目　　录

国际ISO/IEC 标准 42001

第 一 版
2023-12

信息技术——人工智能——管理系统

参考编号 ISO/IEC 42001：2023(E)

受版权保护的文档

国际版权局
CP 401 • Ch. de Blandonnet 8
CH-1214 日内瓦游标卡尺
电话：+41 22 749 01 11
电子邮箱： copyright@iso.org
网站：www.iso.org

在瑞士出版

目录

页

前言

国际标准化组织（ISO）与国际电工委员会（IEC）共同构建了全球标准化体系。作为ISO或IEC成员的各国机构，通过各自组织设立的技术委员会参与国际标准制定，这些委员会专门负责特定技术领域的标准化工作。ISO与IEC的技术委员会在共同关注的领域开展协作。此外，其他国际组织（包括政府机构和非政府组织）在与ISO及IEC保持联络后，也会参与相关标准化工作。

ISO/IEC指令第1部分描述了本文件编制和后续维护所用的程序，特别是不同类型的文件需要不同的批准标准，本文件是按照ISO/IEC指令第2部分的编辑规则起草的（参见www.iso.org/directives或www.iec.ch/members_experts/refdocs）。

ISO与IEC提醒注意，本文件的实施可能涉及使用(a)项专利。ISO与IEC不对相关专利权利的有效性、适用性或证据真实性表明立场。截至本文件发布时，ISO与IEC尚未收到可能需要实施本文件的(a)项专利通知。但需特别提醒实施方，这些信息可能并非最新动态，相关数据可从www.iso.org/patents和https://patents.iec.ch的专利数据库获取。ISO与IEC不对识别任何或全部此类专利权利承担法律责任。

本文件中使用的任何商品名称均为为方便用户而提供的信息，不构成认可。

有关标准的自愿性质、与合格评定相关的ISO特定术语和表达的含义以及ISO对世界贸易组织（WTO）技术性贸易壁垒（TBT）原则的遵守情况的信息，请参阅www.iso.org/iso/foreword.html。在IEC中，参见www.iec.ch/understanding-standards。

本文件由联合技术委员会ISO/IEC JTC 1，*信息技术*，小组委员会SC 42，*人工智能*编制。

有关此文档的任何反馈或问题应直接发送至用户所在国的标准化机构。可在www.iso.org/members.html和www.iec.ch/national-committees上找到这些机构的完整列表。

介绍

人工智能（AI）正日益广泛地应用于所有采用信息技术的领域，并有望成为主要的经济驱动力之一。这一趋势的后果是，某些应用在未来几年可能引发社会挑战。

本文件旨在帮助组织负责任地履行其在人工智能系统方面的作用（例如使用、开发、监控或提供利用人工智能的产品或服务）。人工智能可能引发特定考量因素，例如：

— 人工智能（AI）用于自动决策时，若以不透明且不可解释的方式运作，可能需要超出传统IT系统管理范畴的专门管理措施。

— 采用数据分析、洞察力和机器学习而非人工编码逻辑来设计系统，既拓展了人工智能系统的应用前景，又改变了这类系统开发、论证和部署的方式。

— 持续学习的AI系统在使用过程中会改变其行为。需要特别考虑如何确保其行为变化时仍能继续负责任地使用。

本文件为组织在建立、实施、维护及持续改进人工智能管理体系方面提供了规范要求。各组织在应用这些要求时，应重点关注人工智能特有的功能特性。对于人工智能的某些特性——例如持续学习与优化的能力，或缺乏透明度与可解释性——若其引发的额外顾虑超过传统工作方式的考量，可能需要采取不同的保障措施。采用人工智能管理体系来扩展现有管理架构，是组织层面需要做出的战略决策。

该组织的需求与目标、流程、规模与结构， 以及各利益相关方的期望，均影响着人工智能管理体系的建立与实施。另一组影响因素包括人工智能的多种应用场景， 以及在治理机制与创新之间寻求适当平衡的必要性。各组织可选择采用基于风险的方法来实施这些要求， 以确保针对组织范围内的特定人工智能用例、服务或产品实施适当级别的控制。所有这些影响因素均需定期变更并进行审查。

人工智能管理系统需与组织流程及整体管理体系深度融合。在流程设计、信息系统构建及管控措施制定时，应充分考量人工智能相关具体问题。此类管理流程的关键范例包括：

— 确定组织目标、利益相关方参与及组织政策；

— 管理风险与机遇；

— 管理人工智能系统可信度相关问题的流程，包括安全、公平、透明、数据质量及人工智能系统全生命周期质量；

— 用于管理供应商、合作伙伴及第三方的流程，这些第三方为组织提供或开发人工智能系统。

本文件为实施相关控制措施以支持此类流程提供了指导原则。

本文件不提供管理流程的具体指导。组织可综合运用通用框架、其他国际标准及自身经验，实施适用于特定人工智能应用场景、产品或服务范围内的关键流程，如风险管理、生命周期管理和数据质量管理。

符合本文件要求的组织可生成其在人工智能系统相关职责中责任与问责的证据。

本文件中要求的呈现顺序并不反映其重要性，亦不暗示实施顺序。所列项目仅为参考目的。

与其他管理体系标准的兼容性

本文件采用统一结构框架（包括相同的条款编号、条款标题、文本内容及通用术语与核心定义），旨在提升管理体系标准（MSS）之间的协调一致性。人工智能管理体系针对组织使用人工智能时产生的问题与风险，制定了专门的管理要求。这种统一标准的实施方式，有助于确保与质量、安全、保密及隐私等相关管理体系标准的兼容性与一致性。

