and when it is needed;

b) it is adequately protected (e.g. from loss of confidentiality, improper use or loss of integrity).

For the control of documented information, the organization shall address the following activities, as applicable:

— distribution, access, retrieval and use;

— storage and preservation, including preservation of legibility;

— control of changes (e.g. version control);

— retention and disposition.

Documented information of external origin determined by the organization to be necessary for the planning and operation of the AI management system shall be identified as appropriate and controlled.

NOTE Access can imply a decision regarding the permission to view the documented information only, or the permission and authority to view and change the documented information.

8 Operation

8.1 Operational planning and control

The organization shall plan, implement and control the processes needed to meet requirements, and to implement the actions determined in Clause 6, by:

— establishing criteria for the processes;

— implementing control of the processes in accordance with the criteria.

The organization shall implement the controls determined according to 6.1.3 that are related to the operation of the AI management system (e.g. AI system development and usage life cycle related controls).

The effectiveness of these controls shall be monitored and corrective actions shall be considered if the intended results are not achieved. Annex A lists reference controls and Annex B provides implementation guidance for them.

Documented information shall be available to the extent necessary to have confidence that the processes have been carried out as planned.

The organization shall control planned changes and review the consequences of unintended changes, taking action to mitigate any adverse effects, as necessary.

The organization shall ensure that externally provided processes, products or services that are relevant to the AI management system are controlled.

8.2 AI risk assessment

The organization shall perform AI risk assessments in accordance with 6.1.2 at planned intervals or when significant changes are proposed or occur.

The organization shall retain documented information of the results of all AI risk assessments.

8.3 AI risk treatment

The organization shall implement the AI risk treatment plan according to 6.1.3 and verify its effectiveness.

When risk assessments identify new risks that require treatment, a risk treatment process in accordance with 6.1.3 shall be performed for these risks.

When risk treatment options as defined by the risk treatment plan are not effective, these treatment options shall be reviewed and revalidated following the risk treatment process according to 6.1.3 and the risk treatment plan shall be updated.

The organization shall retain documented information of the results of all AI risk treatments.

8.4 AI system impact assessment

The organization shall perform AI system impact assessments according to 6.1.4 at planned intervals or when significant changes are proposed to occur.

The organization shall retain documented information of the results of all AI system impact assessments.

9 Performance evaluation

9.1 Monitoring, measurement, analysis and evaluation

The organization shall determine:

— what needs to be monitored and measured;

— the methods for monitoring, measurement, analysis and evaluation, as applicable, to ensure valid results;

— when the monitoring and measuring shall be performed;

— when the results from monitoring and measurement shall be analysed and evaluated.

Documented information shall be available as evidence of the results.

The organization shall evaluate the performance and the effectiveness of the AI management system.

9.2 Internal audit

9.2.1 General

The organization shall conduct internal audits at planned intervals to provide information on whether the AI management system:

a) conforms to:

1) the organization's own requirements for its AI management system;

2) the requirements of this document;

b) is effectively implemented and maintained.

9.2.2 Internal audit programme

The organization shall plan, establish, implement and maintain (an) audit programme(s), including the frequency, methods, responsibilities, planning requirements and reporting.

When establishing the internal audit programme(s), the organization shall consider the importance of the processes concerned and the results of previous audits.

The organization shall:

a) define the audit objectives, criteria and scope for each audit;

b) select auditors and conduct audits to ensure objectivity and the impartiality of the audit process;

c) ensure that the results of audits are reported to relevant managers.

Documented information shall be available as evidence of the implementation of the audit programme(s) and the audit results.

9.3 Management review

9.3.1 General

Top management shall review the organization's AI management system, at planned intervals, to ensure its continuing suitability, adequacy and effectiveness.

9.3.2 Management review inputs

The management review shall include:

a) the status of actions from previous management reviews;

b) changes in external and internal issues that are relevant to the AI management system;

c) changes in needs and expectations of interested parties that are relevant to the AI management system;

d) information on the AI management system performance, including trends in:

   1) nonconformities and corrective actions;

   2) monitoring and measurement results;

   3) audit results;

e) opportunities for continual improvement.

9.3.3 Management review results

The results of the management review shall include decisions related to continual improvement opportunities and any need for changes to the AI management system.

Documented information shall be available as evidence of the results of management reviews.

10 Improvement

10.1 Continual improvement

The organization shall continually improve the suitability, adequacy and effectiveness of the AI management system.

10.2 Nonconformity and corrective action

When a nonconformity occurs, the organization shall:

a) react to the nonconformity and as applicable:
   1) take action to control and correct it;
   2) deal with the consequences;

b) evaluate the need for action to eliminate the cause(s) of the nonconformity, so that it does not recur or occur elsewhere, by:
   1) reviewing the nonconformity;
   2) determining the causes of the nonconformity;
   3) determining if similar nonconformities exist or can potentially occur;

c) implement any action needed;

d) review the effectiveness of any corrective action taken;

e) make changes to the AI management system, if necessary.

Corrective actions shall be appropriate to the effects of the nonconformities encountered.

Documented information shall be available as evidence of:

— the nature of the nonconformities and any subsequent actions taken;

— the results of any corrective action.

Annex A
(normative)

Reference control objectives and controls

A.1 General

The controls detailed in Table A.1 provide the organization with a reference for meeting organizational objectives and addressing risks related to the design and operation of AI systems. Not all the control objectives and controls listed in Table A.1 are required to be used, and the organization can design and implement their own controls (see 6.1.3).

Annex B provides implementation guidance for all the controls listed in Table A.1.

Table A.1 — Control objectives and controls

A.2 Policies related to AI		
Objective: To provide management direction and support for AI systems according to business requirements.		
	Topic	Control
A.2.2	AI policy	The organization shall document a policy for the development or use of AI systems.
A.2.3	Alignment with other organizational policies	The organization shall determine where other policies can be affected by or apply to, the organization's objectives with respect to AI systems.
A.2.4	Review of the AI policy	The AI policy shall be reviewed at planned intervals or additionally as needed to ensure its continuing suitability, adequacy and effectiveness.
A.3 Internal organization		
Objective: To establish accountability within the organization to uphold its responsible approach for the implementation, operation and management of AI systems.		
	Topic	Control
A.3.2	AI roles and responsibilities	Roles and responsibilities for AI shall be defined and allocated according to the needs of the organization.
A.3.3	Reporting of concerns	The organization shall define and put in place a process to report concerns about the organization's role with respect to an AI system throughout its life cycle.
A.4 Resources for AI systems		
Objective: To ensure that the organization accounts for the resources (including AI system components and assets) of the AI system in order to fully understand and address risks and impacts.		
	Topic	Control
A.4.2	Resource documentation	The organization shall identify and document relevant resources required for the activities at given AI system life cycle stages and other AI-related activities relevant for the organization.
A.4.3	Data resources	As part of resource identification, the organization shall document information about the data resources utilized for the AI system.
A.4.4	Tooling resources	As part of resource identification, the organization shall document information about the tooling resources utilized for the AI system.

