


	DIN EN ISO/IEC 23894	
ICS 35.240.01	<div>Einsprüche bis 2023-12-06</div> <div>Entwurf</div> <div><p>Informationstechnik – Künstliche Intelligenz – Leitlinien für Risikomanagement (ISO/IEC 23894:2023); Deutsche und Englische Fassung prEN ISO 23894:2023</p><p>Information technology – Artificial intelligence – Guidance on risk management (ISO/IEC 23894:2023); German and English version prEN ISO 23894:2023</p><p>Technologies de l’information – Intelligence artificielle – Recommandations relatives au management du risque (ISO/IEC 23894:2023); Version allemande et anglaise prEN ISO 23894:2023</p><p>Anwendungswarnvermerk</p><p>Dieser Norm-Entwurf mit Erscheinungsdatum 2023-10-06 wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.</p><p>Weil die beabsichtigte Norm von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfs besonders zu vereinbaren.</p><p>Stellungnahmen werden erbeten</p><ul style="list-style-type: none">— vorzugsweise online im Norm-Entwurfs-Portal von DIN unter www.din.de/go/entwuerfe bzw. für Norm-Entwürfe der DKE auch im Norm-Entwurfs-Portal der DKE unter www.entwuerfe.normenbibliothek.de, sofern dort wiedergegeben;— oder als Datei per E-Mail an nia@din.de möglichst in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/go/stellungnahmen-norm-entwuerfe oder für Stellungnahmen zu Norm-Entwürfen der DKE unter www.dke.de/stellungnahme abgerufen werden;— oder in Papierform an den DIN-Normenausschuss Informationstechnik und Anwendungen (NIA), 10772 Berlin oder Am DIN-Platz, Burggrafenstr. 6, 10787 Berlin.<p>Es wird gebeten, mit den Kommentaren zu diesem Entwurf jegliche relevanten Patentrechte, die bekannt sind, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.</p></div> <div>Gesamtumfang 70 Seiten</div> <div>DIN-Normenausschuss Informationstechnik und Anwendungen (NIA)</div>	

Nationales Vorwort

Der Text von ISO/IEC 23894:2023 wurde vom Technischen Komitee ISO/IEC JTC 1 „Information technology“ der Internationalen Organisation für Normung (ISO) erarbeitet und als prEN ISO/IEC 23894:2023 durch das Technische Komitee CEN/CLC/JTC 21 „Künstliche Intelligenz“ übernommen, dessen Sekretariat von DS (Dänemark) gehalten wird.

Das zuständige nationale Normungsgremium ist der Gemeinschaftsarbeitsausschuss NA 043-01-42 GA „DIN/DKE Gemeinschaftsarbeitsausschuss Künstliche Intelligenz“ im DIN-Normenausschuss Informationstechnik und Anwendungen (NIA).

Um Zweifelsfälle in der Übersetzung auszuschließen, ist die englische Originalfassung beigelegt. Die Nutzungsbedingungen für den deutschen Text des Norm-Entwurfes gelten gleichermaßen auch für den englischen Text.

Für die in diesem Dokument zitierten Dokumente wird im Folgenden auf die entsprechenden deutschen Dokumente hingewiesen:

ISO 31000:2018	siehe	DIN ISO 31000:2018-10
ISO 26000:2010	siehe	DIN EN ISO 26000:2021-04
ISO/IEC 29134:2017	siehe	DIN EN ISO/IEC 29134:2020-09

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11

Nationaler Anhang NA (informativ)

Literaturhinweise

DIN ISO 31000:2018-10, *Risikomanagement — Leitlinien (ISO 31000:2018)*

DIN EN ISO 26000:2021-04, *Leitfaden zur gesellschaftlichen Verantwortung (ISO 26000:2010); Deutsche Fassung EN ISO 26000:2020*

DIN EN ISO/IEC 29134:2020-09, *Informationstechnik — Sicherheitsverfahren — Leitlinien für die Datenschutz-Folgenabschätzung (ISO/IEC 29134:2017); Deutsche Fassung EN ISO/IEC 29134:2020*

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11

- Leerseite -

- Entwurf -

Oktober 2023

prEN ISO/IEC 23894

**Informationstechnik – Künstliche Intelligenz – Leitlinien für Risikomanagement
(ISO/IEC 23894:2023)**

Information technology – Artificial intelligence – Guidance on risk management (ISO/IEC 23894:2023)

Technologies de l'information – Intelligence artificielle – Recommandations relatives au management du risque (ISO/IEC 23894:2023)

Nachfolgedokument: DIN EN ISO/IEC 23894 (in Vorbereitung/in preparation/en préparation) (DE30101214)

Inhalt

	Seite
Europäisches Vorwort	4
Vorwort	5
Einleitung	6
1 Anwendungsbereich	7
2 Normative Verweisungen	7
3 Begriffe	7
4 Grundsätze des KI-Risikomanagements	7
5 Rahmenwerk	11
5.1 Allgemeines	11
5.2 Führung und Verpflichtung	11
5.3 Integration	12
5.4 Gestaltung	12
5.4.1 Verstehen der Organisation und ihres Kontextes	12
5.4.2 Artikulieren der Risikomanagementverpflichtung	15
5.4.3 Zuweisung von organisatorischen Rollen, Befugnissen, Verantwortlichkeiten und Rechenschaftspflichten	15
5.4.4 Zuordnung von Ressourcen	15
5.4.5 Einrichten der Kommunikation und Konsultation	15
5.5 Implementierung	15
5.6 Bewertung	16
5.7 Verbesserung	16
5.7.1 Anpassen	16
5.7.2 Fortlaufendes Verbessern	16
6 Risikomanagementprozess	16
6.1 Allgemeines	16
6.2 Kommunikation und Konsultation	16
6.3 Anwendungsbereich, Kontext und Kriterien	16
6.3.1 Allgemeines	16
6.3.2 Festlegen des Anwendungsbereichs	17
6.3.3 Externer und interner Kontext	17
6.3.4 Festlegen von Risikokriterien	18
6.4 Risikobeurteilung	18
6.4.1 Allgemeines	18
6.4.2 Risikoidentifikation	19
6.4.3 Risikoanalyse	21
6.4.4 Risikobewertung	23
6.5 Risikobehandlung	23
6.5.1 Allgemeines	23
6.5.2 Auswahl von Maßnahmen zur Risikobehandlung	23
6.5.3 Erstellen und Implementieren von Plänen zur Risikobehandlung	24
6.6 Überwachen und Überprüfen	24
6.7 Aufzeichnen und Berichten	24
Anhang A (informativ) Ziele	26
A.1 Allgemeines	26
A.2 Verantwortlichkeit	26
A.3 KI-Expertise	26
A.4 Verfügbarkeit und Qualität von Trainings- und Testdaten	26
A.5 Auswirkung auf die Umwelt	27
A.6 Fairness	27
A.7 Instandhaltungsfreundlichkeit	27
A.8 Datenschutz	27

A.9 Robustheit 28

A.10 Sicherheit 28

A.11 Sicherheit 28

A.12 Transparenz und Erklärbarkeit 28

Anhang B (informativ) Risikoquellen 29

B.1 Allgemeines 29

B.2 Komplexität der Umgebung 29

B.3 Fehlende Transparenz und Erklärbarkeit 29

B.4 Automatisierungsgrad 30

B.5 Risikoquellen in Bezug auf maschinelles Lernen 30

B.6 System-Hardwareprobleme 30

B.7 System-Lebenszyklusprobleme 31

B.8 Technologische Reife 31

Anhang C (informativ) Risikomanagement und Lebenszyklus von KI-Systemen 32

Literaturhinweise 36

Tabellen

Tabelle 1 — Anwendung von Risikomanagementgrundsätzen auf künstliche Intelligenz 8

Tabelle 2 — Zu prüfende Punkte bei der Festlegung des externen Organisationskontextes 12

Tabelle 3 — Zu prüfende Punkte bei der Festlegung des internen Kontextes einer Organisation 13

Tabelle 4 — Zusätzliche Anleitungen zum Festlegen von Risikokriterien 18

Tabelle C.1 — Risikomanagement und Lebenszyklus von KI-Systemen 32

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Europäisches Vorwort

Der Text von ISO/IEC 23894:2023 wurde vom Technischen Komitee ISO/IEC JTC 1 „Information technology“ der Internationalen Organisation für Normung (ISO) erarbeitet und als prEN ISO/IEC 23894:2023 durch das Technische Komitee CEN/CLC/JTC 21 „Künstliche Intelligenz“ übernommen, dessen Sekretariat von DS gehalten wird.

Dieses Dokument ist derzeit zur CEN-Umfrage vorgelegt.

Anerkennungsnotiz

Der Text von ISO/IEC 23894:2023 wurde von CEN als prEN ISO/IEC 23894:2023 ohne irgendeine Abänderung genehmigt.

Vorwort

ISO (die Internationale Organisation für Normung) und IEC (die Internationale Elektrotechnische Kommission) bilden das auf die weltweite Normung spezialisierte System. Nationale Normungsorganisationen, die Mitglieder von ISO oder IEC sind, beteiligen sich an der Entwicklung von Internationalen Normen in Technischen Komitees, die von der jeweiligen Organisation eingerichtet wurden, um spezifische Gebiete technischer Aktivitäten zu behandeln. Auf Gebieten von beiderseitigem Interesse arbeiten die Technischen Komitees von ISO und IEC zusammen. Weitere internationale staatliche und nichtstaatliche Organisationen, die in engem Kontakt mit ISO und IEC stehen, nehmen ebenfalls an der Arbeit teil.

Die Verfahren, die bei der Entwicklung dieses Dokuments angewendet wurden und die für die weitere Pflege vorgesehen sind, werden in den ISO/IEC-Directives, Teil 1 beschrieben. Im Besonderen sollten die für die verschiedenen ISO-Dokumentenarten notwendigen Annahmekriterien beachtet werden. Dieses Dokument wurde in Übereinstimmung mit den Gestaltungsregeln der ISO/IEC-Directives, Teil 2 erarbeitet (siehe www.iso.org/directives oder www.iec.ch/members_experts/refdocs).

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. ISO und IEC sind nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren. Details zu allen während der Entwicklung des Dokuments identifizierten Patentrechten finden sich in der Einleitung und/oder in der ISO-Liste der erhaltenen Patenterklärungen (siehe www.iso.org/patents) oder in der IEC-Liste der erhaltenen Patenterklärungen (siehe <http://patents.iec.ch>).

Jeder in diesem Dokument verwendete Handelsname dient nur zur Unterrichtung der Anwender und bedeutet keine Anerkennung.

Für eine Erläuterung des freiwilligen Charakters von Normen, der Bedeutung ISO-spezifischer Begriffe und Ausdrücke in Bezug auf Konformitätsbewertungen sowie Informationen darüber, wie ISO die Grundsätze der Welthandelsorganisation (WTO, en: World Trade Organization) hinsichtlich technischer Handelshemmnisse (TBT, en: Technical Barriers to Trade) berücksichtigt, siehe www.iso.org/iso/foreword.html. In der IEC, siehe www.iec.ch/understanding-standards.

Dieses Dokument wurde vom gemeinsamen Technischen Komitee ISO/IEC JTC 1, *Information technology*, Unterkomitee SC 42, *Artificial intelligence*, erarbeitet.

Rückmeldungen oder Fragen zu diesem Dokument sollten an das jeweilige nationale Normungsinstitut des Anwenders gerichtet werden. Eine vollständige Auflistung dieser Institute ist unter www.iso.org/members.html und www.iec.ch/national-committees zu finden.

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11 prEN ISO/IEC 23894:2023 (D)

Einleitung

Der Zweck des Risikomanagements besteht darin, Werte zu schaffen und zu bewahren. Es verbessert die Leistung, regt Innovationen an und unterstützt das Erreichen von Zielen.

Dieses Dokument ist für die Verwendung in Verbindung mit ISO 31000:2018 vorgesehen. Immer wenn die Leitlinien in diesem Dokument über die Anleitungen in ISO 31000:2018 hinausgehen, wird auf die entsprechenden Abschnitte in ISO 31000:2018 verwiesen, gegebenenfalls gefolgt von einer KI-spezifischen Leitlinie. Um die Zusammenhänge zwischen diesem Dokument und ISO 31000:2018 besser zu verdeutlichen, wird die Abschnittsstruktur von ISO 31000:2018 in diesem Dokument gespiegelt und falls erforderlich durch Unterabschnitte ergänzt.

Dieses Dokument ist in drei Hauptteile untergliedert:

Abschnitt 4: Grundlagen – In diesem Abschnitt werden die zugrundeliegenden Grundsätze des Risikomanagements beschrieben. Der Einsatz von KI erfordert spezifische Erwägungen bezüglich einiger dieser Grundsätze, wie in ISO 31000:2018, Abschnitt 4, beschrieben.

Abschnitt 5: Rahmenwerk – Der Zweck des Risikomanagement-Rahmenwerks ist die Unterstützung der Organisation bei der Integration von Risikomanagement in ihre bedeutende Aktivitäten und Funktionen. Spezifische Aspekte der Entwicklung, der Bereitstellung und des Anbietens sowie der Nutzung von KI-Systemen sind in ISO 31000:2018, Abschnitt 5, beschrieben.

Abschnitt 6: Prozesse – Risikomanagementprozesse umfassen die systematische Anwendung von Managementrichtlinien, -verfahren und -praktiken auf die Tätigkeiten des Kommunizierens, Abstimmens und Festlegens des Kontextes sowie die Bewertung, Behandlung, Überwachung, Überprüfung, Dokumentation und Kommunikation von Risiken. Eine Anpassung dieser Prozesse an KI ist in ISO 31000:2018, Abschnitt 6, beschrieben.

Allgemeine KI-bezogene Ziele und Risikoquellen sind in Anhang A und Anhang B beschrieben. Anhang C enthält ein Beispiel für die Zuordnung zwischen Risikomanagementprozessen und dem Lebenszyklus eines KI-Systems.

1 Anwendungsbereich

Dieses Dokument enthält Anleitungen, wie Organisationen, die Produkte, Systeme und Dienstleistungen entwickeln, einsetzen oder nutzen, bei denen Künstliche Intelligenz (KI) zum Einsatz kommt, mit den spezifischen, mit KI verbundenen Risiken umgehen können. Die Anleitungen haben auch zum Ziel, Organisationen dabei zu unterstützen, das Risikomanagement in ihre KI-bezogene Aktivitäten und Funktionen zu integrieren. Darüber hinaus werden Prozesse für die wirksame Implementierung und Integration des KI-Risikomanagements beschrieben.

Die Anwendung dieser Anleitungen kann für jede Organisation und deren Kontext maßgeschneidert werden.

2 Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

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 Begriffe

Für die Anwendung dieses Dokuments gelten die Begriffe nach ISO 31000:2018, ISO/IEC 22989:2022 and ISO Guide 73:2009.

ISO und IEC stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

- ISO Online Browsing Platform: verfügbar unter <https://www.iso.org/obp>
- IEC Electropedia: verfügbar unter <https://www.electropedia.org/>

4 Grundsätze des KI-Risikomanagements

Das Risikomanagement sollte auf die Bedürfnisse der Organisation mithilfe einer integrierten, strukturierten und umfassenden Herangehensweise eingehen. Leitgrundsätze ermöglichen es einer Organisation, Prioritäten zu identifizieren und Entscheidungen darüber zu treffen, wie die Auswirkungen von Unsicherheiten auf ihre Ziele zu behandeln sind. Diese Grundsätze gelten für alle Organisationsebenen und Ziele, ob strategisch oder betrieblich.

Systeme und Prozesse setzen üblicherweise eine Kombination verschiedener Technologien und Funktionen für spezifische Anwendungsfälle in unterschiedlichen Umgebungen ein. Das Risikomanagement sollte das Gesamtsystem mit all seinen Technologien und Funktionalitäten und die Auswirkungen auf Umgebung und Stakeholder berücksichtigen.

KI-Systeme können neu auftretende Risiken mit positiven oder negativen Auswirkungen auf die Ziele einer Organisation verursachen oder die Wahrscheinlichkeit des Auftretens vorhandener Risiken ändern. Sie können auch eine spezifische Berücksichtigung durch die Organisation erfordern. In diesem Dokument sind zusätzliche Anleitungen zu Grundsätzen, Rahmenwerk und Prozessen des Risikomanagements beschrieben, die eine Organisation implementieren kann.

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

ANMERKUNG Die Definitionen des Wortes „Risiko“ unterscheiden sich wesentlich in verschiedenen Internationalen Normen. In ISO 31000:2018 und verbundenen internationalen Normen bezieht sich „Risiko“ auf eine negative oder positive Abweichung von den Zielen. In einigen anderen Internationalen Normen bezieht sich „Risiko“ nur auf potenzielle negative Ergebnisse, beispielsweise auf sicherheitsbezogene Bedenken. Dieser unterschiedliche Fokus kann beim Versuch, einen den Anforderungen entsprechenden Risikomanagementprozess zu verstehen und ordnungsgemäß zu implementieren, häufig für Verwirrung sorgen.

In ISO 31000:2018, Abschnitt 4 sind einige allgemeine Grundsätze des Risikomanagements festgelegt. Zusätzlich zu den Anleitungen in ISO 31000:2018, Abschnitt 4 enthält Tabelle 1 weitere Anleitungen, wie diese Grundsätze bei Bedarf anzuwenden sind.