信息技术——人工智能——管理系统

1 范围

本文件规定了在组织背景下建立、实施、维护及持续改进人工智能（AI）管理体系的要求与指导原则。

本文件适用于提供或使用人工智能系统产品及服务的组织，旨在帮助其负责任地开发、提供或使用人工智能系统，以实现组织目标，满足相关要求、利益相关方义务及预期。

本文件适用于任何组织，无论其规模、类型和性质如何，只要该组织提供或使用利用人工智能系统的产品或服务。

2 规范性参考文献

本文引用下列文件时，其部分内容或全部内容均构成本文件的要求。对于标注日期的引用，仅引用的版本适用；对于未标注日期的引用，适用所引用文件的最新版本（包括任何修订）。

ISO/IEC 22989：2022，*信息技术——人工智能——人工智能概念与术语*

3 术语和定义

就本文件而言，适用ISO/IEC 22989及下述条款中给出的术语与定义。

ISO和IEC维护术语数据库，用于标准化工作，地址如下：

- ISO在线浏览平台：可在https://www.iso.org/obp获取

— IEC Electropedia：可在https://www.electropedia.org/获取

3.1
组织
具有自己的职能、责任、权限和关系以实现其*目标*的个人或群体（3.6）

注1：组织概念包括但不限于以下形式：个体经营者、公司、法人实体、企业、机构、合伙企业、慈善组织或非营利机构，无论是否注册成立，无论公私性质。

注2：如果组织是较大实体的一部分，则术语“组织”仅指较大实体中在*人工智能管理系统*范围内的部分（3.4）。

3.2
有关的当事人
人或*组织*（3.1）可能受决策或活动影响、被其影响或自认为受影响的组织

条目注释1：ISO/IEC 22989：2022标准第5.19节对人工智能相关利益方进行了概述。

3.3
高管理层
组织（3.1）最高层的指挥和控制者

条目注释1：最高管理层有权在组织内部授权并提供资源。

注2：如果*管理体系*（3.4）的范围仅涵盖组织的一部分，则最高管理层指指导和控制该组织部分的人员。

3.4
管理系统
组织（3.1）为制定*政策*（3.5）和 *目标*（3.6）， 以及实现这些目标的*过程*（3.8）而相互关联或相互作用的要素的集合

条目注释1：管理系统可针对单一学科或多个学科。

条目注释2：管理体系要素包括组织结构、职责分工、规划与运营。

3.5
政策
*组织*的*最高管理层*正式表达的*意图和方向*（3.1）

3.6
客观的
预期结果

条目注释1： 目标可分为战略、战术或操作层面。

注2： 目标可以涉及不同学科（如财务、健康和安全以及环境）。例如， 目标可以是整个组织的，也可以是针对项目、产品或*流程*的（3.8）。

条目注释3： 目标可通过其他方式表述，例如作为预期结果、 目的、操作标准、人工智能目标，或使用具有相似含义的其他词汇（如aim、goal或target）。

注4：在*管理系统*（3.4）方面，*组织*（3.1）根据与人工智能*政策*（3.5）一致，设定人工智能目标，以实现特定结果。

3.7
风险
不确定性效应

条目注释1：效应是指与预期值（正向或负向）的偏离。

条目注释2：不确定性是指与某一事件、其后果或可能性相关的、理解或知识方面信息的不足状态，即使这种不足是部分的。

条目注释3：风险通常通过潜在事件（如ISO指南73所定义）及其后果（如ISO指南73所定义）或两者的组合来表征。

条目注释4：风险通常以事件后果（包括环境变化）与发生可能性（如ISO指南73所定义）的组合形式表示。

3.8
过程
一组相互关联或交互的活动，通过使用或转化输入来实现结果

注1：一个过程的结果是否称为输出、产品或服务取决于参考的上下文。

3.9
能力
运用知识与技能达成预期目标的能力

3.10
有据信息
组织（3.1）必须控制和维护的信息以及所包含的信息介质

条目注释1：记录的信息可以采用任何格式和媒介形式，并且可以来自任何来源。

条目注释2：记录的信息可参考：

管理制度（3.4），包括相关*流程*（3.8）；

— 为组织运作而创建的信息（文档）；

— 成果证据（记录）。

3.11
表演
可测结果

条目注释1：性能可涉及定量或定性结果。

注2：业绩可涉及管理活动、*流程*（3.8）、产品、服务、系统或*组织*（3.1）。

注3：在本文件中，业绩包括使用人工智能系统取得的成果和与人工智能*管理系统*相关的成果（3.4）。该术语的正确解释可从其使用语境中明确得出。

3.12
持续改进
增强*性能*的重复活动（3.11）

3.13
有效
计划活动的实现程度与计划结果达成情况

3.14
要求
已声明、通常默示或具有法律约束力的需要或期望

注1：“一般含意”是指*组织*（3.1）和*有关方面*（3.2）的惯例或通常做法是，所考虑的需要或期望是含意的。

条目注2：规定的要求是已陈述的要求，例如在*文件信息*（3.10）中。

3.15
遵从
要求（3.14）的履行

3.16
新教教徒
*要求*未获满足（3.14）

3.17
校正动作
消除*不符合原因*（3.16）并防止再次发生

3.18
审计
获取证据并进行客观评价以确定达到审计准则的程度的*系统和独立过程*（3.8）

条目注释1：审计可为内部审计（第一方）或外部审计（第二方或第三方），亦可为复合审计（整合两个或多个学科）。

注2：内部审计由*组织*（3.1）本身或由外部一方代表其进行。

注3：条目中“审计证据”与“审计准则”的定义见ISO 19011标准。

3.19
量度
过程（3.8）以确定一个值

3.20
监视
确定系统、*进程*（3.8）或活动的状态

条目注释1：为确定状态，可能需要进行检查、监督或严格观察。

3.21
统治
维持和/或修改风险的措施（3.7）

条目注释1：控制措施包括但不限于任何维持和/或改变风险的过程、政策、设备、实践或其他条件和/或行动。

注2：控制措施未必总能产生预期或假定的调节效应。

[来源：ISO 31000：2018,3.8，修订—— <risk> 作为应用领域增加]