Table A.1 *(continued)*

A.4.5	System and computing resources	As part of resource identification, the organization shall document information about the system and computing resources utilized for the AI system.
A.4.6	Human resources	As part of resource identification, the organization shall document information about the human resources and their competences utilized for the development, deployment, operation, change management, maintenance, transfer and decommissioning, as well as verification and integration of the AI system.
A.5 Assessing impacts of AI systems		
Objective: To assess AI system impacts to individuals or groups of individuals, or both, and societies affected by the AI system throughout its life cycle.		
	Topic	Control
A.5.2	AI system impact assessment process	The organization shall establish a process to assess the potential consequences for individuals or groups of individuals, or both, and societies that can result from the AI system throughout its life cycle.
A.5.3	Documentation of AI system impact assessments	The organization shall document the results of AI system impact assessments and retain results for a defined period.
A.5.4	Assessing AI system impact on individuals or groups of individuals	The organization shall assess and document the potential impacts of AI systems to individuals or groups of individuals throughout the system's life cycle.
A.5.5	Assessing societal impacts of AI systems	The organization shall assess and document the potential societal impacts of their AI systems throughout their life cycle.
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.
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.

Table A.1 *(continued)*

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.
A.7 Data for AI systems		
Objective: To ensure that the organization understands the role and impacts of data in AI systems in the application and development, provision or use of AI systems throughout their life cycles.		
	Topic	Control
A.7.2	Data for development and enhancement of AI system	The organization shall define, document and implement data management processes related to the development of AI systems.
A.7.3	Acquisition of data	The organization shall determine and document details about the acquisition and selection of the data used in AI systems.
A.7.4	Quality of data for AI systems	The organization shall define and document requirements for data quality and ensure that data used to develop and operate the AI system meet those requirements.
A.7.5	Data provenance	The organization shall define and document a process for recording the provenance of data used in its AI systems over the life cycles of the data and the AI system.
A.7.6	Data preparation	The organization shall define and document its criteria for selecting data preparations and the data preparation methods to be used.
A.8 Information for interested parties of AI systems		
Objective: To ensure that relevant interested parties have the necessary information to understand and assess the risks and their impacts (both positive and negative).		
	Topic	Control
A.8.2	System documentation and information for users	The organization shall determine and provide the necessary information to users of the AI system.
A.8.3	External reporting	The organization shall provide capabilities for interested parties to report adverse impacts of the AI system.
A.8.4	Communication of incidents	The organization shall determine and document a plan for communicating incidents to users of the AI system.
A.8.5	Information for interested parties	The organization shall determine and document their obligations to reporting information about the AI system to interested parties.
A.9 Use of AI systems		
Objective: To ensure that the organization uses AI systems responsibly and per organizational policies.		
	Topic	Control
A.9.2	Processes for responsible use of AI systems	The organization shall define and document the processes for the responsible use of AI systems.
A.9.3	Objectives for responsible use of AI system	The organization shall identify and document objectives to guide the responsible use of AI systems.

Table A.1 *(continued)*

A.9.4	Intended use of the AI system	The organization shall ensure that the AI system is used according to the intended uses of the AI system and its accompanying documentation.
A.10 Third-party and customer relationships		
Objective: To ensure that the organization understands its responsibilities and remains accountable, and risks are appropriately apportioned when third parties are involved at any stage of the AI system life cycle.		
	Topic	Control
A.10.2	Allocating responsibilities	The organization shall ensure that responsibilities within their AI system life cycle are allocated between the organization, its partners, suppliers, customers and third parties.
A.10.3	Suppliers	The organization shall establish a process to ensure that its usage of services, products or materials provided by suppliers aligns with the organization's approach to the responsible development and use of AI systems.
A.10.4	Customers	The organization shall ensure that its responsible approach to the development and use of AI systems considers their customer expectations and needs.

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.

B.2 Policies related to AI

B.2.1 Objective

To provide management direction and support for AI systems according to business requirements.

B.2.2 AI policy

Control

The organization should document a policy for the development or use of AI systems.

Implementation guidance

The AI policy should be informed by:

— business strategy;

— organizational values and culture and the amount of risk the organization is willing to pursue or retain;

— the level of risk posed by the AI systems;

— legal requirements, including contracts;

— the risk environment of the organization;

— impact to relevant interested parties (see 6.1.4).

The AI policy should include (in addition to requirements in 5.2):

— principles that guide all activities of the organization related to AI;

— processes for handling deviations and exceptions to policy.

The AI policy should consider topic-specific aspects where necessary to provide additional guidance or provide cross-references to other policies dealing with these aspects. Examples of such topics include:

— AI resources and assets;

— AI system impact assessments [see 6.1.4];

— AI system development.

Relevant policies should guide the development, purchase, operation and use of AI systems.

B.2.3 Alignment with other organizational policies

Control

The organization should determine where other policies can be affected by or apply to, the organization's objectives with respect to AI systems.

Implementation guidance

Many domains intersect with AI, including quality, security, safety and privacy. The organization should consider a thorough analysis to determine whether and where current policies can necessarily intersect and either update those policies if updates are required or include provisions in the AI policy.

Other information

The policies that the governing body sets on behalf of the organization should inform the AI policy. ISO/IEC 38507 provides guidance for members of the governing body of an organization to enable and govern the AI system throughout its life cycle.

B.2.4 Review of the AI policy

Control

The AI policy should be reviewed at planned intervals or additionally as needed to ensure its continuing suitability, adequacy and effectiveness.

Implementation guidance

A role approved by management should be responsible for the development, review and evaluation of the AI policy, or the components within. The review should include assessing opportunities for improvement of the organization's policies and approach to managing AI systems in response to changes to the organizational environment, business circumstances, legal conditions or technical environment.

The review of AI policy should take the results of management reviews into account.

B.3 Internal organization

B.3.1 Objective

To establish accountability within the organization to uphold its responsible approach for the implementation, operation and management of AI systems.

B.3.2 AI roles and responsibilities

Control

Roles and responsibilities for AI should be defined and allocated according to the needs of the organization.

Implementation guidance

Defining roles and responsibilities is critical for ensuring accountability throughout the organization for its role with respect to the AI system throughout its life cycle. The organization should consider AI policies, AI objectives and identified risks when assigning roles and responsibilities, in order to ensure that all relevant areas are covered. The organization can prioritize how the roles and responsibilities are assigned. Examples of areas that can require defined roles and responsibilities can include:

— risk management;

— AI system impact assessments;

— asset and resource management;

— security;

— safety;

— privacy;

— development;

— performance;

— human oversight;

— supplier relationships;

— demonstrate its ability to consistently fulfil legal requirements;

— data quality management (during the whole life cycle).

Responsibilities of the various roles should be defined to the level appropriate for the individuals to perform their duties.

B.3.3 Reporting of concerns

Control

The organization should define and put in place a process to report concerns about the organization's role with respect to an AI system throughout its life cycle.

Implementation guidance

The reporting mechanism should fulfil the following functions:

a) options for confidentiality or anonymity or both;

b) available and promoted to employed and contracted persons;

c) staffed with qualified persons;

d) stipulates appropriate investigation and resolution powers for the persons referred to in c);

e) provides for mechanisms to report and to escalate to management in a timely manner;

f) provides for effective protection from reprisals for both the persons concerned with reporting and investigation (e.g. by allowing reports to be made anonymously and confidentially);

g) provides reports according to 4.4 and, if appropriate, e); while maintaining confidentiality and anonymity in a), and respecting general business confidentiality considerations;

h) provides response mechanisms within an appropriate time frame.