Tabelle 1 — Anwendung von Risikomanagementgrundsätzen auf künstliche Intelligenz

	Grundsatz	Beschreibung (nach ISO 31000:2018, Abschnitt 4)	Auswirkungen auf die Entwicklung und Nutzung von KI
a)	integriert	Das Risikomanagement ist ein integraler Bestandteil aller Aktivitäten einer Organisation.	Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen.
b)	strukturiert und umfassend	Ein strukturierter und umfassender Risikomanagementansatz trägt zu konsistenten und vergleichbaren Ergebnissen bei.	Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen.
c)	maßgeschneidert	Das Rahmenwerk und die Prozesse des Risikomanagements sind an den externen und internen Kontext einer Organisation angepasst und diesem angemessen sowie mit den Zielen der Organisation verbunden.	Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen.

Tabelle 1 (fortgesetzt)

	Grundsatz	Beschreibung (nach ISO 31000:2018, Abschnitt 4)	Auswirkungen auf die Entwicklung und Nutzung von KI
d)	einbeziehend	Die angemessene und rechtzeitige Beteiligung von Stakeholdern ermöglicht die Berücksichtigung ihrer Kenntnisse, Ansichten und Wahrnehmungen. Dies führt zu verbesserter Erkenntnis und einem fundierten Risikomanagement.	<p>Aufgrund der potenziell weitreichenden Auswirkungen von KI auf Stakeholder ist es wichtig, dass Organisationen in Dialog mit verschiedenen internen und externen Gruppe treten, um sowohl Schäden als auch Vorteile zu kommunizieren und um Rückmeldungen und Erkenntnisse in den Risikomanagementprozess einfließen zu lassen.</p> <p>Organisationen sollten sich auch bewusst sein, dass durch die Anwendung von KI-Systemen zusätzliche Stakeholder involviert werden können.</p> <p>Die Kenntnisse, Ansichten und Wahrnehmungen von Stakeholdern sind unter anderem in folgenden Bereichen von Vorteil:</p> <ul style="list-style-type: none">— Insbesondere das maschinelle Lernen (ML) benötigt für die Erfüllung seiner Ziele häufig geeignete Datensätze. Stakeholder können dazu beitragen, Risiken in Bezug auf die Datenerfassung, die Verarbeitungsvorgänge, die Quelle und Art der Daten und die Nutzung von Daten für bestimmte Situationen oder Ausreißer bei den Datensubjekten zu identifizieren.— Die Komplexität von KI-Technologien verursacht Herausforderungen bezüglich der Transparenz und Erklärbarkeit von KI-Systemen. Die Vielfalt von KI-Technologien verstärkt diese Herausforderungen durch Merkmale wie beispielsweise die unterschiedlichen Arten von Datenmodalitäten, KI-Modell-Topologien und Transparenz- und Berichtsmechanismen, die entsprechend den Bedürfnissen der Stakeholder ausgewählt werden sollten. Stakeholder können dazu beitragen, die Ziele zu identifizieren und die Mittel zur Verbesserung der Transparenz und Erklärbarkeit von KI-Systemen zu beschreiben. In bestimmten Fällen können diese Ziele und Mittel für den Anwendungsfall und die verschiedenen beteiligten Stakeholder verallgemeinert werden. In anderen Fällen kann die Aufteilung der Transparenz-Rahmenwerke und der Berichtsmechanismen durch die Stakeholder je nach Anwendungsfall an relevante Personen (z. B. „Aufsichtsbehörden“, „Geschäftsinhaber“, „Modell-Risikobewerter“) angepasst werden.— Der Einsatz von KI-Systemen für die automatisierte Entscheidungsfindung kann interne und externe Stakeholder direkt betreffen. Diese Stakeholder können beispielsweise ihre Ansichten und Wahrnehmungen dazu äußern, ob möglicherweise eine Überwachung durch Menschen erforderlich ist. Stakeholder können dazu beitragen, Fairnesskriterien zu definieren und zu identifizieren, worin Voreingenommenheit bei der Ausführung von KI-Systemen besteht.

Nachfolgedokument: DIN EN ISO/IEC 23894 (in Vorbereitung/in preparation/en préparation) (DE30101214)

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Tabelle 1 (fortgesetzt)

	Grundsatz	Beschreibung (nach ISO 31000:2018, Abschnitt 4)	Auswirkungen auf die Entwicklung und Nutzung von KI
e)	dynamisch	Risiken können auftreten, sich verändern oder verschwinden, wenn sich der externe und interne Kontext einer Organisation verändert. Diese Veränderungen und Ereignisse werden durch das Risikomanagement in angemessener Weise und rechtzeitig vorhergesehen, erkannt, bestätigt und behandelt.	<p>Zur Implementierung der in ISO 31000:2018 bereitgestellten Anleitungen sollten Organisationen organisatorische Strukturen schaffen und Maßnahmen festlegen, um Probleme und Chancen bei aufkommenden Risiken, Trends, Technologien, Einsatzbereichen und Akteuren im Zusammenhang mit KI-Systemen zu identifizieren.</p> <p>Das dynamische Risikomanagement ist aus den folgenden Gründen bei KI-Systemen besonders wichtig:</p> <ul style="list-style-type: none">— KI-Systeme sind durch kontinuierliches Lernen, Verfeinern, Bewerten und Validieren von Natur aus dynamisch. Darüber hinaus haben manche KI-Systeme die Fähigkeit, sich auf Grundlage dieses Kreislaufs anzupassen und zu optimieren und damit selbstständig dynamische Änderungen zu erzeugen.— Die Kundenerwartungen an KI-Systeme sind hoch und können sich potenziell schnell ändern, wenn die Systeme selbst dies tun.— Rechtliche und behördliche Anforderungen im Zusammenhang mit KI ändern sich häufig und werden oft aktualisiert. <p>Auch die Integration in Managementsysteme für Qualität, ökologische Fußabdrücke, Sicherheit, Gesundheitswesen, rechtliche oder unternehmerische Verantwortung oder eine Kombination dieser Themen, die von der Organisation unterhalten werden, kann berücksichtigt werden, um Risiken für die Organisation, Einzelpersonen und Gesellschaften im Zusammenhang mit KI besser zu verstehen und zu behandeln.</p>
f)	beste verfügbare Information	Die Eingaben in das Risikomanagement basieren auf historischen und aktuellen Informationen sowie zukünftigen Erwartungen. Das Risikomanagement berücksichtigt ausdrücklich alle Einschränkungen und Unsicherheiten, die mit solchen Informationen und Erwartungen verbunden sind. Informationen sollten zeitgerecht, verständlich und den relevanten Stakeholdern verfügbar sein.	<p>Berücksichtigt man die Erwartung, dass KI die Art beeinflusst, wie Einzelpersonen mit Technologie interagieren und darauf reagieren, dann ist es für an der Entwicklung von KI-Systemen beteiligten Organisationen ratsam, relevante verfügbare Informationen über den weiteren Einsatz der von ihnen entwickelten KI-Systeme zu verfolgen, während Benutzer von KI-Systemen während der gesamten Lebensdauer dieser Systeme Aufzeichnungen über deren Einsatz führen können.</p> <p>Da es sich bei KI um eine neu aufkommende Technologie handelt, die sich ständig weiterentwickelt, können historische Informationen nur begrenzt verfügbar sein und zukünftige Erwartungen können sich schnell ändern. Organisationen sollten dies berücksichtigen.</p> <p>Wenn überhaupt, sollte die interne Nutzung von KI-Systemen erwogen werden. Die Nachverfolgung der Anwendung von KI-Systemen durch Kunden und externe Benutzer kann durch geistige Eigentumsrechte und vertragliche oder marktspezifische Einschränkungen beschränkt sein. Solche Einschränkungen sollten durch den KI-Risikomanagementprozess erfasst und aktualisiert werden, wenn die Geschäftsbedingungen eine Neubewertung erfordern.</p>

Tabelle 1 (fortgesetzt)

	Grundsatz	Beschreibung (nach ISO 31000:2018, Abschnitt 4)	Auswirkungen auf die Entwicklung und Nutzung von KI
g)	menschliche und kulturelle Faktoren	Menschliches Verhalten und Kultur haben einen wesentlichen Einfluss auf alle Aspekte des Risikomanagements auf allen Ebenen und in jeder Phase.	Organisationen, die an der Gestaltung, Entwicklung oder Auslieferung von KI-Systemen oder einer Kombination davon beteiligt sind, sollten die menschliche und kulturelle Umgebung beobachten, in der sie sich bewegen. Organisationen sollten ihren Fokus darauf richten, zu identifizieren, wie KI-Systeme oder -Komponenten mit vorhandenen gesellschaftlichen Mustern zusammenwirken, wodurch Auswirkungen auf gerechte Ergebnisse, Privatsphäre, Meinungsfreiheit, Fairness, Sicherheit, Beschäftigungsverhältnisse, die Umwelt und die Menschenrechte im Allgemeinen entstehen können.
h)	fortlaufende Verbesserung	Das Risikomanagement wird durch Lernen und Erfahrung fortlaufend verbessert.	Die Identifikation von bisher unbekannten Risiken in Bezug auf den Einsatz von KI-Systemen sollte beim fortlaufenden Verbesserungsprozess berücksichtigt werden. Organisationen, die an der Gestaltung, Entwicklung oder Auslieferung von KI-Systemen oder -Systemkomponenten oder einer Kombination davon beteiligt sind, sollten das KI-Ökosystem in Bezug auf Leistungserfolge, Defizite und gewonnene Erkenntnisse beobachten und ständig neue KI-Forschungsergebnisse und -Techniken (Möglichkeiten zur Verbesserung) verfolgen.

5 Rahmenwerk

5.1 Allgemeines

Der Zweck des Risikomanagement-Rahmenwerks besteht darin, die Organisation dabei zu unterstützen, das Risikomanagement in deren bedeutende Aktivitäten und Funktionen zu integrieren. Es gilt die Anleitung in ISO 31000:2018, 5.1.

Risikomanagement umfasst die Erfassung der für eine Organisation relevanten Informationen, um Entscheidungen zu treffen und auf Risiken einzugehen. Das Steuerungsgremium legt die Gesamtrisikobereitschaft und die organisatorischen Ziele fest, delegiert aber den Entscheidungsfindungsprozess für die Identifizierung, Beurteilung und den Umgang mit Risiken auf Leitungsfunktionen innerhalb der Organisation.

ISO/IEC 38507 [1] beschreibt zusätzliche Erwägungen der Organisationsleitung bezüglich Entwicklung, Kauf oder Einsatz von KI-Systemen. Zu solchen Erwägungen zählen neue Chancen, potenzielle Änderungen der Risikobereitschaft sowie neue Leitungsgrundsätze, um den verantwortungsvollen Einsatz von KI durch die Organisation sicherzustellen. Dies kann zusammen mit den in diesem Dokument beschriebenen Risikomanagementprozessen eingesetzt werden, um die in ISO 31000:2018, 5.2 beschriebene dynamische und sich schrittweise wiederholende organisatorische Integration zu unterstützen.

5.2 Führung und Verpflichtung

Es gilt die Anleitung in ISO 31000:2018, 5.2.

Zusätzlich zu der in ISO 31000:2018, 5.2 bereitgestellten Anleitung gilt Folgendes:

Aufgrund der besonderen Wichtigkeit von Vertrauen und Verantwortlichkeit bei der Entwicklung und Nutzung von KI sollte die oberste Leitung prüfen, auf welche Weise Stakeholder über Grundsätze und Erklärungen zu KI-Risiken und -Risikomanagement informiert werden. Der Nachweis einer entsprechenden Führung und Ver-

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Die Verpflichtung kann äußerst wichtig sein, um sicherzustellen, dass Stakeholder Vertrauen in die verantwortliche Entwicklung und Nutzung von KI haben.

Die Organisation sollte daher die Veröffentlichung von Erklärungen in Bezug auf ihre Verpflichtung zum KI-Risikomanagement erwägen, um das Vertrauen von Stakeholdern in den Einsatz von KI zu erhöhen.

Die oberste Leitung sollte sich auch darüber im Klaren sein, dass spezialisierte Ressourcen für den Umgang mit KI-Risiken erforderlich sein können, und diese Ressourcen entsprechend bereitstellen.

5.3 Integration

Es gilt die Anleitung in ISO 31000:2018, 5.3.

5.4 Gestaltung

5.4.1 Verstehen der Organisation und ihres Kontextes

Es gilt die Anleitung in ISO 31000:2018, 5.4.1.

Zusätzlich zu der in ISO 31000:2018, 5.4.1 bereitgestellten Anleitung enthält Tabelle 2 weitere Faktoren, die beim Verstehen des externen Organisationskontextes zu berücksichtigen sind.

Tabelle 2 — Zu prüfende Punkte bei der Festlegung des externen Organisationskontextes

Allgemeine, in ISO 31000:2018, 5.4.1 bereitgestellte Anleitung	Zusätzliche Anleitung für Organisationen, die sich mit KI befassen
Organisationen sollten mindestens die folgenden Elemente ihres externen Kontextes prüfen:	Organisationen sollten zusätzlich unter anderem die folgenden Elemente prüfen:
— Soziale, kulturelle, politische, rechtliche, behördliche, finanzielle, technologische, wirtschaftliche und umweltbezogene Faktoren, seien sie internationaler, nationaler, regionaler oder lokaler Art;	— Maßgebliche gesetzliche Belange, einschließlich solcher, die sich spezifisch mit KI befassen. — Leitlinien zur ethischen Nutzung und Gestaltung von KI- und automatisierten Systemen, die von regierungsnahen Gruppen, Aufsichtsbehörden, Normungsorganisationen, der Zivilgesellschaft, Hochschulen und Wirtschaftsverbänden herausgegeben wurden. — KI betreffende, domänenspezifische Leitlinien und Rahmenwerke.
— Wesentliche Schlüsselfaktoren und Trends, welche die Ziele der Organisation beeinflussen;	— Technologische Trends und Fortschritte in den verschiedenen KI-Bereichen. — Gesellschaftliche und politische Auswirkungen beim Einsatz von KI-Systemen, einschließlich sozialwissenschaftliche Leitfäden.
— Beziehungen, Wahrnehmung, Werte, Bedürfnisse und Erwartungen externer Stakeholder;	— Wahrnehmung von Stakeholdern, die durch Probleme wie beispielsweise fehlende Transparenz von KI-Systemen (auch als Undurchsichtigkeit bezeichnet) oder voreingenommene KI-Systeme beeinflusst sein kann. — Erwartungen von Stakeholdern in Bezug auf die Verfügbarkeit spezifischer KI-basierter Lösungen und die Verfahren, wie KI-Modelle verfügbar gemacht werden (z. B. durch Benutzerschnittstellen, Software-Entwicklungskits).

Tabelle 2 (fortgesetzt)

Allgemeine, in ISO 31000:2018, 5.4.1 bereitgestellte Anleitung	Zusätzliche Anleitung für Organisationen, die sich mit KI befassen
— Vertragliche Beziehungen und Verpflichtungen;	— Wie der Einsatz von KI und speziell von KI-Systemen, die mit kontinuierlichem Lernen arbeiten, die Fähigkeit der Organisation zum Einhalten vertraglicher Verpflichtungen und Garantien beeinflussen kann. Daraus folgt, dass Organisationen den Geltungsbereich entsprechender Verträge genau prüfen sollten. — Vertragliche Beziehungen während der Gestaltung und Erstellung von KI-Systemen und -Dienstleistungen. Beispielsweise sollten die Eigentumsverhältnisse und Nutzungsrechte von Test- und Trainingsdaten geprüft werden, wenn diese von Dritten bereitgestellt werden.
— Die Komplexität der Netzwerke und Abhängigkeiten;	— Die Nutzung von KI kann die Komplexität von Netzwerken und Abhängigkeiten erhöhen.
— (Anleitungen, die über ISO 31000:2018 hinausgehen).	— KI-Systeme können bestehende Systeme ersetzen, und in diesem Fall kann eine Beurteilung der Risikovorteile und der Risikoübertragung eines KI-Systems im Vergleich zum bestehenden System erfolgen, wobei sicherheits- und umweltbezogene, technische und finanzielle Probleme in Verbindung mit der Implementierung des KI-Systems geprüft werden.

Zusätzlich zu der in ISO 31000:2018, 5.4.1 bereitgestellten Anleitung enthält Tabelle 3 weitere Faktoren, die für das Verstehen des internen Kontextes einer Organisation zu berücksichtigen sind.

Tabelle 3 — Zu prüfende Punkte bei der Festlegung des internen Kontextes einer Organisation

Allgemeine, in ISO 31000:2018, 5.4.1 bereitgestellte Anleitung	Zusätzliche Anleitung für Organisationen, die sich mit KI befassen
Organisationen sollten mindestens die folgenden Elemente ihres internen Kontextes prüfen:	Organisationen sollten zusätzlich unter anderem die folgenden Elemente prüfen:
— Vision, Mission und Werte;	— Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen
— Leitung, Organisationsstruktur, Rollen und Rechenschaftspflichten;	— Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen
— Strategie, Ziele und Grundsätze;	— Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen
— Kultur der Organisation;	— Die Auswirkungen, die ein KI-System auf die Kultur der Organisation haben kann, wenn Verantwortlichkeiten, Rollen und Aufgaben verschoben oder neu eingeführt werden.
— Von der Organisation übernommene Normen, Leitlinien und Modelle;	— Alle zusätzlichen internationalen, regionalen, nationalen und lokalen Normen und Leitlinien, die durch den Einsatz von KI-Systemen auferlegt werden.

