NORME INTERNATIONALE

ISO/IEC 23894

Première édition
2023-02

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

Information technology — Artificial intelligence — Guidance on risk management

Numéro de référence
ISO/IEC 23894:2023(F)

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
Fax: +41 22 749 09 47
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos

L'ISO (Organisation internationale de normalisation) et l'IEC (Commission électrotechnique internationale) forment le système spécialisé de la normalisation mondiale. Les organismes nationaux membres de l'ISO ou de l'IEC participent au développement de Normes internationales par l'intermédiaire des comités techniques créés par l'organisation concernée afin de s'occuper des domaines particuliers de l'activité technique. Les comités techniques de l'ISO et de l'IEC collaborent dans des domaines d'intérêt commun. D'autres organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO et l'IEC, participent également aux travaux.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ~~ISO~~. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives ou www.iec.ch/members_experts/refdocs).

L'attention est attirée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO et l'IEC ne sauraient être tenues pour responsables de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets) ou dans la liste des déclarations de brevets reçues par l'IEC (voir https://patents.iec.ch).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir www.iso.org/iso/avant-propos. Pour l'IEC, voir www.iec.ch/understanding-standards.

Le présent document a été élaboré par le comité technique mixte ISO/IEC JTC 1, *Technologies de l'information*, sous-comité SC 42, *Intelligence artificielle.*

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve à l'adresse www.iso.org/members.html et www.iec.ch/national-committees.

Introduction

La finalité du management du risque est la création et la préservation de la valeur. Il améliore la performance, favorise l'innovation et contribue à l'atteinte des objectifs.

Le présent document est destiné à être utilisé conjointement avec l'ISO 31000:2018. Lorsque le présent document étend les recommandations données dans l'ISO 31000:2018, une référence appropriée aux articles de l'ISO 31000:2018 est donnée et suivie de recommandations spécifiques à l'IA, le cas échéant. Pour rendre la relation entre le présent document et l'ISO 31000:2018 plus explicite, la structure des articles de l'ISO 31000:2018 est reflétée dans le présent document et amendée par des paragraphes si nécessaire.

Le présent document est divisé en trois parties principales:

Article 4: Principes – Le présent article décrit les principes sous-jacents du management du risque. L'utilisation de l'IA exige des considérations spécifiques eu égard à certains de ces principes, tels que décrits dans l'Article 4 de l'ISO 31000:2018.

Article 5: Cadre organisationnel – La finalité du cadre organisationnel de management du risque est d'aider l'organisme à intégrer le management du risque dans les activités et les fonctions significatives. Des aspects propres au développement, à la mise à disposition, à l'offre ou à l'utilisation de systèmes d'IA sont décrits à l'Article 5 de l'ISO 31000:2018.

Article 6: Processus – Le processus de management du risque implique l'application systématique de politiques, de procédures et de pratiques aux activités de communication et de consultation, d'établissement du contexte et d'appréciation, de traitement, de suivi, de revue, d'enregistrement et de compte rendu du risque. Une spécialisation desdits processus relatifs à l'IA est décrite dans l'Article 6 de l'ISO 31000:2018.

Des objectifs et sources de risques courants liés à l'IA sont fournis à l'Annexe A et l'Annexe B. L'Annexe C fournit un exemple de cartographie entre le processus de management du risque et le cycle de vie d'un système d'IA.

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

1 Domaine d'application

Le présent document fournit des recommandations relatives à la manière dont les organismes qui développent, produisent, déploient ou utilisent des produits, systèmes et services faisant appel à l'intelligence artificielle (IA) peuvent gérer le risque spécifiquement lié à l'IA. Ces recommandations visent également à assister les organismes dans l'intégration du management du risque à leurs activités et fonctions liées à l'IA. Le présent document décrit en outre des processus pour la mise en œuvre et l'intégration efficaces du management du risque lié à l'IA.

L'application de ces recommandations peut être adaptée à n'importe quel organisme et à son contexte.