NOTE The organization can utilize existing reporting mechanisms as part of this process.

Other information

In addition to the implementation guidance provided in this clause, the organization should further consider ISO 37002.

B.4 Resources for AI systems

B.4.1 Objective

To ensure that the organization accounts for the resources (including AI system components and assets) of the AI system in order to fully understand and address risks and impacts.

B.4.2 Resource documentation

Control

The organization should identify and document relevant resources required for the activities at given AI system life cycle stages and other AI-related activities relevant for the organization.

Implementation guidance

Documentation of resources of the AI system is critical for understanding risks, as well as potential AI system impacts (both positive and negative) to individuals or groups of individuals, or both, and societies. The documentation of such resources (which can utilize, for instance, data flow diagrams or system architecture diagrams) can inform the AI system impact assessments (see 8.5).

Resources can include, but are not limited to:

— AI system components;

— data resources, i.e. data used at any stage in the AI system life cycle;

— tooling resources (e.g. AI algorithms, models or tools);

— system and computing resources (e.g. hardware to develop and run AI models, storage for data and tooling resources);

— human resources, i.e. people with the necessary expertise (e.g. for the development, sales, training, operation and maintenance of the AI system) in relation to the organization's role throughout the AI system life cycle.

Resources can be provided by the organization itself, by its customers or by third parties.

Other information

Documentation of resources can also help to determine if resources are available and, if they are not available, the organization should revise the design specification of the AI system or its deployment requirements.

B.4.3 Data resources

Control

As part of resource identification, the organization should document information about the data resources utilized for the AI system.

Implementation guidance

Documentation on data should include, but is not limited to, the following topics:

— the provenance of the data;

— the date that the data were last updated or modified (e.g. date tag in metadata);

— for machine learning, the categories of data (e.g. training, validation, test and production data);

— categories of data (e.g. as defined in ISO/IEC 19944-1);

— process for labelling data;

— intended use of the data;

— quality of data (e.g. as described in the ISO/IEC 5259 series[2]);

— applicable data retention and disposal policies;

— known or potential bias issues in the data;

— data preparation.

B.4.4 Tooling resources

Control

As part of resource identification, the organization should document information about the tooling resources utilized for the AI system.

Implementation guidance

Tooling resources for an AI system and particularly for machine learning, can include but are not limited to:

— algorithm types and machine learning models;

— data conditioning tools or processes;

— optimization methods;

— evaluation methods;

— provisioning tools for resources;

— tools to aid model development;

— software and hardware for AI system design, development and deployment.

Other information

ISO/IEC 23053 provides detailed guidance on the types, methods and approaches for various tooling resources for machine learning.

B.4.5 System and computing resources

Control

As part of resource identification, the organization should document information about the system and computing resources utilized for the AI system.

---

2) Under preparation. Stage at the time of publication: ISO/IEC DIS 5259-1:2023, ISO/IEC DIS 5259-2:2023, ISO/IEC DIS 5259-3:2023, ISO/IEC DIS 5259-4:2023, ISO/IEC CD 5259-5:2023.

Implementation guidance

Information about system and computing resources for an AI system can include but is not limited to:

— resource requirements of the AI system (i.e. to help ensure the system can run on constrained resource devices);

— where the system and computing resources are located (e.g. on-premises, cloud computing or edge computing);

— processing resources (including network and storage);

— the impact of the hardware used to run the AI system workloads (e.g. the impact to the environment either through use or the manufacturing of the hardware or cost of using the hardware).

The organization should consider that different resources can be required to allow continual improvement of AI systems. Development, deployment and operation of the system can have different system needs and requirements.

NOTE ISO/IEC 22989 describes various system resource considerations.

B.4.6 Human resources

Control

As part of resource identification, the organization should document information about the human resources and their competences utilized for the development, deployment, operation, change management, maintenance, transfer and decommissioning, as well as verification and integration of the AI system.

Implementation guidance

The organization should consider the need for diverse expertise and include the types of roles necessary for the system. For example, the organization can include specific demographic groups related to data sets used to train machine learning models, if their inclusion is a necessary component of the system design. Necessary human resources can include but are not limited to:

— data scientists;

— roles related to human oversight of AI systems;

— experts on trustworthiness topics such as safety, security and privacy;

— AI researchers and specialists, and domain experts relevant to the AI systems.

Different resources can be necessary at different stages of the AI system life cycle.

B.5 Assessing impacts of AI systems

B.5.1 Objective

To assess AI system impacts to individuals or groups of individuals, or both, and societies affected by the AI system throughout its life cycle.

B.5.2 AI system impact assessment process

Control

The organization should establish a process to assess the potential consequences for individuals or groups of individuals, or both, and societies that can result from the AI system throughout its life cycle.

Implementation guidance

Because AI systems potentially generate significant impact to individuals, groups of individuals , or both, and societies, the organization that provides and uses such systems should, based on the intended purpose and use of these systems, assess the potential impacts of these systems on these groups.

The organization should consider whether an AI system affects:

— the legal position or life opportunities of individuals;

— the physical or psychological well-being of individuals;

— universal human rights;

— societies.

The organization's procedures should include, but are not limited to:

a) circumstances under which an AI system impact assessment should be performed, which can include, but are not limited to:

   1) criticality of the intended purpose and context in which the AI system is used or any significant changes to these;

   2) complexity of AI technology and the level of automation of AI systems or any significant changes to that;

   3) sensitivity of data types and sources processed by the AI system or any significant changes to that;

b) elements that are part of the AI system impact assessment process, which can include:

   1) identification (e.g. sources, events and outcomes);

   2) analysis (e.g. consequences and likelihood);

   3) evaluation (e.g. acceptance decisions and prioritization);

   4) treatment (e.g. mitigation measures);

   5) documentation, reporting and communication (see 7.4, 7.5 and B.3.3);

c) who performs the AI system impact assessment;

d) how the AI system impact assessment can be utilized [e.g. how it can inform the design or use of the system (see B.6 and B.9), whether it can trigger reviews and approvals];

e) individuals and societies that are potentially impacted based on the system's intended purpose, use and characteristics (e.g. assessment for individuals, groups of individuals or societies).

Impact assessment should take various aspects of the AI system into account, including the data used for the development of the AI system, the AI technologies used and the functionality of the overall system.

The processes can vary based on the role of the organization and the domain of AI application and depending on the specific disciplines for which the impact is assessed (e.g. security, privacy and safety).

Other information

For some disciplines or organizations, detailed consideration of the impact on individuals or groups of individuals, or both, and societies is part of risk management, particularly in disciplines such as information security, safety and environmental management. The organization should determine

if discipline-specific impact assessments performed as part of such a risk management process sufficiently integrate AI considerations for those specific aspects (e.g. privacy).

NOTE ISO/IEC 23894 describes how an organization can perform impact analyses for the organization itself, along with individuals or groups of individuals, or both, and societies, as part of an overall risk management process.

B.5.3 Documentation of AI system impact assessments

Control

The organization should document the results of AI system impact assessments and retain results for a defined period.