3.22
主管团体
对组织绩效与合规性负责的个人或群体

条目注释1：并非所有组织（尤其是小型组织）都设有独立于最高管理层的管理机构。

条目注释2：管理机构可包括但不限于董事会、董事会委员会、监事会、受托人或监管机构。

[来源：ISO/IEC 38500：2015,2.9，修订版——新增条目注释。]

3.23
信息安全
保密性、完整性和信息可用性的维护

注1：其他特性，如真实性、可问责性、不可抵赖性和可靠性，也可涉及。

[来源：ISO/IEC 27000：2018,3.28]

3.24
人工智能系统影响评估
指组织在开发、提供或使用人工智能产品或服务时，通过正式且有据可查的流程，对个人、群体或两者以及社会所产生的影响进行识别、评估和应对。

3.25
数据质量

数据满足组织特定上下文数据要求的特性[来源：ISO/IEC 5259-1：—1)3.4]

3.26
适用性声明
所有必要*控制措施*的文件记录（ 3.23）以及控制措施的纳入或排除依据

注1：组织可能不需要附录A列出的所有控制措施，甚至可能超出附录A列出的控制措施，由组织自己建立额外的控制措施。

注2：所有已识别风险均应由组织按照本文件要求进行记录。所有已识别风险及为应对这些风险而建立的风险管理措施（控制措施）均应反映在适用性声明中。

4 组织背景

4.1 理解组织及其背景

该组织应确定与其宗旨相关且影响其实现人工智能管理体系预期结果(s)能力的内外部问题。

该组织应确定气候变化是否属于相关议题。

组织应考虑其开发、提供或使用的人工智能系统（AI）的预期用途。组织应明确其在这些AI系统中的角色。

注1 为理解组织架构及其背景，组织可确定其相对于人工智能系统（AI系统）的角色。这些角色可包括但不限于以下一项或多项：

— 人工智能服务提供商，包括人工智能平台提供商、人工智能产品或服务提供商；

— 人工智能生产者，包括人工智能开发人员、设计人员、操作人员、测试与评估人员、部署人员、人为因素专家、领域专家、影响评估人员、采购人员、治理与监督人员；

— AI客户，包括AI用户；

— 人工智能合作伙伴，包括人工智能系统集成商和数据提供商；

— AI主体，包括数据主体及其他主体；

- 相关当局，包括决策者和监管者。

ISO/IEC 22989标准对这些角色进行了详细描述。此外，美国国家标准与技术研究院（NIST）的人工智能风险管理框架中也阐述了角色类型及其与人工智能系统生命周期的关系。[29]该组织的角色可确定本文件中要求和控制的适用性及适用范围。

注2 本条款需解决的内外部问题，可能因组织职能、管辖范围及其对人工智能管理系统预期目标实现能力的影响而异。具体包括但不限于：

a) 外部环境相关考量因素，例如：

1) 适用的法律要求，包括禁止使用人工智能；

2) 监管机构制定的政策、指南及决策，这些文件对人工智能系统开发与应用过程中法律要求的解释或执行具有影响；

1) 正在准备中。出版时的阶段为ISO/IEC DIS 5259-1：2023。

3) 与预期用途及人工智能系统使用相关的激励或后果；

4) 关于人工智能开发与使用的文化、传统、价值观、规范及伦理；

5) 利用人工智能系统开发新产品与服务的竞争格局及发展趋势；

b) 内部上下文相关考量因素，例如：

1) 组织背景、治理、 目标（见6.2）、政策和程序；

2) 合同义务

3) 拟开发或使用的人工智能系统的预期用途。

注3 角色确定可以通过组织处理的数据类别相关的义务形成（例如，在处理 PII 时，个人身份信息（PII）处理者或 PII 控制者）。有关 PII 及相关角色，请参阅ISO/IEC 29100 。角色也可以根据人工智能系统的特定法律要求来确定。

4.2 了解相关方的需求和期望

该组织应确定：

— 与人工智能管理系统相关的利益相关方；

— 相关利益方的法定要求；

——这些要求中哪些将通过人工智能管理系统得到解决。

注：相关利益方可提出与气候变化相关的诉求。

4.3 人工智能管理系统的范围界定

该组织应确定人工智能管理系统的边界和适用性，以确立其范围。

在确定该范围时，组织应考虑：

——4.1中提到的外部和内部问题；

—4.2节所述要求。

该范围应以书面形式提供。

人工智能管理体系的范围应涵盖组织在该文件要求下的各项活动，包括管理体系、领导力、规划、支持、运行、绩效、评估、改进、控制及目标。

4.4 人工智能管理系统

该组织应根据本文件要求，建立、实施、维护、持续改进并记录人工智能管理体系，包括所需流程及其相互作用。

5 领导力

5.1 领导力和承诺

最高管理层应通过以下方式展现对人工智能管理体系的领导力与承诺：

——确保制定人工智能政策（见5.2）和人工智能目标（见6.2），并与组织的战略方向相一致；

— 确保将人工智能管理系统需求整合至组织的业务流程中；

— 确保人工智能管理系统所需资源可用；

— 传达有效人工智能管理的重要性，并强调遵守人工智能管理体系要求；

— 确保人工智能管理系统实现其预期目标；

— 指导和支持相关人员为人工智能管理系统（AI管理系统）的有效性做出贡献；

——促进持续改进；

— 支持其他相关角色，以展示其在各自职责范围内的领导力。

注1本文件中对“业务”的引用可广义理解为指组织存在目的的核心活动。

注2：在组织内部建立、鼓励并示范一种负责任地使用、开发和管理人工智能系统的文化，可作为高层管理团队展现承诺与领导力的重要方式。通过领导力确保对这种负责任方法的认知与遵守，并支持人工智能管理系统，将有助于该系统的成功实施。

