INTERNATIONAL STANDARD

ISO/IEC 23894

First edition
2023-02

Information technology — Artificial intelligence — Guidance on risk management

Technologies de l'information — Intelligence artificielle — Recommandations relatives au management du risque

Reference number
ISO/IEC 23894:2023(E)

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

Published in Switzerland

Contents

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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see https://patents.iec.ch).

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

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

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

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

Introduction

The purpose of risk management is the creation and protection of value. It improves performance, encourages innovation and supports the achievement of objectives.

This document is intended to be used in connection with ISO 31000:2018. Whenever this document extends the guidance given in ISO 31000:2018, an appropriate reference to the clauses of ISO 31000:2018 is made followed by AI-specific guidance, if applicable. To make the relationship between this document and ISO 31000:2018 more explicit, the clause structure of ISO 31000:2018 is mirrored in this document and amended by sub-clauses if needed.

This document is divided into three main parts:

Clause 4: Principles – This clause describes the underlying principles of risk management. The use of AI requires specific considerations with regard to some of these principles as described in ISO 31000:2018, Clause 4.

Clause 5: Framework – The purpose of the risk management framework is to assist the organization in integrating risk management into significant activities and functions. Aspects specific to the development, provisioning or offering, or use of AI systems are described in ISO 31000:2018, Clause 5.

Clause 6: Processes – Risk management processes involve the systematic application of policies, procedures and practices to the activities of communicating and consulting, establishing the context, and assessing, treating, monitoring, reviewing, recording and reporting risk. A specialization of such processes to AI is described in ISO 31000:2018, Clause 6.

Common AI-related objectives and risk sources are provided in Annex A and Annex B. Annex C provides an example mapping between the risk management processes and an AI system life cycle.

Information technology — Artificial intelligence — Guidance on risk management

1 Scope

This document provides guidance on how organizations that develop, produce, deploy or use products, systems and services that utilize artificial intelligence (AI) can manage risk specifically related to AI. The guidance also aims to assist organizations to integrate risk management into their AI-related activities and functions. It moreover describes processes for the effective implementation and integration of AI risk management.

The application of this guidance can be customized to any organization and its context.

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 31000:2018, *Risk management — Guidelines*

ISO Guide 73:2009, *Risk management — Vocabulary*

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 31000:2018, ISO/IEC 22989:2022 and ISO Guide 73:2009 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/

4 Principles of AI risk management

Risk management should address the needs of the organization using an integrated, structured and comprehensive approach. Guiding principles allow an organization to identify priorities and make decisions on how to manage the effects of uncertainty on its objectives. These principles apply to all organizational levels and objectives, whether strategic or operational.

Systems and processes usually deploy a combination of various technologies and functionalities in various environments, for specific use cases. Risk management should take into account the whole system, with all its technologies and functionalities, and its impact on the environment and stakeholders.

AI systems can introduce new or emergent risks for an organization, with positive or negative consequences on objectives, or changes in the likelihood of existing risks. They also can necessitate specific consideration by the organization. Additional guidance for the risk management principles, framework and processes an organization can implement is provided by this document.

NOTE Different International Standards have significantly different definitions of the word “risk.” In ISO 31000:2018 and related International Standards, “risk” involves a negative or positive deviation from the objectives. In some other International Standards, “risk” involves potential negative outcomes only, for example, safety-related concerns. This difference in focus can often cause confusion when trying to understand and properly implement a conformant risk management process.

ISO 31000:2018, Clause 4 defines several generic principles for risk management. In addition to guidance in ISO 31000:2018, Clause 4, Table 1 provides further guidance on how to apply such principles where necessary.

Table 1 — Risk management principles applied to artificial intelligence

	Principle	**Description (as given in ISO 31000:2018, Clause 4)**	**Implications for the development and use of AI**
a)	Integrated	Risk management is an integral part of all organizational activities.	No specific guidance beyond ISO 31000:2018.
b)	Structured and comprehensive	A structured and comprehensive approach to risk management contributes to consistent and comparable results.	No specific guidance beyond ISO 31000:2018.
c)	Customized	The risk management framework and process are customized and proportionate to the organization’s external and internal context related to its objectives.	No specific guidance beyond ISO 31000:2018.