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Tabelle 3 (fortgesetzt)

Allgemeine, in ISO 31000:2018, 5.4.1 bereitgestellte Anleitung	Zusätzliche Anleitung für Organisationen, die sich mit KI befassen
<ul style="list-style-type: none">— Fähigkeiten im Sinne von Ressourcen und Kenntnissen (z. B. Kapital, Zeit, Menschen, geistiges Eigentum, Prozesse, Systeme und Technologien);	<ul style="list-style-type: none">— Die zusätzlichen Risiken für das in der Organisation vorhandene Wissen bezüglich Transparenz und Erklärbarkeit von KI-Systemen.— Durch die Nutzung von KI-Systemen kann sich die Anzahl menschlicher Ressourcen ändern, die für die Umsetzung einer bestimmten Fähigkeit erforderlich sind, oder die Art der erforderlichen Ressourcen kann sich ändern, beispielsweise in Form niedrigerer Qualifikation oder des Verlusts von Expertise, wenn die menschliche Entscheidungsfindung zunehmend durch KI-Systeme gestützt wird.— Die spezifischen Kenntnisse über KI-Technologien und Datenwissenschaft, die für die Entwicklung und Nutzung von KI-Systemen erforderlich sind.— Die Verfügbarkeit von KI-Werkzeugen, -Plattformen und -Bibliotheken kann die Entwicklung von KI-Systemen ermöglichen, ohne dass ein umfassendes Verständnis der Technologie, ihrer Einschränkungen und potenziellen Fallstricke vorhanden ist.— Die Möglichkeit, dass KI zu Problemen und Chancen im Zusammenhang mit dem geistiges Eigentum an spezifischen KI-Systemen führen kann. Organisationen sollten überprüfen, über welches eigene geistige Eigentum sie in diesem Bereich verfügen und wie geistige Eigentumsrechte die Transparenz, Sicherheit und Fähigkeit zur Zusammenarbeit mit Stakeholdern beeinflussen kann, um zu klären, ob und welche Schritte eingeleitet werden sollten.
<ul style="list-style-type: none">— Daten, Informationssysteme und Informationsflüsse;	<ul style="list-style-type: none">— KI-Systeme können zur Automatisierung, Optimierung und Erweiterung der Datenverarbeitung eingesetzt werden.— KI-Systeme nutzen Daten und können daher zusätzliche Anforderungen an die Qualität und Vollständigkeit von Daten und Informationen stellen.
<ul style="list-style-type: none">— Beziehungen zu internen Stakeholdern unter Berücksichtigung ihrer Wahrnehmungen und Werte;	<ul style="list-style-type: none">— Die Wahrnehmung von Stakeholdern, die durch Probleme wie beispielsweise fehlende Transparenz von KI-Systemen oder voreingenommene KI-Systeme beeinflusst sein kann.— Die Bedürfnisse und Erwartungen von Stakeholdern können in höherem Maß durch spezifische KI-Systeme erfüllt werden.— Die Notwendigkeit, Stakeholder über die Fähigkeiten, Fehlermöglichkeiten und den Umgang mit Fehlern von KI-Systemen aufzuklären.

Tabelle 3 (fortgesetzt)

Allgemeine, in ISO 31000:2018, 5.4.1 bereitgestellte Anleitung	Zusätzliche Anleitung für Organisationen, die sich mit KI befassen
— Vertragliche Beziehungen und Verpflichtungen;	<ul style="list-style-type: none">— Die Wahrnehmung von Stakeholdern, die durch unterschiedliche Herausforderungen im Zusammenhang mit KI-Systemen beeinflusst sein kann, wie beispielsweise potenziell fehlende Transparenz und Unfairness.— Die Bedürfnisse und Erwartungen von Stakeholdern können durch spezifische KI-Systeme erfüllt werden.— Die Notwendigkeit, Stakeholder über die Fähigkeiten, Fehlermöglichkeiten und den Umgang mit Fehlern von KI-Systemen aufzuklären.— Die Erwartungen von Stakeholdern zum Schutz personenbezogener Daten und zu grundsätzlichen individuellen und kollektiven Rechten und Freiheiten.
— Gegenseitige Abhängigkeiten und Verbindungen.	— Die Nutzung von KI-Systemen kann die Komplexität von gegenseitigen Abhängigkeiten und Verbindungen erhöhen.

Zusätzlich zur Anleitung in ISO 31000:2018, 5.4.1 sollten Organisationen berücksichtigen, dass für die Nutzung von KI-Systemen eine zusätzliche spezialisierte Ausbildung erforderlich sein kann.

5.4.2 Artikulieren der Risikomanagementverpflichtung

Es gilt die Anleitung in ISO 31000:2018, 5.4.2.

5.4.3 Zuweisung von organisatorischen Rollen, Befugnissen, Verantwortlichkeiten und Rechenschaftspflichten

Es gilt die Anleitung in ISO 31000:2018, 5.4.3.

Zusätzlich zur Anleitung in ISO 31000:2018, 5.4.3 sollten die oberste Leitung und Aufsichtsbehörden, wo angezeigt, die erforderlichen Ressourcen bereitstellen und Personen identifizieren, die:

- die Befugnis zur Behandlung von KI-Risiken haben;
- die Verantwortung für die Einrichtung und Überwachung von Prozessen zur Behandlung von KI-Risiken haben.

5.4.4 Zuordnung von Ressourcen

Es gilt die Anleitung in ISO 31000:2018, 5.4.4.

5.4.5 Einrichten der Kommunikation und Konsultation

Es gilt die Anleitung in ISO 31000:2018, 5.4.5.

5.5 Implementierung

Es gilt die Anleitung in ISO 31000:2018, 5.5.

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11 prEN ISO/IEC 23894:2023 (D)

5.6 Bewertung

Es gilt die Anleitung in ISO 31000:2018, 5.6.

5.7 Verbesserung

5.7.1 Anpassen

Es gilt die Anleitung in ISO 31000:2018, 5.7.1.

5.7.2 Fortlaufendes Verbessern

Es gilt die Anleitung in ISO 31000:2018, 5.7.2.

6 Risikomanagementprozess

6.1 Allgemeines

Es gilt die Anleitung in ISO 31000:2018, 6.1.

Organisationen sollten einen risikobasierten Ansatz zum Identifizieren, Beurteilen und Verstehen von KI-Risiken implementieren, denen sie ausgesetzt sind, und sie sollten entsprechend der Risikohöhe geeignete Behandlungsmaßnahmen ergreifen. Der Erfolg des gesamten KI-Risikomanagementprozesses einer Organisation beruht auf der Identifikation, Festlegung und erfolgreichen Implementierung von eng gefassten Risikomanagementprozessen auf strategischer, betrieblicher, Programm- und Projektebene. Aufgrund von Bedenken, die unter anderem durch die potenzielle Komplexität, fehlende Transparenz und Unvorhersehbarkeit mancher KI-basierter Technologien aufkommen können, sollte den Risikomanagementprozessen auf der KI-System-Projektebene besondere Aufmerksamkeit gewidmet werden. Diese Prozesse auf Systemprojektebene sollten an den Zielen der Organisation ausgerichtet sein, und es sollte ein Informationsaustausch mit anderen Risikomanagementebenen stattfinden. Beispielsweise sollten Eskalationen und gewonnene Erkenntnisse auf KI-Projektebene auch in höhere Ebenen, wie die strategische, betriebliche und Programmebene (und, sofern zutreffend, weitere Ebenen) einfließen.

Anwendungsbereich, Kontext und Kriterien von Risikomanagementprozessen auf Projektebene werden direkt durch die Lebenszyklusphasen des KI-Systems beeinflusst, die im Projektanwendungsrahmen liegen. In Anhang C sind mögliche Zusammenhänge zwischen einem Risikomanagementprozess auf Projektebene und dem Lebenszyklus eines KI-Systems (wie in ISO/IEC 22989:2022 festgelegt) beschrieben.

6.2 Kommunikation und Konsultation

Es gilt die Anleitung in ISO 31000:2018, 6.2.

Die Anzahl der durch ein KI-System betroffenen Stakeholder kann höher sein als anfänglich angenommen, und es können bisher nicht berücksichtigte externe Stakeholder sowie andere Teile einer Gesellschaft betroffen sein.

6.3 Anwendungsbereich, Kontext und Kriterien

6.3.1 Allgemeines

Es gilt die Anleitung in ISO 31000:2018, 6.3.1.

Zusätzlich zur Anleitung in ISO 31000:2018, 6.3.1 sollten in Organisationen, die KI einsetzen, der Anwendungsbereich des KI-Risikomanagements, der Kontext des KI-Risikomanagementprozesses und die Kriterien zur Bewertung der Risikohöhe für die Unterstützung von Entscheidungsfindungsprozessen erweitert werden, um zu ermitteln, wo in der Organisation KI-Systeme entwickelt oder eingesetzt werden. Eine solche Auflistung zur

Entwicklung und Nutzung von KI sollte dokumentiert und in den Risikomanagementprozess der Organisation aufgenommen werden.

6.3.2 Festlegen des Anwendungsbereichs

Es gilt die Anleitung in ISO 31000:2018, 6.3.2.

Der Anwendungsbereich sollte die spezifischen Aufgaben und Verantwortungsbereiche der unterschiedlichen Ebenen einer Organisation berücksichtigen. Darüber hinaus sollten Ziel und Zweck der von der Organisation entwickelten oder genutzten KI-Systeme berücksichtigt werden.

6.3.3 Externer und interner Kontext

Es gilt die Anleitung in ISO 31000:2018, 6.3.3.

Aufgrund der großen potenziellen Auswirkungen von KI-Systemen sollte die Organisation dem Stakeholder-Umfeld bei der Festlegung und Einrichtung des Kontextes eines Risikomanagementprozesses besondere Aufmerksamkeit widmen.

Es sollte erwogen werden, eine Liste von unter anderem den folgenden Stakeholdern zu erstellen:

- die (eigene) Organisation;
- Kunden, Partner und Dritte;
- Lieferanten;
- Endbenutzer;
- Regulierungsbehörden;
- zivile Organisationen;
- Einzelpersonen;
- betroffene Gemeinden;
- Gesellschaften.

Weiterhin sind folgende Punkte beim externen und internen Kontext zu berücksichtigen:

- ob die KI-Systeme Menschen schädigen oder wichtige Dienstleistungen verhindern können (die bei Unterbrechung Leben, Gesundheit oder persönliche Sicherheit gefährden würden), ob sie Menschenrechte verletzen können (z. B. durch unfaire und voreingenommene automatisierte Entscheidungsfindung) oder zu Umweltschäden beitragen können;
- externe und interne Erwartungen bezüglich der gesellschaftlichen Verantwortung einer Organisation;
- externe und interne Erwartungen bezüglich der Umweltverantwortung einer Organisation.

Die Leitlinien nach ISO 26000:2010 [2], in denen Aspekte der sozialen Verantwortung beschrieben sind, sollten als Rahmenwerk für das Verstehen und Behandeln von Risiken Anwendung finden, insbesondere bei Kernthemen wie Organisationsleitung, Menschenrechten, Arbeitsverfahren, Umwelt, fairen Betriebsverfahren, Verbraucherbelangen und der gesellschaftlichen Einbeziehung und Entwicklung.

ANMERKUNG Weitere Hintergrundinformationen zur Vertrauenswürdigkeit sind in ISO/IEC TR 24028:2020 [3] angegeben.

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

6.3.4 Festlegen von Risikokriterien

Es gilt die Anleitung in ISO 31000:2018, 6.3.4.

Zusätzlich zur Anleitung in ISO 31000:2018, 6.3.4 sind in Tabelle 4 weitere Leitlinien zu Faktoren aufgeführt, die beim Festlegen von Risikokriterien zu berücksichtigen sind:

Tabelle 4 — Zusätzliche Anleitungen zum Festlegen von Risikokriterien

Erwägungen bei der Festlegung von Risikokriterien nach ISO 31000:2018, 6.3.4	Zusätzliche Erwägungen im Zusammenhang mit der Entwicklung und Nutzung von KI-Systemen
<ul style="list-style-type: none">— Natur und Art der Unsicherheiten, die die Ergebnisse und Ziele (materiell und immateriell) beeinflussen können;— Wie Auswirkungen (positiv und negativ) und Wahrscheinlichkeit festgelegt und gemessen werden;	<ul style="list-style-type: none">— Organisationen sollten sinnvolle Schritte unternehmen, um Unsicherheiten in allen Teilen des KI-Systems zu verstehen, einschließlich verwendete Daten, Software, mathematische Modelle, physikalische Erweiterungen und „Human-in-the-Loop“-Aspekte des Systems (wie beispielsweise entsprechende menschliche Aktivitäten während der Datenerfassung und -kennzeichnung).
<ul style="list-style-type: none">— Zeitliche Faktoren;	<ul style="list-style-type: none">— Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen
<ul style="list-style-type: none">— Konsistenz bei der Anwendung von Messungen;	<ul style="list-style-type: none">— Organisationen sollten sich bewusst sein, dass KI eine sich schnell entwickelnde Technologiedomäne ist. Messverfahren sollten ständig bezüglich ihrer Effektivität und Eignung für die eingesetzten KI-Systeme überprüft werden.
<ul style="list-style-type: none">— Art und Weise, wie die Risikohöhe zu bestimmen ist;	<ul style="list-style-type: none">— Organisationen sollten einen konsistenten Ansatz zur Bestimmung der Risikohöhe verfolgen. Der Ansatz sollte die potenziellen Auswirkungen von KI-Systemen auf unterschiedliche, mit KI in Zusammenhang stehende Ziele widerspiegeln (siehe Anhang A).
<ul style="list-style-type: none">— Wie Kombinationen und Abfolgen mehrerer Risiken berücksichtigt werden;	<ul style="list-style-type: none">— Keine spezifischen Anleitungen, die über ISO 31000:2018 hinausgehen
<ul style="list-style-type: none">— Die Kapazität der Organisation.	<ul style="list-style-type: none">— Bei der Entscheidung über die Risikobereitschaft der Organisation in Bezug auf KI sollten die KI-Kapazität, der Kenntnisstand und die Fähigkeit, erkannte KI-Risiken zu mindern, berücksichtigt werden.

6.4 Risikobeurteilung

6.4.1 Allgemeines

Es gilt die Anleitung in ISO 31000:2018, 6.4.1.

KI-Risiken sollten identifiziert, bemessen oder qualitativ beschrieben sowie in Bezug auf für die Organisation relevante Risikokriterien und Ziele priorisiert werden. Anhang B enthält einen beispielhaften Katalog von KI-bezogenen Risikoquellen. Ein solcher beispielhafter Katalog kann nicht als umfassend angesehen werden. Allerdings zeigt die Erfahrung, welchen Nutzen die Verwendung eines solchen Katalogs als Basis für Organisationen hat, die eine Risikobeurteilung erstmalig durchführen oder die das KI-Risikomanagement in bestehende Managementstrukturen integrieren. Der Katalog dient diesen Organisationen als dokumentierte Grundlinie.

Organisationen, die sich mit der Entwicklung, Bereitstellung oder Anwendung von KI-Systemen befassen, sollten daher ihre Risikobeurteilungsaktivitäten nach dem Lebenszyklus des Systems ausrichten. Für unterschiedliche Phasen des Systemlebenszyklus können unterschiedliche Risikobeurteilungsverfahren gelten.

6.4.2 Risikoidentifikation

6.4.2.1 Allgemeines

Es gilt die Anleitung in ISO 31000:2018, 6.4.2.

6.4.2.2 Identifikation von Vermögensgegenständen und ihrem Wert

Die Organisation sollte Vermögensgegenstände identifizieren, die mit der Gestaltung und Nutzung von KI zusammenhängen und die in den Anwendungsbereich des in 6.3.2 definierten Risikomanagementprozesses fallen. Das Verständnis, welche Vermögensgegenstände in den Anwendungsbereich fallen und wie entscheidend oder wertvoll diese Vermögensgegenstände sind, ist ein integraler Bestandteil der Beurteilung von Auswirkungen. Es sollten sowohl der Wert als auch die Art des Vermögensgegenstandes (materiell oder immateriell) berücksichtigt werden. In Bezug auf die Entwicklung und Nutzung von KI sollten unter anderem Vermögensgegenstände in den folgenden Kontexten berücksichtigt werden:

- Vermögensgegenstände und ihr Wert für die Organisation:
 - Zu den materiellen Vermögensgegenständen können Daten, Modelle und das KI-System selbst zählen.
 - Zu den immateriellen Vermögensgegenständen können Ruf und Vertrauen zählen.
- Vermögensgegenstände und ihr Wert für Einzelpersonen:
 - Zu den materiellen Vermögensgegenständen können die personenbezogenen Daten von Einzelpersonen zählen,
 - Zu den immateriellen Vermögensgegenständen können der Schutz personenbezogener Daten und die Gesundheit und Sicherheit von Einzelpersonen zählen.
- Vermögensgegenstände und ihr Wert für Gemeinschaften und Gesellschaften:
 - Zu den materiellen Vermögensgegenständen kann die Umwelt zählen,
 - Immaterielle Vermögensgegenstände sind vermutlich eher wertebasiert, wie beispielsweise sozio-kulturelle Überzeugungen, gemeinschaftliches Wissen, Zugang zu Bildung und Gleichheit.