5.2 AI政策

最高管理层应制定人工智能政策，该政策应：

a) 符合该组织的宗旨；

b) 提供了一套设定人工智能目标的框架（参见6.2）；

c) 包括承诺满足适用要求；

d) 包括对人工智能管理系统持续改进的承诺。

人工智能政策应：

- 以书面形式提供；

— 参考其他组织政策；

— 在组织内部进行沟通；

— 适当时向有关各方提供。

建立AI政策的控制目标和控制措施见表A.1的A.2部分。这些控制措施的实施指南见B.2。

注：ISO/IEC 38507标准中提供了组织在制定人工智能政策时需考虑的事项。

5.3 角色、职责和权限

最高管理层应确保相关职位的职责与权限在组织内部得到明确分配与传达。

最高管理层应明确以下职责与权限：

a) 确保人工智能管理系统符合本文件的要求；

b) 向最高管理层汇报人工智能管理系统运行情况。

注：定义和分配角色与职责的对照见A.3.2表A.1。
本控制措施的实施指南见B.3.2。

6 规划

6.1 应对风险和机遇的行动

6.1.1 概要

在规划人工智能管理系统时，组织应考虑4.1中提到的问题和4.2中提到的要求，并确定需要解决的风险和机会，以：

— 保证人工智能管理系统能够实现其预期目标；

- 预防或减少不良反应；

— 实现持续改进。

该组织应建立并维护支持以下工作的AI风险标准：

—区分可接受与不可接受的风险；

- 开展人工智能风险评估；

— 开展人工智能风险处置；

评估人工智能风险影响。

注1：关于确定组织愿意承担或保留的风险量级与类型的相关考量，详见ISO/IEC 38507和ISO/IEC 23894标准。

该组织应根据以下内容确定风险与机遇：

— 人工智能系统的领域与应用场景；

- 预期用途；

——4.1中描述的外部和内部环境。

注2：人工智能管理系统可涵盖多个AI系统。此种情况下，需针对每个AI系统或AI系统组别分别确定其应用机会与用途。

该组织应制定以下计划：

a) 应对这些风险和机遇的行动；

b) 如何：

1) 将这些措施整合并落实到其人工智能管理系统流程中；

2) 评估这些措施的有效性。

该组织应保留有关为识别和应对人工智能风险及人工智能机遇所采取行动的书面记录信息。

注3：关于开发、提供或使用人工智能产品、系统和服务的组织如何实施风险管理的指南，详见ISO/IEC 23894标准。

注4 组织及其活动的背景可能对组织的风险管理活动产生影响。

注5 不同行业对风险的定义方式及其风险管理构想可能存在差异。3.7本文对风险的定义采用广泛适用的视角，可适配附录D所列各行业。但无论如何，作为风险评估的重要环节，企业首先需要根据自身实际情况制定风险认知框架。具体可参考人工智能系统开发及应用所在行业的标准定义，例如ISO/IEC Guide 51中关于风险的界定。

6.1.2 人工智能风险评估

该组织应制定并建立人工智能风险评估流程，该流程应：

a) 遵循并符合人工智能政策（参见5.2)和人工智能目标（见6.2）；

注：在评估6.1.2d) 1) 部分的后果时，组织可采用人工智能系统。
如6.1.4所述的冲击评估。

b) 其设计旨在使重复的人工智能风险评估能够产生一致、有效且可比的结果；

c) 识别有助于或阻碍实现其人工智能目标的风险；

d) 分析人工智能对以下方面构成的风险：

1) 评估已识别风险若实际发生，对组织、个人及社会可能造成的潜在影响；

2) 在适用情况下，评估已识别风险的实际可能性；

3) 确定风险等级；

e) 评估人工智能对以下方面的影响：

1) 将风险分析结果与风险标准进行比较（参见6.1.1）；

2) 优先评估风险以确定风险处理方案。

该组织应保留有关人工智能风险评估流程的书面记录。

6.1.3 AI风险处理

考虑到风险评估结果，组织应定义一个人工智能风险处理流程，以：

a) 选择适当的人工智能风险治疗方案；

b) 确定实施所选的AI风险处理方案所需的所有控制措施，并将这些控制措施与附录A中的控制措施进行比较，以验证是否遗漏了必要的控制措施；

注1附录A提供了满足组织目标和解决风险的参考控制措施
与人工智能系统的设计和使用相关。

c) 考虑附件A中与实施人工智能风险处理方案相关的控制措施；

d) 确定是否需要除附录A中的控制措施之外的其他控制措施，以实施所有风险处理方案；

e) 考虑附录B中关于实施b)和c)项所定控制措施的指导；

注2：所选控制措施中已隐含控制目标。组织可从附录A中选取合适的控制目标及措施组合。附录A所列控制措施并非穷尽性方案，可能需要补充其他控制目标或措施。若需超出附录A范围的其他控制措施，组织可自行设计或从现有资源中选取。若适用，人工智能风险管理可整合至其他管理体系中。

f) 出具适用性声明，其中包含必要的控制措施[参见b)、c)和d)]，并说明控制措施的纳入与排除依据。排除依据可包括：风险评估认为该控制措施非必要时，或其未被适用的外部要求所要求（或受其例外规定约束）时。

