INTERNATIONAL STANDARD

# ISO/IEC 42001

First edition
2023-12

# Information technology — Artificial intelligence — Management system

Reference number
ISO/IEC 42001:2023(E)

© ISO/IEC 2023

## COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2023

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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



© ISO/IEC 2023 – All rights reserved

# Contents

Page



© ISO/IEC 2023 – All rights reserved

© ISO/IEC 2023 – All rights reserved

# 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.

© ISO/IEC 2023 – All rights reserved

# 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.


© ISO/IEC 2023 – All rights reserved

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.

ISO27001-2013 信息技术 安全技术 信息安全管理体系内审员培训
https://www.pinzhi.org/forum.php?mod=viewthread&tid=69586

ISO/IEC-27000:2016 信息技术-安全技术信息安全管理体系-概述和
https://www.pinzhi.org/forum.php?mod=viewthread&tid=58973

ISO 20000-1:2018《信息技术 服务管理 第一部分 服务管理体系要求》
https://www.pinzhi.org/forum.php?mod=viewthread&tid=65825

GB/T 22080-2016《信息安全管理体系 要求》
https://www.pinzhi.org/forum.php?mod=viewthread&tid=72037

华为信息安全管理体系考察表 Information Security System Audit
https://www.pinzhi.org/forum.php?mod=viewthread&tid=71380

ISO/IEC 20000-1《信息技术服务管理》和ISO/IEC 27001《信息安全管理体系》
https://www.pinzhi.org/forum.php?mod=forumdisplay&fid=79

GB/T 41271-2022 《生产过程质量控制 通信一致性测试方法》
https://www.pinzhi.org/forum.php?mod=viewthread&tid=78372

ISO/IEC 27701-2019《隐私信息管理体系》【中文译本】
https://www.pinzhi.org/forum.php?mod=viewthread&tid=70961

ISO/IEC 27701:2019《隐私信息管理体系标准》
https://www.pinzhi.org/forum.php?mod=viewthread&tid=62551

ISO27701-2019手册程序文件表单全套文件 (373页 Word文档)
https://www.pinzhi.org/forum.php?mod=viewthread&tid=83870

ISO/IEC 29100:2011 Security techniques - Privacy framework 标准
https://www.pinzhi.org/forum.php?mod=viewthread&tid=72491

ISO/IEC 20000-1《信息技术服务管理》和ISO/IEC 27001《信息安全管理体系》
https://www.pinzhi.org/forum.php?mod=forumdisplay&fid=79

© ISO/IEC 2023 – All rights reserved

**INTERNATIONAL STANDARD** **ISO/IEC 42001:2023(E)**

# 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.

© ISO/IEC 2023 – All rights reserved

**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.


© ISO/IEC 2023 – All rights reserved

**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

© ISO/IEC 2023 – All rights reserved

**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


© ISO/IEC 2023 – All rights reserved

**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.

© ISO/IEC 2023 – All rights reserved

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.


© ISO/IEC 2023 – All rights reserved

# 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.

© ISO/IEC 2023 – All rights reserved

### 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.


© ISO/IEC 2023 – All rights reserved

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;

© ISO/IEC 2023 – All rights reserved

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);


© ISO/IEC 2023 – All rights reserved

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.

© ISO/IEC 2023 – All rights reserved

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.


© ISO/IEC 2023 – All rights reserved

### 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.

© ISO/IEC 2023 – All rights reserved

### 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.


© ISO/IEC 2023 – All rights reserved

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.

© ISO/IEC 2023 – All rights reserved

## 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.


© ISO/IEC 2023 – All rights reserved

# 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. |

© ISO/IEC 2023 – All rights reserved

**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. |


© ISO/IEC 2023 – All rights reserved

**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. |

© ISO/IEC 2023 – All rights reserved

**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. |


© ISO/IEC 2023 – All rights reserved

# 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;

© ISO/IEC 2023 – All rights reserved

— 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.


© ISO/IEC 2023 – All rights reserved

**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.

© ISO/IEC 2023 – All rights reserved

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 B.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.


© ISO/IEC 2023 – All rights reserved

**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.

© ISO/IEC 2023 – All rights reserved

**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.


© ISO/IEC 2023 – All rights reserved

**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

© ISO/IEC 2023 – All rights reserved

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;


© ISO/IEC 2023 – All rights reserved

— 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

© ISO/IEC 2023 – All rights reserved

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.


© ISO/IEC 2023 – All rights reserved

**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.

© ISO/IEC 2023 – All rights reserved

**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;


© ISO/IEC 2023 – All rights reserved

— 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.

© ISO/IEC 2023 – All rights reserved

**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 2023 – All rights reserved

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;

© ISO/IEC 2023 – All rights reserved

— 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.


© ISO/IEC 2023 – All rights reserved

**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);

© ISO/IEC 2023 – All rights reserved

— 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-


© ISO/IEC 2023 – All rights reserved

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;

© ISO/IEC 2023 – All rights reserved

— 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;


© ISO/IEC 2023 – All rights reserved

— 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;

© ISO/IEC 2023 – All rights reserved

— 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


© ISO/IEC 2023 – All rights reserved

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

© ISO/IEC 2023 – All rights reserved

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.


© ISO/IEC 2023 – All rights reserved

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.

© ISO/IEC 2023 – All rights reserved

# 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.


© ISO/IEC 2023 – All rights reserved

### 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).

© ISO/IEC 2023 – All rights reserved

### 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.


© ISO/IEC 2023 – All rights reserved

# 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

© ISO/IEC 2023 – All rights reserved

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.


© ISO/IEC 2023 – All rights reserved

# 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*

© ISO/IEC 2023 – All rights reserved

[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


© ISO/IEC 2023 – All rights reserved

*This page deliberately left blank*

# 国际标准

ISO/IEC
42001
第1版
2023-12

# 人工智能 管理体系

# Artificial intelligence - Management system


ISO/IEC 42001:2023
© ISO/IEC 2023

# 前　言

国际标准化组织(ISO)是由各国标准化团体(ISO 成员团体)组成的世界性的联合会。制定国际标准工作通常由 ISO 的技术委员会完成。各成员团体若对某技术委员会确定的项目感兴趣，均有权参加该委员会的工作。与 ISO 保持联系的各国际组织(官方的或非官方的)也可参加有关工作。ISO 与国际电工委员会(IEC)在电工技术标准化方面保持密切合作的关系。

制定本标准及其后续标准维护的程序在ISO/IEC指引 第1部分均有描述。应特别注意用于各不同类别ISO文件批准准则。本标准根据 ISO/IEC导则第2部分的规则起草(见 www. iso. org/directives或www. iec. ch/members_experts/refdocs) 。

本标准中的某些内容有可能涉及一些专利权问题，对此应引起注意。ISO不负责识别任何这样的专利权问题。在标准制定期间识别的专利权细节将出现在引言/或收到的ISO专利权声明清单中（www.iso.org/patents）。

本标准中使用的任何商品名称仅为方便用户而提供的信息， 并不构成认可。

ISO与合格评定相关的特定术语和表述含义的解释以及ISO遵循的世界贸易组织（WTO）贸易技术壁垒（TBT）原则关信息访问以下URLL：www.iso.org/iso/foreword.html.。在 IEC 中， 请参见www.iec. ch/understanding-standards

本标准由 ISO/IEC JTC 1，联合技术委员会，信息技术 SC 42人工智能分委员会编写。

关于本标准的任何反馈或或者疑问都应直接向用户的国家标准机构提出。 完整的国家标准机构列表可访问www.iso.org/members.html 以及 www.iec.ch/national-committees 获取。

# 引　言

人工智能（Al）正越来越多地应用于利用信息技术的各个行业，并有望成为主要的经济驱动因素之一。这一趋势的一个后果是，某些应用可能在未来几年引发社会挑战。

本文件旨在帮助各组织负责任地履行其在人工智能体系方面的职责(如使用、开发、监视或提供利用人工智能的产品或服务）。人工智能可能会产生一些具体的考虑因素，例如

——使用 Al 进行自动决策，有时是以不透明和无法理解的方式进行的，这就需要在传统的 IT 系统管理之外进行专门的管理；

——利用数据分析、洞察力和机器学习，而不是人为编码的逻辑来设计系统，既增加了 Al 系统的应用机会，也改变了开发、论证和部署此类系统的方式；

——进行持续学习的人工智能系统在使用过程中会改变其行为。它们需要特别考虑以确保在行为不断变化下继续负责任地使用。

本文件提供了在组织范围内建立、实施、保持和持续改进 Al 管理体系的要求。组织应将其要求的应用重点放在 Al 独有的特征上。Al 的某些特征（如持续学习和改进的能力或缺乏透明度或可理解性），如果与传统的任务执行方式相比引起额外的隐忧，则需要采取不同的防护措施。

组织的需求和目标、过程、规模和结构以及各相关方的期望都会影响人工智能管理体系的建立和实施。影响人工智能管理体系建立和实施的另一组因素是人工智能的众多用例，以及在治理机制和创新之间取得适当平衡的必要性。组织可选择使用基于风险的方法来应用这些要求，以确保对组织范围内的特定人工智能用例、服务或产品实施适当级别的控制。预计所有这些影响因素都会发生变化，并不时进行评审。

人工智能管理系统应与组织的过程和整体管理结构进行整合。在设计过程时应考虑与人工智能有关的具体问题，包括信息体系和控制措施、 在设计过程、信息体系和控制措施时，应考虑与人工智能有关的具体问题。这些管理过程的重要实例包括：

——组织目标、相关方的参与和组织方针的确定；

——风险和机遇的管理；

——管理与人工智能体系可信度有关的问题的过程，如人工智能系统的安全、安全、公平、透明、数据质量和整个生命周期的质量；

——管理供方、合作伙伴和为组织提供或开发 人工智能系统的供方、合作伙伴和第三方的管理过程。

本文件提供了部署适用控制措施的指导原则，以支持这些过程。

本文件避免了对管理过程的具体指导。组织可将公认的框架、其他国际标准和自身经验相结合，实施适用于范围内特定人工智能用例、产品或服务的关键过程，如风险管理、生命周期管理和数据质量管理。

符合本文件要求的组织可提供证据，证实其在AI体系方面的责任和义务。

本文件中提出要求的顺序并不反映其重要性，也不意味着实施要求的顺序。列举清单项目仅供参考。

**与其他管理体系标准的兼容性**

本文件采用统一结构（相同的条款编号、条款标题、文本和通用术语和核心定义），以加强管理系统之间的一致性标准（MSS）。人工智能管理体系提供了管理问题的特定要求以及在组织中使用AI所产生的风险。这种通用方法有助于实施其他管理体系标准并与之保持一致，例如与质量、安全、安保和隐私相关的标准。

# 人工智能　管理体系

## 1　范围

本文件规定了在组织范围内建立、实施、保持和持续改进人工智能管理体系的要求并提供了指导。

本文件适用于提供或使用人工智能系统的产品或服务的组织。本文件旨在帮助组织负责任地开发、提供或使用人工智能系统，以实现其目标，并满足与相关方相关的适用法规要求、义务和期望。

本文件适用于任何提供或使用人工智能系统的产品或服务的组织，无论其规模、类型和性质如何。

## 2　规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅注日期对应的版本适用于本文件；不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

ISO/IEC 22989:2022 信息技术　人工智能　人工智能概念和术语

## 3　术语和定义

ISO/IEC 22989:2022界定的及下列术语和定义适用于本文件。

3.1

**组织　organization**

为实现目标（3.6），由职责、权限和相互关系构成自身功能的一个人或一组人。

注 1：组织的概念包括，但不限于个体经营者、公司、集团公司、商行、企事业单位、权力机构、合伙企业、慈善机构或研究机构，或上述组织的部分或组合，无论是否具有法人资格，公有或私有。

注 2：如果组织是大型实体的某个组成部分，那么，术语“组织”仅指在人工智能管理体系(3.4)范围内的这个组成部分。

3.2

**相关方　interested party**

能够影响决策、被影响或认为自己受到决策或活动影响的人或组织(3.1)