Bezüglich der Bewertung von Vermögensgegenständen und des Zusammenhangs mit Auswirkungen siehe 6.4.2.6 und 6.4.3.2.

ANMERKUNG Die Verwendung des Wortes „Vermögensgegenstand“ in Zusammenhang mit den veranschaulichenden Beispielen in diesem Abschnitt hat keine rechtliche Bedeutung.

6.4.2.3 Identifikation von Risikoquellen

Die Organisation sollte eine Liste von Risikoquellen im Zusammenhang mit der KI-Entwicklung, -Anwendung oder beidem innerhalb des definierten Anwendungsbereichs aufstellen.

Risikoquellen können unter anderem innerhalb der folgenden Bereiche identifiziert werden:

- Organisation;
- Prozesse und Verfahren;

- Entwurf -**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)**

- Managementroutinen;
- Personal;
- physikalische Umgebung;
- Daten;
- KI-Systemkonfiguration;
- Einsatzumgebung;
- Hardware, Software, Netzwerkressourcen und Dienste;
- Abhängigkeit von externen Parteien.

Beispiele für Risikoquellen im Zusammenhang mit KI sind in Anhang B zu finden.

6.4.2.4 Identifikation von potenziellen Ereignissen und Ergebnissen

Die Organisation sollte potenzielle Ereignisse identifizieren, die im Zusammenhang mit der KI-Entwicklung oder -Anwendung stehen und die unterschiedliche materielle oder immaterielle Auswirkungen haben können.

Die Ereignisse können mit einer oder mehreren der folgenden Verfahren und Quellen identifiziert werden:

- veröffentlichte Normen;
- veröffentlichte technische Spezifikationen;
- veröffentlichte technische Berichte;
- veröffentlichte wissenschaftliche Abhandlungen;
- Marktdaten zu ähnlichen, bereits eingesetzten Systemen oder Anwendungen;
- Berichte über Vorfälle bei ähnlichen, bereits eingesetzten Systemen oder Anwendungen;
- Feldversuche;
- Gebrauchstauglichkeitsstudien;
- Ergebnisse geeigneter Untersuchungen;
- Berichte von Stakeholdern;
- Befragungen und Berichte von internen oder externen Experten;
- Simulationen.

6.4.2.5 Identifikation von Steuerungsmaßnahmen

Die Organisation sollte die relevanten Steuerungsmaßnahmen für die KI-Entwicklung, -Anwendung oder beides identifizieren. Steuerungsmaßnahmen sollten während der Risikomanagementaktivitäten identifiziert und dokumentiert werden (in internen Systemen, Verfahren, Auditberichten usw.).

Steuerungsmaßnahmen können eingesetzt werden, um das Gesamtrisiko durch Minderung von Risikoquellen, Ereignissen und Ergebnissen positiv zu beeinflussen.

Die Wirksamkeit der identifizierten Steuerungsmaßnahmen sollte ebenfalls betrachtet werden, insbesondere die Unwirksamkeit von Maßnahmen.

6.4.2.6 Identifikation von Auswirkungen

Die Organisation sollte als Teil der KI-Risikobeurteilung Risikoquellen, Ereignisse oder Ergebnisse, die zu Risiken führen können, identifizieren. Sie sollte auch alle Auswirkungen auf die Organisation selbst, auf Einzelpersonen, Gemeinschaften, Gruppen und Gesellschaften identifizieren. Organisationen sollten besonders darauf achten, alle Unterschiede zwischen den Gruppen zu identifizieren, die die Vorteile der Technologie erfahren, und den Gruppen, die negative Auswirkungen erfahren.

Die Auswirkungen auf die Organisation unterscheiden sich notwendigerweise von den Auswirkungen auf Einzelpersonen und Gesellschaften. Auswirkungen auf Organisationen können unter anderem folgende sein:

- Zeit für Untersuchungen und Reparaturen;
- gewonnene und verlorene (Arbeits-)Zeit;
- gewonnene oder verlorene Chancen;
- Bedrohungen der Gesundheit oder Sicherheit von Einzelpersonen;
- Finanzaufwand für spezifische Fachkenntnisse zur Schadensbeseitigung;
- Einstellung, Zufriedenheit und Bindung von Mitarbeitern;
- Ansehen, Reputation und Kulanz;
- Straf- und Bußgelder;
- Rechtsstreit mit Kunden.

Je nach Kontext können die Auswirkungen auf Einzelpersonen und Gesellschaften allgemeinerer Natur sein; in diesem Fall kann die Organisation möglicherweise die Auswirkung auf Einzelpersonen oder Gesellschaften nicht genau abschätzen.

Anstelle jede Art von Auswirkung zu spezifizieren, kann in erster Linie betrachtet werden, wie entscheidend die Auswirkungen allgemein (beispielsweise bei Einzelpersonen auf den Schutz personenbezogener Daten, Fairness, Menschenrechte usw. oder bei Gesellschaften auf die Umwelt) sind.

Die genauen Auswirkungen können vom Kontext abhängen, in dem die Organisation tätig ist, und von den Bereichen, für die das KI-System entwickelt und eingesetzt wird.

ANMERKUNG 1 Auswirkungen können positiv oder negativ sein. Beides kann von der Organisation bei der Beurteilung von Auswirkungen auf die Organisation, Einzelpersonen und Gesellschaften berücksichtigt werden.

ANMERKUNG 2 Auswirkungen auf Einzelpersonen und Gesellschaften können üblicherweise auch Auswirkungen auf die Organisation zur Folge haben. Beispielsweise kann ein Sicherheitsvorfall, der den Benutzer eines Produkts der Organisation betrifft, zu Haftungsansprüchen gegen die Organisation führen und ihren Ruf und den Produktabsatz negativ beeinflussen.

6.4.3 Risikoanalyse

6.4.3.1 Allgemeines

Es gilt die Anleitung in ISO 31000:2018, 6.4.3.

- Entwurf -**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)**

Der Ansatz der Analyse sollte im Einklang mit den Risikokriterien stehen, die als Teil der Kontextfestlegung entwickelt wurden (siehe 6.3).

6.4.3.2 Beurteilung von Auswirkungen

Bei der Beurteilung der in der Risikobeurteilung identifizierten Auswirkungen sollte die Organisation zwischen der Beurteilung der geschäftlichen Auswirkungen, der Beurteilung der Auswirkungen auf Einzelpersonen und der Beurteilung der gesellschaftlichen Auswirkungen unterscheiden.

Bei der Analyse der geschäftlichen Auswirkungen sollte der Grad festgelegt werden, zu dem die Organisation betroffen ist, und es sollten unter anderem die folgenden Elemente berücksichtigt werden:

- Bedeutung der Auswirkung;
- materielle und immaterielle Auswirkungen;
- Kriterien zur Festlegung der Gesamtauswirkung (wie in 6.3.4 festgelegt).

Bei der Analyse der Auswirkungen auf Einzelpersonen sollte der Grad festgelegt werden, zu dem Einzelpersonen durch die KI-Entwicklung, -Anwendung oder beides betroffen sein können. Es sollten unter anderem die folgenden Elemente berücksichtigt werden:

- die Art der verwendeten Daten von Einzelpersonen;
- die beabsichtigte Auswirkung der Entwicklung oder der Anwendung von KI;
- die potenzielle Auswirkung von Voreingenommenheit auf Einzelpersonen;
- die potenzielle Auswirkung auf Grundrechte, die zu materiellen und immateriellen Schäden für Einzelpersonen führen kann;
- die potenzielle Auswirkung auf die Fairness gegenüber Einzelpersonen;
- die Sicherheit von Einzelpersonen;
- Schutzmaßnahmen und Steuerungsmaßnahmen zur Minderung unerwünschter Voreingenommenheit und Unfairness;
- das rechtliche und kulturelle Umfeld der Einzelperson (dieses kann die Festlegung der relativen Auswirkung beeinflussen).

Bei Analysen der gesellschaftlichen Auswirkungen sollte der Grad festgelegt werden, zu dem eine Gesellschaft durch die KI-Entwicklung, -Anwendung oder beides betroffen sein kann. Es sollten unter anderem die folgenden Elemente berücksichtigt werden:

- der Umfang der Auswirkungen auf die Gesellschaft (die Reichweite des KI-Systems bei verschiedenen Populationen), einschließlich, von wem das System genutzt wird oder für wen es gestaltet wurde (so kann beispielsweise die Nutzung durch Regierungen eine Gesellschaft potenziell stärker betreffen als eine private Nutzung);
- wie ein KI-System die sozialen und kulturellen Werte verschiedener betroffener Gruppen beeinflusst (einschließlich spezifischer Arten, wie das KI-System bereits vorhandene Muster der Schädigung verschiedener sozialer Gruppen verstärkt oder mindert).

6.4.3.3 Beurteilung der Wahrscheinlichkeit

Gegebenenfalls sollte die Organisation die Wahrscheinlichkeit des Eintretens von Ereignissen und Ergebnissen, die zu Risiken führen, beurteilen. Die Wahrscheinlichkeit kann auf qualitativer oder quantitativer Ebene

bestimmt werden und sollte den unter 6.3.4 ermittelten Kriterien entsprechen. Die Wahrscheinlichkeit kann unter anderem durch die folgenden Faktoren begründet und beeinflusst sein:

- Art, Bedeutung und Anzahl der Risikoquellen;
- Häufigkeit, Schweregrad und Durchdringungsgrad von Bedrohungen;
- interne Faktoren, wie beispielsweise der betriebliche Erfolg von Grundsätzen und Verfahren und die Motivation interner Akteure;
- externe Faktoren, wie die Geografie und andere soziale, wirtschaftliche und Umweltinteressen;
- Erfolg (Minderung) oder Unwirksamkeit von Steuerungsmaßnahmen (siehe 6.4.2.5).

Organisationen sollten Wahrscheinlichkeitsberechnungen nur durchführen, wenn sie anwendbar und nützlich sind, um festzustellen, wo Risikobehandlungen durchzuführen sind. Die Entscheidungsfindung auf Basis von Wahrscheinlichkeiten kann wesentliche technische, ökonomische und heuristische Probleme verursachen, insbesondere, wenn die Wahrscheinlichkeit entweder nicht berechnet werden kann oder eine große Fehlerspanne aufweist.

6.4.4 Risikobewertung

Es gilt die Anleitung in ISO 31000:2018, 6.4.4.

6.5 Risikobehandlung

6.5.1 Allgemeines

Es gilt die Anleitung in ISO 31000:2018, 6.5.1.

6.5.2 Auswahl von Maßnahmen zur Risikobehandlung

Es gilt die Anleitung in ISO 31000:2018, 6.5.2.

Die von der Organisation festgelegten Möglichkeiten der Risikobehandlung sollten so gestaltet sein, dass sie negative Auswirkungen von Risiken auf eine annehmbare Höhe reduzieren und die Wahrscheinlichkeit erhöhen, dass positive Ergebnisse erreicht werden können. Falls die erforderliche Verringerung negativer Ergebnisse nicht durch die Anwendung verschiedener Risikobehandlungsoptionen erreicht werden kann, sollte die Organisation für die Restrisiken eine Risiko-Nutzen-Analyse durchführen.

Nach ISO 31000:2018, 6.5.2 sollte die Organisation Folgendes in Erwägung ziehen:

- Vermeiden des Risikos, indem entschieden wird, die Tätigkeit, die Anlass zu dem Risiko gibt, nicht zu beginnen oder fortzusetzen;
- Eingehen oder Erhöhen des Risikos zur Nutzung einer Chance;
- Beseitigen der Risikoursache;
- Verändern der Wahrscheinlichkeit;
- Verändern der Auswirkungen;
- gemeinsames Tragen des Risikos (z. B. durch Verträge oder den Abschluss von Versicherungen);
- Beibehalten des Risikos auf Grundlage einer fundierten Entscheidung.

- Entwurf -**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)****6.5.3 Erstellen und Implementieren von Plänen zur Risikobehandlung**

Es gilt die Anleitung in ISO 31000:2018, 6.5.3.

Nachdem der Plan zur Risikobehandlung dokumentiert wurde, sollten die in 6.5.2 ausgewählten Maßnahmen zur Risikobehandlung implementiert werden.

Die Implementierung jeder Maßnahme zur Risikobehandlung und ihre Wirksamkeit sollten nach 6.7 verifiziert und dokumentiert werden.

6.6 Überwachen und Überprüfen

Es gilt die Anleitung in ISO 31000:2018, 6.6.

6.7 Aufzeichnen und Berichten

Es gilt die Anleitung in ISO 31000:2018, 6.7.

Die Organisation sollte während und nach der Implementierungsphase ein System für die Sammlung und Verifizierung von Informationen über das Produkt oder ähnliche Produkte erstellen, dokumentieren und pflegen. Die Organisation sollte auch öffentlich zugängliche Informationen für ähnliche auf dem Markt verfügbare Systeme sammeln und prüfen.

Diese Informationen sollten dann in Bezug auf eine mögliche Relevanz für die Vertrauenswürdigkeit des KI-Systems beurteilt werden. Insbesondere sollte bei der Bewertung beurteilt werden, ob bisher nicht erkannte Risiken vorhanden sind oder ob in der Vergangenheit beurteilte Risiken nicht mehr annehmbar sind. Diese Informationen können in den KI-Risikomanagementprozess der Organisation zur Anpassung der Ziele, Anwendungsfälle oder gewonnenen Erkenntnisse einfließen.

Wenn eine dieser Bedingungen zutrifft, sollten Organisationen Folgendes durchführen:

- die Auswirkungen auf die bisherigen Risikomanagementaktivitäten beurteilen und die Ergebnisse dieser Beurteilung wieder in den Risikomanagementprozess einfließen lassen.
- eine Überprüfung der Risikomanagementaktivitäten für das KI-System durchführen. Wenn die Möglichkeit besteht, dass das Restrisiko oder dessen Annehmbarkeit sich geändert hat, sollten die Auswirkungen auf vorhandene Risikokontrollmaßnahmen bewertet werden.

Die Ergebnisse dieser Bewertung sollten dokumentiert werden. Die Risikomanagementdokumentation sollte die Nachverfolgbarkeit jedes identifizierten Risikos über alle Risikomanagementprozesse hinweg ermöglichen. Die Dokumentationen können eine gemeinsame Vorlage nutzen, auf die sich die Organisation geeinigt hat.

Zusätzlich zur Dokumentation von Anwendungsbereich, Kontext und Kriterien (siehe 6.3), Risikobeurteilung (siehe 6.4) und Risikobehandlung (siehe 6.5) sollten die Aufzeichnungen mindestens die folgenden Informationen enthalten:

- eine Beschreibung und Identifikation des analysierten Systems;
- die angewandte Methodik;
- eine Beschreibung des beabsichtigten Gebrauchs des KI-Systems;
- die Identität der Person(en) und der Organisation, durch die die Risikobeurteilung erfolgte;
- die Festlegungen und das Datum der Risikobeurteilung;
- den Veröffentlichungsstatus der Risikobeurteilung;

- Entwurf -

**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)**

— ob und zu welchem Grad die Ziele erreicht wurden.

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Anhang A **(informativ)**

Ziele

A.1 Allgemeines

Bei der Identifizierung der Risiken von KI-Systemen sollten verschiedene Ziele in Bezug auf KI betrachtet werden, abhängig von der Art des betrachteten Systems und seinem Anwendungskontext. Die in Bezug auf KI zu berücksichtigenden Ziele umfassen unter anderem die in Abschnitt A.2 bis Abschnitt A.12 beschriebenen Ziele.

A.2 Verantwortlichkeit

Der Begriff Verantwortlichkeit bezieht sich sowohl auf ein Merkmal von Organisationen als auch auf eine Systemeigenschaft:

- Die Verantwortlichkeit von Organisationen bedeutet, dass die Organisation die Verantwortung für Entscheidungen und Aktionen übernimmt, indem sie sie erläutert und für diese gegenüber dem Steuerungsgremium, den Behörden und im weiteren Sinne den Stakeholdern verantwortlich ist.
- Die Systemverantwortlichkeit bezieht sich auf die Rückverfolgbarkeit von Entscheidungen und Aktionen einer Entität bis zu dieser Entität.

Durch die Nutzung von KI können sich vorhandene Rahmenwerke für Berechtigungen ändern. Wenn bisher Aktionen von Personen durchgeführt wurden, die dafür zur Verantwortung gezogen werden konnten, können solche Aktionen jetzt vollständig oder teilweise von KI-Systemen durchgeführt werden. Wer in diesem Fall die Verantwortung tragen würde, ist Thema von fortlaufenden Überlegungen der Regulierungsbehörden. Entwickler und Benutzer von KI-Systemen sollten die entsprechenden rechtlichen Anforderungen in den einzelnen Ländern, in denen das System vermarktet und eingesetzt wird, kennen.