注3 该组织可提供书面依据，说明排除任何控制目标的合理性
通用的或针对特定的人工智能系统，无论是附录A所列的系统还是该组织自己建立的系统。

g) 制定人工智能风险处理方案。

该组织应获得指定管理层对人工智能风险处理计划及剩余人工智能风险接受的批准。必要控制措施应包括：

—与6.2目标一致；

- 以书面形式提供；

— 在组织内部进行沟通；

— 适当时向有关各方提供。

该组织应保留有关人工智能风险处理流程的书面记录。

6.1.4 人工智能系统影响评估

该组织应制定评估流程，用以评估人工智能系统开发、提供或使用可能对个人或群体（或两者）以及社会造成的潜在影响。

人工智能系统影响评估应确定人工智能系统的部署、预期用途及可预见的误用对个人或群体（或两者）以及社会可能产生的潜在后果。

人工智能系统影响评估应考虑该系统部署的具体技术与社会背景， 以及适用的司法管辖区。

人工智能系统影响评估的结果应予以记录。在适当情况下，可将系统影响评估结果提供给组织所定义的相关利益方。

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

注：在某些情况下（如安全或隐私关键的人工智能系统），组织可要求将特定学科的人工智能系统影响评估（例如安全、隐私或安全影响）作为组织整体风险管理活动的一部分。

6.2 人工智能目标及其实现规划

该组织应在相关职能和层级上制定人工智能目标。

人工智能目标应：

a) 与AI政策保持一致（参见5.2）；

b) 可测量的（如果可行的话）；

c) 考虑适用要求；

d) 监测；

e) 传达；

f) 酌情予以更新；

g) 可作为文件化信息提供。

在规划如何实现其人工智能目标时，该组织应确定：

— 将要采取什么措施；

— 需要哪些资源；

— 谁将负责；

— 完成时间；

——将如何评估结果。

注：附录C提供了与风险管理相关的AI目标的非排他性列表。为负责任开发和使用AI系统而制定的目标和实现这些目标的措施的控制目标和控制措施见表A.1的A.6.1和A.9.3 。对这些控制措施的实施指南见B.6.1和B.9.3。