注 1：ISO/IEC 22989:2022, 5.19 提供了人工智能相关方的概述。

3.3

**最高管理者 top management**

在最高层指挥并控制组织(3.1)的一个人或一组人。

注 1：最高管理者有权在组织内部授权和提供资源。

注 2：如果管理体系(3.4)的范围仅覆盖组织的某个组织部分，那么最高管理者是指挥和控制该部分的一个人或一组人。

3.4

**管理体系 management system**

组织(3.1)为确立方针(3.5)和目标(3.6)以及实现这些目标的过程(3.8)所形成的相互关联或相互作用的一组要件。

注 1：一个管理体系可能针对一个或几个主题。

注 2：管理体系要件包括组织的结构、岗位和职责、策划和运行。

3.5

**方针 policy**

由最高管理者(3.3)正式表述的组织(3.1)的意图和方向。

3.6

**目标 objective**

要实现的结果。

注 1：目标可能是战略性的、战术性的或运行的。

注 2：目标可能涉及不同的主题(如财务、健康和安全、环境)。它们可能存在于不同层面，诸如组织整体层面或项目、产品或过程(3.8)层面。

注 3：目标能够用其他方式表述，如：预期的结果、宗旨、运行准则，人工智能目标或使用其他有类似含义的词(如：终点或指标)。

注 4：在人工智能管理体系(3.4)中，组织(3.1)设定的人工智能目标与人工智能方针(3.5)保持一致，以实现特定的结果。

3.7

**风险 risk**

不确定性的影响。

注 1：影响是对预期的偏离——积极的或消极的。

注 2：不确定性是指对某一事件、其后果或可能性缺乏相关信息、理解或知识的状态，甚至是部分缺乏。

注 3：风险通常被描述为潜在事件(如 ISO 指南 73 所定义)和后果(如 ISO 指南 73 所定义)，或两者的组合。

注 4：风险通常用事件的后果(包括环境的变化)和相关的发生可能性(如 ISO 指南 73 中定义)的组合来表示。

3.8

**过程 process**

使用或转化输入以实现结果的一组相互关联或相互作用的活动。

注 1：某个过程的结果是称为输出，还是称为产品或服务，取决于相关语境。

3.9

**能力 competence**

应用知识和技能实现预期结果的本领。

3.10

**成文信息 documented information**

组织(3.1)需要控制和维护的信息及其载体。

注 1：成文信息能够以任何形式和载体存在，且来源不限。

注 2：成文信息可能涉及：

—管理体系(3.4)，包括相关过程(3.8)；

—为组织运行而创建的信息(文件)；

—实现的结果的证据（记录）。

3.11

**绩效 performance**

可测量的结果。

注 1:绩效可能涉及定量的或定性的结果。

注 2:绩效可能与活动、过程（3.8）、产品、服务、体系或组织（3.1）的管理有关。

注 3:在本文件中，绩效既指使用人工智能系统取得的结果，也指与人工智能管理体系相关的结果（3.4）。该术语的正确解释可从其使用的环境中得出。

3.12

**持续改进 continual improvement**

提高绩效（3.11）的循环活动。

3.13

**有效性 effectiveness**

完成策划的活动和实现策划的结果的程度。

3.14

**要求/需求 requirement**

规定的、不言而喻的或有义务履行的需求或期望。

注 1：不言而喻的或有义务履行的需要或期望是指需求。其中，“不言而喻”是指组织（3.1）和相关方（3.2）的惯例或一般做法，不言而喻的需求或期望是不用说就明白的。

注 2：规定的需要或期望是指要求，也就是符合 GB/T 1.1 中定义的要求，即表达声明符合该文件需要满足的客观可证实的准则。

3.15

**符合 conformity**

满足要求(3.14)。

3.16

**不符合　nonconformity**

未满足要求(3.14)。

3.17

**纠正措施　corrective action**

为消除不符合(3.16)的原因并防止再次发生而采取的措施。

3.18

**审核　audit**

获取审核证据并对其进行客观评价，以确定审核准则满足程度所进行的系统的、独立的过程（3.8）。

**注 1：**注 1：审核可能为内部（第一方）审核或外部（第二方或第三方）审核，也可能为多体系审核（合并两个或多个主题）。

**注 2：**注 2：内部审核由组织（3.1）自行实施或代表组织的外部机构实施。

**注 3：**注 3：“审核证据”和“审核准则”的定义见 ISO 19011。

3.19

**测量　measurement**

确定数值的过程（3.8）。

3.20

**监视　monitoring**

确定体系、过程(3.8)或活动的状态。

**注 1：**确定状态可能需要检查、监督或严格观察。

3.21

**控制　control**

<风险>维持和/或修改风险（3.7）的措施。

**注 1：**控制包括但不限于任何维持和/或修改风险的过程、政策、设备、实践或其他条件和/或措施。

**注 2：**控制可能并不总是发挥预期或假定的修改效果。

[来源:ISO 31000:2018, 3.8，<风险>作为应用领域。]

3.22

**治理机构　governing body**

对组织的绩效和一致性负责的人或一组人。

**注 1：**并非所有组织，特别是小型组织，都有一个独立于最高管理者的理事机构。

**注 2：**治理机构可以包括但不限于董事会、董事会委员会、监事会、受托人或监事。

[来源:ISO/IEC 38500:2015, 2.9，修改-增加了条目注释。]

3.23

**信息安全　information security**

保持信息的保密性、完整性和可用性。

注 1：其他属性，如真实性、可问责性、不可否认性和可靠性也可能涉及。

［来源:ISO/IEC 27000:2018, 3.28，修改-增加了对条目的注释。］

3.24

**人工智能体系影响评估　AI system impact assessment**

由开发、提供或使用人工智能产品或服务的组织识别、评估和解决对个人(或个人群体)和社会的影响的正式的、文件化的过程。

3.25

**数据质量　data quality**

符合组织在特定情况下数据需求的一种数据特征。

［来源:ISO/IEC 5259-1, 3.4］

3.26

**适用性声明　statement of applicability**

所有必要控制（3.23），以及包括或排除控制的理由的文件化。

注 1：组织可能不需要本文件附件 A 中列出的所有控制，甚至可能超出附件 A 中所列的控制，并由组织自己制定额外的控制。

注 2：所有已识别的风险，组织应按本文件的要求形成文件。所有已识别的风险和为解决这些风险而制定的风险管理措施(控制)应反映在适用性声明中。

## 4　组织环境

### 4.1　理解组织及其环境

组织应确定与其宗旨相关的，并影响其实现人工智能管理体系预期结果的能力的内部和外部因素。

组织应考虑组织开发、提供或使用的人工智能系统的预期目的。组织应确定其在人工智能系统中的角色。

为了了解组织及其环境，组织确定其相对于人工智能系统的角色可能会有所帮助。这些角色可以包括但不限于以下一种或多种：

——人工智能提供商，包括人工智能平台提供商、人工智能产品或服务提供商；

——工智能生产者，包括人工智能开发者、人工智能设计师、人工智能运营商、人工智能测试和评 估人员、人工智能部署人员、人工智能人为因素专业人员、领域专家、人工智能影响评估人员、采购人 员、人工智能治理和监督专业人员；

——人工智能客户，包括人工智能用户；

——人工智能合作伙伴，包括人工智能系统集成商和数据提供商；

——人工智能主体，包括数据主体和其他主体；　—相关监管机构，包括政策制定者和监管机构。

ISO/IEC 22989提供了这些角色的详细描述。此外，角色类型及其与人工智能系统生命周期的关系也在NIST 人工智能风险管理框架中进行了描述[28]。组织的角色可以确定本文件中的要求和控制的适用性和适用性程度。

注 1：根据本条款要解决的外部和内部因素可能因组织的角色和管辖权及其对其实现人工智能管理体系预期结果的能力的影响而有所不同。这些可以包括但不限于：

a) 外部环境相关的考虑，如：

1)适用的法律要求，包括禁止使用人工智能；

2)监管机构的政策、指导方针和决定对人工智能系统开发和使用中法律要求的解释或执行产生影响；

3)与人工智能系统的预期目的和使用相关的激励或后果；

4)人工智能发展和使用方面的文化、传统、价值观、规范和伦理；

5)使用人工智能系统的新产品和服务的竞争格局和趋势。

b) 内部环境相关的考虑，如：

1)组织环境、治理、 目标(见6.2)、方针和程序；

2)合同义务；

3)拟开发或使用人工智能系统的预期目的。

注 2：角色的确定可以通过与组织处理的数据类别相关的义务来确定(例如，在处理PII 时，PII 处理者或PII控制者)。有关PII和相关角色，请参阅 ISO/IEC 29100。角色也可以通过特定于人工智能系统的法律要求来了解。

### 4.2 理解相关方的需求和期望

组织应确定：

——与人工智能管理体系有关的相关方；

——这些相关方的有关要求；

——哪些要求将通过人工智能管理体系予以解决。

### 4.3 确定人工智能管理体系的范围

组织应确定人工智能管理体系的边界和适用性，以确定其范围。

组织应根据以下内容确定人工智能管理体系的范围：

——4.1 提及的内部和外部因素；

——4.2 提及的要求。

范围应作为成文信息可获取。

人工智能管理体系的范围应根据本文件对人工智能管理体系、领导、策划、支持、运行、绩效、评价、改进、控制和目标的要求确定组织的活动。

### 4.4 人工智能管理体系

组织应按照本文件的要求，建立、实施、保持、持续改进人工智能管理体系，包括所需的过程及其相互作用，并形成文件。

## 5 领导作用

### 5.1 领导作用和承诺

最高管理者应通过以下方式展示对人工智能管理体系的领导作用和承诺：

——确保制定人工智能方针（见 5.2）和人工智能目标（见 6.2），并与组织的战略方向保持一致；

——确保将人工智能管理体系的要求融入组织的业务过程；

——确保具备人工智能管理体系所需的资源；

——沟通有效的人工智能管理和符合人工智能管理体系要求的重要性；

——确保人工智能管理体系实现预期效果；

——指导和支持有关人员促进人工智能管理体系的有效性；

——推动持续改进；

——支持其他相关人员在其职责范围内发挥领导作用。

**注 1**：本文件中提及的"业务"可作广义解释，指那些与组织存在目的相关的核心活动。

**注 2**：在组织内部建立、鼓励和构建一种文化，以负责任的方式使用、开发和治理人工智能系统，这是最高管理者承诺和领导力的重要体现。确保意识到并遵守这种负责任的方法，并通过领导支持有助于人工智能管理体系的成功。

### 5.2 人工智能方针方针

最高管理者应制定一项人工智能方针方针，方针应：

a) 适合于组织的宗旨；

b) 为制定人工智能目标提供一个框架（见 6.2）；

c) 包括满足适用要求的承诺；

d) 包括对持续改进人工智能管理体系的承诺。

人工智能方针应：

——作为文件资料提供；

——引用其他相关的组织政策；

——在组织内得到沟通；

——适宜时，可为有关相关方所获取。

A.2 提供了制定人工智能方针的控制目标和控制。这些控制的实施指南见 B.2。

**注 1**：ISO/IEC 38507 提供了组织在制定人工智能方针时的注意事项。

### 5.3 岗位、职责和权限

最高管理者应确保在组织内部分配并沟通相关岗位的职责和权限。

最高管理者应分配职责和权限，以便：

a) 确保人工智能管理体系符合本文件的要求；

b) 向最高管理者报告人工智能管理体系的绩效。

附件 A（A.3.2）提供了定义和分配角色与责任的控制。B.3.2 提供了该控制的实施指南。

## 6 策划

### 6.1 确定风险和机遇的措施

### 6.1.1 通则

在对人工智能管理体系进行策划时，组织应考虑4.1中提及的因素和4.2中提及的要求，并确定需要应对的风险和机遇，以便：

——确保人工智能管理体系能够实现其预期结果；

——防止或减少不利影响；

——实现持续改进。

组织应建立和保持人工智能风险准则，以支持以下工作：

——区分可接受与不可接受的风险；

——进行人工智能风险评估；

——实施人工智能风险应对；

——评估人工智能风险的影响。

**注1**：ISO/IEC 38507 和 ISO/IEC 23894 中提供了确定组织愿意追求或保留的风险数量和类型的考虑因素。

组织应根据以下因素确定风险和机遇：

——人工智能系统的领域和应用背景；

——预期用途；

——4.1 中提及的外部和内部因素。

**注2**：在人工智能管理体系的范围内可考虑不止一个人工智能系统。在这种情况下，应针对每个人工智能系统或人工智能系统组确定机会和用途。

组织应策划：

a) 应对这些风险和机遇的措施；

b) 如何做：

1）将这些行动融入其人工智能管理体系过程并加以实施；

2）评估这些行动的有效性。

组织应保留为识别和应对人工智能风险和机会而采取的措施的成文信息。

**注3**：关于如何为开发、提供或使用人工智能产品、系统和服务的组织实施风险管理的指导见 ISO/IEC 23894。

**注4**：组织及其活动的背景会对组织的风险管理活动产生影响。

**注5**：不同部门和行业对风险的定义以及风险管理的设想可能有所不同。 3.7 中对风险的规范性定义允许对风险有一个广泛的认识，以适应任何部门，如条款 D.1 中提到的部门。 在任何情况下，作为风险评估的一部分，组织的职责是首先采用适合其环境的风险观。这可以包括通过人工智能系统为之开发和使用的部门所使用的定义来看待风险，如 ISO/IEC 指南 51 中的定义。