Implementation guidance

The documentation can be helpful in determining information that should be communicated to users and other relevant interested parties.

AI system impact assessments should be retained and updated, as needed, in alignment with the elements of an AI system impact assessment documented in B.5.2. Retention periods can follow organization retention schedules or be informed by legal requirements or other requirements.

Items that the organization should consider documenting can include, but are not limited to:

— the intended use of the AI system and any reasonable foreseeable misuse of the AI system;

— positive and negative impacts of the AI system to the relevant individuals or groups of individuals, or both, and societies;

— predictable failures, their potential impacts and measures taken to mitigate them;

— relevant demographic groups the system is applicable to;

— complexity of the system;

— the role of humans in relationships with system, including human oversight capabilities, processes and tools, available to avoid negative impacts;

— employment and staff skilling.

B.5.4 Assessing AI system impact on individuals or groups of individuals

Control

The organization should assess and document the potential impacts of AI systems to individuals or groups of individuals throughout the system's life cycle.

Implementation guidance

When assessing the impacts on individuals or groups of individuals, or both, and societies, the organization should consider its governance principles, AI policies and objectives. Individuals using the AI system or whose PII are processed by the AI system, can have expectations related to the trustworthiness of the AI system. Specific protection needs of groups such as children, impaired persons, elderly persons and workers should be taken into account. The organization should evaluate these expectations and consider the means to address them as part of the system impact assessment.

Depending on the scope of AI system purpose and use, areas of impact to consider as part of the assessment can include, but are not limited to:

— fairness;

— accountability;

— transparency and explainability;

— security and privacy;

— safety and health;

— financial consequences;

— accessibility;

— human rights.

Other information

Where necessary, the organization should consult experts (e.g. researchers, subject matter experts and users) to obtain a full understanding of potential impacts of the AI system on individuals or groups of individuals, or both, and societies.

B.5.5 Assessing societal impacts of AI systems

Control

The organization should assess and document the potential societal impacts of their AI systems throughout their life cycle.

Implementation guidance

Societal impacts can vary widely depending on the organization's context and the types of AI systems. The societal impacts of AI systems can be both beneficial and detrimental. Examples of these potential societal impacts can include:

— environment sustainability (including the impacts on natural resources and greenhouse gas emissions);

— economic (including access to financial services, employment opportunities, taxes, trade and commerce);

— government (including legislative processes, misinformation for political gain, national security and criminal justice systems);

— health and safety (including access to healthcare, medical diagnosis and treatment, and potential physical and psychological harms);

— norms, traditions, culture and values (including misinformation that leads to biases or harms to individuals or groups of individuals, or both, and societies).

Other information

Development and use of AI systems can be computationally intensive with related impacts to environmental sustainability (e.g. greenhouse gas emissions due to increased power usage, impacts on water, land, flora and fauna). Likewise, AI systems can be used to improve the environmental sustainability of other systems (e.g. reduce greenhouse gas emissions related to buildings and transportation). The organization should consider the impacts of its AI systems in the context of its overall environmental sustainability goals and strategies.

The organization should consider how its AI systems can be misused to create societal harms and how they can be used to address historical harms. For example, can AI systems prevent access to financial services such as loans, grants, insurance and investments and likewise can AI systems improve access to these instruments?

AI systems have been used to influence the outcomes of elections and to create misinformation (e.g. deepfakes in digital media) that can lead to political and social unrest. Government's use of AI systems for criminal-justice purposes has exposed the risk of biases to societies, individuals or groups of

individuals. The organization should analyse how actors can misuse AI systems and how the AI systems can reinforce unwanted historical social biases.

AI systems can be used to diagnose and treat illnesses and to determine qualifications for health benefits. AI systems are also deployed in scenarios where malfunctions can result in death or injury to humans (e.g. self-driving automobiles, human-machine teaming). The organization should consider both the positive and negative outcomes when using AI systems, such as in health and safety related scenarios.

NOTE ISO/IEC TR 24368 provides a high-level overview of ethical and societal concerns related to AI systems and applications.

B.6 AI system life cycle

B.6.1 Management guidance for AI system development

B.6.1.1 Objective

To ensure that the organization identifies and documents objectives and implements processes for the responsible design and development of AI systems.

B.6.1.2 Objectives for responsible development of AI system

Control

The organization should identify and document objectives to guide the responsible development of AI systems, and take those objectives into account and integrate measures to achieve them in the development life cycle.

Implementation guidance

The organization should identify objectives (see 6.2) that affect the AI system design and development processes. These objectives should be taken into account in the design and development processes. For example, if an organization defines "fairness" as one objective, this should be incorporated in the requirements specification, data acquisition, data conditioning, model training, verification and validation, etc. The organization should provide requirements and guidelines as necessary to ensure that measures are integrated into the various stages (e.g. the requirement to use a specific testing tool or method to address unfairness or unwanted bias) to achieve such objectives.

Other information

AI techniques are being used to augment security measures such as threat prediction detection and prevention of security attacks. This is an application of AI techniques that can be used to reinforce security measures to protect both AI systems and conventional non-AI based software systems. Annex C provides examples of organizational objectives for managing risk, which can be useful in determining the objectives for AI system development.

B.6.1.3 Processes for responsible design and development of AI systems

Control

The organization should define and document the specific processes for the responsible design and development of the AI system.

Implementation guidance

Responsible development for AI system processes should include consideration of, without limitation, the following:

— life cycle stages (a generic AI system life cycle model is provided by ISO/IEC 22989, but the organization can specify their own life cycle stages);

— testing requirements and planned means for testing;

— human oversight requirements, including processes and tools, especially when the AI system can impact natural persons;

— at what stages AI system impact assessments should be performed;

— training data expectations and rules (e.g. what data can be used, approved data suppliers and labelling);

— expertise (subject matter domain or other) required or training for developers of AI systems or both;

— release criteria;

— approvals and sign-offs necessary at various stages;

— change control;

— usability and controllability;

— engagement of interested parties.

The specific design and development processes depend on the functionality and the AI technologies that are intended to be used for the AI system.

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.

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.

B.6.2.3 Documentation of AI system design and development

Control

The organization should document the AI system design and development based on organizational objectives, documented requirements and specification criteria.

Implementation guidance

There are many design choices necessary for an AI system, including, but not limited to:

— machine learning approach (e.g. supervised vs. unsupervised);

— learning algorithm and type of machine learning model utilized;

— how the model is intended to be trained and which data quality (see B.7);

— evaluation and refinement of models;

— hardware and software components;

— security threats considered throughout the AI system life cycle; security threats specific to AI systems include data poisoning, model stealing or model inversion attacks;

— interface and presentation of outputs;

— how humans can interact with the system;

— interoperability and portability considerations.

There can be multiple iterations between design and development, but documentation on the stage should be maintained and a final system architecture documentation should be available.

Other information

For more information about human-centred design for interactive systems, see ISO 9241-210.

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.

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;
- 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.

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.

B.6.2.5 AI system deployment

Control

The organization should document a deployment plan and ensure that appropriate requirements are met prior to deployment.