Table 1 *(continued)*

	Principle	Description (as given in ISO 31000:2018, Clause 4)	Implications for the development and use of AI
d)	Inclusive	Appropriate and timely involvement of stakeholders enables their knowledge, views and perceptions to be considered. This results in improved awareness and informed risk management.	Because of the potentially far-reaching impacts of AI to stakeholders, it is important that organizations seek dialog with diverse internal and external groups, both to communicate harms and benefits, and to incorporate feedback and awareness into the risk management process. Organizations should also be aware that the use of AI systems can introduce additional stakeholders. The areas in which the knowledge, views and perceptions of stakeholders are of benefit include but are not restricted to: — Machine learning (ML) in particular often relies on the set of data appropriate to fulfil its objectives. Stakeholders can help in the identification of risks regarding the data collection, the processing operations, the source and type of data, and the use of the data for particular situations or where the data subjects can be outliers. — The complexity of AI technologies creates challenges related to transparency and explainability of AI systems. The diversity of AI technologies further drives these challenges due to characteristics such as multiple types of data modalities, AI model topologies, and transparency and reporting mechanisms that should be selected per stakeholders' needs. Stakeholders can help to identify the goals and describe the means for enhancing transparency and explainability of AI systems. In certain cases, these goals and means can be generalized across the use case and different stakeholders involved. In other cases, stakeholder segmentation of transparency frameworks and reporting mechanisms can be tailored to relevant personas (e.g. "regulators", "business owners", "model risk evaluators") per the use case. — Using AI systems for automated decision-making can directly affect internal and external stakeholders. Such stakeholders can provide their views and perceptions concerning, for example, where human oversight can be needed. Stakeholders can help in defining fairness criteria and also help to identify what constitutes bias in the working of the AI system.

Table 1 *(continued)*

	Principle	Description (as given in ISO 31000:2018, Clause 4)	Implications for the development and use of AI
e)	Dynamic	Risks can emerge, change or disappear as an organization's external and internal context changes. Risk management anticipates, detects, acknowledges and responds to those changes and events in an appropriate and timely manner.	To implement the guidance provided by ISO 31000:2018, organizations should establish organizational structures and measures to identify issues and opportunities related to emerging risks, trends, technologies, uses and actors related to AI systems. Dynamic risk management is particularly important for AI systems because: — The nature of AI systems is itself dynamic, due to continuous learning, refining, evaluating, and validating. Additionally, some AI systems have the ability to adapt and optimize based on this loop, creating dynamic changes on their own. — Customer expectations around AI systems are high and can potentially change quickly as the systems themselves do. — Legal and regulatory requirements related to AI are frequently changing and being updated. Integration with the management systems on quality, environmental footprints, safety, healthcare, legal or corporate responsibility, or any combination of these maintained by the organization, can also be considered to further understand and manage AI-related risks to the organization, individuals and societies.
f)	Best available information	The inputs to risk management are based on historical and current information, as well as on future expectations. Risk management explicitly takes into account any limitations and uncertainties associated with such information and expectations. Information should be timely, clear and available to relevant stakeholders.	Taking into account the expectation that AI affects the way individuals interact with and react to technology, it is advisable for organizations engaged in the development of AI systems to keep track of relevant information available regarding the further uses of the AI systems that they developed, while users of AI systems can maintain records of the uses of those systems throughout the entire lifetime of the AI system. As AI is an emerging technology and constantly evolving, historical information can be limited, and future expectations can change quickly. Organizations should take this into account. The internal use of AI systems should be considered, if any. Tracking the use of AI systems by customers and external users can be limited by intellectual property, contractual or market-specific restrictions. Such restrictions should be captured in the AI risk management process and updated when business conditions warrant revisiting.