### 6.1.2 人工智能风险评估

组织应规定并建立人工智能风险评估过程，该过程应：

a) 引用并符合人工智能方针（见 5.2）和人工智能目标（见 6.2）；

**注**：在评估作为 6.1.2 d) 1) 部分的后果时，组织可利用 6.1.4 所述的人工智能系统影响评估。

b) 在设计上使重复的人工智能风险评估能够产生一致、有效和可比较的结果；

c) 识别有助于或阻碍实现人工智能目标的风险；

d) 分析人工智能风险，以便：

1）评估如果确定的风险成为现实，将对组织、个人和社会造成的潜在后果；

2）酌情评估已识别风险的现实可能性；

3）确定风险等级。

e） 评估人工智能风险，以便：

1）将风险分析结果与风险准则（见 6.1.1）进行比较；

2）对评估的风险进行优先排序，以便进行风险应对。

组织应保留有关人工智能风险评估过程的成文信息。

### 6.1.3 人工智能风险应对

考虑到风险评估结果，组织应确定人工智能风险应对过程，以便：

a） 选择适当的人工智能风险应对方案；

b） 确定实施所选人工智能风险应对方案所必需的所有控制，并将这些控制与附件 A 中的控制进行比较，以核实没有遗漏任何必要的控制。

**注1：** 附件 A 提供了实现组织目标和处理与人工智能系统的设计和使用有关的风险的参考控制。

c） 考虑附件 A 中与实施人工智能风险应对方案相关的控制；

**注2：** 组织可参考附件 B，了解 b）和 c）中确定的控制的实施指南。

d） 确定除了附件 A 中的控制外，是否还需要其他控制，以实施所有风险应对备选方案；

控制目标隐含在所选择的控制中。组织可根据需要选择附件 A 中列出的控制目标和控制。附件 A 中的控制并非详尽无遗，可能还需要额外的控制目标和控制。如果需要附件 A 以外的不同或额外控制，组织可设计此类控制或从现有来源获取。如适用，人工智能风险管理可融入其他管理体系。

e） 编制一份适用性声明，其中包含必要的控制[见 b)、c)和 d)]，并说明纳入和排除控制的理由。排除的理由可包括风险评估认为不需要的控制，以及适用的外部要求不需要（或属于例外情况）的控制。

**注3：** 组织可提供文件证明排除一般或特定人工智能系统控制目标的理由，不论是附件 A 中列出的还是组织自己制定的。

f） 制定人工智能风险应对办法；

获得指定管理层对人工智能风险应对计划和接受残余人工智能风险的批准。 必要的控制应：

——与 6.2 中的目标相一致；

——作为文件信息提供；

——在组织内得到沟通；

——适宜时，可为有关相关方所获取。

组织应保留有关人工智能风险应对过程的成文信息。

### 6.1.4 人工智能系统影响评估

组织应确定过程，用于评估开发、提供或使用人工智能系统可能对个人和社会造成的潜在后果。

人工智能系统影响评估应确定人工智能系统的部署、预期使用和可预见的滥用对个人和社会造成的潜在后果。

影响评估应考虑到部署人工智能系统的具体技术和社会背景以及适用的管辖范围。

系统影响评估的结果应形成成文信息。在适当情况下，可将系统影响评估的结果提供给组织规定的相关 利益方。

**注1：** 组织应在风险评估中考虑人工智能系统影响评估结果（见 6.1.2）。A.5 提供了评估人工智能系统影响的控制。

**注2：** 在某些情况下（如对安全或隐私至关重要的人工智能系统），组织可要求进行特定学科的人工智能系统影响评估（如物理安全、隐私或信息安全影响），作为组织整体风险管理活动的一部分。

### 6.2 人工智能目标及其实现的策划

组织应在相关职能和层次制定人工智能目标。

人工智能目标应：

a) 符合人工智能方针（见 5.2）；

b) 可测量（如可行）；

c) 考虑到适用的要求；

d) 予以监视；

e) 予以沟通；

f) 适时更新；

g) 作为文件信息提供。

在规划如何实现其人工智能目标时，组织应确定：

——要做什么；

——需要什么资源；

——由谁负责；

——何时完成；

——如何评价结果。

**注：附件C提供了与风险管理有关的人工智能目标的非排他性清单。附件 A（A.6.1 和 A.9.3）提供了确定负责任地开发和使用人工智能系统的控制目标和控制。B.6.1 和 B.9.3 提供了这些控制的实施指南。**

### 6.3 变更的策划

当组织确定有必要更改人工智能管理体系时，变更应按所策划的方式实施。

## 7 支持

### 7.1 资源

组织应确定并提供建立、实施、保持和持续改进人工智能管理体系所需的资源。

人工智能资源的控制目标和控制方法见A.4。B.4提供了这些控制的实施指南。

### 7.2 能力

组织应：

——确定在其控制下从事影响其人工智能绩效的工作的人员的必要能力；

——基于适当的教育、培训或经验，确保这些人员胜任的；

——适用时，采取措施以获得所需的能力，并评价措施的有效性。

保留适当的成文信息，作为人员能力的证据。

B.4.6提供了关于人力资源的实施指导，包括考虑必要的专门知识。

注 1：适用的措施可包括，例如:对现有员工提供培训、指导或重新分配；或雇用或承包有能力的人。

### 7.3 意识

组织应确保在其控制下工作的人员知晓：

——人工智能方针(见 5.2)；

——他们对人工智能管理体系有效性的贡献，包括提高人工智能绩效的益处；

——不符合人工智能管理体系要求的后果。

### 7.4 沟通

组织应确定与人工智能管理体系相关的内部和外部沟通，包括：

——沟通什么；

——何时沟通；

——与谁沟通；

——如何沟通。

### 7.5 成文信息

#### 7.5.1 通则

组织的人工智能管理体系应包括：

a) 本文件要求的形成文件的信息；

b) 组织确定的人工智能管理体系有效性所必需的形成文件的信息。

**注：**人工智能管理体系成文信息的程度可能因组织而异，原因如下：

——组织的规模，以及活动、过程、产品和服务的类型；

——过程及其相互作用的复杂程度；

——人员的能力。

#### 7.5.2 创建和更新成文信息

在创建和更新成文信息时，组织应确保适当的：

a) 标识和说明（如标题、日期、作者、索引编号）；

b) 形式（如语言、软件版本、图表）和载体（如纸质的、电子的）；

c) 评审和批准，以保持适宜性和充分性。

#### 7.5.3 成文信息的控制

应控制人工智能管理体系和本文件要求的成文信息，以确保其：

a) 在需要的场合和时机，均可获得并适用；

b) 予以妥善保护（如：防止泄密、不当使用或缺失）。

为控制成文信息，适用时，组织应进行下列活动：

——分发、访问、检索和使用；

——存储和防护，包括保持可读性；

——更改控制（如版本控制）；

——保留和处置。

对于组织所确定的策划和运行人工智能管理体系所必需的来自外部的成文信息，组织应进行适当识别，并予以控制。

注 1：对于成文信息的“访问”可能意味着仅允许查阅或者意味着允许查阅和并授权修改。

## 8 运行

### 8.1 运行的策划和控制

组织应策划、实施和控制满足要求和实施第6条确定的措施所需的过程，通过：

——建立过程准则；

——根据准则实施过程控制。

组织应实施根据6.2.3确定的与人工智能管理体系运行相关的控制(如与人工智能系统开发和使用生命周期相关的控制)。

应监视这些控制的有效性，如果未能实现预期结果，应考虑采取纠正措施。附件A列出了参考控制，附件B提供了实施指南。

在必要的范围和程度上，确定并保持、保留成文信息，以确信过程已经按策划进行；

组织应控制策划的变更，评审非预期变更的后果，必要时，采取措施减轻不利影响。

组织应确保与人工智能管理体系相关的外部提供的过程、产品或服务得到控制。

### 8.2 人工智能风险评估

组织应按照6.1.2的规定策划的时间间隔内或在提出重大变更时，进行人工智能风险评估。

组织应保留所有人工智能风险评估结果的成文信息。

### 8.3 人工智能风险应对

组织应按6.1.3实施人工智能风险应对计划，并验证其有效性。

当风险评估识别出需要应对的新风险时，应对这些风险执行6.1.3的风险应对过程。

当风险应对计划确定的风险应对方案无效时，应按照6.1.3的风险应对过程对这些风险应对方案进行评审和重新验证，并更新风险应对计划。

组织应保留所有人工智能风险应对结果的成文信息。

### 8.4 人工智能系统影响评估

组织应根据6.1.4按照计划的时间间隔或在提出重大变更时执行人工智能系统影响评估。

组织应保留所有人工智能系统影响评估结果的成文信息。

## 9 绩效评价

### 9.1 监视、测量、分析和评价

组织应确定：

——需要监视和测量什么；

——需要用什么方法进行监视、测量、分析和评价，以确保结果有效；

——何时实施监视和测量；

——何时对监视和测量的结果进行分析和评价。

成文信息应作为结果证据可获取。

组织应评价人工智能管理体系的绩效和有效性。

## 9.2 内部审核

### 9.2.1 通则

组织应按计划的时间间隔进行内部审核，以提供人工智能管理体系是否：

a) 符合：

1) 组织自身对人工智能管理体系的要求；

2) 本文件的要求；

b) 有效地实施和保持。

### 9.2.2 内部审核程序

组织应策划、建立、实施和保持审核方案，包括审核频率、方法、职责、策划要求和报告。

在制定内部审核方案时，组织应考虑相关过程的重要性和以往审核的结果。

组织应：

a) 确定每次审核的审核目标、准则和范围；

b) 选择审核员并实施审核，确保审核过程的客观性和公正性；

c) 确保将审核结果报告给相关管理者。

应提供形成文件的信息，作为审核方案实施和审核结果的证据。

## 9.3 管理评审

### 9.3.1 通则

最高管理者应在策划的时间间隔内对组织的人工智能管理体系进行评审，以确保人工智能管理体系持续的适宜性、充分性和有效性。

### 9.3.2 管理评审输入

管理评审应包括：

a) 以往管理评审所采取措施的状况；

b) 与人工智能管理体系相关的外部和内部因素的变化；

c) 与人工智能管理体系相关的相关方的需求和期望的变化；

d) 人工智能管理体系绩效信息，包括以下方面的趋势：

1）不符合及纠正措施；

2）监视和测量结果；

3）审核结果；

e) 持续改进的机会。

### 9.3.3 管理评审输出

管理评审的输出应包括持续改进的机会，以及变更人工智能管理体系的任何需要的决定。

成文信息应作为管理评审结果证据可获取。

# 10 改进

## 10.1 持续改进

组织应持续改进人工智能管理体系的适宜性、充分性和有效性。

**注 1**：适宜性指管理体系适合于组织的目标、运行、文化和业务体系的程度。

**注 2**：充分性指管理体系满足适用性要求的程度。

**注 3**：有效性指完成策划的活动和实现策划的结果的程度。

### 10.2 不符合和纠正措施

发生不符合时，组织应：

a）对不符合做出反应，并且如适用：

1）采取控制和纠正措施；

2）处置后果；

b）通过以下活动评价采取措施的需要，以消除产生不符合的原因，避免其再次发生或在其他地方发生；

1）评审不符合；

2）确定产生不符合的原因；

3）确定是否存在或可能发生类似的不符合；

c）实施任何所需的措施；

d）评审所采取的任何纠正措施的有效性；e）如必要，变更人工智能管理体系。

纠正措施应与不符合产生的影响相适应。

成文信息应作为以下事项的证据可获取：

——不符合的性质和所采取的任何后续措施；

——任何纠正措施的结果。

# 附 录 A
（规范性）
参考控制目标与控制

## A.1 概述

表A.1详细列出的控制为组织提供了一个参考，以实现组织目标和应对与人工智能系统的设计和运行有关的风险。并非表A.1中列出的所有控制目标和控制都必须使用，组织可以设计和实施自己的控制（见6.1.3）。