6.3 变更规划

当组织确定需要对人工智能管理系统进行变更时，应以计划化的方式实施变更。

7 支持

7.1 资源

该组织应确定并提供建立、实施、维护及持续改进人工智能管理体系所需的资源。

注：AI资源的控制目标和控制措施见A.4表A.1 。这些控制措施的实施指南见B.4条款。

7.2 能力

该组织应：

— 确定受其控制且影响其人工智能性能的工作人员所需具备的胜任能力；

— 确保这些人员具备适当教育、培训或经验，从而具备胜任能力；

在适用情况下，采取行动以获取必要能力，并评估所采取行动的有效性。

应提供适当且有文件记录的信息作为能力证明。

注1人力资源的实施指导，包括对必要专门知识的考虑，见B.4.6。

注2 适用措施可包括：例如：为现有雇员提供培训、指导或重新分配岗位；或聘用或聘用具备相应能力的人员。

7.3 意识

在组织控制下工作的人应知悉：

——人工智能政策（见5.2）；

— 他们对人工智能管理系统有效性的贡献，包括人工智能性能提升带来的益处；

— 不符合人工智能管理系统要求的后果。

7.4 通讯

该组织应确定与人工智能管理体系相关的内外部沟通，包括：

— 它将传达什么；

- 何时沟通；

— 与谁沟通；

——如何沟通。

7.5 有据可查的信息

7.5.1 概要

该组织的人工智能管理系统应包括：

a) 本文件要求的书面信息；

b) 经组织确认为确保人工智能管理系统有效运行所必需的书面信息。

注：由于以下原因，不同组织对人工智能（AI）管理系统的记录信息范围可能存在差异：

— 组织规模及其活动、流程、产品与服务的类型；

— 过程及其相互作用的复杂性；

——人的能力。

7.5.2 创建和更新文件信息

在创建和更新文件化信息时，组织应确保符合以下要求：

— 标识与描述（例如标题、 日期、作者或参考编号）；

— 格式（如语言、软件版本、图形）和媒介（如纸质、电子）；

- 审查和批准适用性和充分性。

7.5.3 文件信息的控制

应对人工智能管理系统和本文件要求的记录信息进行控制，以确保：

a) 它在需要时和需要的地点可用且适合使用；

b) 其受到充分保护（例如防止机密性丧失、不当使用或完整性破坏）。

为控制已记录信息，组织应针对适用情况开展以下活动：

— 分发、访问、检索和使用；

— 存储与保存，包括可读性保存；

— 变更控制（如版本控制）；

— 保留与处置。

组织确定为人工智能管理系统规划与运行所必需的外部来源记录信息，应予以适当识别与管控。

注：访问权限可能仅指查看已记录信息的权限，或同时包含查看及修改已记录信息的权限与权限。

8 操作

8.1 业务规划与控制

该组织应规划、实施并控制满足要求所需的过程，并执行所确定的行动。第6条，通过：

- 制定流程标准；

— 根据标准实施对流程的控制。

组织应实施根据6.1.3确定的与人工智能管理体系运行相关的控制（例如，人工智能系统开发和使用生命周期相关控制）。

应监测这些控制措施的有效性，如果预期结果未实现，则应考虑采取纠正措施。附录A列出了参考控制措施，附录B提供了实施指南。

应提供必要的书面记录信息，以确保流程按计划执行。

该组织应控制计划变更，并审查非预期变更的后果，必要时采取措施减轻任何不利影响。

该组织应确保对外提供的与人工智能管理体系相关的流程、产品或服务受到控制。

8.2 人工智能风险评估

该组织应按照相关规定开展人工智能风险评估6.1.2在计划的时间间隔或当提出或发生重大变更时。

该组织应保存所有人工智能风险评估结果的书面记录。

8.3 AI风险处理

组织应按照6.1.3实施AI风险处置计划，并验证其有效性。

当风险评估发现需要治疗的新风险时，应按照6.1.3对这些风险执行风险治疗过程。

如果风险治疗方案定义的风险治疗方案无效，则应根据6.1.3对这些治疗方案进行审查和重新确认，并更新风险治疗方案。

该组织应保存所有人工智能风险处理结果的书面记录。

8.4 人工智能系统影响评估

组织应按照6.1.4规定的计划时间间隔或当拟进行重大变更时，执行人工智能系统影响评估。

该组织应保存所有人工智能系统影响评估结果的书面记录。

9 绩效评价

9.1 监测、测量、分析和评价

该组织应确定：

— 需要监测和测量的；

— 监测、测量、分析和评估的方法，视情况而定，以确保结果有效；

— 监测与测量应执行的时间；

— 监测与测量结果应进行分析和评估。

记录信息应作为结果的证据提供。

该组织应评估人工智能管理系统的性能与有效性。

9.2 内部审计

9.2.1 概要

该组织应按计划间隔开展内部审核， 以提供关于人工智能管理系统是否符合要求的信息：

a) 符合：

1) 该组织对自身人工智能管理系统的具体要求；

2) 本文件的要求；

b) 有效实施并持续维护。

9.2.2 内部审计方案

该组织应规划、建立、实施并维护（一项）或多项审计计划，包括频率、方法、职责、规划要求及报告。

在制定内部审计计划时，组织应充分考虑相关流程的重要性及既往审计结果。

该组织应：

a) 明确每次审计的目标、标准及范围；

b) 选择审计师并开展审计，以确保审计过程的客观性和公正性；

c) 确保审计结果及时上报至相关管理层。

记录信息应作为审计计划（或多个审计计划）实施及审计结果的证明材料。

9.3 管理审查

9.3.1 概要

最高管理层应按计划定期审查组织的人工智能管理体系， 以确保其持续适用性、充分性和有效性。

9.3.2 管理评审输入

管理评审应包括：

a) 既往管理评审中采取措施的执行情况；

b) 与人工智能管理系统相关的内外部问题的变化；

c) 与人工智能管理系统相关的利益相关方需求和期望的变化；

d) 人工智能管理系统性能信息，包括以下趋势：

   1) 不符合项及纠正措施；

   2) 监测和测量结果；

   3) 审计结果；

e) 持续改进的机会。

9.3.3 管理评审结果

管理评审的结果应包括与持续改进机会相关的决策， 以及是否需要对人工智能管理体系进行变更。

记录信息应作为管理评审结果的证据提供。

10 改善

10.1 持续改进

该组织应持续改进人工智能管理系统的适用性、充分性及有效性。

10.2 不符合项及纠正措施

当发生不符合项时，组织应：

a) 针对不符合项作出反应，并在适用情况下：

   1) 采取行动控制和纠正；

   2) 处理后果；

b) 评估是否需要采取行动消除不符合项的原因，以防止其再次发生或在其他地方出现，具体方法如下：

   1) 审查不符合项；

   2) 确定不符合项的原因；

   3) 确定是否存在或可能发生类似的不符合项；

c) 采取一切必要措施；

d) 评估所采取的任何纠正措施的有效性；

e) 必要时对AI管理系统进行修改。

纠正措施应与所发现的不符合项的影响相适应。

记录信息应作为以下情况的证据：

— 不合格项的性质及后续采取的措施；

— 任何纠正措施的结果。

附件A

（规范性）参考控制目标和控制措施

A.1 概要

表A.1中详述的控制措施为组织提供了实现组织目标和解决与人工智能系统设计和运行相关风险的参考。并非表A.1中列出的所有控制目标和控制措施都必须使用，组织可以设计和实施自己的控制措施（见6.1.3）。

附录B提供了表A.1所列所有控制措施的实施指南。

表A.1 —— 控制目标与控制措施

A.2 与人工智能相关的政策		
目的：根据业务需求为人工智能系统提供管理指导和支持。		
	主题	控制
A.2.2	AI政策	该组织应制定并记录关于人工智能系统开发或使用的政策。
A.2.3	与其他组织政策的协调	该组织应确定其他政策可能影响或适用于其人工智能系统目标的范围。
A.2.4	人工智能政策综述	应按计划时间间隔或根据需要额外审查人工智能政策，以确保其持续适用性、充分性和有效性。
A.3 内部组织		
目的：在组织内部建立问责机制，以确保其对人工智能系统实施、运行和管理的负责任态度。		
	主题	控制
A.3.2	AI角色与职责	应根据组织需求明确并分配人工智能（AI）的角色与职责。
A.3.3	报告关注事项	该组织应定义并建立一套流程，用于在整个生命周期内报告关于组织在人工智能系统中角色的关切。
A.4 AI系统资源		
目的：确保组织对人工智能系统资源（包括AI系统组件及资产）进行核算，以全面理解并应对相关风险与影响。		
	主题	控制
A.4.2	资源文件	该组织应识别并记录特定人工智能系统生命周期阶段活动所需的相关资源，以及其他与组织相关的AI相关活动。
A.4.3	数据资源	作为资源识别的一部分，该组织应记录用于人工智能系统的数据资源相关信息。
A.4.4	工具资源	作为资源识别的一部分，组织应记录用于人工智能系统（AI系统）的工具资源相关信息。