Implementation guidance

AI systems can be developed in various environments and deployed in others (such as developed on premises and deployed using cloud computing) and the organization should take these differences into account for the deployment plan. The organization should also consider whether components are deployed separately (e.g. software and model can be deployed independently). Additionally, the organization should have a set of requirements to be met prior to release and deployment (sometimes referred to as "release criteria"). This can include verification and validation measures that are to be passed, performance metrics that are to be met, user testing to be completed, as well as management approvals and sign-offs to be obtained. The deployment plan should take into account the perspectives of and impacts to relevant interested parties.

B.6.2.6 AI system operation and monitoring

Control

The organization should 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.

Implementation guidance

Each minimum activity for operation and monitoring can take account of various considerations. For example:

— System and performance monitoring can include monitoring for general errors and failures, as well as for whether the system is performing as expected with production data. Technical performance criteria can include success rates in resolving problems or in achieving tasks, or confidence rates. Other criteria can be related to meeting commitment or expectation and needs of interested parties, including, for example, ongoing monitoring to ensure compliance with customer requirements or applicable legal requirements.

— Some deployed AI systems evolve their performance as a result of ML, where production data and output data are used to further train the ML model. Where continuous learning is used, the organization should monitor the performance of the AI system to ensure that it continues to meet its design goals and operates on production data as intended.

— The performance of some AI systems can change even if such systems do not use continuous learning, usually due to concept or data drift in production data. In such cases, monitoring can identify the need for retraining to ensure that the AI system continues to meet its design goals and operates on production data as intended. More information can be found in ISO/IEC 23053.

— Repairs can include responses to errors and failures in the system. The organization should have processes in place for the response and repair of these issues. Additionally, updates can be necessary as the system evolves or as critical issues are identified, or as the result of externally identified issues (e.g. non-compliance with customer expectations or legal requirement). There should be processes in place for updating the system including components affected, update schedule, information to users on what is included in the update.

— System updates can also include changes in the system operations, new or modified intended uses, or other changes in system functionality. The organization should have procedures in place to address operational changes, including communication to users.

— Support for the system can be internal, external or both, depending on the needs of the organization and how the system was acquired. Support processes should consider how users can contact the appropriate help, how issues and incidents are reported, support service level agreements and metrics.

— Where AI systems are being used for purposes other than those for which they were designed or in ways that were not anticipated, the appropriateness of such uses should be considered.

— AI-specific information security threats related to the AI systems applied and developed by the organization should be identified. AI-specific information security threats include, but are not limited to data poisoning, model stealing and model inversion attacks.

Other information

The organization should consider operational performance that can affect interested parties and consider this when designing and determining performance criteria.

Performance criteria for AI systems in operation should be determined by the task under consideration, such as classification, regression, ranking, clustering or dimensionality reduction.

Performance criteria can include statistical aspects such as error rates and processing duration. For each criterion, the organization should identify all relevant metrics as well as interdependences between metrics. For each metric, the organization should consider acceptable values based on, for example, domain expert's recommendations and analysis of expectations of interested parties relative to existing non-AI practices.

For example, an organization can determine that the F_1 score is an appropriate performance metric based on its assessment of the impact of false positives and false negatives, as described in

ISO/IEC TS 4213. The organization can then establish an F_1 value that the AI system is expected to meet. It should be evaluated if these issues can be handled by existing measures. If that is not the case, changes to existing measures should be considered or additional measures should be defined to detect and handle these issues.

The organization should consider the performance of non-AI systems or processes in operation and use them as potentially relevant context when establishing performance criteria.

The organization should additionally ensure that the means and processes used to evaluate the AI system, including, where applicable, the selection and management of evaluation data, improve the completeness and the reliability in assessment of its performance with respect to the defined criteria.

Development of performance assessment methodologies can be based on criteria, metrics and values. These should inform the amount of data and the types of processes used in the assessment and the roles and expertise of personnel that carries out the assessment.

Performance assessment methodologies should reflect attributes and characteristics of operation and use as closely as possible to ensure that assessment results are useful and relevant. Some aspects of performance assessment can require controlled introduction of erroneous or spurious data or processes to assess impact on performance.

The quality model in ISO/IEC 25059 can be used to define performance criteria.

B.6.2.7 AI system technical documentation

Control

The organization should 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.

Implementation guidance

The AI system technical documentation can include, but is not limited to the following elements:

— a general description of the AI system including its intended purpose;

— usage instructions;

— technical assumptions about its deployment and operation (run-time environment, related software and hardware capabilities, assumptions made on data, etc.);

— technical limitations (e.g. acceptable error rates, accuracy, reliability, robustness);

— monitoring capabilities and functions that allow users or operators to influence the system operation.

Documentation elements related to all AI system life cycle stages (as defined in ISO/IEC 22989) can include, but are not limited to:

— design and system architecture specification;

— design choices made and quality measures taken during the system development process;

— information about the data used during system development;

— assumptions made and quality measures taken on data quality (e.g. assumed statistical distributions);

— management activities (e.g. risk management) taken during development or operation of the AI system;

— verification and validation records;

— changes made to the AI system when it is in operation;

— impact assessment documentation as described in B.5.

The organization should document technical information related to the responsible operation of the AI system. This can include, but is not limited to:

— documenting a plan for managing failures. This can include for example, the need to describe a rollback plan for the AI system, turning off features of the AI system, an update process or a plan for notifying customers, users, etc. of changes to the AI system, updated information on system failures and how these can be mitigated;

— documenting processes for monitoring the health of the AI system (i.e. the AI system operates as intended and within its normal operating margins, also referred to as observability) and processes for addressing AI system failures;

— documenting standard operating procedures for the AI system, including which events should be monitored and how event logs are prioritized and reviewed. It can also include how to investigate failures and the prevention of failures;

— documenting the roles of personnel responsible for operation of the AI system as well as those responsible for accountability of the system use, especially in relation to handling the effects of AI system failures or managing updates to the AI system;

— documenting system updates like changes in the system operations, new or modified intended uses, or other changes in system functionality.

The organization should have procedures in place to address operational changes including communication to users and internal evaluations on the type of change.

Documentation should be up to date and accurate. Documentation should be approved by the relevant management within the organization.

When provided as part of the user documentation, the controls provided in Table A.1 should be taken into account.

B.6.2.8 AI system recording of event logs

Control

The organization should 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.

Implementation guidance

The organization should ensure logging for AI systems it deploys to automatically collect and record event logs related to certain events that occur during operation. Such logging can include but is not limited to:

— traceability of the AI system's functionality to ensure that the AI system is operating as intended;

— detection of the AI system's performance outside of the AI system's intended operating conditions that can result in undesirable performance on production data or impacts to relevant interested parties through monitoring of the operation of the AI system.

AI system event logs can include information, such as the time and date each time the AI system is used, the production data on which the AI system operates on, the outputs that fall out of the range of the intended operation of the AI system, etc.

Event logs should be kept for as long as required for the intended use of the AI system and within the data retention policies of the organization. Legal requirements related to data retention can apply.

Other information

Some AI systems, such as biometric identification systems, can have additional logging requirements depending on jurisdiction. Organizations should be aware of these requirements.

B.7 Data for AI systems

B.7.1 Objective

To ensure that the organization understands the role and impacts of data in AI systems in the application and development, provision or use of AI systems throughout their life cycles.

B.7.2 Data for development and enhancement of AI system

Control

The organization should define, document and implement data management processes related to the development of AI systems.