附件B提供了表A.1所列所有控制的实施指南。

表A.1-控制目标和控制

| | | |
|---|---|---|
| **A.2 与人工智能相关的政策** | | |
| 目标：根据业务需求，为人工智能系统提供管理指导和支持。 | | |
| | 主题 | 控制 |
| A.2.2 | 人工智能方针 | 组织应记录有关开发或使用人工智能系统的政策。 |
| A.2.3 | 与其他组织方针保持一致 | 组织应确定其他方针的哪些方面受到组织在人工智能系统方面目标的影响或适用于该目标。 |
| A.2.4 | 人工智能方针的评审 | 人工智能方针应按计划的时间间隔进行评审，或根据需要进行额外评审，以确保其持续的适宜性、充分性和有效性。 |
| **A.3 内部组织** | | |
| 目标：在组织内部建立问责制，坚持以负责任的方式实施、运行和管理人工智能系统。 | | |
| | 主题 | 控制 |
| A.3.2 | 人工智能的岗位和职责 | 应根据组织的需要确定和分配人工智能的岗位和职责。 |
| A.3.3 | 报告关切 | 组织应制定并落实一套流程，以便组织中的员工报告与人工智能系统全生命周期有关的组织中的岗位的关切。 |
| **A.4 人工智能系统的资源** | | |
| 目标：确保组织考虑人工智能系统的资源（包括人工智能系统组件和资产），以充分理解并应对风险和影响。 | | |
| | 主题 | 控制 |
| A.4.2 | 资源记录 | 组织应识别并记录特定人工智能系统生命周期阶段的活动以及与组织相关的其他人工智能相关活动所需的相关资源。 |

| | | |
|---|---|---|
| A.4.3 | 数据资源 | 作为资源识别的一部分，该组织应记录有关人工智能系统所使用的数据资源的信息。 |
| A.4.4 | 工具资源 | 作为资源识别的一部分，组织应记录有关人工智能系统所使用的工具资源的信息。 |
| A.4.5 | 系统和计算资源 | 作为资源识别的一部分，该组织应记录有关人工智能系统所使用的系统和计算资源的信息。 |
| A.4.6 | 人力资源 | 作为资源识别的一部分，组织应记录开发、部署、运行、变更管理、维护、转让和退役以及验证和集成人工智能系统所使用的人力资源及其能力的相关信息。 |
| A.5 **评估人工智能系统的影响** | | |
| 目标：评估人工智能系统在整个生命周期内对相关方的影响。 | | |
| | 主题 | 控制 |
| A.5.2 | 人工智能系统影响评估过程 | 组织应建立一个流程，以评估人工智能系统在其整个生命周期内可能对个人和社会造成的潜在后果。 |
| A.5.3 | 人工智能系统影响评估记录 | 组织应记录人工智能系统影响评估的结果，并在规定期限内保留结果。 |
| A.5.4 | 评估人工智能系统对个人和群体的影响 | 组织应评估并记录人工智能系统在整个系统生命周期内对个人或群体的潜在影响。 |
| A.5.5 | 评估人工智能系统的社会影响 | 组织应评估并记录其人工智能系统在整个生命周期中可能产生的社会影响。 |
| A.6 **人工智能系统生命周期** | | |
| A.6.1 人工智能系统开发管理指南 | | |
| 目标：确保组织识别和记录目标并实施负责任的人工智能系统设计和开发过程。 | | |
| | 主题 | 控制 |
| A.6.1.2 | 负责任地开发人工智能系统的目标 | 组织应识别并记录目标，以指导可信赖人工智能系统的开发，并在开发生命周期中考虑这些目标，并融入实现这些目标的措施。 |
| A.6.1.3 | 可信赖人工智能系统设计和开发流程 | 组织应定义并记录负责任的人工智能系统设计和开发的具体流程。 |
| A.6.2 人工智能系统生命周期 | | |
| 目标：定义人工智能系统生命周期各阶段的标准和要求。 | | |
| | 主题 | 控制 |
| A.6.2.2 | 人工智能系统的要求和规范 | 组织应规定并记录新的人工智能系统或对现有系统 |

| | | |
|---|---|---|
| | | 的重大改进的要求。 |

| | | |
|---|---|---|
| A.6.2.3 | 人工智能系统的设计和开发记录 | 组织应根据组织目标、文件化的要求和规范标准，记录人工智能系统的设计和开发。 |
| A.6.2.4 | 人工智能系统的验证和确认 | 组织应定义并记录人工智能系统的验证和确认措施，并规定其使用标准。 |
| A.6.2.5 | 人工智能系统的部署 | 组织应记录部署计划，并确保在部署前满足适当的要求。 |
| A.6.2.6 | 人工智能系统的运行和监视 | 组织应确定并记录人工智能系统持续运行的必要要素。至少应包括系统和性能监视、维修、更新和支持。 |
| A.6.2.7 | 人工智能系统的技术记录 | 组织应确定用户、合作伙伴、监管机构等各类相关方需要的人工智能系统技术记录，并以适当形式向他们提供技术文件。 |
| A.6.2.8 | 人工智能系统的事件记录日志 | 组织应确定在人工智能系统生命周期的哪些阶段应启用事件日志记录，但至少应在人工智能系统使用时启用。 |
| **A.7 人工智能系统的数据** | | |
| 目标：确保组织理解人工智能系统中的数据在人工智能系统的生命周期中的应用和开发、提供或使用中的作用和影响。 | | |
| | 主题 | 控制 |
| A.7.2 | 用于开发和增强人工智能系统的数据 | 组织应定义、记录和实施与开发人工智能系统有关的数据管理流程。 |
| A.7.3 | 数据采集 | 组织应确定并记录人工智能系统所用数据的采集和选择细节。 |
| A.7.4 | 人工智能系统的数据质量 | 组织应定义和记录数据质量要求，并确保用于开发和运行人工智能系统的数据符合这些要求。 |
| A.7.5 | 数据来源 | 组织应定义并记录一个流程，用于记录人工智能系统中使用的数据在数据和人工智能系统生命周期中的来源。 |
| A.7.6 | 数据准备 | 组织应定义并记录其选择数据准备及要使用的数据准备方法的准则。 |
| **A.8 人工智能系统相关方的信息** | | |
| 目标：确保有关各方掌握必要的信息，以了解和评估风险及其影响（正面和负面）。 | | |
| | 主题 | 控制 |
| A.8.2 | 为用户提供的系统文件和信息 | 组织应确定并向系统用户提供必要的信息。 |

| | | |
|---|---|---|
| A.8.3 | 外部报告 | 组织应为有关各方提供报告系统负面影响的能力。 |
| A.8.4 | 事故通报 | 组织应确定并记录向系统用户通报事故的计划。 |
| A.8.5 | 向有关各方提供信息 | 组织应确定并记录向有关各方报告人工智能系统信息的义务。 |
| **A.9 人工智能系统的使用** | | |
| 目标：确保组织负责任地并按照组织的方针使用人工智能系统。 | | |
| | 主题 | 控制 |
| A.9.2 | 负责任地使用人工智能系统的流程 | 组织应定义并记录负责任地使用人工智能系统的过程。 |
| A.9.3 | 负责任地使用人工智能系统的目标 | 组织应识别并记录指导可问责地使用人工智能系统的目标。 |
| A.9.4 | 人工智能系统的预期用途 | 组织应确保按照人工智能系统及其附带文件的预期用途使用人工智能系统。 |
| **A.10 第三方及客户关系** | | |
| 目标：确保组织了解其责任并承担责任，并在人工智能系统生命周期的任何阶段涉及第三方时适当分摊风险。 | | |
| | 主题 | 控制 |
| A.10.2 | 分配职责 | 组织应确保在其人工智能系统生命周期内的责任被分配在组织、其合作伙伴、供方、客户和第三方之间。 |
| A.10.3 | 供方 | 组织应建立一套流程，确保其对供方提供的服务、产品或材料的使用符合组织负责任地开发和使用人工智能系统的方针。 |
| A.10.4 | 客户 | 组织应确保其开发和使用人工智能系统的负责任方法考虑到客户的期望和需求。 |

# 附 录 B
（规范性）
人工智能控制的实施指南

## B.1 概述

本附件中记录的实施指南与附件A中列出的控制有关。它提供了支持实施附件A中列出的控制和实现控制目标的信息，但组织不必在适用性声明中记录或证明融入或不融入附件B中的实施指南（见 6.1.3）。

实施指南并非在所有情况下都适用或充分，也并非总能满足组织的具体控制要求。组织可根据其具体要求和风险应对需要，扩展或修改实施指南，或自行定义控制的实施。

附件B可作为在本文件定义的人工智能管理体系中确定和实施人工智能风险应对控制的指南。除本附件所列控制外，还可确定其他组织上和技术上的控制（见 6.1.3 中的人工智能系统管理风险应对）。本附件可视为制定组织特定控制实施的起点。

## B.2 与人工智能相关的方针

### B.2.1 目标

根据业务需求，为人工智能系统提供管理指导和支持。

### B.2.2 人工智能方针

**控制**

组织应制定开发或使用人工智能系统的方针。

**实施指南**

人工智能方针应参考以下内容：

— 业务战略；
— 组织的价值观和文化， 以及组织愿意承担或保留的风险程度；
— 人工智能系统带来的风险程度；
— 法律要求，包括合同；
— 组织的风险环境；
— 对相关利益方的影响（见 6.1.4）。

人工智能方针应包括（除5.2中的要求外）：

— 指导组织与人工智能有关的所有活动的原则；
— 处理偏离方针和例外情况的程序。

必要时，人工智能方针应考虑特定主题方面，以提供更多指导或提供与涉及这些方面的其他方针的交叉参考。

此类主题的例子包括：

— 人工智能资源和资产；
— 人工智能系统影响评估(见 6.1.4)；
— 人工智能系统开发。

相关方针应指导人工智能系统的开发、购买、运行和使用。

### B.2.3 与其他组织方针保持一致

**控制**

组织应确定其他方针的哪些方面受到组织在人工智能系统方面目标的影响或适用于该目标。

**实施指南**

许多领域都与人工智能有交叉，包括质量、信息安全、物理安全和隐私。组织应考虑进行全面分析，以确定当前方针是否以及在哪些方面存在必然交叉，并在需要更新时更新这些方针，或在人工智能方针中融入相关规定。

**其他信息**

治理机构代表组织制定的方针应为人工智能方针提供参考。ISO/IEC 38507为治理管理机构的成员提供了指导，以便在人工智能系统的整个生命周期内启用和治理人工智能系统。

### B.2.4 人工智能方针的评审

**控制**

人工智能方针应按计划的时间间隔进行评审，或根据需要进行额外评审，以确保其持续的适宜性、充分性和有效性。

**实施指南**

应由管理层批准的一个角色负责人工智能方针或其组成部分的制定、审查和评估。评审应包括根据组织环境、业务情况、法律条件或技术环境的变化，评估机会以改进组织的人工智能系统管理方针和方法的机会。

人工智能方针的评审应考虑到管理评审的结果。

## B.3 内部组织

### B.3.1 目标

在组织内部建立问责制，坚持以负责任的方式实施、运行和管理人工智能系统。

### B.3.2 人工智能的岗位和职责

**控制**

应根据组织的需要确定和分配人工智能的岗位和职责。

**实施指南**

定义岗位和职责对于确保整个组织中与人工智能系统全生命周期有关的岗位可问责至关重要。在分配角色和责任时，组织应考虑人工智能方针、人工智能目标和已识别的风险，以确保涵盖所有相关领域。组织可优先考虑如何分配岗位和职责。需要明确角色和职责的示例包括：

— 风险管理；
— 人工智能系统影响评估；
— 资产和资源管理；
— 信息安全；
— 物理安全；
— 隐私；
— 开发；
— 绩效；
— 人员监督；
— 供方关系；
— 证明其有能力始终如一地满足法律要求；

— 数据质量管理（整个生命周期）。

各种岗位的职责应界定到适合个人履行其职责的程度。

### B.3.3 报告关切

**控制**

组织应制定并落实一套流程，以便组织中的员工报告与人工智能系统全生命周期有关的组织中的岗位的关切。

**实施指南**

报告机制应具备以下功能：

a) 可选择保密或匿名或两者兼有；

b) 向受雇人员和合同人员提供并宣传；

c) 配备合格的工作人员

d) 为 b)中提到的人员规定适当的调查和解决权力；

e) 规定及时向管理层报告和上报的机制；

f) 为报告和调查相关人员提供有效保护，使其免遭报复（如允许匿名和保密报告）；

g) 根据 4.4 和（如适用）e)提供报告；同时保持 a)中的保密性和匿名性，并尊重一般的商业机密考虑因素

h) 在适当的时限内提供回应机制。

注 组织可利用现有的报告机制作为该流程的一部分。

**其他信息**

除本条款提供的实施指南外，组织还应进一步考虑ISO 37002。

## B.4 人工智能系统的资源

### B.4.1 目标

确保组织考虑人工智能系统的资源（包括人工智能系统组件和资产）进行核算，以充分了解和应对风险和影响。

### B.4.2 资源记录

**控制**

组织应确定并记录特定人工智能系统生命周期阶段活动所需的相关资源，以及与组织相关的其他人工智能相关活动。

**实施指南**

记录人工智能系统的资源对于了解风险以及人工智能系统对个人和社会的潜在影响（包括正面和负面影响）至关重要。记录此类资源（可利用数据流图或系统架构图等）可为人工智能系统影响评估提供信息（见 B.5）。