A.3 KI-Expertise

KI-Systeme und ihre Entwicklung unterscheiden sich von Softwarelösungen, die keine KI einsetzen. Es wird eine Auswahl entsprechender Spezialisten mit interdisziplinären Fachkenntnissen und Expertise bei der Beurteilung, Entwicklung und dem Einsatz von KI-Systemen benötigt. Organisationen sollten sicherstellen, dass Personen mit einer solchen Expertise an der Entwicklung und Spezifikation von KI-Systemen beteiligt sind.

Auch bei den Endbenutzern von KI-Systemen sollte KI-Expertise vorhanden sein. Die Benutzer sollten ein ausreichendes Verständnis haben, wie das KI-System funktioniert, und befähigt sein, falsche Entscheidungen oder Ausgaben zu erkennen und zu umgehen.

A.4 Verfügbarkeit und Qualität von Trainings- und Testdaten

KI-Systeme, die auf ML basieren, benötigen Trainings- und Testdaten, mit denen die Systeme trainiert und in Bezug auf das beabsichtigte Verhalten verifiziert werden. Im Einsatz befindliche KI-Systeme arbeiten mit Produktionsdaten. Datentyp und Datenqualität der Trainings-, Test- und Produktionsdaten sollten an das beabsichtigte Verhalten angepasst sein.

Trainings- und Testdaten sollten auf ihre Aktualität und Relevanz für den beabsichtigten Zweck geprüft werden. Die benötigte Menge an Trainings- und Testdaten kann je nach beabsichtigter Funktionalität und Komplexität der Umgebung variieren. Die Trainings- und Testdaten sollten ausreichend unterschiedliche Merkmale aufweisen, damit das KI-System über eine starke Prognosefähigkeit verfügt. Darüber hinaus sollte die Konsistenz von Trainings- und Testdaten sichergestellt werden, wobei gegebenenfalls unabhängige Datensätze verwendet werden.

Es ist möglich, dass Trainings- und Testdaten im Unternehmen nicht verfügbar sind und extern beschafft werden. Auch in diesem Fall sollte die Datenqualität sichergestellt werden.

A.5 Auswirkung auf die Umwelt

Die Nutzung von KI kann aus Umweltsicht Auswirkungen haben. Die Nutzung von KI kann positive Auswirkungen auf die Umwelt haben. So kann ein KI-System beispielsweise genutzt werden, um Stickoxide bei Gasturbinen zu verringern. Die Nutzung von KI kann aufgrund der extensiven Ressourcennutzung; auch negative Auswirkungen auf die Umwelt haben. So erfordert beispielsweise die Trainingsphase mancher KI-Systeme Rechnerressourcen und kann erhebliche Mengen an elektrischer Energie verbrauchen. Diese Auswirkungen auf die Umwelt sollten berücksichtigt werden.

A.6 Fairness

Die Nutzung von KI-Systemen für die automatisierte Entscheidungsfindung kann unfair gegenüber bestimmten Personen oder Personengruppen sein. Unfaire Ergebnisse haben eine Reihe von Ursachen, beispielsweise Voreingenommenheit von Zielfunktionen, nicht ausgewogene Datensätze und menschliche Voreingenommenheit bei den Trainingsdaten und bei Rückmeldungen an Systeme. Unfairness kann auch durch Voreingenommenheit beim Produktkonzept, den Problemformulierungen oder der Auswahl entstehen, wann und wo KI-Systeme eingesetzt werden.

Weitere Informationen zu Voreingenommenheit und Fairness von KI-Systemen siehe ISO/IEC TR 24027 [4].

A.7 Instandhaltungsfreundlichkeit

Bei der Instandhaltungsfreundlichkeit geht es um die Fähigkeit der Organisation, Änderungen des KI-Systems durchzuführen, um Fehler zu korrigieren oder es an neue Anforderungen anzupassen. Weil auf ML basierende KI-Systeme trainiert werden und keinen regelbasierten Ansatz verfolgen, sollten die Instandhaltungsfreundlichkeit eines KI-Systems und deren Auswirkungen untersucht werden.

A.8 Datenschutz

Beim Datenschutz geht es um die Fähigkeit von Einzelpersonen, zu steuern, welche ihrer Informationen erfasst, gespeichert und verarbeitet werden können und durch wen die Informationen veröffentlicht werden können. Wie in ISO/IEC TR 24028:2020 [3] erläutert, sind „viele KI-Techniken (z. B. tiefes Lernen) stark von Big Data abhängig, da ihre Genauigkeit von der Menge der genutzten Daten abhängt. Der Missbrauch oder die Veröffentlichung mancher Daten, insbesondere personenbezogener und sensibler Daten (z. B. Gesundheitsakten), kann auf die Datensubjekte schädliche Auswirkungen haben. Daher hat sich der Schutz der Privatsphäre zu einem wichtigen Problem für die Big-Data-Analyse und KI entwickelt.“

Es sollte gründlich überlegt werden, ob aus einem KI-System sensible personenbezogene Daten abgeleitet werden können. Bei KI-Systemen gehört zum Schutz personenbezogener Daten der Schutz der für die Erstellung und den Betrieb des KI-Systems verwendeten Daten, um sicherzustellen, dass das KI-System nicht für den unberechtigten Zugriff auf Daten genutzt werden kann, und der Zugriffsschutz für Modelle, die für eine Einzelperson personalisiert wurden oder mit deren Hilfe Informationen oder Merkmale solcher Einzelpersonen abgeleitet werden können.

Die unberechtigte Sammlung, Verwendung und Veröffentlichung personenbezogener Daten kann auch direkte Auswirkungen auf grundlegende Menschenrechte haben, wie beispielsweise Diskriminierung und Rede- und Informationsfreiheit. Auch sollten Auswirkungen auf ethische Grundsätze bezüglich des Respektierens menschlicher Werte und der Menschenwürde berücksichtigt werden.

ANMERKUNG Eine Datenschutz-Folgenabschätzung (siehe ISO/IEC 29134:2017 [5]), oft auch als Datenschutz-Risikobeurteilung bezeichnet, ist ein nützliches Werkzeug zur Handhabung der Risiken, die sich bei der Verwendung personenbezogener Daten während der Datenerfassung, des Trainings von KI-Systemen und der Nutzung von KI-Systemen ergeben.

- Entwurf -**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)****A.9 Robustheit**

Bei der Robustheit geht es um die Fähigkeit eines Systems, sein Leistungsniveau unter verschiedenen Nutzungsbedingungen aufrechtzuerhalten. Der Grad, zu dem ein KI-System oder eine zugehörige Komponente bei ungültigen Eingaben oder unter belastenden Umgebungsbedingungen korrekt funktionieren kann, sollte ebenso betrachtet werden wie dessen Fähigkeit, Messwerte und Ergebnisse zu reproduzieren.

Die Robustheit stellt im Kontext von KI-Systemen neue Herausforderungen dar. Neuronale Netzwerkarchitekturen stellen eine besondere Herausforderung dar, da sie schwer verständlich sind und aufgrund ihrer nichtlinearen Natur manchmal ein unerwartetes Verhalten zeigen. Die Charakterisierung neuronaler Netzwerke ist ein offener Forschungsbereich, und es gibt sowohl bei Test- wie auch bei Verifizierungsansätzen Einschränkungen.

Weitere Informationen zur Robustheit neuronaler Netzwerke siehe ISO/IEC TR 24029-1 [6].

A.10 Sicherheit

Die Nutzung von KI-Systemen kann zu neuen Sicherheitsbedrohungen führen. Bei Sicherheit geht es um die Erwartung, dass ein System unter festgelegten Bedingungen nicht zu einem Zustand führt, bei dem menschliches Leben, die Gesundheit, das Eigentum oder die Umwelt gefährdet werden. Die Anwendung von KI-Systemen in automatisierten Fahrzeugen, Fertigungsgeräten und Robotern kann zu Sicherheitsrisiken führen. Auf KI-Systeme in bestimmten Domänen (z. B. Maschinenkonstruktion, Transport, medizinische Geräte) sollten spezifische Normen für diese Domänen in Betracht gezogen werden.

Weitere Informationen zu funktionaler Sicherheit von KI-Systemen siehe ISO/IEC TR 5469¹ [7].

A.11 Sicherheit

Das Informationssicherheits-Risikomanagement ist in ISO/IEC 27005:2022 [8] definiert. Im Kontext von KI und insbesondere bei KI-Systemen, die auf ML-Ansätzen basieren, sollten neben den klassischen Informations- und Systemsicherheitsbedenken einige neue, in ISO/IEC TR 24028:2020 [3] beschriebene Probleme berücksichtigt werden, wie beispielsweise Data Poisoning Attacks, Adversarial Attacks und Model Stealing Attacks.

A.12 Transparenz und Erklärbarkeit

Transparenz bezieht sich sowohl auf Merkmale von Organisationen, die KI-Systeme betreiben, als auch auf die Systeme selbst. Organisationen sind manchmal transparent in Bezug auf die Anwendung solcher Systeme, die Nutzung der erfassten Daten (wie beispielsweise Verbraucher- und Benutzerdaten, öffentliche Daten und andere erfasste Datensätze), die Maßnahmen zum Management von KI-Systemen, das Verstehen und Kontrollieren von Risiken usw. Transparenz von KI-Systemen bedeutet, dass Stakeholder angemessene Informationen über ein System erhalten (z. B. Fähigkeiten und Einschränkungen), damit sie die Entwicklung, den Betrieb und die Anwendung von KI-Systemen im Hinblick auf ihre eigenen Ziele beurteilen können. Bei der Erklärbarkeit von KI-Systemen geht es um die Fähigkeit, nachzuvollziehen und zu verstehen, wie die Ergebnisse eines bestimmten Systems erzeugt wurden.

¹ In Vorbereitung. Stand zum Zeitpunkt der Veröffentlichung: ISO/IEC DTR 5469:2022.

Anhang B (informativ)

Risikoquellen

B.1 Allgemeines

Bei der Identifizierung der Risiken von KI-Systemen sollten verschiedene Risikoquellen berücksichtigt werden, abhängig von der Art des betrachteten Systems und seinem Anwendungskontext. Die zu berücksichtigenden Risikoquellen umfassen unter anderem die in Abschnitt B.2 bis Abschnitt B.8 beschriebenen Möglichkeiten.

B.2 Komplexität der Umgebung

Die Komplexität der Umgebung [9] eines KI-Systems bestimmt die Bandbreite möglicher Situationen, für deren Unterstützung ein KI-System in seinem betrieblichen Kontext vorgesehen ist.

Bestimmte KI-Technologien wie ML sind besonders für die Handhabung komplexer Umgebungen geeignet und werden daher häufig für Systeme verwendet, die in komplexen Umgebungen eingesetzt werden, beispielsweise beim automatisierten Fahren. Es ist aber eine große Herausforderung, während des Gestaltungs- und Entwicklungsprozesses alle entsprechenden Situationen zu identifizieren, die das System handhaben soll, und zu erkennen, ob die Trainings- und Testdaten all diese Situationen abdecken.

Daher können komplexe Umgebungen im Vergleich zu einfachen Umgebungen zusätzliche Risiken bergen. Besondere Überlegungen sollten der Bestimmung des Grads gelten, zu dem das KI-System verstanden wird:

- Ein vollständiges Verständnis der Umgebung, wobei das KI-System auf alle möglicherweise auftretenden Umgebungszustände vorbereitet ist, ermöglicht eine bessere Risikokontrolle, ist aber nur bei einfachen, vorhersehbaren oder gesteuerten Umgebungen möglich.
- Wenn aufgrund der hohen Komplexität oder Unsicherheit der Umgebung nur ein teilweises Verständnis möglich ist, so dass das KI-System nicht alle möglichen Zustände der Umgebung vorhersehen kann (beispielsweise beim autonomen Fahren), dann kann nicht davon ausgegangen werden, dass alle relevanten Situationen berücksichtigt wurden. Dies kann zu einem Unsicherheitsgrad führen, der eine Risikoquelle darstellt, und sollte bei der Gestaltung solcher Systeme berücksichtigt werden.

B.3 Fehlende Transparenz und Erklärbarkeit

Bei Transparenz geht es um das Kommunizieren entsprechender Tätigkeiten und Entscheidungen einer Organisation (z. B. Grundsätze, Prozesse) und entsprechender Informationen über ein KI-System (z. B. Fähigkeiten, Leistung, Einschränkungen, Gestaltungsentscheidungen, Algorithmen, Trainings- und Testdaten, Verifizierungs- und Validierungsprozesse und Ergebnisse) an die entsprechenden Stakeholder. Dadurch kann es Stakeholdern ermöglicht werden, die Entwicklung, den Betrieb und die Anwendung von KI-Systemen im Hinblick auf ihre Erwartungen zu beurteilen. Welcher Informationstyp und -grad angemessen ist, hängt stark von den Stakeholdern, dem Anwendungsfall, dem Systemtyp und den gesetzlichen Anforderungen ab. Wenn Organisationen den betroffenen Stakeholdern keine angemessenen Informationen zur Verfügung stellen können, kann dies die Vertrauenswürdigkeit und die Zurechenbarkeit der Organisation und des KI-Systems negativ beeinflussen.

Erklärbarkeit ist die Eigenschaft eines KI-Systems, wichtige, entscheidungsrelevante Faktoren in einer für Menschen verständlichen Weise auszudrücken. ML-Modelle können ein Verhalten zeigen, das durch Untersuchung des Modells oder des für das Trainieren verwendeten Algorithmus schwer zu verstehen ist, insbesondere bei Deep Learning. Wenn solche wichtigen Faktoren nicht ausgedrückt werden können, hat dies einen negativen Einfluss auf die Validierung des KI-Systems und das menschliche Vertrauen in das System, da nicht klar ist, warum das System eine Entscheidung getroffen hat und ob es in allen Fällen die richtige Entscheidung treffen

- Entwurf -**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)**

kann. Diese Unsicherheit kann viele Risiken verursachen, die starke Auswirkungen auf allgemeine Ziele wie Vertrauenswürdigkeit und Rechenschaftspflicht sowie spezifische Ziele wie Sicherheit, Fairness und Robustheit haben. Die Erklärbarkeit ist daher nicht nur für Stakeholder als Teil der Transparenz von KI-Systemen von Bedeutung, sondern auch für die Organisation selbst und ihre eigene Validierung und Verifizierung des KI-Systems.

Übermäßige Transparenz und Erklärbarkeit kann ebenfalls zu Risiken in Bezug auf den Schutz personenbezogener Daten, Sicherheit, Vertraulichkeitsanforderungen und geistiges Eigentum führen.

B.4 Automatisierungsgrad

KI-Systeme können mit unterschiedlichen Automatisierungsgraden arbeiten. Diese reichen von keiner Automatisierung, wobei der Bediener das System vollständig steuert, bis hin zum vollständig automatisierten System. KI-Systeme sind häufig automatisierte Systeme. Je nach spezifischem Anwendungsfall können die automatisierten Entscheidungen solcher Systeme Auswirkungen auf verschiedene Interessengebiete, wie beispielsweise Sicherheit und Fairness, haben.

Bei einem Automatisierungsgrad, bei dem bei Bedarf ein externer Agent verfügbar sein muss, kann die Übergabe vom System zum Agenten eine Risikoquelle darstellen (z. B. Zeitbeschränkungen, Aufmerksamkeit des Agenten).

Weitere Informationen zu Automatisierungsgraden siehe ISO/IEC 22989:2022, 5.2.

B.5 Risikoquellen in Bezug auf maschinelles Lernen

Viele Vorteile von KI hängen mit ML und dessen Unterbereichen zusammen, wie beispielsweise tiefes Lernen (Deep Learning). Das Verhalten von ML-Systemen ist nicht nur stark von den eingesetzten Algorithmen abhängig, sondern auch von den Daten, mit denen die ML-Modelle trainiert werden. Daraus ergeben sich unter anderem die folgenden möglichen Auswirkungen auf die KI-Merkmale:

- Datenqualität: Die Qualität der Trainings- und Testdaten hat einen unmittelbaren Einfluss auf die Funktionalität des Systems. Eine nicht ausreichende Datenqualität kann verschiedene Ziele, wie beispielsweise Fairness, Sicherheit und Robustheit, beeinträchtigen.
- Bei KI-Systemen, die ML einsetzen, sind die Datenerfassungsprozesse eine besonders schwer zu diagnostizierende und erkennende Risikoquelle. Beispiele:
 - Daten können nicht repräsentativ für die Anwendbarkeitsdomäne sein, was zu Risiken in Bezug auf geschäftliche Ziele führt.
 - Die Datengewinnung und -speicherung kann bedeutende ethische und rechtliche Risiken nach sich ziehen. Wird der Datenerfassungsprozess nicht abgesichert, dann kann dies zu Risiken durch Adversarial Attacks, Data Poisoning Attacks oder andere Formen der Manipulation führen.
- Kontinuierlich lernende KI-Systeme erreichen normalerweise eine Verbesserung des Systems aufgrund der entstehenden Produktionsdaten, aber es kann auch eine Verschärfung von Risiken auftreten, da sich das Verhalten im Betrieb auf eine bei der Inbetriebnahme nicht vorhergesehene Art ändern kann.