表A.1（*续*）

A.4.5	系统和计算资源	作为资源识别的一部分，该组织应记录有关用于人工智能系统的系统和计算资源的信息。
A.4.6	人力资源	在资源识别过程中，组织需完整记录以下人力资源及其专业能力：涉及人工智能系统开发、部署、运行、变更管理、维护、迁移与退役，以及验证与集成等全流程的人员配置。
A.5 评估人工智能系统的影响		
目的：评估人工智能系统在其整个生命周期中对个体或群体（或两者）以及受该系统影响的社会所产生的影响。		
	主题	控制
A.5.2	AI系统影响评估过程	该组织应建立一套流程，用以评估人工智能系统在其整个生命周期中可能对个人或群体（或两者）以及社会产生的潜在影响。
A.5.3	人工智能系统影响评估的文档记录	组织应记录人工智能系统影响评估的结果，并在确定的周期内保留结果。
A.5.4	评估人工智能系统对个人或群体的影响	该组织应评估并记录人工智能系统在整个系统生命周期中对个人或群体可能产生的影响。
A.5.5	评估人工智能系统对社会的影响	该组织应评估并记录其人工智能系统在整个生命周期中的潜在社会影响。
A.6 人工智能系统生命周期		
A.6.1 人工智能系统开发的管理指南		
目的：确保组织识别并记录目标，并实施负责人工智能系统设计与开发的流程。		
	主题	控制
A.6.1.2	人工智能系统责任开发目标	该组织应识别并记录指导负责开发人工智能系统的目标，并在开发生命周期中考虑这些目标并整合实现措施。
A.6.1.3	负责任的人工智能系统设计与开发流程	该组织应明确并记录人工智能系统设计与开发的具体流程。
A.6.2 人工智能系统生命周期		
目的：明确人工智能系统生命周期各阶段的标准与要求。		
	主题	控制
A.6.2.2	人工智能系统需求与规范	该组织应明确并记录对新人工智能系统或现有系统重大改进的要求。
A.6.2.3	人工智能系统设计与开发的文档记录	该组织应基于组织目标、书面要求及规范标准，对人工智能系统的设计与开发进行文档化记录。
A.6.2.4	AI系统验证与验证	该组织应定义并记录人工智能系统的验证与确认措施，并明确其使用标准。
A.6.2.5	人工智能系统部署	该组织应制定部署计划并确保在部署前满足相关要求。

表A.1（*续*）

A.6.2.6	AI系统操作与监控	该组织应定义并记录人工智能系统持续运行所需的必要要素。至少应包括系统与性能监控、维修、更新及支持。
A.6.2.7	人工智能系统技术文档	该组织应明确为用户、合作伙伴、监管机构等各相关方所需的人工智能系统技术文档类别，并以适当形式提供相关技术文档。
A.6.2.8	事件日志的AI系统记录	该组织应确定在人工智能系统生命周期的哪些阶段应启用事件日志记录，但至少应在人工智能系统使用期间启用。
A.7 AI系统数据		
目的**确保组织理解数据在人工智能系统应用和开发、提供或使用人工智能系统整个生命周期中的作用和影响。**		
	主题	控制
A.7.2	人工智能系统开发与增强的数据	该组织应定义、记录并实施与人工智能系统开发相关的数据管理流程。
A.7.3	数据获取	该组织应确定并记录人工智能系统所用数据的获取与选择细节。
A.7.4	人工智能系统数据质量	该组织应定义并记录数据质量要求，并确保用于开发和运行人工智能系统的数据符合这些要求。
A.7.5	数据来源	该组织应定义并记录一套流程，用于在其人工智能系统数据及系统生命周期内追踪所用数据的来源。
A.7.6	数据准备	组织应定义并记录其选择数据准备和使用数据准备方法的标准。
A.8 人工智能系统相关方须知		
目的：确保相关利益方掌握必要信息，以充分理解并评估风险及其影响（包括积极与消极两方面）。		
	主题	控制
A.8.2	用户系统文档与信息	该组织应确定并向人工智能系统用户提供必要信息。
A.8.3	对外报告	该组织应为相关方提供报告人工智能系统不良影响的能力。
A.8.4	事件通报	该组织应制定并记录一份向人工智能系统用户传达事件的计划。
A.8.5	有关各方的信息	该组织应确定并记录其向利益相关方报告人工智能系统信息的义务。
A.9 使用AI系统		
目的：确保组织负责任地使用人工智能系统，并符合组织政策。		
	主题	控制
A.9.2	人工智能系统负责任使用流程	该组织应定义并记录人工智能系统负责任使用的流程。
A.9.3	负责任使用人工智能系统的目标	该组织应识别并记录目标，以指导人工智能系统的负责任使用。

表A.1（*续*）

A.9.4	AI系统的预期用途	该组织应确保人工智能系统按照其预期用途及随附文件进行使用。
A.10 与第三方及客户的关系		
目的：确保组织在人工智能系统生命周期的任何阶段涉及第三方时，均能明确自身责任、保持问责制，并合理分配风险。		
	主题	控制
A.10.2	分配职责	该组织应确保其人工智能系统生命周期内的职责在组织、合作伙伴、供应商、客户及第三方之间进行分配。
A.10.3	供应商	该组织应建立相应流程，确保其对供应商所提供服务、产品或材料的使用符合该组织在人工智能系统负责任开发与应用方面的方针。
A.10.4	客户	该组织应确保其对人工智能系统开发与使用的负责任方法，充分考虑客户的期望与需求。