资源可包括但不限于：

— 人工智能系统组件；

— 数据资源，即在人工智能系统生命周期的任何阶段使用的数据；

— 工具资源（如人工智能算法、模型或工具）；

— 系统和计算资源（如开发和运行人工智能模型的硬件、数据存储和工具资源）；

— 人力资源，即与组织在整个人工智能系统生命周期中的作用有关的、具有必要专业知识的人员（如开发、销售、培训、运行和维护人工智能系统）。

资源可以由组织本身、客户或第三方提供。

**其他信息**

记录资源也有助于确定是否有可用资源，如果没有可用资源，组织应修改人工智能系统的设计规范或其部署要求。

### B.4.3 数据资源

**控制**

作为资源识别的一部分，组织应记录有关人工智能系统所使用的数据资源的信息。

**实施指南**

数据文档应包括但不限于以下内容：

— 数据的出处；
— 数据最后更新或修改的日期（如元数据中的日期标签）；
— 对于机器学习，数据类别（如训练、验证、测试和生产数据）；
— 数据类别（如 ISO/IEC 19944-1 中定义的类别）；
— 标注数据的过程；
— 数据的预期用途；
— 数据的质量（如 ISO/IEC 5259 系列中的描述）；
— 适用的数据保留和处置政策；
— 数据中已知或潜在的偏差问题；
— 数据准备。

### B.4.4 工具资源

**控制**

作为资源识别的一部分，组织应记录有关人工智能系统所使用的工具资源的信息。

**实施指南**

人工智能系统特别是机器学习的工具资源可包括但不限于

— 算法类型和机器学习模型；
— 数据调节工具或流程
— 优化方法
— 评估方法；
— 资源配置工具；
— 辅助模型开发的工具；
— 用于人工智能系统设计、开发和部署的软件和硬件；

**其他信息**

ISO/IEC 23053 为机器学习各种工具资源的类型、方法和途径提供了详细指导。

### B.4.5 系统和计算资源

**控制**

作为资源识别的一部分，该组织应记录有关人工智能系统所使用的系统和计算资源的信息。

**实施指南**

人工智能系统的系统和计算资源信息可包括但不限于以下内容：

— 人工智能系统的资源要求（即帮助确保系统可在资源有限的设备上运行）；

— 系统和计算资源的位置（如本地、云计算或边缘计算）；
— 处理资源（包括网络和存储）；
— 用于运行人工智能系统工作负载的硬件的影响（例如，通过使用或制造硬件或使用硬件的成本对环境造成的影响）。

组织应考虑可能需要不同的资源，以便不断改进人工智能系统。系统的开发、部署和运行可能会有不同的系统需求和要求。

注 ISO/IEC 22989描述了各种系统资源考虑因素。

### B.4.6 人力资源

**控制**

作为资源识别的一部分，组织应记录开发、部署、运行、变更管理、维护、转让和退役以及验证和集成人工智能系统所使用的人力资源及其能力的相关信息。

**实施指南**

组织应考虑对不同专业知识的需求，并融入系统所需的角色类型。例如，如果与用于训练机器学习模型的数据集相关的特定人口群体是系统设计的必要组成部分，则组织可将其融入其中。必要的人力资源包括但不限于：

— 数据科学家；
— 与人工智能系统人工监督相关的角色；
— 信息安全、物理安全和隐私等可信赖领域的专家；
— 人工智能研究人员和专家，以及与人工智能系统相关的领域专家。

在人工智能系统生命周期的不同阶段，可能需要不同的资源。

## B.5 评估人工智能系统的影响

### B.5.1 目标

评估人工智能系统在其整个生命周期内对受其影响的个人和社会的影响。

### B.5.2 人工智能系统影响评估过程

**控制**

组织应建立一个流程，以评估人工智能系统在其整个生命周期内可能对个人和社会造成的潜在后果。

**实施指南**

由于人工智能系统可能对个人、个人群体和社会产生重大影响，提供和使用此类系统的组织应根据这些系统的预期目的和用途，评估这些系统对这些群体的潜在影响。

组织应考虑人工智能系统是否会影响以下方面：

— 个人的法律地位或生活机会；
— 个人的身心健康；
— 普遍人权；
— 社会。

组织的流程应包括但不限于：

a) 应进行人工智能系统影响评估的情况，包括但不限于：
   1) 使用人工智能系统的预期目的和背景的关键性，或这些方面的任何重大变化；

2) 人工智能技术的复杂性和人工智能系统的自动化程度，或这方面的任何重大变化；
3) 人工智能系统处理的数据类型和来源的敏感性或任何重大变化。

b) 作为人工智能系统影响评估过程一部分的要素，可包括：
1) 识别（如来源、事件和结果）；
2) 分析（如后果和可能性）
3) 评估（如接受决定和优先级）；
4) 处理（如缓解措施）；
5) 记录、报告和沟通（见 7.4、7.5 和 B.3.3）；

c) 由谁进行人工智能系统影响评估；

d) 如何利用人工智能系统影响评估[例如，如何为系统的设计或使用提供信息（见 B.6 和 B.9），是否可以触发审查和批准]；

e) 根据系统的预期目的、用途和特点，可能受到影响的个人和社会（例如，针对个人、个人群体或社会的评估）。

影响评估应考虑到人工智能系统的各个方面，包括用于开发人工智能系统的数据、使用的人工智能技术和整个系统的功能。

根据组织的角色、人工智能应用领域以及影响评估的具体学科（如信息安全、隐私和物理安全），流程可能会有所不同。

**其他信息**

对于某些学科或组织来说，详细考虑对个人和社会的影响是风险管理的一部分，特别是在信息安全、物理安全和环境管理等学科。组织应确定，作为此类风险管理过程一部分的特定学科影响评估是否充分融入了对这些特定方面（如隐私）的人工智能考虑。

注 ISO/IEC 23894 描述了组织如何对组织本身以及个人和社会进行影响分析，作为整体风险管理流程的一部分。

### B.5.3 人工智能系统影响评估记录

**控制**

组织应记录人工智能系统影响评估的结果，并在规定期限内保留结果。

**实施指南**

这些文件有助于确定应传达给用户和其他有关方面的信息。

人工智能系统影响评估应按照B.5.2中记录的人工智能系统影响评估要素保留并在需要时更新。保留期限可遵循组织保留时间表，或根据法律要求或其他要求确定。

组织应考虑记录的项目包括但不限于：

— 人工智能系统的预期用途和可合理预见的滥用；
— 人工智能系统对相关个人和社会的积极和消极影响；
— 可预测的故障、其潜在影响以及为减轻故障而采取的措施；
— 系统适用的相关人口群体；
— 系统的复杂性；
— 人在系统关系中的作用，包括人的监督能力、程序和工具，可用于避免负面影响；
— 就业和员工技能。

### B.5.4 评估人工智能系统对个人和群体的影响

**控制**

组织应评估并记录人工智能系统在整个系统生命周期内对个人或群体的潜在影响。

**实施指南**

在评估对个人的影响时，组织应考虑其治理原则、人工智能方针和目标。使用人工智能系统的个人 或其 PII 被人工智能系统处理的个人，可能对人工智能系统的可信度抱有期望。应考虑到儿童、残疾 人、老年人和工人等群体的特殊保护需求。作为系统影响评估的一部分，组织应评估这些期望，并考虑 解决这些期望的方法。

根据人工智能系统的目的和使用范围，作为评估一部分需要考虑的影响领域可以包括但不限于以下方面：

— 公平性；
— 问责制；
— 透明度和可解释性；
— 安全和隐私；
— 安全与健康；
— 财务后果；
— 可获取性；
— 人权。

**其他信息**

必要时，组织应咨询专家(如研究人员、主题专家和用户)，以充分了解体系对个人的潜在影响。

### B.5.5 评估人工智能系统的社会影响

**控制**

组织应评估并记录其人工智能系统在整个生命周期中可能产生的社会影响。

**实施指南**

社会影响因组织背景和人工智能系统类型的不同而大相径庭。人工智能系统的社会影响既可能是有益的，也可能是有害的。这些潜在社会影响的例子包括

— 环境可持续性（包括对自然资源和温室气体排放的影响）；
— 经济（包括获得金融服务、就业机会、税收、贸易和商业）；
— 政府（包括立法程序、为政治利益提供错误信息、国家安全和刑事司法系统）；
— 健康与安全（包括获得医疗保健、医疗诊断和治疗，以及潜在的身心伤害）；
— 规范、传统、文化和价值观（包括导致偏见或对个人造成伤害的错误信息）。

**其他信息**

人工智能系统的开发和使用可能是计算密集型的，会对环境可持续性产生相关影响（例如，因用电量增加而产生的温室气体排放，对水、土地、动植物的影响）。同样，人工智能系统也可用于改善其他系统的环境可持续性（如减少与建筑和运输有关的温室气体排放）。组织应结合其总体环境可持续性目标和战略，考虑其人工智能系统的影响。

组织应考虑如何预防滥用人工智能系统造成社会危害，以及如何利用人工智能系统解决历史遗留问题。例如，人工智能系统能否阻止人们获得贷款、赠款、保险和投资等金融服务，同样，人工智能系统能否改善人们获得这些工具的途径？

人工智能系统已被用于影响选举结果和制造错误信息（如数字媒体中的深度伪造），从而导致政治和社会动荡。政府将人工智能系统用于刑事司法目的，暴露了对个人和群体的偏见风险。本组织应分析不良行为者如何滥用人工智能系统， 以及人工智能系统如何强化不受欢迎的历史性社会偏见。

人工智能系统可用于诊断和治疗疾病，并确定享受健康福利的资格。人工智能系统还可用于发生故障可能导致人员伤亡的场景（如自动驾驶汽车、人机协作）。组织在使用人工智能系统时，如在与健康和安全相关的场景中，应同时考虑积极和消极的结果。

注 ISO/IEC TR 24368提供了与人工智能系统和应用有关的伦理和社会问题的高级概述。

## B.6 人工智能系统生命周期

### B.6.1 人工智能系统开发管理指南

#### B.6.1.1 目标

确保组织识别和记录目标并实施负责任的人工智能系统设计和开发过程。

#### B.6.1.2 负责任地开发人工智能系统的目标

**控制**

组织应识别并记录指导开发可信赖人工智能系统的目标，并在开发生命周期中考虑到这些目标，并融入实现这些目标的措施。

**实施指南**

组织应确定影响人工智能系统设计和开发过程的目标（见 6.2）。这些目标应在设计和开发过程中加以考虑。例如，如果一个组织将 "公平性 "定义为目标之一，则应将其融入需求说明、数据采集、数据调节、模型训练、验证和确认等过程中。组织应提供必要的要求和指南，以确保在各个阶段都有相应的措施（例如，要求使用特定的测试工具或方法来解决不公平或不必要的偏差），从而实现这些目标。

**其他信息**

人工智能技术正被用于增强安全措施，如威胁预测检测和安全攻击预防。这是人工智能技术的一种应用，可用于加强安全措施，保护人工智能系统和传统的非人工智能软件系统。附件C提供了管理风险的组织目标示例，有助于确定人工智能系统开发的目标。

#### B.6.1.3 可信赖人工智能系统设计和开发流程

**控制**

组织应定义并记录负责任的人工智能系统设计和开发的具体流程。

**实施指南**

负责任的人工智能系统流程开发应包括但不限于考虑以下方面：

— 生命周期阶段（ISO/IEC 22989 提供了通用的人工智能系统生命周期模型，但组织可指定自己的生命周期阶段）；
— 测试要求和计划测试手段；
— 人工监督要求，包括流程和工具，特别是当人工智能系统可能对自然人产生影响时；
— 应在哪些阶段进行人工智能系统影响评估；
— 培训数据预期和规则（如可使用哪些数据、经批准的数据供方和标签）；
— 人工智能系统开发人员所需的专业知识（主题领域或其他）或培训，或两者兼而有之；
— 发布标准；
— 各阶段所需的批准和签核；
— 变更控制；
— 可用性和可控性；
— 有关各方的参与。