B.6 System-Hardwareprobleme

Risikoquellen in Bezug auf Hardwareprobleme sind unter anderem:

- Hardwarefehler aufgrund defekter Komponenten. Beispiele sind Kurzschlüsse oder Unterbrechungen einzelner oder mehrerer Speicherzellen, defekte Busleitungen, Abweichungen der Oszillatoren, Haftfehler oder Störschwingungen an Ein- oder Ausgängen von integrierten Schaltungen.

- Entwurf -**E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)**

- „Weiche Fehler“ wie beispielsweise unerwünschte temporäre Statusänderungen von Speicherzellen oder Logikkomponenten, meist hervorgerufen durch hochenergetische Strahlung.
- Die Möglichkeit zur Übertragung trainierter ML-Modelle zwischen verschiedenen Systemen kann aufgrund unterschiedlicher Hardware-Ausstattung der Systeme bezüglich Verarbeitungsleistung, Speicher und verfügbarer entsprechender KI-Hardwarebeschleuniger eingeschränkt sein.
- Netzwerkfehler, Bandbreitenbeschränkungen und erhöhte Latenzzeit aufgrund begrenzter und geteilter Netzwerkressourcen, wenn das KI-System entfernte Verarbeitungs- und Speicherressourcen benötigt.

B.7 System-Lebenszyklusprobleme

Nicht angemessene oder nicht ausreichende Verfahren, Prozesse und Verwendung von KI-Systemen während ihres Lebenszyklus können zu Risiken führen. Beispiele für solche Risiken sind:

- Gestaltung und Entwicklung: Ein fehlerhafter Gestaltungsprozess antizipiert möglicherweise nicht jeden Kontext, in dem das KI-System verwendet wird, wodurch es beim Einsatz in einem solchen Kontext zu unerwarteten Ausfällen kommen kann.
- Verifizierung und Validierung: Ein nicht angemessener Verifizierungs- und Validierungsprozess bei der Veröffentlichung aktualisierter KI-System-Versionen kann zu versehentlichen Regressionen und nicht beabsichtigten Verschlechterungen in Bezug auf Qualität, Zuverlässigkeit oder Sicherheit führen.
- Einsatz: Eine nicht für den Einsatz angemessene Konfiguration kann zu Ressourcenproblemen bei Speicher, Rechenleistung, Massenspeicher, Redundanz oder Lastverteilung führen.
- Wartung, Aktualisierung und Überarbeitung: Ein KI-System, das nicht mehr vom Entwickler unterstützt oder gewartet, aber immer noch eingesetzt wird, kann zu langfristigen Risiken oder Haftungsansprüchen gegenüber der Entwicklungsorganisation führen.
- Wiederverwendung: Ein funktionierendes KI-System kann in einem Kontext eingesetzt werden, für den es ursprünglich nicht gestaltet wurde; dadurch kann es zu Problemen aufgrund der unterschiedlichen Anforderungen der geplanten und tatsächlichen Verwendung kommen. So kann z. B. ein System, das für die Identifizierung von Gesichtern auf Fotos in sozialen Netzwerken gestaltet wurde, eingesetzt werden, um Gesichter von Verdächtigen in Überwachungsaufnahmen zu identifizieren; dies ist eine Anwendung, die einen viel höheren Präzisionsgrad erfordert als der ursprüngliche Anwendungsfall.
- Außerbetriebnahme: Organisationen, die den Einsatz eines bestimmten KI-Systems oder einer auf KI-Technologien basierenden Komponente beenden, können Informationen oder Entscheidungsexpertise verlieren, die durch das außer Betrieb genommene System bereitgestellt wurden. Darüber hinaus kann sich die Art ändern, wie eine Organisation Informationen verarbeitet und Entscheidungen trifft, wenn das außer Betrieb genommene System durch ein anderes ersetzt wird.

B.8 Technologische Reife

Die technologische Reife gibt an, wie ausgereift eine bestimmte Technologie in einem bestimmten Anwendungskontext ist. Weniger ausgereifte Technologien, die bei der Entwicklung und Anwendung von KI-Systemen eingesetzt werden, können zu Risiken führen, die der Organisation nicht bekannt sind oder die sie nur schwer beurteilen kann. Für ausgereifte Technologien kann eine größere Menge an Erfahrungsdaten verfügbar sein, so dass Risiken leichter zu identifizieren und zu beurteilen sind. Allerdings besteht bei ausgereiften Technologien auch das Risiko von Selbstzufriedenheit und technischen Schulden.

Anhang C
(informativ)

Risikomanagement und Lebenszyklus von KI-Systemen

Tabelle C.1 enthält ein Beispiel für die Zuordnung von Risikomanagementprozessen zum Lebenszyklus eines KI-Systems nach ISO/IEC 22989:2022.

Tabelle C.1 — Risikomanagement und Lebenszyklus von KI-Systemen

→ Risiko- management	Rahmenwerk zum KI-Risiko- management (Abschnitt 5)	KI-Risikomanagementprozess (Abschnitt 6)				
		Anwen- dungs- bereich, Kontext und Kriterien	Risiko- beurteilung	Risiko- behandlung	Überwa- chen und Überprüfen	Aufzeichnen und Berichten
KI-System Lebens- zyklus ↓	Aktivitäten auf organisatorischer Ebene im Zusammenhang mit dem Risiko- management	Rückmeldungen der Risikomanagementprozesse des KI-Systems werden empfangen und verarbeitet. Als Ergebnis wird das Risikomanagement-Rahmenwerk der Organisation verbessert, indem ihre Risikomanagementwerkzeuge erweitert und verfeinert werden:				
		Ein Katalog von Risikokriterien.	Ein Katalog potenzieller Risikoquellen. Ein Katalog von Techniken zur Beurteilung und Messung von Risikoquellen.	Ein Katalog bekannter oder implementierter Minderungsmaßnahmen.	Ein Katalog implementierter Techniken zur Überwachung und Steuerung von KI-Systemen.	Ein Katalog etablierter Verfahren und definierter Formate zum Aufzeichnen, Berichten und Teilen von KI-System-Informationen mit internen und externen Stakeholdern.
Anfangsphase	Das Steuerungsgremium untersucht die Ziele des KI-Systems im Kontext der Grundsätze und Werte der Organisation und der Stakeholder. Auf Basis einer Analyse (üblicherweise in mehreren Schichten) wird festgelegt, ob das KI-System machbar ist und das Problem behandelt, das die Organisation lösen möchte.	Der Risikomanagementprozess des KI-Systems und die Risikokriterien des Systems werden durch Anpassung des Risikomanagementrahmenwerks der Organisation erarbeitet.	Spezifische Risikoquellen des KI-Systems werden identifiziert (möglichweise in einem Mehrschichtenansatz) und detailliert beschrieben.	Es wird ein detaillierter Risikobehandlungsplan erstellt. Möglicherweise werden Verfahren für „Machbarkeitsnachweise“ definiert.	Erforderliche Machbarkeitsnachweise werden implementiert, getestet und bewertet.	Die Analyse und ihre Ergebnisse sowie die Empfehlungen werden dokumentiert und der obersten Leitung berichtet.

Tabelle C.1 (fortgesetzt)

→ Risiko- management	Rahmenwerk zum KI-Risiko- management (Abschnitt 5)	KI-Risikomanagementprozess (Abschnitt 6)				
		Anwen- dungs- bereich, Kontext und Kriterien	Risiko- beurteilung	Risiko- behandlung	Überwa- chen und Überprüfen	Aufzeichnen und Berichten
Gestaltung und Entwicklung	Das Steuerungs- gremium beurteilt fortlaufend die Ziele, die Wirksamkeit und die Machbarkeit des Systems auf Grundlage der erhaltenen Rückmeldungen neu.	Möglicher- weise werden die Risikokri- terien des KI-Systems als Ergebnis der Rück- meldungs- berichte geändert.	Die Risiko- beurteilung wird fortlaufend durchgeführt (möglicher- weise auf mehreren Ebenen).	Der Risiko- behand- lungsplan wird im- plementiert. Die Risiko- behandlung und die (ver- bleibende) Risiko- beurteilung werden fortgeführt, bis die festgelegten Risiko- kriterien erfüllt sind.	Während der Tests wird die Verifizie- rung und Validierung des Risiko- behand- lungsplans für die System- komponenten sowie für das Gesamtsys- tem beurteilt und angepasst.	Die Ergebnisse werden dokumentiert und an die entsprechenden Risiko- management- Prozess- aktivitäten zurück- gemeldet. Falls erforderlich, werden die Schluss- folgerungen an die Führungs- ebenen und das Steuerungs- gremium kommuniziert.
Verifizie- rung und Validierung						
Einsatz	Das Steuerungs- gremium beurteilt fortlaufend die Ziele und die Machbarkeit des Systems auf Grundlage der erhaltenen Rück- meldungsberichte neu.	Die Risiko- kriterien des KI-Systems und der Risiko- management- prozess werden an die erforderlichen Konfigura- tions- änderungen angepasst.	Die Risiko- beurteilung wird fortlaufend durchgeführt (möglicher- weise auf mehreren Ebenen).	Der Risiko- behand- lungsplan wird potenziell aufgrund von Konfigurati- onsände- rungen aktualisiert und imple- mentiert. Die Risiko- behandlung und die (ver- bleibende) Risiko- beurteilung werden fortgeführt, bis die festgelegten Risiko- kriterien erfüllt sind.	Der Risiko- behand- lungsplan des KI-Systems wird unter Berücksich- tigung der erforderli- chen Anpassungen neu beurteilt.	

Nachfolgedokument: DIN EN ISO/IEC 23894 (in Vorbereitung/in preparation/en préparation) (DE30101214)

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Tabelle C.1 (fortgesetzt)

→ Risiko- management	Rahmenwerk zum KI-Risiko- management (Abschnitt 5)	KI-Risikomanagementprozess (Abschnitt 6)				
KI-System Lebens- zyklus ↓		Anwen- dungs- bereich, Kontext und Kriterien	Risiko- beurteilung	Risiko- behandlung	Überwa- chen und Überprüfen	Aufzeichnen und Berichten
Betrieb, Überwa- chung	Das Steuerungs- gremium beurteilt fortlaufend die Ziele und die Machbarkeit des Systems auf Grundlage der erhaltenen Rück- meldungsberichte neu.	Möglicher- weise werden die Risikokri- terien des KI-Systems als Ergebnis der Rück- meldungs- berichte geändert.	Der Risiko- beurteilungs- plan des Systems wird potenziell aufgrund von Änderungen der Risiko- kriterien angepasst.	Der Risiko- behand- lungsplan des Systems wird potenziell an Risikoände- rungen aufgrund von Ergebnissen der Risiko- beurteilung angepasst.	Der Risiko- behand- lungsplan für die Systemkom- ponenten wird beurteilt und angepasst.	
Kontinuier- liche Validierung						

Tabelle C.1 (fortgesetzt)

→ Risiko- management	Rahmenwerk zum KI-Risiko- management (Abschnitt 5)	KI-Risikomanagementprozess (Abschnitt 6)				
		Anwen- dungs- bereich, Kontext und Kriterien	Risiko- beurteilung	Risiko- behandlung	Überwa- chen und Überprüfen	Aufzeichnen und Berichten
Neube- wertung	Das Steuerungs- gremium untersucht erneut die Ziele des KI-Systems und deren Verhältnis zu den Grundsätzen und Werten der Organisation und der Stakeholder. Auf Grundlage der Analyse wird entschieden, ob das KI-System machbar ist.	Der Risiko- management- prozess des KI-Systems und die Risiko- kriterien des Systems wer- den in Bezug auf mögliche Änderungen des spezifi- schen Zwecks und Anwend- ungsbereichs des KI-Systems, das Ergebnis der Betriebs- überwachung und neue behördliche Anforde- rungen neu bewertet.	Die Liste der für das KI-System spezifischen vorhandenen Risikoquellen wird auf Relevanz und mögliche Lücken untersucht.	Der Risiko- behand- lungsplan wird möglicher- weise aktualisiert. Die Risiko- behandlung und die (ver- bleibenden) Risikobeur- teilungen werden fortgeführt, bis die festgelegten Risiko- kriterien erfüllt sind.	Der Risiko- behand- lungsplan des KI-Systems wird unter Berücksich- tigung der erforderli- chen Anpassungen neu beurteilt.	
Ausmuste- rung oder Austausch	Das Steuerungs- gremium führt eine neue Untersuchung der Ziele des KI-Systems durch und entscheidet auf Grundlage der Analyse, ob die Ausmusterung oder der Austausch des KI-Systems durchführbar sind.	Der Ausmuste- rungsprozess des KI-System- Risiko- managements- und die Risiko- kriterien der System- ausmusterung werden erarbeitet.	Spezifische Risikoquellen für die Aus- musterung des KI-Systems werden identifiziert und detailliert beschrieben.	Es wird ein detaillierter Risiko- behand- lungsplan erstellt.	Erforderliche Machbar- keits- nachweise werden implemen- tiert, getestet und bewertet.	

- Entwurf -

E DIN EN ISO/IEC 23894:2023-11
prEN ISO/IEC 23894:2023 (D)

Literaturhinweise

- [1] ISO/IEC 38507:2022, *Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations*
- [2] ISO 26000:2010, *Guidance on social responsibility*
- [3] ISO/IEC TR 24028:2020, *Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*
- [4] ISO/IEC TR 24027:2021, *Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*
- [5] ISO/IEC 29134:2017, *Information technology — Security techniques — Guidelines for privacy impact assessment*
- [6] ISO/IEC TR 24029-1:2021, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*
- [7] ISO/IEC TR 5469², *Artificial intelligence — Functional safety and AI systems*
- [8] ISO/IEC 27005:2022, *Information security, cybersecurity and privacy protection — Guidance on managing information security risks*
- [9] RUSSELL, S. J., and NORVIG, P. *Artificial intelligence: a modern approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010

2 In Vorbereitung Stand zum Zeitpunkt der Veröffentlichung: ISO/IEC DTR 5469:2022.

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Principles of AI risk management	1
5 Framework	5
5.1 General	5
5.2 Leadership and commitment	5
5.3 Integration	6
5.4 Design	6
5.4.1 Understanding the organization and its context	6
5.4.2 Articulating risk management commitment	8
5.4.3 Assigning organizational roles, authorities, responsibilities and accountabilities	8
5.4.4 Allocating resources	8
5.4.5 Establishing communication and consultation	8
5.5 Implementation	9
5.6 Evaluation	9
5.7 Improvement	9
5.7.1 Adapting	9
5.7.2 Continually improving	9
6 Risk management process	9
6.1 General	9
6.2 Communication and consultation	9
6.3 Scope, context and criteria	9
6.3.1 General	9
6.3.2 Defining the scope	10
6.3.3 External and internal context	10
6.3.4 Defining risk criteria	10
6.4 Risk assessment	11
6.4.1 General	11
6.4.2 Risk identification	11
6.4.3 Risk analysis	14
6.4.4 Risk evaluation	15
6.5 Risk treatment	15
6.5.1 General	15
6.5.2 Selection of risk treatment options	15
6.5.3 Preparing and implementing risk treatment plans	16
6.6 Monitoring and review	16
6.7 Recording and reporting	16
Annex A (informative) Objectives	18
Annex B (informative) Risk sources	21
Annex C (informative) Risk management and AI system life cycle	24
Bibliography	26

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.

- Entwurf -

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.	<p>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.</p> <p>Organizations should also be aware that the use of AI systems can introduce additional stakeholders.</p> <p>The areas in which the knowledge, views and perceptions of stakeholders are of benefit include but are not restricted to:</p> <ul style="list-style-type: none">— 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.

Nachfolgedokument: DIN EN ISO/IEC 23894 (in Vorbereitung/in preparation/en préparation) (DE30101214)

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.	<p>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.</p> <p>Dynamic risk management is particularly important for AI systems because:</p> <ul style="list-style-type: none">— 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. <p>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.</p>
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.	<p>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.</p> <p>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.</p> <p>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.</p>

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.

Table 2 (continued)

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:
— (guidance beyond ISO 31000:2018).	— An AI system can replace an existing system and, in such a case, an assessment of the risk benefits and risk transfers of an AI system versus the existing system can be undertaken, considering safety, environmental, social, technical and financial issues associated with the implementation of the AI system.

In addition to guidance provided in ISO 31000:2018, 5.4.1, Table 3 lists additional factors to consider when understanding the internal context of an organization.

Table 3 — Consideration when establishing the internal 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 internal context:	Organizations should additionally consider, but not exclusively, the following elements:
— Vision, mission and values;	— No specific guidance beyond ISO 31000:2018
— Governance, organizational structure, roles and accountabilities;	— No specific guidance beyond ISO 31000:2018
— Strategy, objectives and policies;	— No specific guidance beyond ISO 31000:2018
— The organization's culture;	— The effect that an AI system can have on the organization's culture by shifting and introducing new responsibilities, roles and tasks.
— Standards, guidelines and models adopted by the organization;	— Any additional international, regional, national and local standards and guidelines that are imposed by the use of AI systems.
— Capabilities, understood in terms of resources and knowledge (e.g. capital, time, people, intellectual property, processes, systems and technologies);	— The additional risks to organizational knowledge related to transparency and explainability of AI systems. — The use of AI systems can result in changes to the number of human resources needed to realize a certain capability, or in a variation of the type of resources needed, for instance, deskilling or loss of expertise where human decision-making is increasingly supported by AI systems. — The specific knowledge in AI technologies and data science required to develop and use AI systems. — The availability of AI tools, platforms and libraries can enable the development of AI systems without there being a full understanding of the technology, its limitations and potential pitfalls. — The potential for AI to raise issues and opportunities related to intellectual property for specific AI systems. Organizations should consider their own intellectual property in this area and ways that intellectual property can affect transparency, security and the ability to collaborate with stakeholders, to determine whether any steps should be taken.