Implementation guidance

Data management can include various topics such as, but not limited to:

— privacy and security implications due to the use of data, some of which can be sensitive in nature;

— security and safety threats that can arise from data dependent AI system development;

— transparency and explainability aspects including data provenance and the ability to provide an explanation of how data are used for determining an AI system's output if the system requires transparency and explainability;

— representativeness of training data compared to operational domain of use;

— accuracy and integrity of the data.

NOTE Detailed information of AI system life cycle and data management concepts is provided by ISO/IEC 22989.

B.7.3 Acquisition of data

Control

The organization should determine and document details about the acquisition and selection of the data used in AI systems.

Implementation guidance

The organization can need different categories of data from different sources depending on the scope and use of their AI systems. Details for data acquisition can include:

— categories of data needed for the AI system;

— quantity of data needed;

— data sources (e.g. internal, purchased, shared, open data, synthetic);

— characteristics of the data source (e.g. static, streamed, gathered, machine generated);

— data subject demographics and characteristics (e.g. known or potential biases or other systematic errors);

— prior handling of the data (e.g. previous uses, conformity with privacy and security requirements);

— data rights (e.g. PII, copyright);

— associated meta data (e.g. details of data labelling and enhancing);

— provenance of the data.

Other information

The data categories and a structure for the data use in ISO/IEC 19944-1 can be used to document details about data acquisition and use.

B.7.4 Quality of data for AI systems

Control

The organization should define and document requirements for data quality and ensure that data used to develop and operate the AI system meet those requirements.

Implementation guidance

The quality of data used to develop and operate AI systems potentially has significant impacts on the validity of the system's outputs. ISO/IEC 25024 defines data quality as the degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions. For AI systems that use supervised or semi-supervised machine learning, it is important that the quality of training, validation, test and production data are defined, measured and improved to the extent possible, and the organization should ensure that the data are suitable for its intended purpose. The organization should consider the impact of bias on system performance and system fairness and make such adjustments as necessary to the model and data used to improve performance and fairness so they are acceptable for the use case.

Other information

Additional information regarding data quality is available in the ISO/IEC 5259 series[2] on data quality for analytics and ML. Additional information regarding different forms of bias in data used in AI systems is available in ISO/IEC TR 24027.

B.7.5 Data provenance

Control

The organization should define and document a process for recording the provenance of data used in its AI systems over the life cycles of the data and the AI system.

Implementation guidance

According to ISO 8000-2, a record of data provenance can include information about the creation, update, transcription, abstraction, validation and transferring of the control of data. Additionally, data sharing (without transfer of control) and data transformations can be considered under data provenance. Depending on factors such as the source of the data, its content and the context of its use, organizations should consider whether measures to verify the provenance of the data are needed.

B.7.6 Data preparation

Control

The organization shall define and document its criteria for selecting data preparations and the data preparation methods to be used.

Implementation guidance

Data used in an AI system ordinarily needs preparation to make it usable for a given AI task. For example, machine learning algorithms are sometimes intolerant of missing or incorrect entries, non-

normal distribution and widely varying scales. Preparation methods and transforms can be used to increase the quality of the data. Failure to properly prepare the data can potentially lead to AI system errors. Common preparation methods and transformations for data used in AI systems include:

— statistical exploration of the data (e.g. distribution, mean, median, standard deviation, range, stratification, sampling) and statistical metadata (e.g. data documentation initiative (DDI) specification[28]);

— cleaning (i.e. correcting entries, dealing with missing entries);

— imputation (i.e. methods for filling in missing entries);

— normalization;

— scaling;

— labelling of the target variables;

— encoding (e.g. converting categorical variables to numbers).

For a given AI task, the organization should document its criteria for selecting specific data preparation methods and transforms as well as the specific methods and transforms used in the AI task.

NOTE For additional information on data preparation specific to machine learning see the ISO/IEC 5259 series[2] and ISO/IEC 23053.

B.8 Information for interested parties

B.8.1 Objective

To ensure that relevant interested parties have the necessary information to understand and assess the risks and their impacts (both positive and negative).

B.8.2 System documentation and information for users

Control

The organization should determine and provide the necessary information to users of the system.

Implementation guidance

Information about the AI system can include both technical details and instructions, as well as general notifications to users that they are interacting with an AI system, depending on the context. This can also include the system itself, as well as potential outputs of the system (e.g. notifying users that an image is created by AI).

Although AI systems can be complex, it is critical that users are able to understand when they are interacting with an AI system, how the system works. Users also need to understand its intended purpose and intended uses, its potential to cause harm or benefit the user. Some system documentation can necessarily be targeted for more technical uses (e.g. system administrators), and the organization should understand the needs of different interested parties and what understandability can mean to them. The information should also be accessible, both in terms of ease of use in finding it, as well as for users who can need additional accessibility features.

Information that can be provided to users include, but are not limited to:

— purpose of the system;

— that the user is interacting with an AI system;

— how to interact with the system;

— how and when to override the system;

— technical requirements for system operation, including the computational resources needed, and limitations of the system as well as its expected lifetime;

— needs for human oversight;

— information about accuracy and performance;

— relevant information from the impact assessment, including potential benefits and harms, particularly if they are applicable in specific contexts or certain demographic groups (see B.5.2 and B.5.4);

— revisions to claims about the system's benefits;

— updates and changes in how the system works, as well as any necessary maintenance measures, including their frequency;

— contact information;

— educational materials for system use.

Criteria used by the organization to determine whether and what information is to be provided should be documented. Relevant criteria include but are not limited to the intended use and reasonably foreseeable misuse of the AI system, the expertise of the user and specific impact of the AI system.

Information can be provided to users in numerous ways, including documented instructions for use, alerts and other notifications built into the system itself, information on a web page, etc. Depending on which methods the organization uses to provide information, it should validate that the users have access to this information, and that the information provided is complete, up to date and accurate.

B.8.3 External reporting

Control

The organization should provide capabilities for interested parties to report adverse impacts of the system.

Implementation guidance

While the system operation should be monitored for reported issues and failures, the organization should also provide capabilities for users or other external parties to report adverse impacts (e.g. unfairness).

B.8.4 Communication of incidents

Control

The organization should determine and document a plan for communicating incidents to users of the system.

Implementation guidance

Incidents related to the AI system can be specific to the AI system itself, or related to information security or privacy (e.g. a data breach). The organization should understand its obligations around notifying users and other interested party about incidents, depending on the context in which the system operates. For example, an incident with an AI component that is part of a product that affects safety can have different notification requirements than other types of systems. Legal requirements (such as contracts) and regulatory activity can apply, which can specify requirements for:

— types of incidents that must be communicated;

— the timeline for notification;

— whether and which authorities must be notified;

— the details required to be communicated.

The organization can integrate incident response and reporting activities for AI into their broader organizational incident management activities, but should be aware of unique requirements related to AI systems, or individual components of AI systems (e.g. a PII data breach in training data for the system can have different reporting requirements related to privacy).

Other information

ISO/IEC 27001 and ISO/IEC 27701 provide additional details on incident management for security and privacy respectively.

B.8.5 Information for interested parties

Control

The organization should determine and document its obligations to reporting information about the AI system to interested parties.