2 Références normatives

Les documents suivants sont cités dans le texte de sorte qu'ils constituent, pour tout ou partie de leur contenu, des exigences du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

ISO 31000:2018, *Management du risque — Lignes directrices*

Guide ISO 73:2009, *Management du risque — Vocabulaire*

ISO/IEC 22989:2022, *Technologies de l'information — Intelligence artificielle — Concepts et terminologie relatifs à l'intelligence artificielle*

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions donnés dans l'ISO 31000:2018, l'ISO/IEC 22989:2022 et le Guide ISO 73:2009 s'appliquent.

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes:

— ISO Online browsing platform: disponible à l'adresse https://www.iso.org/obp;

— IEC Electropedia: disponible à l'adresse https://www.electropedia.org/.

4 Principes du management du risque lié à l'IA

Il convient que le management du risque aborde les besoins de l'organisme à l'aide d'une approche intégrée, structurée et globale. Des principes directeurs permettent à un organisme d'identifier des priorités et de prendre des décisions sur la façon de gérer les effets de l'incertitude sur ses objectifs. Ces principes s'appliquent à tous les niveaux et objectifs organisationnels, qu'ils soient stratégiques ou opérationnels.

Les systèmes et processus déploient habituellement une combinaison de plusieurs technologies et fonctionnalités dans divers environnements, pour des cas d'utilisation spécifiques. Il convient que le management du risque tienne compte du système dans sa globalité, avec toutes ses technologies, ses fonctionnalités, son impact sur l'environnement et ses parties prenantes.

Les systèmes d'IA peuvent introduire des risques nouveaux ou émergents pour un organisme, avec des conséquences positives ou négatives sur ses objectifs, ou une modification de la vraisemblance des risques existants. Ils peuvent également nécessiter une considération spécifique de la part de l'organisme. Des recommandations additionnelles pour les principes de management du risque, le cadre organisationnel et les processus qu'un organisme peut mettre en œuvre sont fournies dans le présent document.

NOTE La définition du mot «risque» peut différer significativement dans les différentes Normes internationales. Dans l'ISO 31000:2018 et les normes associées, «risque» implique un écart positif ou négatif par rapport à des objectifs. Dans d'autres Normes internationales, «risque» implique des conséquences potentiellement négatives uniquement, par exemple, des préoccupations relatives à la sûreté. Cette différence de perspective peut fréquemment porter à confusion lorsqu'il s'agit d'essayer de comprendre et de mettre en œuvre correctement un processus de management du risque conforme.

L'Article 4 de l'ISO 31000:2018 définit plusieurs principes génériques pour le management du risque. Outre les recommandations données dans l'ISO 31000:2018, Article 4, le Tableau 1 fournit des recommandations supplémentaires relatives à la façon d'appliquer ces principes, le cas échéant.

Tableau 1 — Principes du management du risque appliqués à l'intelligence artificielle

	Principe	Description (telle que donnée dans l'ISO 31000:2018, Article 4)	Implications pour le développement et l'utilisation de l'IA
a)	Intégré	Le management du risque est intégré à toutes les activités de l'organisme.	Aucune recommandation spécifique au-delà de l'ISO 31000:2018.
b)	Structuré et global	Une approche structurée et globale du management du risque contribue à la cohérence de résultats qui peuvent être comparés.	Aucune recommandation spécifique au-delà de l'ISO 31000:2018.
c)	Adapté	Le cadre organisationnel et le processus de management du risque sont adaptés et proportionnés au contexte externe et interne de l'organisme aussi bien qu'à ses objectifs.	Aucune recommandation spécifique au-delà de l'ISO 31000:2018.
d)	Inclusif	L'implication appropriée et au moment opportun des parties prenantes permet de prendre en compte leurs connaissances, leurs opinions et leur perception. Cela conduit à un management du risque mieux éclairé et plus pertinent.	En raison de l'importance possible des impacts de l'IA sur les parties prenantes, il est important que les organismes cherchent à établir un dialogue avec divers groupes internes et externes, afin de communiquer sur les avantages et inconvénients, et d'intégrer leur retour d'information et leur perception au processus de management du risque. Il convient que les organismes soient également conscients que l'utilisation de systèmes d'IA peut impliquer l'introduction de parties prenantes additionnelles. Les domaines dans lesquels les connaissances, les opinions et les perceptions des parties prenantes sont utiles incluent, sans s'y limiter: — En particulier, l'apprentissage automatique (ML, de l'anglais «machine learning») repose souvent sur un ensemble approprié de données pour remplir ses objectifs. Les parties prenantes peuvent ainsi aider à identifier les risques relatifs à la collecte des données, aux opérations de traitement, à la source et au type de données ainsi qu'à l'utilisation des données dans des circonstances particulières ou lorsque les personnes concernées peuvent constituer des exceptions.