Table 3 (continued)

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 internal context:	Organizations should additionally consider, but not exclusively, the following elements:
— Data, information systems and information flows;	— AI systems can be used to automate, optimize and enhance data handling. — As consumers of data, additional quality and completeness constraints on data and information can be imposed by AI systems.
— Relationships with internal stakeholders, taking into account their perceptions and values;	— Stakeholder perception, which can be affected by issues such as lack of transparency of AI systems or biased AI systems. — Stakeholder needs and expectations can be satisfied to a greater extent by specific AI systems. — The need for stakeholders to be educated on capabilities, failure modes and failure management of AI systems.
— Contractual relationships and commitments;	— Stakeholder perception, which can be affected by different challenges associated with AI systems such as potential lack of transparency and unfairness. — Stakeholder needs and expectations can be satisfied by specific AI systems. — The need for stakeholders to be educated on capabilities, failure modes and failure management of AI systems. — Stakeholders' expectations of privacy, and individual and collective fundamental rights and freedoms.
— Interdependencies and interconnections;	— The use of AI systems can increase the complexity of interdependencies and interconnections.

In addition to the guidance provided in ISO 31000:2018, 5.4.1, organizations should consider that the use of AI systems can increase the need for specialized training.

5.4.2 Articulating risk management commitment

The guidance provided in ISO 31000:2018, 5.4.2 applies.

5.4.3 Assigning organizational roles, authorities, responsibilities and accountabilities

The guidance provided in ISO 31000:2018, 5.4.3 applies.

In addition to the guidance of ISO 31000:2018, 5.4.3, top management and oversight bodies, where applicable, should allocate resources and identify individuals:

- with authority to address AI risks;
- with responsibility for establishing and monitoring processes to address AI risks.

5.4.4 Allocating resources

The guidance provided in ISO 31000:2018, 5.4.4 applies.

5.4.5 Establishing communication and consultation

The guidance provided in ISO 31000:2018, 5.4.5 applies.

5.5 Implementation

The guidance provided in ISO 31000:2018, 5.5 applies.

5.6 Evaluation

The guidance provided in ISO 31000:2018, 5.6 applies.

5.7 Improvement

5.7.1 Adapting

The guidance provided in ISO 31000:2018, 5.7.1 applies.

5.7.2 Continually improving

The guidance provided in ISO 31000:2018, 5.7.2 applies.

6 Risk management process

6.1 General

The guidance provided in ISO 31000:2018, 6.1 applies.

Organizations should implement a risk-based approach to identifying, assessing, and understanding the AI risks to which they are exposed and take appropriate treatment measures according to the level of risk. The success of the overall AI risk management process of an organization relies on the identification, establishment and the successful implementation of narrowly scoped risk management processes on strategic, operational, programme and project levels. Due to concerns related but not limited to the potential complexity, lack of transparency and unpredictability of some AI-based technologies, particular consideration should be given to risk management processes at the AI system project level. These system project level processes should be aligned with the organization's objectives and should be both informed by and inform other levels of risk management. For example, escalations and lessons learned at the AI project level should be incorporated at the higher levels, such as the strategic, operational and programme levels, and others as applicable.

The scope, context and criteria of a project-level risk management process are directly affected by the stages of the AI system's life cycle that are in the scope of the project. Annex C shows possible relations between a project-level risk management process and an AI system life cycle (as defined in ISO/IEC 22989:2022).

6.2 Communication and consultation

The guidance provided in ISO 31000:2018, 6.2 applies.

The set of stakeholders that can be affected by AI systems can be larger than initially foreseen, can include otherwise unconsidered external stakeholders and can extend to other parts of a society.

6.3 Scope, context and criteria

6.3.1 General

The guidance provided in ISO 31000:2018, 6.3.1 applies.

In addition to the guidance provided in ISO 31000:2018, 6.3.1, for organizations using AI the scope of the AI risk management, the context of the AI risk management process and the criteria to evaluate the significance of risk to support decision-making processes should be extended to identify where AI

systems are being developed or used in the organization. Such an inventory of AI development and use should be documented and included in the organization's risk management process.

6.3.2 Defining the scope

The guidance provided in ISO 31000:2018, 6.3.2 applies.

The scope should take the specific tasks and responsibilities of the different levels of an organization into account. Moreover, the objectives and purpose of the AI systems developed or used by the organization should be considered.

6.3.3 External and internal context

The guidance provided in ISO 31000:2018, 6.3.3 applies.

Because of the magnitude of potential effects of AI systems, the organization should pay special attention to the environment of its stakeholders when forming and establishing the context of the risk management process.

Care should be taken to consider a list of stakeholders, including, but not limited to:

- the organization (itself);
- customers, partners and third parties;
- suppliers;
- end users;
- regulators;
- civil organizations;
- individuals;
- affected communities;
- societies.

Some other considerations for external and internal context are:

- whether AI systems can harm human beings, deny essential services (which if interrupted would endanger life, health or personal safety) or infringe human rights (e.g. by unfair and biased automated decision-making) or contribute to environmental harm;
- external and internal expectations for the organization's social responsibility;
- external and internal expectations for the organization's environmental responsibility.

The guidelines in ISO 26000:2010^[2] outlining aspects of social responsibility should apply as a framework for understanding and treating risk, particularly on core subjects of organizational governance, human rights, labour practices, the environment, fair operating practices, consumer issues and community involvement and development.

NOTE Further background information on trustworthiness is available in ISO/IEC TR 24028:2020^[3].

6.3.4 Defining risk criteria

The guidance provided in ISO 31000:2018, 6.3.4 applies.

In addition to the guidance provided in ISO 31000:2018, 6.3.4, Table 4 provides additional guidelines on factors to be considered when defining risk criteria:

Table 4 — Additional guidance when defining risk criteria

Considerations for defining risk criteria as provided in ISO 31000:2018, 6.3.4	Additional considerations in the context of the development and use of AI systems
<ul style="list-style-type: none">— The nature and type of uncertainties that can affect outcomes and objectives (both tangible and intangible);— How consequences (both positive and negative) and likelihood will be defined and measured;	<ul style="list-style-type: none">— Organizations should take reasonable steps to understand uncertainty in all parts of the AI system, including the utilized data, software, mathematical models, physical extension, and human-in-the-loop aspects of the system (such as any related human activity during data collection and labelling).
<ul style="list-style-type: none">— Time-related factors;	<ul style="list-style-type: none">— No specific guidance beyond ISO 31000:2018
<ul style="list-style-type: none">— Consistency in the use of measurements;	<ul style="list-style-type: none">— Organizations should be aware that AI is a fast-moving technology domain. Measurement methods should be consistently evaluated according to their effectiveness and appropriateness for the AI systems in use.
<ul style="list-style-type: none">— How the level of risk is to be determined;	<ul style="list-style-type: none">— Organizations should establish a consistent approach to determine the risk level. The approach should reflect the potential impact of AI systems regarding different AI-related objectives (see Annex A).
<ul style="list-style-type: none">— How combinations and sequences of multiple risks will be taken into account;	<ul style="list-style-type: none">— No specific guidance beyond ISO 31000:2018
<ul style="list-style-type: none">— The organization's capacity.	<ul style="list-style-type: none">— The organization's AI capacity, knowledge level and ability to mitigate realized AI risks should be considered when deciding its AI risk appetite.

6.4 Risk assessment

6.4.1 General

The guidance provided in ISO 31000:2018, 6.4.1 applies.

AI risks should be identified, quantified or qualitatively described and prioritized against risk criteria and objectives relevant to the organization. Annex B provides a sample catalogue of AI-related risk sources. Such a sample catalogue cannot be considered comprehensive. However, experience has shown the value of using such a catalogue as base for any organization performing a risk assessment exercise for the first time or integrating AI risk management into existing management structures. The catalogue serves as a documented baseline for these organizations.

Organizations engaged in the development, provisioning or application of AI systems therefore should align their risk assessment activities with the system life cycle. Different methods for risk assessment can apply to different stages of the system life cycle.

6.4.2 Risk identification

6.4.2.1 General

The guidance provided in ISO 31000:2018, 6.4.2 applies.

6.4.2.2 Identification of assets and their value

The organization should identify assets related to the design and use of AI that fall within the scope of the risk management process as defined in 6.3.2. Understanding what assets are within the scope and the relative criticality or value of those assets is integral to assessing the impact. Both the value of the asset and the nature of the asset (tangible or intangible) should be considered. Additionally, in relation

to the development and use of AI, assets should be considered in the context of elements including but not limited to the following:

- Assets of and their value to the organization:
 - Tangible assets can include data, models and the AI system itself.
 - Intangible assets can include reputation and trust.
- Assets of and their value to individuals:
 - Tangible assets can include an individual's personal data,
 - Intangible assets can include privacy, health, and safety of an individual.
- Assets of and their value to communities and societies:
 - Tangible assets can include the environment,
 - Intangible assets are likely more values based, such as socio-cultural beliefs, community knowledge, educational access and equity.

For valuation of assets and the relation to impact, see 6.4.2.6 and 6.4.3.2.

NOTE The use of the word "asset" with the illustrative examples in this clause does not have any legal implications.

6.4.2.3 Identification of risk sources

The organization should identify a list of risk sources related to the development or use of AI, or both, within the defined scope.

Risk sources can be identified within, but not limited to, the following areas:

- organization;
- processes and procedures;
- management routines;
- personnel;
- physical environment;
- data;
- AI system configuration;
- deployment environment;
- hardware, software, network resources and services;
- dependence on external parties.

Examples of AI-related risk sources can be found in Annex B.

6.4.2.4 Identification of potential events and outcomes

The organization should identify potential events that are related to the development or use of AI and can result in a variety of tangible or intangible consequences.

Events can be identified through one or more of the following methods and sources:

- published standards;

- published technical specifications;
- published technical reports;
- published scientific papers;
- market data on similar systems or applications already in use;
- reports of incidents on similar systems or applications already in use;
- field trials;
- usability studies;
- the results of appropriate investigations;
- stakeholder reports;
- interviews with, and reports from, internal or external experts;
- simulations.

6.4.2.5 Identification of controls

The organization should identify controls relevant to either the development or use of AI, or both. Controls should be identified during the risk management activities and documented (in internal systems, procedures, audit reports, etc.).

Controls can be utilized to positively affect the overall risk by mitigating risk sources and events and outcomes.

The operating effectiveness of the identified controls should also be taken into account, particularly control failures.

6.4.2.6 Identification of consequences

As part of AI risk assessment, the organization should identify risk sources, events or outcomes that can lead to risks. It should also identify any consequences to the organization itself, to individuals, communities, groups and societies. Organizations should take particular care to identify any differences between the groups who experience the benefits of the technology and the groups who experience negative consequences.

Consequences to the organization necessarily differ from those to individuals and to societies. Consequences to organizations can include but are not limited to:

- investigation and repair time;
- (work) time gained and lost;
- opportunities gained or lost;
- threats to health or safety of individuals;
- financial costs of specific skills to repair the damage;
- employee recruitment, satisfaction and retention;
- image reputation and goodwill;
- penalties and fines;
- customer litigations.

Depending on the context, consequences to individuals and to societies can be more general, in which case the organization can be unable to estimate exactly what the effect to each individual or to societies is.

Rather than specifying each type of effect, this can be considered generally as with the degree of the criticality of effects (for example, to privacy, fairness, human rights, etc., in the case of an individual, or to the environment in the case of societies) being a key element.

The exact effects can depend on the context in which the organization operates and areas for which the AI system is developed and used.

NOTE 1 Consequences can be positive or negative. The organization can consider both when assessing the consequences to the organization, to individuals and to societies.

NOTE 2 Consequences to individuals and societies usually can also lead to consequences to the organization. For example, a safety incident to a user of a product of the organization can result in liability claims to the organization and negatively impact its reputation and product sales.

6.4.3 Risk analysis

6.4.3.1 General

The guidance provided in ISO 31000:2018, 6.4.3 applies.

The analysis approach should be consistent with the risk criteria developed as part of establishing the context (see 6.3).

6.4.3.2 Assessment of consequences

When assessing the consequences identified in the risk assessment, the organization should distinguish between a business impact assessment, an impact assessment for individuals and a societal impact assessment.

Business impact analyses should determine the degree to which the organization is affected, and consider elements including but not limited to the following:

- criticality of the impact;
- tangible and intangible impacts;
- criteria used to establish the overall impact (as determined in 6.3.4).

Impact analyses for individuals should determine the degree to which an individual can be affected by the development or use of AI by the organization, or both. They should consider elements including but not limited to the following:

- types of data used from the individuals;
- intended impact of the development or use of AI;
- potential bias impact to an individual;
- potential impact on fundamental rights that can result in material and non-material damage to an individual;
- potential fairness impact to an individual;
- safety of an individual;
- protections and mitigating controls around unwanted bias and unfairness;

- jurisdictional and cultural environment of the individual (which can affect how relative impact is determined).

Impact analyses for societies should determine the degree to which societies can be affected by the either development or use of AI by the organization, or both. They should consider elements including but not limited to the following:

- scope of societal impact (how broad is the reach of the AI system into different populations), including who the system is being used by or designed for (for instance, governmental use can potentially impact societies more than private use);
- how an AI system affects social and cultural values held by various affected groups (including specific ways that the AI system amplifies or reduces pre-existing patterns of harm to different social groups).

6.4.3.3 Assessment of likelihood

Where applicable, the organization should assess the likelihood of occurrence of events and outcomes causing risks. Likelihood can be determined on a qualitative or quantitative scale and should align to the criteria established as part of 6.3.4. Likelihood can be informed and affected by (not limited to):

- types, significance, and number of risk sources;
- frequency, severity, and pervasiveness of threats;
- internal factors such as operational success of policies and procedures and motivation of internal actors;
- external factors such as geography and other social, economic and environmental concerns;
- success (mitigation) or failure of controls (see 6.4.2.5).

Organizations should incorporate likelihood calculations only where they are applicable and useful for identifying where to apply risk treatments. There can be significant technical, economic and heuristic issues with decision-making based likelihoods, particularly when the likelihood either can't be calculated or where the calculation has a large margin of error.

6.4.4 Risk evaluation

The guidance provided in ISO 31000:2018, 6.4.4 applies.

6.5 Risk treatment

6.5.1 General

The guidance provided in ISO 31000:2018, 6.5.1 applies.

6.5.2 Selection of risk treatment options

The guidance provided in ISO 31000:2018, 6.5.2 applies.

Risk treatment options defined by the organization should be designed to reduce negative consequences of risks to an acceptable level, and to increase the likelihood that positive outcomes can be achieved. If the required reduction of negative outcomes cannot be achieved by applying different risk treatment options, the organization should carry out a risk-benefit analysis for the residual risks.

In accordance with ISO 31000:2018, 6.5.2 the organization should consider:

- avoiding the risk by deciding not to start or continue with the activity that gives rise to the risk;
- taking or increasing the risk in order to pursue an opportunity;

- removing the risk source;
- changing the likelihood;
- changing the consequences;
- sharing the risk (for instance, through contracts or buying insurance);
- retaining the risk by informed decision.

6.5.3 Preparing and implementing risk treatment plans

The guidance provided in ISO 31000:2018, 6.5.3 applies.

Once the risk treatment plan has been documented, the risk treatment measures selected in 6.5.2 should be implemented.

The implementation of each risk treatment measure and its effectiveness should be verified and recorded according to 6.7.

6.6 Monitoring and review

The guidance provided in ISO 31000:2018, 6.6 applies.

6.7 Recording and reporting

The guidance provided in ISO 31000:2018, 6.7 applies.

The organization should establish, record, and maintain a system for the collection and verification of information on the product or similar products from the implementation and post-implementation phases. The organization should also collect and review publicly available information on similar systems on the market.

This information should then be assessed for possible relevance on the trustworthiness of the AI system. In particular, the evaluation should assess whether previously undetected risks exist or previously assessed risks are no longer acceptable. This information can be fed and factored into the organization's AI risk management process as adjustment of objectives, use cases or lessons learned.

If any of these conditions apply, organizations should perform the following:

- assess the effect on previous risk management activities and feed the results of this assessment back into the risk management process.
- carry out a review of the risk management activities for the AI system. If there is a possibility that the residual risk or their acceptance have changed, the effects on existing risk control measures should be evaluated.

The results of this assessment should be recorded. The risk management record should allow the traceability of each identified risk through all risk management processes. The records can leverage a common template that is agreed upon by the organization.

In addition to the documentation of the scope, context and criteria (see 6.3), risk assessment (see 6.4) and risk treatment (see 6.5), the record should include at least the following information:

- a description and identification of the system that has been analysed;
- methodology applied;
- a description of the intended use of the AI system;
- the identity of the person(s) and organization that carried out the risk assessment;

- the terms of reference and date of the risk assessment;
- the release status of the risk assessment;
- if and to what degree objectives have been met.