Implementation guidance

In some cases, a jurisdiction can require information about the system to be shared with authorities such as regulators. Information can be reported to interested parties such as customers or regulatory authorities within the appropriate timeframe. The information shared can include, for example:

— technical system documentation, including, but not limited, to data sets for training, validation and testing as well as algorithmic choices justifications and verification and validation records;

— risks related to the system;

— results of impact assessments;

— logs and other system records.

The organization should understand their obligations in this respect and ensure that the appropriate information is shared with the correct authorities. Additionally, it is presupposed that the organization is aware of jurisdictional requirements related to information shared with law enforcement authorities.

B.9 Use of AI systems

B.9.1 Objective

To ensure that the organization uses AI systems responsibly and per organizational policies.

B.9.2 Processes for responsible use of AI systems

Control

The organization should define and document the processes for the responsible use of AI systems.

Implementation guidance

Depending on its context, the organization can have many considerations for determining whether to use a particular AI system. Whether the AI system is developed by the organization itself or sourced from a third party, the organization should be clear on what these considerations are and develop policies to address them. Some examples are:

— required approvals;

— cost (including for ongoing monitoring and maintenance);

— approved sourcing requirements;

— legal requirements applicable to the organization.

Where the organization has accepted policies for the use of other systems, assets, etc., these policies can be incorporated if desired.

B.9.3 Objectives for responsible use of AI system

Control

The organization should identify and document objectives to guide the responsible use of AI systems.

Implementation guidance

The organization operating in different contexts can have different expectations and objectives for what constitutes the responsible development of AI systems. Depending on its context, the organization should identify its objectives related to responsible use. Some objectives include:

— fairness;

— accountability;

— transparency;

— explainability;

— reliability;

— safety;

— robustness and redundancy;

— privacy and security;

— accessibility.

Once defined, the organization should implement mechanisms to achieve its objectives within the organization. This can include determining if a third-party solution fulfils the organization's objectives or if an internally developed solution is applicable for the intended use. The organization should determine at which stages of the AI system life cycle meaningful human oversight objectives should be incorporated. This can include:

— involving human reviewers to check the outputs of the AI system, including having authority to override decisions made by the AI system;

— ensuring that human oversight is included if required for acceptable use of the AI system according to instructions or other documentation associated with the intended deployment of the AI system;

— monitoring the performance of the AI system, including the accuracy of the AI system outputs;

— reporting concerns related to the outputs of the AI system and their impact to relevant interested parties;

— reporting concerns with changes in the performance or ability of the AI system to make correct outputs on the production data;

— considering whether automated decision-making is appropriate for a responsible approach to the use of an AI system and the intended use of the AI system.

The need for human oversight can be informed by the AI system impact assessments (see B.5). The personnel involved in human oversight activities related to the AI system should be informed of, trained

and understand the instructions and other documentation to the AI system and the duties they carry out to satisfy human oversight objectives. When reporting performance issues, human oversight can augment automated monitoring.

Other information

Annex C provides examples of organizational objectives for managing risk, which can be useful in determining the objectives for AI system use.

B.9.4 Intended use of the AI system

Control

The organization should ensure that the AI system is used according to the intended uses of the AI system and its accompanying documentation.

Implementation guidance

The AI system should be deployed according to the instructions and other documentation associated with the AI system (see B.8.2). The deployment can require specific resources to support the deployment, including the need to ensure that human oversight is applied as required (see B.9.3). It can be necessary that for acceptable use of the AI system, the data that the AI system is used on aligns with the documentation associated with the AI system to ensure that the AI system performance is accurate.

The operation of the AI system should be monitored (see B.6.2.6). Where the correct deployment of the AI system according to its associated instructions causes concern regarding the impact to relevant interested parties or the organization's legal requirements, the organization should communicate its concerns to the relevant personnel inside the organization as well as to any third-party suppliers of the AI system.

The organization should keep event logs or other documentation related to the deployment and operation of the AI system which can be used to demonstrate that the AI system is being used as intended or to help with communicating concerns related to the intended use of the AI system. The time period during which event logs and other documentation are kept depends on the intended use of the AI system, the organization's data retention policies and relevant legal requirements for data retention.

B.10 Third-party and customer relationships

B.10.1 Objective

To ensure that the organization understands its responsibilities and remains accountable, and risks are appropriately apportioned when third parties are involved at any stage of the AI system life cycle.

B.10.2 Allocating responsibilities

Control

The organization should ensure that responsibilities within their AI system life cycle are allocated between the organization, its partners, suppliers, customers and third parties.

Implementation guidance

In an AI system life cycle, responsibilities can be split between parties providing data, parties providing algorithms and models, parties developing or using the AI system and being accountable with regard to some or all interested parties. The organization should document all parties intervening in the AI system life cycle and their roles and determine their responsibilities.

Where the organization supplies an AI system to a third party, the organization should ensure that it takes a responsible approach to developing the AI system. See the controls and guidance in B.6. The organization should be able to provide the necessary documentation (see B.6.2.7 and B.8.2) for the AI

system to relevant interested parties and to the third party that the organization is supplying the AI system to.

When processed data includes PII, responsibilities are usually split between PII processors and controllers. ISO/IEC 29100 provides further information on PII controllers and PII processors. Where the privacy of PII is to be preserved, controls such as those described in ISO/IEC 27701 should be considered. Based on the organization's and AI system's data processing activities on PII and the organization's role in application and development of the AI system through their life cycle, the organization can take on the role of a PII controller (or joint PII controller), PII processor or both.

B.10.3 Suppliers

Control

The organization should establish a process to ensure that its usage of services, products or materials provided by suppliers aligns with the organization's approach to the responsible development and use of AI systems.

Implementation guidance

Organizations developing or using an AI system can utilize suppliers in a number of ways, from sourcing datasets, machine learning algorithms or models, or other components of a system such as software libraries, to an entire AI system itself for use on its own or as part of another product (e.g. a vehicle).

Organizations should consider different types of suppliers, what they supply, and the varying level of risk this can pose to the system and organization as a whole in determining the selection of suppliers, the requirements placed on those suppliers, and the levels of ongoing monitoring and evaluation needed for the suppliers.

Organizations should document how the AI system and AI system components are integrated into AI systems developed or used by the organization.

Where the organization considers that the AI system or AI system components from a supplier do not perform as intended or can result in impacts to individuals or groups of individuals, or both, and societies that are not aligned with the responsible approach to AI systems taken by the organization, the organization should require the supplier to take corrective actions. The organization can decide to work with the supplier to achieve this objective.

The organization should ensure that the supplier of an AI system delivers appropriate and adequate documentation related to the AI system (see B.6.2.7 and B.8.2).

B.10.4 Customers

Control

The organization should ensure that its responsible approach to the development and use of AI systems considers their customer expectations and needs.

Implementation guidance

The organization should understand customer expectations and needs when it is supplying a product or service related to an AI system (i.e. when it is itself a supplier). These can come in the form of requirements for the product or service itself during a design or engineering phase, or in the form of contractual requirements or general usage agreements. One organization can have many different types of customer relationships, and these can all have different needs and expectations.

The organization should particularly understand the complex nature of supplier and customer relationships and understand where responsibility lies with the provider of the AI system and where it lies with the customer, while still meeting needs and expectations.