具体的设计和开发流程取决于人工智能系统打算使用的功能和人工智能技术。

### B.6.2 人工智能系统生命周期

#### B.6.2.1 目标

定义人工智能系统生命周期各阶段的标准和要求。

#### B.6.2.2 人工智能系统的要求和规范

**控制**

组织应规定并记录新的人工智能系统或对现有系统的重大改进的要求。

**实施指南**

组织应记录开发人工智能系统的理由及其目标。应考虑、记录和理解的一些因素包括：

a) 为什么要开发人工智能系统，例如，是受业务案例、客户要求还是政府政策的驱动；

b) 如何训练模型以及如何实现数据要求。

应明确人工智能系统的要求，并应贯穿整个人工智能系统的生命周期。如果所开发的人工智能系统无法按预期运行，或出现了可用于更改和改进要求的新信息，则应重新审视这些要求。例如，从财务角度看，开发人工智能系统可能变得不可行。

**其他信息**

ISO/IEC 5338:-[正在编写中。 出版时的阶段：ISO/IEC DIS 5338]提供了描述人工智能系统生命周期的流程。有关交互系统以人为本设计的更多信息，请参见 ISO 9241-210。

#### B.6.2.3 人工智能系统的设计和开发记录

**控制**

组织应根据组织目标、文件化的要求和规范标准，记录人工智能系统的设计和开发。

**实施指南**

人工智能系统有许多必要的设计选择，包括但不限于

- 机器学习方法（如监督式与非监督式）；
- 所使用的机器学习模型的学习算法和类型；
- 模型的训练方式和数据质量（见 B.7）；
- 模型的评估和改进；
- 硬件和软件组件；
- 在整个人工智能系统生命周期中考虑的安全威胁；人工智能系统特有的安全威胁包括数据中毒、模型窃取或模型反转攻击；
- 界面和输出展示；
- 人类如何与系统互动；
- 互操作性和可移植性方面的考虑。

在设计和开发之间可能会有多次迭代，但应保留各阶段的文档，并提供最终的系统架构文档。

**其他信息**

有关交互系统以人为本设计的更多信息，请参阅ISO 9241-210。

#### B.6.2.4 人工智能系统的验证和确认

**控制**

组织应定义并记录人工智能系统的验证和确认措施，并规定其使用标准。

**实施指南**

核查和验证措施可包括但不限于：

— 测试方法和工具；

— 测试数据的选择及其在预期使用领域的代表性；

— 发布标准要求。

组织应定义和记录评估标准，如但不限于：

— 评估人工智能系统组件和整个人工智能系统对个人和社会影响风险的计划；

— 评估计划可基于以下因素，例如：

  — 人工智能系统的可靠性和安全性要求，包括人工智能系统性能的可接受误差率；

  — 负责任的人工智能系统开发和使用目标，如 B.6.1.2 和 B.9.3 中的目标；

  — 操作因素，如数据质量、预期用途，包括每个操作因素的可接受范围；

  — 任何可能需要定义更严格操作要素的预期用途，包括不同的操作要素可接受范围或较低的误差率；

— 用于评估根据人工智能系统输出结果作出决定或受制于决定的相关利益方是否能够充分解释人工智能系统输出结果的方法、指导或衡量标准。应根据人工智能系统影响评估的结果确定评估频率；

— 任何可接受的因素，这些因素可能导致无法达到目标的最低性能水平，特别是在评估人工智能系统对个人和社会的影响时（例如，计算机视觉系统图像分辨率低或背景噪声影响语音识别系统）。还应记录处理这些因素导致的人工智能系统性能低下的机制。

人工智能系统应根据记录的评估标准进行评估。

如果人工智能系统不能满足形成成文信息的评估标准，特别是不能满足负责任的人工智能系统开发和使 用目标（见B.6.1.2，B.9.3），组织应重新考虑或管理人工智能系统预期用途的缺陷、其性能要求以及 组织如何有效处理对个人和社会的影响。

注：有关如何处理神经网络稳健性的更多信息，请参阅 ISO/IEC TR 24029-1。

### B.6.2.5 人工智能系统的部署

**控制**

组织应记录部署计划，并确保在部署前满足适当的要求。

**实施指南**

人工智能系统可以在不同的环境中开发，也可以在其他环境中部署（如在本地开发，使用云计算部署），组织在制定部署计划时应考虑到这些差异。他们还应考虑组件是否分开部署（例如，软件和模型可以独立部署）。此外，企业应制定一套发布和部署前必须满足的要求（有时称为“发布标准”）。这可能包括必须通过的验证和确认措施、必须达到的性能指标、必须完成的用户测试，以及必须获得的管理审批和签批。部署计划应考虑到相关利益方的观点和影响。

### B.6.2.6 人工智能系统的运行和监视

**控制**

组织应确定并记录人工智能系统持续运行的必要要素。至少应包括系统和性能监视、维修、更新和支持。

**实施指南**

运行和监测的每项最低活动都可以考虑各种因素。例如：

— 系统和性能监控可包括对一般错误和故障的监控，以及对系统是否按预期运行生产数据的监控。技术性能标准可包括解决问题或完成任务的成功率或信任率。其他标准可能与满足相关方的承诺或期望和需求有关，例如包括持续监控，以确保符合客户要求或适用的法律要求。

— 一些已部署的人工智能系统的性能会随着人工智能的发展而不断提高，其中生产数据和输出数据被用于进一步训练人工智能模型。在使用持续学习的情况下，组织应监控人工智能系统的性能，以确保其持续满足设计目标，并按照预期在生产数据上运行。

— 有些人工智能系统即使不使用持续学习，其性能也会发生变化，这通常是由于生产数据中的概念或数据漂移造成的。在这种情况下，监控可以确定是否需要重新训练，以确保人工智能系统继续实现其设计目标，并按预期在生产数据上运行。更多信息可参见 ISO/IEC 23053。

— 维修可包括对系统中的错误和故障做出响应。组织应制定应对和修复这些问题的流程。此外，随着系统的发展，或随着发现的关键问题减少，或由于外部发现的问题（如不符合客户期望或法律要求），可能有必要进行更新。应制定更新系统的流程，包括受影响的组件、更新时间表、向用户提供的关于更新内容的信息。

— 系统更新还可包括系统操作的变更、新的或修改后的预期用途，或系统功能的其他变更。组织应制定程序来处理操作变更，包括与用户沟通。

— 对系统的支持可以是内部的、外部的或两者兼有，这取决于组织的需求和系统的获取方 式。支持流程应考虑用户如何联系适当的帮助、如何报告问题和事件、支持服务级别协议和衡量标准。

— 如果人工智能系统的使用目的与设计目的不同，或使用的方式未在预料之中，则应考虑此类使用的适当性。

— 应识别与组织应用和开发的人工智能系统相关的人工智能特有的安全威胁。人工智能特有的安全威胁包括但不限于数据中毒、模型窃取和模型反转攻击。

**其他信息**

组织应考虑可能影响相关方的运行性能，并在设计和确定性能标准时考虑这一点。

运行中的人工智能系统的性能标准应根据所考虑的任务来确定，如分类、回归、排序、聚类或降维。

性能标准可包括错误率和处理持续时间等统计方面。对于每项标准，企业都应确定所有相关指标以及指标之间的相互依赖关系。对于每个指标，企业都应根据领域专家的建议和对相关方相对于现有非人工智能实践的期望分析等，考虑可接受的值。

例如，如 ISO/IEC TS 4213 所述，组织可根据其对假阳性和假阴性影响的评估，确定 F1 分数是一个合适的性能指标。然后，组织可以确定人工智能系统应达到的 F1 值。应评估这些问题是否可以通过现有措施来解决。如果不能，则应考虑更改现有措施，或定义其他措施来检测和处理这些问题。

组织应考虑运行中的非人工智能系统或流程的绩效，并在制定绩效标准时 将其作为潜在的相关背景。

此外，组织应确保用于评估人工智能系统的手段和程序，包括酌情选择和管 理评估数据，以提高根据规定标准评估其绩效的完整性和可靠性。

绩效评估方法的制定可以标准、度量和价值为基础。这些标准、指标和价值应能为评估中使用的数据量和程序类型以及评估人员的作用和专业技能提供依据。

绩效评估方法应尽可能地反映运行和使用的属性和特点，以确保评估结果的有用性和相关性。性能评估的某些方面可能需要有控制地引入错误或虚假数据或流程，以评估对性能的影响。

ISO/IEC 25059:-[编写中。出版时的阶段：ISO/IEC DIS 25059。]中的质量模型可用于定义性能标准。

### B.6.2.7 人工智能系统的技术记录

**控制**

组织应确定用户、合作伙伴、监管机构等各类相关方需要的人工智能系统技术记录，并以适当形式向他们提供技术文件。

**实施指南**

人工智能系统技术文件可包括但不限于以下内容：

— 人工智能系统的一般说明，包括其预期目的；
— 使用说明；
— 关于其部署和运行的技术假设（运行环境、相关软件和硬件能力、对数据的假设等）；
— 技术限制（如可接受的错误率、准确性、可靠性、稳健性）；
— 允许用户或操作员影响系统运行的监控能力和功能。

与所有人工智能系统生命周期阶段有关的文档要素（如 ISO/IEC 22989 所定义）可包括但不限于以下内容：

— 设计和系统架构说明；
— 在系统开发过程中做出的设计选择和采取的质量措施；
— 系统开发过程中使用的数据信息；
— 对数据质量所作的假设和采取的质量措施（如假设的统计分布）；
— 人工智能系统开发或运行过程中开展的管理活动（如风险管理）；
— 核实和验证记录；
— 人工智能系统运行时所作的改动；
— B.5 所述的影响评估文件。

组织应记录与负责任地运行人工智能系统有关的技术信息。这可包括但不限于：

— 记录管理先前未知故障的计划。例如，这可能包括需要说明人工智能系统的回滚计划、关闭人工智能系统的功能、更新流程或将人工智能系统的变更通知客户、用户等的计划、系统故障的最新信息以及如何缓解这些故障；
— 记录监测人工智能系统健康状况的程序（即人工智能系统按预期并在正常运行范围内运行，也称为可观察性）和处理人工智能系统故障的程序；
— 记录人工智能系统的标准操作程序，包括应监控哪些事件以及如何优先处理和审查事件日志。还可包括如何调查故障和预防未来故障；
— 记录负责人工智能系统操作的人员和负责系统使用的人员的职责，特别是在处理人工智能系统故障的影响或管理人工智能系统更新方面；
— 记录系统更新，如系统操作的变化、新的或修改后的预期用途，或系统功能的其他变化。
— 组织应制定适当的程序来解决操作上的变更，包括与用户的沟通和对变更类型的内部评估。

文件应及时更新并准确无误。文件应得到组织内相关管理层的批准。

作为用户文档的一部分提供时，应考虑表A.1中提供的控制。

### B.6.2.8 人工智能系统的事件记录日志

**控制**

组织应确定在人工智能系统生命周期的哪些阶段应启用事件日志记录，但至少应在人工智能系统使用时启用。

**实施指南**

组织应确保为其部署的人工智能系统记录日志，以自动收集和记录与操作期间发生的某些事件相关的事件日志。此类日志记录可包括但不限于以下内容：

— 人工智能系统功能的可追溯性，以确保人工智能系统按预期运行；

— 通过监控人工智能系统的运行，检测人工智能系统在预期运行条件之外的性能，这可能会导致生产数据的不良性能或对相关利益方造成影响。

人工智能系统事件日志可包括一些信息，如每次使用人工智能系统的时间和日期、人工智能系统运行的生产数据、超出人工智能系统预期运行范围的输出等。

事件日志的保存时间应根据人工智能系统的预期用途以及组织的数据保存政策和与数据保存相关的法律要求而定。

**其他信息**

某些人工智能系统（如生物识别系统）可能会根据管辖范围而有额外的日志记录要求。各组织应了解这些要求。

## B.7 人工智能系统的数据

### B.7.1 目标

确保组织了解人工智能系统中的数据在人工智能系统的整个生命周期中的应用、开发、提供或使用中的作用和影响。

### B.7.2 用于开发和增强人工智能系统的数据

**控制**

组织应定义、记录和实施与人工智能系统开发相关的数据管理流程。

**实施指南**

— 数据管理可以包括各种控制，包括但不限于：

— 由于使用数据而影响到隐私和安全，其中一些数据在本质上可能是敏感的；

— 依赖于数据的人工智能系统开发可能产生的安全和安全威胁；

— 透明度和可解释性方面，包括数据来源，以及如果系统需要透明度和可解释性，则提供解释数据如何用于确定人工智能系统的输出的能力；

— 训练数据与运行使用领域相比的代表性；

— 数据的准确性和完整性。

注：人工智能系统生命周期和数据管理概念的详细信息由 ISO/IEC 22989 提供。

### B.7.3 数据采集

**控制**

该组织应确定并记录有关在人工智能系统中使用的数据的采集和选择的细节。

**实施指南**

根据其人工智能系统的范围和使用情况，该组织可能需要来自不同来源的不同类别的数据。关于数据采集的详细信息可以包括：

— 人工智能系统所需的数据类别；

— 所需数据量；

— 数据源（例如：内部、购买、共享、开放数据、合成）；

— 数据源的特性（例如，静态、流式、收集、机器生成）；

— 数据主体的人口统计学特征（如已知或潜在的偏差或其他系统错误）；
— 事先处理数据（例如，以前的使用、符合隐私和安全要求）；
— 数据权利（如 PII、版权）；
— 相关的元数据（例如，数据标签和增强的细节）；
— 数据的来源。