Tableau 1 *(suite)*

	Principe	Description (telle que donnée dans l'ISO 31000:2018, Article 4)	Implications pour le développement et l'utilisation de l'IA
			— La complexité des technologies d'IA donne naissance à des défis liés à la transparence et à l'explicabilité des systèmes d'IA. La diversité des technologies d'IA renforce ces défis en raison de caractéristiques telles que les différents types de modalités de données, les topologies des modèles d'IA, ainsi que les mécanismes de transparence et d'établissement de rapports qu'il convient de sélectionner en fonction des besoins des parties prenantes. Les parties prenantes peuvent contribuer à l'identification des buts et à la description des moyens d'améliorer la transparence et à l'explicabilité des systèmes d'IA. Dans certains cas, ces buts et moyens peuvent être généralisés sur l'ensemble du cas d'utilisation et pour les différentes parties prenantes impliquées. Dans d'autres cas, la segmentation des parties prenantes en ce qui concerne les cadres de transparence et des mécanismes de création de rapports peut être adaptée en fonction des différents acteurs (par exemple, «organismes de réglementation», «propriétaires d'entreprise», «évaluateurs de risques du modèle») et du cas d'utilisation. — L'utilisation de systèmes d'IA pour la prise automatisée de décisions peut directement affecter les parties prenantes internes et externes. Ces dernières peuvent fournir leurs opinions et leurs perceptions concernant, par exemple, les situations pouvant nécessiter une supervision humaine. Les parties prenantes peuvent aider à définir des critères d'équité, de même qu'à identifier ce qui constitue un biais dans le fonctionnement du système d'IA.
e)	Dynamique	Des risques peuvent surgir, être modifiés ou disparaître lorsque le contexte externe et interne d'un organisme change. Le management du risque anticipe, détecte, reconnaît et réagit à ces changements et événements en temps voulu et de manière appropriée.	Pour mettre en œuvre les recommandations fournies par l'ISO 31000:2018, il convient que les organismes établissent des structures et mesures organisationnelles afin d'identifier les défis et opportunités liés aux risques émergents, aux tendances, aux technologies, aux utilisations et aux acteurs liés aux systèmes d'IA. Le management dynamique du risque est particulièrement important pour les systèmes d'IA, car:

Tableau 1 *(suite)*

	Principe	Description (telle que donnée dans l'ISO 31000:2018, Article 4)	Implications pour le développement et l'utilisation de l'IA
			— La nature des systèmes d'IA est elle-même dynamique, en raison du caractère continu de l'apprentissage, de l'affinage, de l'évaluation et de la validation. En outre, certains systèmes d'IA sont en mesure de s'adapter et de s'optimiser grâce à cette boucle, ce qui entraîne des évolutions dynamiques en tant que telles. — Les attentes des clients concernant les systèmes d'IA sont élevées et peuvent évoluer rapidement, tout comme les systèmes eux-mêmes. — Les exigences juridiques et réglementaires relatives à l'IA font fréquemment l'objet de modifications et de mises à jour. L'intégration à des systèmes de management relatifs à la qualité, à l'empreinte environnementale, à la sûreté, aux soins de santé, à la responsabilité juridique ou sociétale, ou à plusieurs de ces systèmes maintenus par l'organisme peut également être prise en compte pour mieux comprendre et gérer les risques liés à l'IA pour l'organisme, les individus et les sociétés.
f)	Meilleure information disponible	Les données d'entrée du management du risque sont fondées sur des informations historiques et actuelles ainsi que sur les attentes futures. Le management du risque tient compte explicitement de toutes limites et incertitudes associées à ces informations et attentes. Il convient que les informations soient disponibles à temps, claires et accessibles aux parties prenantes pertinentes.	Compte tenu de l'attente concernant l'impact probable de l'IA sur la façon dont les individus interagissent avec la technologie, et y réagissent, il est recommandé pour les organismes engagés dans le développement de systèmes d'IA de garder une trace des informations pertinentes disponibles concernant les autres utilisations des systèmes d'IA qu'ils ont développés, et pour les utilisateurs de systèmes d'IA de tenir à jour des enregistrements des utilisations de ces systèmes tout au long de l'intégralité du cycle de vie du système d'IA. L'IA étant une technologie émergente et en constante évolution, les informations historiques peuvent être limitées et les attentes futures peuvent évoluer rapidement. Il convient que les organismes en tiennent compte. Il convient de tenir compte de l'utilisation en interne de systèmes d'IA, le cas échéant. Assurer le suivi de l'utilisation des systèmes d'IA par des clients et des utilisateurs externes peut être limité par des restrictions de propriété intellectuelle, contractuelles ou propres au marché. Il convient que de telles restrictions soient prises en compte dans le processus de management du risque lié à l'IA et mises à jour lorsque les conditions d'activité le justifient.

Tableau 1 *(suite)*

	Principe	Description (telle que donnée dans l'ISO 31000:2018, Article 4)	Implications pour le développement et l'utilisation de l'IA
g)	Facteurs humains et culturels	Le comportement humain et la culture influent de manière significative sur tous les aspects du management du risque à chaque niveau et à chaque étape.	Il convient que les organismes engagés dans la conception, le développement ou le déploiement de systèmes d'IA, ou dans plusieurs de ces mécanismes, surveillent le paysage humain et culturel dans lequel ils évoluent. Il convient que les organismes se concentrent sur l'identification de la façon dont les systèmes ou composants d'IA interagissent avec les modèles sociétaux existants, eu égard à l'obtention de conséquences équitables, au respect de la vie privée, à la liberté d'expression, à l'équité, à la sûreté, à la sécurité, à l'emploi, à l'environnement et aux droits de l'homme dans leur ensemble.
h)	Amélioration continue	Le management du risque est amélioré en continu par l'apprentissage et l'expérience.	Il convient que l'identification des risques auparavant inconnus liés à l'utilisation de systèmes d'IA soit prise en compte dans le processus d'amélioration continue. Il convient que les organismes engagés dans la conception, le développement ou le déploiement de systèmes ou de composants de systèmes d'IA, ou d'une combinaison de ces derniers, surveillent l'écosystème d'IA eu égard aux réussites relatives à la performance, aux lacunes et aux enseignements tirés, et restent conscients des nouvelles conclusions de recherche et techniques d'IA (opportunités d'amélioration).

5 Cadre organisationnel

5.1 Généralités

La finalité du cadre organisationnel de management du risque est d'aider l'organisme à intégrer le management du risque dans les activités et les fonctions significatives. Les recommandations données dans l'ISO 31000:2018, 5.1, s'appliquent.

Le management du risque implique de réunir des informations pertinentes pour qu'un organisme prenne des décisions et aborde le risque. Tandis que l'organe de gouvernance définit l'appétit global pour le risque et les objectifs organisationnels, il délègue le processus de prise de décision relatif à l'identification, à l'appréciation et au traitement du risque à la direction au sein de l'organisme.