Table 1 *(continued)*

	Principle	Description (as given in ISO 31000:2018, Clause 4)	Implications for the development and use of AI
g)	Human and cultural factors	Human behaviour and culture significantly influence all aspects of risk management at each level and stage.	Organizations engaged in the design, development or deployment of AI systems, or any combination of these, should monitor the human and cultural landscape in which they are situated. Organizations should focus on identifying how AI systems or components interact with pre-existing societal patterns that can lead to impacts on equitable outcomes, privacy, freedom of expression, fairness, safety, security, employment, the environment, and human rights broadly.
h)	Continual improvement	Risk management is continually improved through learning and experience.	The identification of previously unknown risks related to the use of AI systems should be considered in the continual improvement process. Organizations engaged in the design, development or deployment of AI systems or system components, or any combination of these, should monitor the AI ecosystem for Performance successes, shortcomings and lessons learned, and maintain awareness of new AI research findings and techniques (opportunities for improvement).

5 Framework

5.1 General

The purpose of the risk management framework is to assist the organization in integrating risk management into significant activities and functions. The guidance provided in ISO 31000:2018, 5.1 applies.

Risk management involves assembling relevant information for an organization to make decisions and address risk. While the governing body defines the overall risk appetite and organizational objectives, it delegates the decision-making process of identifying, assessing and treating risk to management within the organization.

ISO/IEC 38507[1] describes additional governance considerations for the organization regarding the development, purchase or use of an AI system. Such considerations include new opportunities, potential changes to the risk appetite as well as new governance policies to ensure the responsible use of AI by the organization. It can be used in combination with the risk management processes described in this document to help guide the dynamic and iterative organizational integration described in ISO 31000:2018, 5.2.

5.2 Leadership and commitment

The guidance provided in ISO 31000:2018, 5.2 applies.

In addition to the guidance provided in ISO 31000:2018, 5.2 the following applies:

Due to the particular importance of trust and accountability related to the development and use of AI, top management should consider how policies and statements related to AI risks and risk management are communicated to stakeholders. Demonstrating this level of leadership and commitment can be critical for ensuring that stakeholders have confidence that AI is being developed and used responsibly.

The organization should therefore consider issuing statements related to its commitment to AI risk management to increase confidence of their stakeholders on their use of AI.

Top management should also be aware of the specialized resources that can be needed to manage AI risk, and allocate those resources appropriately.

5.3 Integration

The guidance provided in ISO 31000:2018, 5.3 applies.

5.4 Design

5.4.1 Understanding the organization and its context

The guidance provided in ISO 31000:2018, 5.4.1 applies.

In addition to guidance provided in ISO 31000:2018, 5.4.1, Table 2 lists additional factors to consider when understanding the external context of an organization.

Table 2 — Consideration when establishing the external context of an organization

Generic guidance provided by ISO 31000:2018, 5.4.1	**Additional guidance for organizations engaged in AI**
Organizations should consider at least the following elements of their external context:	Organizations should additionally consider, but not exclusively, the following elements:
— The social, cultural, political, legal, regulatory, financial, technological, economic and environmental factors, whether international, national, regional or local;	— Relevant legal requirements, including those specifically relating to AI. — Guidelines on ethical use and design of AI and automated systems issued by government-related groups, regulators, standardization bodies, civil society, academia and industry associations. — Domain-specific guidelines and frameworks related to AI.
— Key drivers and trends affecting the objectives of the organization;	— Technology trends and advancements in the various areas of AI. — Societal and political implications of the deployment of AI systems, including guidance from social sciences.
— External stakeholders' relationships, perceptions, values, needs and expectations;	— Stakeholder perceptions, which can be affected by issues such as lack of transparency (also referred to as opaqueness) of AI systems or biased AI systems. — Stakeholder expectations on the availability of specific AI-based solutions and the means by which the AI models are made available (e.g. through a user interface, software development kit).
— Contractual relationships and commitments;	— How the use of AI, especially AI systems using continuous learning, can affect the ability of the organization to meet contractual obligations and guarantees. Consequently, organizations should carefully consider the scope of relevant contracts. — Contractual relationships during the design and production of AI systems and services. For example, ownership and usage rights of test and training data should be considered when provided by third parties.
— The complexity of networks and dependencies;	— The use of AI can increase the complexity of networks and dependencies.