**其他信息**

ISO/IEC 19944-1中的数据类别和数据使用的结构可用于记录有关数据采集和使用的细节。

### B.7.4 人工智能系统的数据质量

**控制**

组织应定义和记录数据质量要求，并确保用于开发和运行人工智能系统的数据符合这些要求。

**实施指南**

用于开发和操作人工智能系统的数据质量可能会对系统输出的有效性产生重大影响 。ISO/IEC 25024将数据质量定义为在规定条件下使用时，数据特征满足规定和隐含需求的程度。对于使用监督或半监督机器学习的人工智能系统，尽可能对训练、验证、测试和生产数据的质量进行定义、测量和改进是很重要的，组织宜确保数据能达到其预期目的。组织应考虑偏差对系统性能和系统公平性的影响，并对用于提高性能和公平性的模型和数据做出必要的调整，以便它们能够被用例所接受。

**其他信息**

关于数据质量的其他信息可在ISO/IEC 5259关于分析和ML数据质量的系列中获得。关于在人工智能系统中使用的数据中不同形式的偏差的其他信息可参考ISO/IEC TR 24027。

### B.7.5 数据来源

**控制**

该组织应该定义并记录一个流程，以记录其人工智能系统在数据和生命周期中使用的数据的来源。

**实施指南**

根据ISO 8000-2，数据来源的记录可以包括关于数据控制的创建、更新、转录、抽象、验证和传输的信息。此外，还可以在数据来源下考虑数据共享（不转移控制）和数据转换。根据数据来源、数据内容和使用上下文等因素，组织应该考虑是否需要采取验证数据来源的措施。

### B.7.6 数据准备

**控制**

组织应定义并记录其选择数据准备及要使用的数据准备方法的准则。

**实施指南**

在人工智能系统中使用的数据通常需要做好准备，以使其可用于给定的人工智能任务。例如，机器学习算法有时不能容忍缺失或不正确的条目、非正态分布和广泛变化的尺度。制备方法和变换可以用来提高数据的质量。如果未能正确准备好数据，可能会导致人工智能系统错误。在人工智能系统中使用的数据的常用准备方法和转换包括：

— 对数据的统计探索（例如分布、平均值、中位数、标准差、范围、分层、抽样）和统计元数据（例如数据文档计划（DDI）规范[27]）；
— 清理（即纠正条目，处理缺失的条目）；
— 估算（即用于填写缺失条目的方法）；
— 规范化；

— 缩放比例；
— 添加目标变量的标签；
— 编码（例如，将分类变量转换为数字）。

对于给定的人工智能任务，应该记录其选择特定数据准备方法和转换的标准，以及在人工智能任务中使用的特定方法和转换。

注：关于机器学习特有的数据准备的更多信息，请参见ISO/IEC 5259系列和ISO/IEC 23053。

## B.8 人工智能系统相关方的信息

### B.1 目标

确保有关各方掌握必要的信息，以了解和评估风险及其影响（正面和负面）。

### B.8.2 为用户提供的系统文件和信息

**控制**

组织应确定并向系统的用户提供必要的信息。

**实施指南**

关于人工智能系统的信息可以包括技术细节和说明，以及对用户正在与人工智能系统交互的一般通知，这取决于环境。这还可以包括系统本身，以及系统的潜在输出（例如，通知用户一个图像是由人工智能创建的）。

虽然人工智能系统可能很复杂，但用户能够理解他们何时与人工智能系统交互，以及该系统的工作原理是至关重要的。用户还需要了解其预期目的和预期用途，以及其对用户造成伤害或受益的可能性。一些系统文档必须用于更多的技术用途（例如系统管理员），组织应该了解不同感兴趣方的需求以及可理解性对他们意味着什么。这些信息也应该是可访问的，无论是在查找它的易用性方面，还是对于那些可能需要额外的可访问性功能的用户方面。

可提供给用户的信息包括但不限于：

— 系统的用途；
— 用户正在与一个人工智能系统进行交互；
— 如何与系统进行交互；
— 如何以及何时覆盖该系统；
— 系统运行的技术要求，包括所需的计算资源，以及系统的限制及其预期寿命；
— 对人类监督的需求；
— 关于准确性和性能的信息；
— 来自影响评估的相关信息，包括潜在的好处和危害，特别是如果它们适用于特定的情况或某些人口群体(见 B.5.2 和 B.5.4)；
— 对系统的好处的修正；
— 更新和更改系统的工作方式，以及任何必要的维护措施，包括其频率；
— 联系方式
— 供系统使用的教育材料。

组织用来决定是否提供什么信息的标准应该被记录下来。相关标准包括但不限于人工智能系统的预期用途和合理可预见的误用、用户的专业知识和人工智能系统的具体影响。

信息可以以多种方式提供给用户，包括有记录的使用说明、系统本身内置的警报和其他通知、网页上的信息等。根据组织使用的提供信息的方法，它应该验证用户是否可以访问这些信息，并且所提供的信息是否是完整的、最新的和准确的。

### B.8.3 外部报告

**控制**

组织应为相关方提供报告系统负面影响的能力。

**实施指南**

虽然应监测系统运行是否存在报告的问题和故障，但组织还应为用户或其他外部各方提供报告不利影响（例如，不公平）的能力。

### B.8.4 事故通报

**控制**

组织应确定并记录向系统用户沟通事故的计划。

**实施指南**

与人工智能系统有关的事件可以是人工智能系统本身的特定事件，也可以是与信息安全或隐私有关的事件（如数据泄露）。组织应根据系统运行的具体情况，了解其在通知用户和其他相关方有关事件方面的义务。例如，作为影响安全的产品一部分的人工智能组件发生事故时，其通知要求可能与其他类型的系统不同。法律要求（如合同）和监管活动可以适用，它们可以明确规定以下要求：

— 必须通报的事件类型；
— 通知的时限
— 是否必须通知有关当局以及通知哪些当局；
— 必须通报的详细信息。

组织可将人工智能的事件响应和报告活动整合到更广泛的组织事件管理活动中，但应注意与人工智能系统或人工智能系统单个组件相关的独特要求（例如，系统训练数据中的PII数据泄露可能会有与隐私相关的不同报告要求）。

**其他信息**

ISO/IEC 27001和ISO/IEC 27701分别提供了有关安全和隐私事件管理的更多详细信息。

### B.8.5 向有关各方提供信息

**控制**

该组织应确定并记录其向相关各方报告有关人工智能系统的信息的义务。

**实施指南**

在某些情况下，一个司法管辖区可能会要求与监管机构等权威机构共享有关该系统的信息。信息可以在适当的时间框架内报告给相关方，如客户或监管当局。共享的信息可以包括，例如：

— 技术系统文件，包括但不限于，用于训练、验证和测试的数据集，以及算法选择的理由和验证和验证记录；
— 与本系统相关的风险；
— 影响评估结果；
— 日志和其他系统记录。

本组织应了解其在这方面的义务，并确保与正确的当局分享适当的信息。此外，该组织知道与执法当局共享的信息有关的司法要求。

## B.9 人工智能系统的使用

### B.9.1 目标

确保组织负责任地和按照组织的政策使用人工智能系统。

### B.9.2 系统的文档和信息

**控制**

该组织应该定义并记录了负责任地使用人工智能系统的过程。

**实施指南**

根据其上下文，该组织在决定是否使用特定的人工智能系统时可以有许多考虑因素。无论人工智能系统是由组织本身开发的还是来自第三方，组织都应该清楚这些考虑是什么，并制定政策来解决这些问题。例如：

— 要求的批准；
— 成本（包括持续的监测和维护费用）；
— 批准的采购要求；
— 适用于本组织的法律要求。

如果组织已经接受了使用其他系统、资产等的策略，那么如果需要，可以合并这些策略。

### B.9.3 负责任地使用人工智能系统的目标

**控制**

该组织应确定并记录目标，以指导负责任地使用人工智能系统。

**实施指南**

在不同的环境下运作的组织可能对人工智能系统的负责任的发展有不同的期望和目标。根据其上下文，组织应确定其与可信使用相关的目标。一些目标包括：

— 公正；
— 可问责；
— 透明；
— 可解释性；
— 可靠；
— 安全；
— 鲁棒性和冗余性；
— 隐私和安全；
— 无障碍。

一旦定义，组织应该实施机制来实现组织内的目标。这可以包括确定一个第三方解决方案是否满足了组织的目标，或者一个内部开发的解决方案是否适用于预期的用途。该组织应确定在人工智能系统生命周期的哪个阶段应融入有意义的人类监督目标。这可以包括：

— 让人类审查员检查人工智能系统的输出，包括有权推翻人工智能系统做出的决策；
— 如果需要根据与预期部署人工智能系统相关的指示或其他文件使用人工智能系统，确保包括人工监督；

— 监控人工智能系统的性能，包括人工智能系统输出的准确性；
— 报告有关人工智能系统输出的关切及其对相关利害关系方的影响；
— 报告人工智能系统正确输出生产数据的性能或能力的变化；
— 考虑自动决策是否适合于一种负责任的方法来使用人工智能系统和人工智能系统的预期使用。

通过影响评估可以得知人类监督的必要性（见B.5）。参与与人工智能系统相关的人工监督活动的人员应被告知、培训和理解有关人工智能系统的指示和其他文件，以及它们为满足人类监督目标所执行的职责。在报告性能问题时，人工监督可以增强自动监控。

**其他信息**

附件C提供了管理风险的组织目标的例子，这可能有助于确定人工智能系统使用的目标。

### B.9.4 工智能系统的预期用途

**控制**

该组织应确保人工智能系统根据人工智能系统的预期使用及其附带文件进行使用。

**实施指南**

人工智能系统应根据说明和其他与人工智能系统相关的文档进行部署(见B.8.2)。部署可能需要特定的资源来支持部署，包括确保根据需要应用人工监督的需要（见B.9.3).为了可以使用人工智能系统，人工智能系统使用的数据与人工智能系统相关的文件一致， 以确保人工智能系统性能准确。

应监控人工智能系统的运行情况（见B.6.2.6)。如果根据相关说明正确部署的人工智能系统对利益相关方或组织的法律要求造成影响，组织应将其对影响的关切传达给组织内部的相关人员以及人工智能系统的任何第三方供方。

该组织应保存与人工智能系统的部署和操作相关的事件日志或其他文档，这些文档可用于证明人工智能系统正在按预期使用，或帮助沟通与预期使用相关的问题。事件日志和其他文档的保存时间取决于人工智能系统的预期用途、组织的数据保留政策以及数据保留的相关法律要求。

## B.10 第三方及客户关系

### B.10.1 目标

确保组织理解其责任并保持责任，当第三方参与到人工智能系统生命周期的任何阶段时，风险就会得到适当的分配。

### B.10.2 分配责任

**控制**

组织应确保在其人工智能系统生命周期内的责任被分配在组织、其合作伙伴、供方、客户和第三方之间。

**实施指南**

在人工智能系统的生命周期中，责任可以分为提供数据的各方、提供算法和模型的各方、开发或使用人工智能系统的各方，并对部分或所有相关各方负责。该组织应记录干预人工智能系统生命周期的所有各方及其角色，并确定其责任。

当该组织向第三方提供人工智能系统时，该组织应确保其采取负责任的方法来开发人工智能系统。请参阅 B.6 中的控制和实施指南。组织应能够向相关方及该组织供应人工智能系统的第三方提供 人工智能系统所需的文件（见B.6.2.7和B.8.2）。

当处理后的数据包括PII时，职责通常在PII处理器和控制器之间分配。ISO/IEC 29100提供了关于PII控制器和PII处理器的进一步信息。如果要保留PII的隐私，则应考虑诸如ISO/IEC 27701中所描述的实施指南。基于组织和人工智能系统的数据处理活动

当处理后的数据包括PII时，职责通常在PII处理者和控制者之间分配。ISO/IEC 29100提供了关于PII控制者和PII处理者的进一步信息。如果要保留PII的隐私，则应考虑诸如ISO/IEC 27701中所描述的控制。根据组织和人工智能系统对 PII 的数据处理活动以及组织在人工智能系统整个生命周期的应用和开发中的角色，组织可以承担 PII 控制者（或联合PII 控制者）、PII 处理者，或两者均承担。