For example, the organization can identify risks related to the use of its AI products and services by the customer and can decide to treat the identified risks by giving appropriate information to its customer, so that the customer can then treat the corresponding risks.

As an example of appropriate information, when an AI system is valid for a certain domain of use, the limits of the domain should be communicated to the customer. See B.6.2.7 and B.8.2.

Annex C
(informative)

Potential AI-related organizational objectives and risk sources

C.1 General

This annex outlines potential organizational objectives, risk sources and descriptions that can be considered by the organization when managing risks. This annex is not intended to be exhaustive or applicable for every organization. The organization should determine the objectives and risk sources that are relevant. ISO/IEC 23894 provides more detailed information on these objectives and risk sources, and their relationship to risk management. Evaluation of AI systems, initially, regularly and when warranted, provides evidence that an AI system is being assessed against organizational objectives.

C.2 Objectives

C.2.1 Accountability

The use of AI can change existing accountability frameworks. Where previously persons would be held accountable for their actions, their actions can now be supported by or based on the use of an AI system.

C.2.2 AI expertise

A selection of dedicated specialists with interdisciplinary skill sets and expertise in assessing, developing and deploying AI systems is needed.

C.2.3 Availability and quality of training and test data

AI systems based on ML need training, validation and test data in order to train and verify the systems for the intended behaviour.

C.2.4 Environmental impact

The use of AI can have positive and negative impacts on the environment.

C.2.5 Fairness

The inappropriate application of AI systems for automated decision-making can be unfair to specific persons or groups of persons.

C.2.6 Maintainability

Maintainability is related to the ability of the organization to handle modifications of the AI system in order to correct defects or adjust to new requirements.

C.2.7 Privacy

The misuse or disclosure of personal and sensitive data (e.g. health records) can have harmful effects on data subjects.

C.2.8 Robustness

In AI, robustness properties demonstrate the ability (or inability) of the system to have comparable performance on new data as on the data on which it was trained or the data of typical operations.

C.2.9 Safety

Safety relates to the expectation that a system does not, under defined conditions, lead to a state in which human life, health, property or the environment is endangered.

C.2.10 Security

In the context of AI and in particular with regard to AI systems based on ML approaches, new security issues should be considered beyond classical information and system security concerns.

C.2.11 Transparency and explainability

Transparency relates both to characteristics of an organization operating AI systems and to those systems themselves. Explainability relates to explanations of important factors influencing the AI system results that are provided to interested parties in a way understandable to humans.

C.3 Risk sources

C.3.1 Complexity of environment

When AI systems operate in complex environments, where the range of situation is broad, there can be uncertainty on the performance and therefore a source of risk (e.g. complex environment of autonomous driving).

C.3.2 Lack of transparency and explainability

The inability to provide appropriate information to interested parties can be a source of risk (i.e. in terms of trustworthiness and accountability of the organization).

C.3.3 Level of automation

The level of automation can have an impact on various areas of concerns, such as safety, fairness or security.

C.3.4 Risk sources related to machine learning

The quality of data used for ML and the process used to collect data can be sources of risk, as they can impact objectives such as safety and robustness (e.g. due to issues in data quality or data poisoning).

C.3.5 System hardware issues

Risk sources related to hardware include hardware errors based on defective components or transferring trained ML models between different systems.

C.3.6 System life cycle issues

Sources of risk can appear over the entire AI system life cycle (e.g. flaws in design, inadequate deployment, lack of maintenance, issues with decommissioning).

C.3.7 Technology readiness

Risk sources can be related to less mature technology due to unknown factors (e.g. system limitations and boundary conditions, performance drift), but also due to the more mature technology due to technology complacency.

Annex D
(informative)

Use of the AI management system across domains or sectors

D.1 General

This management system is applicable to any organization developing, providing or using products or services that utilize an AI system. Therefore, it is applicable potentially to a great variety of products and services, in different sectors, which are subject to obligations, good practices, expectations or contractual commitment towards interested parties. Examples of sectors are:

— health;

— defence;

— transport;

— finance;

— employment;

— energy.

Various organizational objectives (see Annex C for possible objectives) can be considered for the responsible development and use of an AI system. This document provides requirements and guidance from an AI technology specific view. For several of the potential objectives, generic or sector-specific management system standards exist. These management system standards consider the objective usually from a technology neutral point of view, while the AI management system provides AI technology specific considerations.

AI systems consist not only of components using AI technology, but can use a variety of technologies and components. Responsible development and use of an AI system therefore requires taking into account not only AI-specific considerations, but also the system as a whole with all the technologies and components that are used. Even for the AI technology specific part, other aspects besides AI-specific considerations should be taken into account. For example, as AI is an information processing technology, information security applies generally to it. Objectives such as safety, security, privacy and environmental impact should be managed holistically and not separately for AI and the other components of the system. Integration of the AI management system with generic or sector-specific management system standards for relevant topics is therefore essential for responsible development and use of an AI system.

D.2 Integration of AI management system with other management system standards

When providing or using AI systems, the organization can have objectives or obligations related to aspects which are topics of other management system standards. These can include, for example, security, privacy, quality, respectively topics covered in ISO/IEC 27001, ISO/IEC 27701 and ISO 9001.

When providing, using or developing AI systems, potential relevant generic management system standards, but not limited to that, are:

— ISO/IEC 27001: In most contexts, security is key to achieving the objectives of the organization with the AI system. The way an organization pursues security objectives depends on its context and its own policies. If an organization identifies the need to implement an AI management system

and to address security objectives in a similar thorough and systematic way, it can implement an information security management system in conformity with ISO/IEC 27001. Given that both ISO/IEC 27001 and the AI management systems use the high-level structure, their integrated use is facilitated and of great benefit for the organization. In this case, the way to implement controls which (partly) relate to information security in this document (see B.6.1.2) can be integrated with the organization's implementation of ISO/IEC 27001.

— ISO/IEC 27701: In many context and application domains, PIIs are processed by AI systems. The organization can then comply with the applicable obligations for privacy and with its own policies and objectives. Similarly, as for ISO/IEC 27001, the organization can benefit from the integration of ISO/IEC 27701 with the AI management system. Privacy-related objectives and controls of the AI management system (see B.2.3 and B.5.4) can be integrated with the organization's implementation of ISO/IEC 27701.

— ISO 9001: For many organizations, conformity to ISO 9001 is a key sign that they are customer-oriented and genuinely concerned about internal effectiveness. Independent conformity assessment to ISO 9001 facilitates business across organizations and inspires customer confidence in products or services. The level of customer's confidence in an organization or AI system can be highly reinforced when an AI management system is implemented jointly with ISO 9001 when AI technologies are involved. The AI management system can be complementary to the ISO 9001 requirements (risk management, software development, supply chain coherence, etc.) in helping the organization meet its objectives.

Besides the generic management system standards mentioned above, an AI management system can also be used jointly with a management system dedicated to a sector. For example, both ISO 22000 and an AI management system are relevant for an AI system that is used for food production, preparation and logistics. Another example is ISO 13485. The implementation of an AI management system can support requirements related to medical device software in ISO 13485 or requirements from other International Standards from the medical sector such as IEC 62304.

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