L'ISO/IEC 38507[1] décrit des considérations additionnelles relatives à la gouvernance pour l'organisme en ce qui concerne le développement, l'achat ou l'utilisation d'un système d'IA. De telles considérations incluent les nouvelles opportunités, les modifications potentielles relatives à l'appétit pour le risque ainsi que les nouvelles politiques de gouvernance pour garantir l'utilisation responsable de l'IA par l'organisme. Elles peuvent être utilisées conjointement avec le processus de management du risque décrit dans le présent document pour contribuer à guider l'intégration organisationnelle dynamique et itérative décrite dans l'ISO 31000:2018, 5.2.

5.2 Leadership et engagement

Les recommandations données dans l'ISO 31000:2018, 5.2, s'appliquent.

Outre les recommandations fournies dans l'ISO 31000:2018, 5.2, les éléments suivants s'appliquent:

En raison de l'importance particulière de la confiance et de la redevabilité liées au développement et à l'utilisation de l'IA, il convient que la direction tienne compte de la façon dont les politiques et énoncés

relatifs aux risques de l'IA et au management du risque sont communiqués aux parties prenantes. Faire preuve de ce niveau de leadership et d'engagement peut être critique pour garantir que les parties prenantes ont confiance dans le fait que l'IA est développée et utilisée de manière responsable.

Il convient que l'organisme considère par conséquent la publication d'énoncés relatifs à son engagement vis-à-vis du management du risque lié à l'IA pour accroître la confiance de ses parties prenantes dans son utilisation de l'IA.

Il convient que la direction soit également consciente des ressources spécialisées qui peuvent être nécessaires pour gérer le risque lié à l'IA et qu'elle affecte ces ressources de manière appropriée.

5.3 Intégration

Les recommandations données dans l'ISO 31000:2018, 5.3, s'appliquent.

5.4 Conception

5.4.1 Compréhension de l'organisme et de son contexte

Les recommandations données dans l'ISO 31000:2018, 5.4.1, s'appliquent.

Outre les recommandations données dans l'ISO 31000:2018, 5.4.1, le Tableau 2 énumère les facteurs additionnels à prendre en compte lors de la compréhension du contexte externe d'un organisme.

Tableau 2 — Considération lors de l'établissement du contexte externe d'un organisme

Recommandations génériques données par l'ISO 31000:2018, 5.4.1 Il convient que les organismes considèrent au moins les éléments suivants de leur contexte externe:	**Recommandations additionnelles pour les organismes engagés dans l'IA** Il convient que les organismes considèrent également, sans s'y limiter, les éléments suivants:
— facteurs sociaux, culturels, politiques, légaux, réglementaires, financiers, technologiques, économiques et environnementaux, au niveau international, national, régional ou local;	— exigences légales pertinentes, y compris celles qui sont spécifiquement relatives à l'IA; — lignes directrices sur l'utilisation et la conception éthiques de systèmes automatisés et d'IA publiées par des groupes liés au gouvernement, des organismes de régulation, des organismes de normalisation, des sociétés civiles, des académies et des associations industrielles; — lignes directrices et cadres organisationnels spécifiques au domaine pour l'IA;
— moteurs et tendances clés ayant une incidence sur les objectifs de l'organisme;	— tendances et avancées technologiques dans les différents domaines de l'IA; — implications sociétales et politiques du déploiement de systèmes d'IA, y compris les recommandations issues des sciences sociales;
— relations avec les parties prenantes externes, leurs perceptions, leurs valeurs, leurs besoins et leurs attentes;	— perceptions des parties prenantes, qui peuvent être affectées par des enjeux tels que le manque de transparence (également opacité) des systèmes d'IA ou des systèmes d'IA biaisés; — attentes des parties prenantes sur la disponibilité de solutions spécifiques basées sur l'IA et les moyens par lesquels les modèles d'IA sont mis à disposition (par exemple, par l'interface utilisateur ou par un kit de développement logiciel);