### B.10.3 供方

**控制**

组织应建立一个过程，以确保其对供方提供的服务、产品或材料的使用与组织对负责任地开发和使用人工智能系统的方法相一致。

**实施指南**

组织开发或使用一个人工智能系统可以利用供方在多种方式，从采购数据集，机器学习算法或模型，或其他组件的系统，如软件库，整个人工智能系统本身使用自己或作为另一个产品的一部分（如车辆）。

组织在选择供方，确定对这些供方的要求以及确定对这些供方持续监控和评估的级别时，应 考虑不同类型的供方、其提供的产品以及其可能给整个系统和组织带来的不同程度的风险。

各组织应记录这些人工智能系统和人工智能系统组件如何集成到该组织开发或使用的人工智能系统中。

如果组织认为供方提供的人工智能系统或人工智能系统组件未按预期运行，或其对个人和社会造成的影响与组织对人工智能系统采取的负责任的方法不一致，则组织应要求供方采取纠正措施。组织可以决定与供方合作以实现这一目标。

组织应确保人工智能系统的供方提供与人工智能系统相关的适当且充分的文件（见 B.6.2.7和B.8.2）。

### B.10.4 客户

**控制**

该组织应确保其负责开发和使用人工智能系统的负责任的方法考虑到其客户的期望和需求。

**实施指南**

当组织提供与人工智能系统相关的产品或服务时（即当它本身是一个供方时），组织应该了解客户的期望和需求。这些要求可以以在设计或工程阶段对产品或服务本身的要求的形式出现，也可以以合同要求或一般使用协议的形式出现。一个组织可以有许多不同类型的客户关系，而这些客户关系都可以有不同的需求和期望。

组织应该特别了解供方和客户关系的复杂性质，并了解人工智能系统的供方的责任，以及客户 的责任，同时仍然满足需求和期望。

例如，该组织可以识别与客户使用其人工智能产品和服务相关的风险，并可以通过向其客户提供适当的信息来决定处理已识别的风险，这样客户就可以处理相应的风险。

作为适当信息的一个例子，当一个人工智能系统对某个使用域有效时，该域的限制应该传达给客户。见B.6.2.7和B.8.2。

# 附 录C
（资料性）
# 潜在的与人工智能相关的组织目标和风险来源

## C.1 通用

本附件概述了组织在管理风险时可以考虑的潜在组织目标、风险来源和说明。本附件并非打算详尽或适用于每个组织。组织应确定相关的目标和风险来源。ISO/IEC 23894提供了关于这些目标和风险来源及其与风险管理的关系的更详细的信息。对人工智能系统的评估，最初，定期和必要时，提供证据，表明人工智能系统正在根据组织目标进行评估。

## C.2 目标

### C.2.1 责任

人工智能的使用可以改变现有的问责框架。以前的行为追责制，他们的行动现在可以由人工智能系统支持或基于使用人工智能系统。

### C.2.2 人工智能专业知识

需要挑选出具有在评估、开发和部署人工智能系统方面的跨学科技能和专业知识的专业专家。

### C.2.3 培训和测试数据的可用性和质量

基于机器学习的人工智能系统需要训练、验证和测试数据， 以便训练和验证系统的预期行为。

### C.2.4 环境影响

人工智能的使用可能会对环境产生积极的和消极的影响。

### C.2.5 公平性

人工智能系统在自动决策方面的不当应用可能对特定的个人或个人群体不公平。

### C.2.6 可维护性

可维护性与组织处理人工智能系统的修改，以纠正缺陷或调整新的需求的能力有关。

### C.2.7 隐私性

滥用或披露个人和敏感数据（如健康记录）可能对数据主体产生有害影响。

### C.2.8 鲁棒性

在人工智能中，鲁棒性特性表明系统在新数据上能够（或无法）在新数据上具有与经过训练的数据或典型操作数据相当的性能。

### C.2.9 安全性

安全涉及到期望一个系统在规定的条件下不会导致人类的生命、健康、财产或环境受到威胁的状态。

### C.2.10 安全性

在人工智能的背景下，特别是关于基于机器学习方法的人工智能系统，应该考虑新的安全问题，超越经典的信息和系统安全问题

### C.2.11 透明度和可解释性

透明度既涉及操作人工智能系统的组织的特征，也涉及这些系统本身。可解释性与对影响人工智能系统结果的重要因素的解释有关，这些因素以一种人类可以理解的方式提供给相关方。

## C.3 风险来源

### C.3.1 环境复杂性

当人工智能系统在复杂的环境中运行时，情况的范围很广，性能可能存在不确定性，因此存在风险来源（例如， 自动驾驶的复杂环境）。

### C.3.2 缺乏透明度和可解释性

无法向有关各方提供适当的信息可能是一个风险的来源（即，在组织的可信赖性和问责制方面）。

### C.3.3 自动化程度

自动化水平可以对各种关注的领域产生影响，如安全、公平或安全。

### C.3.4 与机器学习相关的风险来源

用于机器学习的数据质量和用于收集数据的过程可能是一个风险来源，因为它可能影响诸如安全性和稳健性等目标（例如，由于数据质量问题或数据中毒）。

### C.3.5 硬件系统问题

与硬件相关的风险来源包括基于有缺陷的组件的硬件错误或在不同系统之间传输训练过的ML模型。

### C.3.6 系统生命周期问题

风险来源可能出现在整个人工智能系统的生命周期中（例如，设计缺陷、部署不足、缺乏维护、退役问题）。

### C.3.7 技术准备

风险源可能与未知因素（如系统限制和边界条件、性能漂移）导致的较不成熟技术有关，但也由于技术自满而导致的更成熟技术有关。

# 附 录D
（资料性）
# 跨领域或部门使用人工智能管理体系

## D.1 通用

该管理体系适用于任何开发、提供或使用人工智能系统的产品或服务的组织。因此，它可能适用于不同部门的各种产品和服务，这些产品和服务都符合对有关各方的义务、 良好实践、期望或合同承诺。行业的例子有：

— 健康
— 国防
— 交通工具
— 资金
— 就业
— 能源

为了负责任地开发和使用人工智能系统，可以考虑各种组织目标（可能的目标见附件C） 。本文档提供了来自人工智能技术特定视角的需求和指导。对于一些潜在的目标，存在着通用的或特定于部门的管理体系标准。这些管理体系标准通常从技术中立的角度来考虑目标，而人工智能管理体系则提供了人工智能技术的具体考虑。

人工智能系统不仅由使用人工智能技术的组件组成，而且可以使用各种技术和组件。因此，对于负责任的开发和使用人工智能系统，不仅应考虑到特定于人工智能的因素，而且应将该系统视为一个整体，包括所使用的所有技术和组件。即使是人工智能技术的特定部分，除了人工智能特定的考虑外，还应该考虑其他方面。例如， 由于人工智能是一种信息处理技术，信息安全一般也适用于它。

诸如安全、安全、隐私和环境影响等目标应该对人工智能和系统的其他组件进行全面管理，而不是单独管理。因此，将人工智能管理体系与相关主题的通用或特定部门的管理体系标准相结合，对于负责任的开发和使用人工智能系统至关重要。

## D.2 人工智能管理体系与其他管理体系标准的集成

— 当提供或使用人工智能时，组织可以有与其他管理体系标准主题的各个方面相关的目标或义务。这些可以包括安全性、隐私、质量，分别是 ISO/IEC 27001、ISO/IEC 27701 和 ISO 9001 中涵盖的主题。

在提供、使用或开发人工智能时，但不限于这些潜在的相关通用管理体系标准包括：

— ISO/IEC 27001：在大多数情况下，安全是通过人工智能系统实现组织目标的关键。一个组织追求安全目标的方式取决于其上下文和其自己的策略。如果一个组织确定需要实施一个人工智能管理体系，并以类似的彻底和系统的方式解决安全目标，它就可以实施一个符合 ISO/IEC 27001 的信息安全管理体系。鉴于 ISO/IEC 27001 和人工智能管理体系都使用了高级结构，它们易于集成使用，并为组织提供了极大的好处。在这种情况下，实施本文件（见 B.6.1.2）中（部分）与信息安全相关的控制的方式可以与该组织实施的 ISO/IEC 27001 相结合。

— ISO/IEC 27701：在许多上下文和应用领域中，PII 是由人工智能系统处理的。然后，该组织可以遵守适用的隐私义务和它自己的政策和目标。类似地，对于 ISO/IEC 27001，该组织可以从 ISO/IEC 27701 与人工智能管理体系的集成中获益。人工智能管理体系的隐私相关目标和控制(见 B.2.3 和 B.5.4)可与本组织实施的 ISO/IEC 27701 相集成。

— ISO 9001：对于许多组织来说，符合 ISO 9001 是一个关键的标志，表明他们以客户为导向，真正关注内部有效性。对 ISO 9001 进行的独立合格性评估促进了跨组织的业务，并激发了客户对产品或服务的信心。当涉及到人工智能技术时，当一个人工智能管理体系与 ISO 9001 联合实施时，客户对一个组织或人工智能系统的信心水平可以得到高度增强。在帮助该组织实现其目标方面，人工智能管理体系可以作为 ISO 9001 要求的补充（风险管理、软件开发、供应链一致性等）。。

除了上面提到的通用管理体系标准外，人工智能管理体系还可以与专门用于一个部门的管理体系联合使用。例如，ISO 22000和人工智能管理体系都与用于食品生产、准备和物流的人工智能系统相关。另一个例子是ISO 13485。人工智能管理体系的实现可以支持ISO 13485中与医疗设备软件相关的要求，或来自医疗行业的其他国际标准的要求，如IEC 62304。

# 参 考 文 献


[1] ISO 8000-2 数据质量 第 2 部分：术语

[2] ISO 9001 质量管理体系 要求

[3] ISO 9241-210 人与系统交互的人体工程学 第 210 部分： 以人为本的交互系统设计 交互系统

[4] ISO 13485 医疗器械 质量管理体系 监管要求

[5] ISO 22000 食品安全管理体系 对食品链中任何组织的要求

[6] IEC 62304 医疗器械软件 软件生命周期过程

[7] ISO/IEC 指南 51 安全方面 融入标准的指南

[8] ISO/IEC TS 4213 信息技术 人工智能 机器学习分类性能评估

[9] ISO/IEC 5259（所有部分） 分析和机器学习（ML）的数据质量

[10] ISO/IEC 5338 信息技术 人工智能 人工智能系统生命周期流程

[11] ISO/IEC 17065 合格评定 对产品、过程和服务认证机构的要求

[12] ISO/IEC 19944-1 云计算和分布式平台 数据流、数据类别和数据使用 第 1 部分：基础

[13] ISO/IEC 23053 使用机器学习（ML）的人工智能（AI）系统框架

[14] ISO/IEC 23894 信息技术 人工智能 风险管理指南

[15] ISO/IEC TR 24027 信息技术 人工智能（AI） 人工智能系统偏差和人工智能辅助决策

[16] ISO/IEC TR 24029-1 人工智能（AI） 神经网络鲁棒性评估 第 1 部分：概述

[17] ISO/IEC TR 24368 信息技术 人工智能 伦理和社会问题概述

[18] ISO/IEC 25024 系统和软件工程 系统和软件质量要求与评估（SQuaRE） 数据质量的测量

[19] ISO/IEC 25059 软件工程 系统和软件质量要求与评估（SQuaRE）人工智能系统质量模型

[20] ISO/IEC 27000:2018 信息技术 安全技术 信息安全管理体系--概述和术语

[21] ISO/IEC 27701 安全技术 针对隐私信息管理的 ISO/IEC 27001 和 ISO/IEC 27002 扩展 要求和准则

[22] ISO/IEC 27001 信息安全、网络安全和隐私保护 信息安全管理体系 要求

[23] ISO/IEC 29100 信息技术 安全技术 隐私框架

[24] ISO 31000:2018 风险管理 指南

[25] ISO 37002 举报管理体系 指南

[26] ISO/IEC 38500:2015 信息技术 组织的信息技术治理

[27] ISO/IEC 38507 信息技术 信息技术治理 组织使用人工智能的治理影响

[28] 生命周期 D.D.I.3.3，2020-04-15。数据文档倡议（DDI）联盟。[查看日期 2022-02-19]。见： https:// ddialliance .org/ Specification/ DDI -Lifecycle/ 3 .3/

[29] 风险框架 N.I.S.T.-A.I. 1.0, 2023-01-26. 美国国家技术研究院 (NIST) [查看日期 2023-04-17] https:// www .nist .gov/ itl/ ai -risk -management -framework