Annex A

(informative)

Objectives

A.1 General

When identifying risks of AI systems, various AI-related objectives should be taken into account, depending on the nature of the system under consideration and its application context. AI-related objectives to consider include but are not limited to the objectives described in Clauses A.2 to A.12.

A.2 Accountability

Accountability refers both to a characteristic of organizations and to a system property:

- Organizational accountability means that an organization takes responsibility for its decisions and actions by explaining them and being answerable for them to the governing body, to legal authorities and more broadly to stakeholders.
- System accountability relates to being able to trace the decisions and actions of an entity to that entity.

The use of AI can change existing accountability frameworks. Where previously persons performed actions for which they would be held accountable, such actions can now be fully or partially performed by AI systems. Who would be accountable in this case is an ongoing consideration by regulators. Developers and users of AI systems should be aware of the related legislation in the countries where the system is brought onto the market and used.

A.3 AI expertise

AI systems and their development are different from non-AI software solutions. A selection of dedicated specialists with inter-disciplinary skillsets and expertise in assessing, developing and deploying AI systems is needed. Organizations should ensure that people with such expertise are engaged in the development and specification of AI systems.

Expertise of AI should extend to the end users of AI systems. Users should have sufficient understanding of how the AI system functions and are empowered to detect and override erroneous decisions or outputs.

A.4 Availability and quality of training and test data

AI systems based on ML need training and test data in order to train and verify the systems for the intended behaviour. The deployed AI system operates on production data. The training, test and production data should be fit to the intended behaviour with respect to data type and quality.

Training and test data should be validated for their currency and relevance for the intended purpose. The amount of training and test data required can vary based on the intended functionality and complexity of the environment. The training and test data should have sufficiently diverse features in order to provide strong predictive power for the AI system. Furthermore, consistency should be ensured across training and test data, while using independent datasets when applicable.

It is possible that training and test data is not available in the company and is sourced externally. Data quality should be ensured also in that case.

A.5 Environmental impact

The use of AI can cause effects from an environmental point of view. The use of AI can have positive effects on the environment. For example, an AI system can be used to reduce nitrogen oxide in a gas turbine. The use of AI can also have a negative effect on the environment due to the extensive use of resources. For example, the training phase of some AI systems requires computing resources and can consume substantial amounts of electrical power. These impacts on the environment should be considered.

A.6 Fairness

The use of AI systems for automated decision-making can be unfair to specific persons or groups of persons. Unfair outcomes have a number of causes such as bias in objective functions, imbalanced data sets, and human biases in training data and in providing feedback to systems. Unfairness can also be caused by a bias issue in the product concept, the problem formulation or choices about when and where to deploy AI systems.

For further information on bias and fairness in AI systems, see ISO/IEC TR 24027^[4].

A.7 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. Because AI systems based on ML are trained and do not follow a rule-based approach, the maintainability of an AI system and its implications should be investigated.

A.8 Privacy

Privacy is related to the ability of individuals to control or influence what information related to them can be collected, stored and processed, and by whom that information can be disclosed. As explained in ISO/IEC TR 24028:2020^[3], "many AI techniques (e.g. deep learning) highly depend on big data since their accuracy relies on the amount of data they use. The misuse or disclosure of some data, particularly personal and sensitive data (e.g. health records) can have harmful effects on data subjects. Thus, privacy protection has become a major concern in big data analysis and AI."

Consideration should be taken to determine if an AI system can infer sensitive personal data. For AI systems, protecting privacy includes protecting the data used for building and operating the AI system, ensuring that the AI system cannot be used to give unwarranted access to its data, and protecting access to models personalized for an individual or that can be used to infer information or characteristics of similar individuals.

Improper collections, uses and disclosures of personal information can also have direct impacts on fundamental human rights such as discrimination and freedom of expression and information. Impacts on ethic principles in terms of respect of human values, and human dignity should also be considered.

NOTE A data protection impact assessment (see ISO/IEC 29134:2017^[5]), often referred to as a privacy impact assessment, is a useful tool for managing the risks related to the use of personal data during the collection of data, training of an AI system, and use of an AI system.

A.9 Robustness

Robustness is related to the ability of a system to maintain its level of performance under the various circumstances of its usage. The degree to which an AI system or related component can function correctly in the presence of invalid inputs or stressful environmental conditions should be taken into consideration as well as the ability to reproduce measures and results.

Robustness poses new challenges in the context of AI systems. Neural network architectures represent a specific challenge as they are both hard to explain and sometimes have unexpected behaviour due to their nonlinear nature. Characterizing the robustness of neural networks is an open area of research, and there are limitations to both testing and verification approaches.

For further information on robustness of neural networks, see ISO/IEC TR 24029-1^[6].

A.10 Safety

The use of AI systems can introduce new safety threats. 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. Use of AI systems in automated vehicles, manufacturing devices, and robots can introduce risks related to safety. Specific standards for particular application domains (e.g. the design of machinery, transport, medical devices) should be taken into account for AI systems in these domains.

For further information on functional safety in AI systems, see ISO/IEC TR 5469¹⁾ [7].

A.11 Security

Information security risk management is defined in ISO/IEC 27005:2022^[8]. In the context of AI, and in particular with regard to AI systems based on ML approaches, several new issues such as data poisoning, adversarial attacks and model stealing as described in ISO/IEC TR 24028:2020^[3] should be considered beyond classical information and system security concerns.

A.12 Transparency and explainability

Transparency relates both to characteristics of an organization operating AI systems, and to those systems themselves. Organizations are sometimes transparent on how they apply such systems, how they use collected data (such as consumer and user data, public data, other collected data sets), which measures they put in place to manage AI systems, understand and control their risks, etc. Transparency of AI systems is to provide appropriate information about a system (e.g. capabilities and limitations) to stakeholders to enable them to assess development, operation and use of AI systems against their objectives. AI system explainability relates to an ability to rationalize and help to understand how, for a specific system, its outcome has been generated.

1) Under preparation. Stage at the time of publication: ISO/IEC DTR 5469:2022.

Annex B (informative)

Risk sources

B.1 General

When identifying risks of AI systems, various risks sources should be taken into account depending on the nature of the system under consideration and its application context. Risk sources to consider include but are not limited to the issues and opportunities described in Clauses B.2 to B.8.

B.2 Complexity of environment

The complexity of the environment^[9] of an AI system determines the range of potential situations an AI system is intended to support in its operational context.

Certain AI technologies like ML are specifically suited to handle complex environments and are therefore often used for systems used for complex environments like automated driving. A great challenge however is to identify in the design and development process all relevant situations that the system is expected to handle and that the training and test data cover all these situations.

Hence, complex environments can result in additional risks relative to simple environments. Special consideration should be given to determining the degree to which the AI system environment is understood:

- Complete understanding of environment that is only possible for simple predictable or controlled environments, such that the AI system is prepared for all possible states of the environment that it can encounter, allows for better risk control.
- In case of partial understanding due to high complexity or uncertainty of the environment, such that the AI system cannot forecast all possible states of the environment (for instance, autonomous driving), it cannot be assumed that all relevant situations are considered. This can result in a level of uncertainty, which is a source of risk, and should be taken into account when designing such systems.

B.3 Lack of transparency and explainability

Transparency is about communicating appropriate activities and decisions of an organization (e.g. policies, processes) and appropriate information about an AI system (e.g. capabilities, performance, limitations, design choices, algorithms, training and test data, verification and validation processes and results) to relevant stakeholders. This can enable stakeholders to assess development, operation and use of AI systems against their expectations. The kind and level of information that is appropriate strongly depends on the stakeholders, use case, system type and legislative requirements. If organizations are unable to provide the appropriate information to the relevant stakeholders, it can negatively affect trustworthiness and accountability of the organization and AI system.

Explainability is the property of an AI system that the important factors influencing a decision can be expressed in a way that humans can understand. An ML model can have behaviour that is difficult to understand by inspection of the model or the algorithm used to train it, especially in the case of deep learning. If such important factors cannot be expressed, validation of the AI system and the trust of humans in the system are negatively affected as it is not clear why the system has made a decision and if it can make the correct decision in all cases. This uncertainty can result in many risks and strongly effect general objectives such as trustworthiness and accountability, and specific objectives such as

safety, security, fairness and robustness. Explainability is therefore not only relevant for stakeholders as part of AI system transparency but also for the organization itself for its own validation and verification of the AI system.

Excessive transparency and explainability can also lead to risks in relation to privacy, security, confidentiality requirements and intellectual properties.

B.4 Level of automation

AI systems can operate with different levels of automation. They can range from no automation where an operator fully controls the system to fully automated systems. AI systems are often automated systems. Depending on the specific use case, the automated decisions of such systems can have an effect on various areas of concern such as safety, fairness or security.

For a level of automation where an external agent must be ready when necessary, the handover from the system to the agent can be a risk source (e.g. time constraints, attention of the agent).

For further information on levels of automation, see ISO/IEC 22989:2022, 5.2.

B.5 Risk sources related to machine learning

Many advances in AI are related to ML and subfields thereof such as deep learning. The behaviour of ML systems is critically dependent not just on the algorithms in use but also on the data on which the ML models are trained. Therefore, possible effects on AI characteristics include:

- Data quality: The quality of training and test data directly affects the functionality of the system. Inadequate data quality can affect various objectives such as fairness, safety and robustness.
- For AI systems utilizing ML, the processes used to collect data are a source of risks that are especially hard to diagnose and detect. For example:
 - Data can become unrepresentative of the domain of application, leading to risks to business objectives.
 - Data sourcing and storage can incur significant ethical and legal risks. Failing to secure the data collection process can lead to risks from adversarial attacks, data poisoning or other manipulation.
- Continuous learning AI systems intends to improve the systems on the basis of the evolving production data, at the same time can exacerbate risk as they can change their behaviour during use in a way that was not expected when it was brought into use.

B.6 System hardware issues

Risk sources related to hardware issues include, but are not limited to:

- Hardware errors based on defective components. Examples are short circuits or interruptions of single or multiple memory cells, defective bus lines, drifting oscillators, stuck-at faults or parasitic oscillations at the inputs or outputs of integrated circuits.
- Soft errors such as unwanted temporary state changes of memory cells or logic components, mostly caused by high energy radiation.
- Transferring trained ML models between different systems can be constrained due to differing hardware capabilities of the systems in terms of processing power, memory and the availability of dedicated AI hardware accelerators.
- When an AI system requires remote processing and storage, network errors, bandwidth restrictions and increased latency due to the limited and shared nature of network resources.

B.7 System life cycle issues

Inappropriate or insufficient methods, processes and also usage of an AI system along its life cycle can lead to risks. Examples of such risks are:

- Design and development: A flawed design process can fail to anticipate the contexts in which the AI system is used, causing it to fail unexpectedly when used in these contexts.
- Verification and validation: An inadequate verification and validation process for releasing updated versions of the AI system can lead to accidental regressions or unintended deterioration or degradation in quality, reliability or safety.
- Deployment: An inadequate deployment configuration can lead to resource problems related to memory, compute, network, storage, redundancy or load balancing.
- Maintenance, update and revision: An AI system no longer supported or maintained by the developer but still in use can present long-term risks or liability to the developing organization.
- Reuse: A functioning AI system can be used in a context for which it was not originally designed, causing problems due to differing requirements between the designed and actual use. For example, a system designed for identifying faces in photos shared on a social network can be used to attempt to identify faces of criminal suspects in surveillance footage, an application that requires a much higher degree of precision than the original use case.
- Decommissioning: Organizations that terminate the use of a certain AI system or a component based on AI technologies can lose information or decision expertise that have been provided by the decommissioned system. Moreover, if another system is used to replace the decommissioned one, the way an organization processes information and makes decisions can change.

B.8 Technology readiness

Technology readiness indicates how mature a given technology is in a given application context. Less mature technologies used in the development and application of AI systems can impose risks that are unknown to the organization or are hard to assess. For mature technologies a larger variety of experience data can be available, making risks easier to identify and to assess. However, there is also a risk of complacency and technical debt if technologies are mature.

Annex C (informative)

Risk management and AI system life cycle

Table C.1 shows an example of a mapping between the risk management processes and an AI system life cycle as defined in ISO/IEC 22989:2022.

Table C.1 — Risk management and AI system life cycle

→ Risk management	AI risk management framework (Clause 5)	AI risk management process (Clause 6)				
AI system Life cycle ↓		Scope, context and criteria	Risk assessment	Risk treatment	Monitoring and review	Recording and reporting
Organizational level activities related to risk management	Governing body sets directions for AI risk management.	Feedback reports from AI systems’ risk management processes are being received and processed.				
	Top management commits. High-level risk management appetite and general criteria are established.	As a result, the organizational risk management framework is being improved by extending and refining of the organization’s risk management tools:				
		A catalogue of risk criteria.	A catalogue of potential risk sources. A catalogue of techniques for risk sources’ assessment and measurements.	A catalogue of known or implemented mitigation measures.	A catalogue of known or implemented techniques for monitoring and controlling AI systems.	A catalogue of established methods and defined formats for tracing, recording, reporting, and sharing the information about AI systems with internal and external stakeholders.
Inception	Governing body examines the AI system objectives in the context of the organization’s and the stakeholders’ principles and values, Based on a (typically multi-layer) analysis, determines whether the AI system is feasible and addresses the problem the organization seeks to solve.	The AI system risk management process and the system’s risk criteria are established through customization of the organization’s risk management framework.	Risk sources specific to the AI system are identified (potentially in a multi-layered manner) and described in detail.	A detailed risk treatment plan is established. Potentially, “proof of concept” methods are defined.	Necessary “proof of concept” methods are implemented, tested and evaluated.	The analysis with its results and the recommendation are recorded and communicated to the top management.
Design and development	Governing body continually re-assesses the objectives, the efficacy and the feasibility of the system based on received feedback reports.	Potentially, the AI system risk criteria is modified as a result of the feedback reports.	The risk assessment is performed continuously (potentially on multiple layers).	The risk treatment plan is implemented. The risk treatment and the (remaining) risks assessment continue until the established risk criteria are met.	During the testing, verification and validation the risk treatment plan for the system’s components as well as for the whole system is assessed and adjusted.	The results are recorded and fed back to the relevant risk management process activities. As necessary, the conclusions are communicated to the management chain and to the governance body.
Verification and validation						

Table C.1 (continued)

→ Risk management AI system Life cycle ↓	AI risk management framework (Clause 5)	AI risk management process (Clause 6)				
		Scope, context and criteria	Risk assessment	Risk treatment	Monitoring and review	Recording and reporting
Deployment	Governing body continually re-assesses the objectives and the feasibility of the system based on received feedback reports.	The AI system risk criteria and the risk management process are adjusted for the necessary "configuration" changes.	The risk assessment is performed continuously (potentially on multiple layers).	The risk treatment plan is potentially updated due to "configuration" changes and implemented. The risk treatment and the (remaining) risks assessment continue until the established risk criteria are met.	The AI system's risk treatment plan is being re-assessed to allow for necessary adjustments.	
Operation, monitoring Continuous validation	Governing body continually re-assesses the objectives and the feasibility of the system based on received feedback reports.	Potentially, the AI system risk criteria is modified as a result of the feedback reports.	The system's risk assessment plan is potentially adjusted for risk criteria changes.	The system's risk treatment plan is potentially adjusted for risk changes in risk assessment outcomes.	The risk treatment plan for the system's components is assessed and adjusted.	
Re-evaluation	Governing body re-examines the AI system objectives and their relation to the organization's and the stakeholders' principles and values, Based on the analysis, determines whether the AI system is feasible.	The AI system risk management process and the system's risk criteria are re-evaluated against any potential changes to the specific purpose and scope of the AI system, outcome of operation monitoring and new regulatory requirements	The list of existing risk sources specific to the AI system are examined for relevance and any possible gaps.	The risk treatment plan is potentially updated. The risk treatment and the (remaining) risks assessments continue until the established risk criteria are met.	The AI system's risk treatment plan is being re-assessed to allow for necessary adjustments.	
Retirement or replacement Triggers a new risk management process with new objectives, risks and their mitigation.	Governing body re-examines the AI system objectives based on the analysis, determines whether the AI system retirement or replacement is feasible.	The AI system risk management retirement process and the system's retirement risk criteria are established.	Risk sources specific to the AI system retirement are identified and described in detail.	Detailed risk treatment plan is established.	Necessary "proof of concept" methods are implemented, tested and evaluated.	

Bibliography

- [1] ISO/IEC 38507:2022, *Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations*
- [2] ISO 26000:2010, *Guidance on social responsibility*
- [3] ISO/IEC TR 24028:2020, *Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*
- [4] ISO/IEC TR 24027:2021, *Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*
- [5] ISO/IEC 29134:2017, *Information technology — Security techniques — Guidelines for privacy impact assessment*
- [6] ISO/IEC TR 24029-1:2021, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*
- [7] ISO/IEC TR 5469²⁾, *Artificial intelligence — Functional safety and AI systems*
- [8] ISO/IEC 27005:2022, *Information security, cybersecurity and privacy protection — Guidance on managing information security risks*
- [9] RUSSELL S. J., NORVIG P., *Artificial intelligence: a modern approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010

2) Under preparation. Stage at the time of publication: ISO/IEC DTR 5469:2022.