附录B

人工智能控制措施（规范性）实施指南

B.1 概要

本附录中记录的实施指南与表A.1列出的控制有关。它提供信息以支持表A.1中列出的控制的实施，并达到控制目标，但组织无需在适用性声明中记录或证明实施指南的纳入或排除（参见6.1.3）。

实施指南并非在所有情况下都适用或充分，且未必完全符合组织的特定控制要求。组织可根据自身具体需求及风险处理要求，对实施指南进行扩展或修改，或自行制定控制措施的实施方案。

本附录将作为指导文件，用于确定和实施本文件中定义的人工智能管理系统中的人工智能风险处理控制措施。除本附录所列控制措施外，还可确定其他组织和技术控制措施（参见6.1.3）本附录可作为制定组织特定控制措施实施方案的起点。

B.2 与人工智能相关的政策

B.2.1 目标

根据业务需求为人工智能系统提供管理指导与技术支持。

B.2.2 人工智能政策

统治

该组织应制定并记录人工智能系统开发或使用的政策。

实施指南

人工智能政策应由以下因素指导：

— 商业战略；

— 组织价值观与文化，以及组织愿意承担或保留的风险程度；

——人工智能系统所构成的风险等级；

— 法律要求，包括合同；

— 组织的风险环境；

— 对相关利益方的影响（见6.1.4）。

除了5.2中的要求外，人工智能政策还应包括：

— 指导组织所有与人工智能相关活动的原则；

— 处理政策偏差与例外情况的流程。

在必要时，人工智能政策应考虑特定主题方面， 以提供额外指导或与其他涉及这些方面的政策进行交叉引用。此类主题的示例包括：

- AI资源和资产；

-人工智能系统影响评估（参见6.1.4）；

—人工智能系统开发。

相关政策应指导人工智能系统的开发、采购、运行及使用。

B.2.3 与其他组织政策的一致性

统治

该组织应确定其他政策可能影响或适用于其人工智能系统目标的范围。

实施指南

人工智能与多个领域存在交叉，包括质量、安全、安全性和隐私保护。组织应进行全面分析， 以确定现行政策是否及在哪些方面必然存在交叉，并在必要时更新相关政策，或在人工智能政策中纳入相关条款。

其他信息

管理机构代表组织制定的政策应为人工智能政策提供依据。ISO/IEC 38507为组织管理机构成员提供了指导，以实现并管理人工智能系统在其整个生命周期中的应用。

B.2.4 对人工智能政策的审查

统治

应按计划时间间隔或根据需要额外审查人工智能政策， 以确保其持续适用性、充分性和有效性。

实施指南

管理层应指定专人负责人工智能政策及其组成部分的制定、审查与评估工作。审查内容需涵盖：针对组织环境、业务状况、法律条件或技术环境的变化，评估改进组织政策及人工智能系统管理方法的可行性。

人工智能政策的审查应考虑管理审查的结果。

B.3 内部组织

B.3.1 目标

在组织内部建立问责机制， 以确保其在人工智能系统的实施、运营和管理方面采取负责任的态度。

B.3.2 人工智能的角色与职责

统治

应根据组织需求明确并分配人工智能（AI）的角色与职责。

实施指南

明确角色与职责是确保组织在人工智能系统全生命周期中履行责任的关键。在分配角色与职责时，组织应综合考虑人工智能政策、 目标及已识别风险， 以确保所有相关领域均被覆盖。组织可优先确定角色与职责的分配顺序。需要明确角色与职责的领域示例包括：

— 风险管理；

- 人工智能系统影响评估；

— 资产和资源管理；

——安全；

- 安全；

——隐私；

— 发展；

— 表现；

- 人类监督；

— 供应商关系；

— 证明其能够持续满足法律要求；

— 数据质量管理（贯穿整个生命周期）。

应根据个人执行其职责的适当程度，界定各种角色的责任。

B.3.3 关注事项的报告

统治

该组织应制定并实施一套流程，用于在整个生命周期内报告关于组织在人工智能系统中角色的关切。

实施指南

报告机制应履行以下职能：

a) 保密或匿名或两者兼有；

b) 向在职人员及合同制人员提供并推广；

c) 配备合格人员；

d) 规定对c)项所指人员的适当调查与处置权限；

e) 规定了及时上报并向管理层升级的机制；

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

g) 根据4.4规定提供报告，如适用，还应提供e）报告；同时在a）中保持机密性和匿名性，并遵守一般商业保密规定；

h) 在适当的时间范围内提供应对机制。

注：该组织可将现有报告机制作为本流程的一部分加以利用。

其他信息

除本条款规定的实施指南外，组织还应考虑ISO 37002标准。

B.4 AI系统资源

B.4.1 目标

为确保组织全面核算人工智能系统的资源（包括系统组件及资产），从而充分理解并应对相关风险与影响。

B.4.2 资源文档

统治

该组织应识别并记录特定人工智能系统生命周期阶段活动所需的相关资源， 以及其他与组织相关的AI相关活动。

实施指南

记录人工智能系统的资源对于了解风险以及人工智能系统对个人或群体或两者以及社会的潜在影响（积极和消极）至关重要。记录此类资源（例如可以利用数据流程图或系统架构图）可以为人工智能系统影响评估提供信息（见B.5）。

资源包括但不限于：

— 人工智能系统组件；

— 数据资源，即人工智能系统生命周期各阶段所使用的数据；

— 工具资源（如人工智能算法、模型或工具）；

— 系统与计算资源（例如用于开发和运行人工智能模型的硬件、数据存储及工具资源）；

— 人力资源，即具备必要专业知识（如人工智能系统开发、销售、培训、运维等）的人员，需与组织在整个人工智能系统生命周期中的角色相匹配。

资源可由组织自身、其客户或第三方提供。

其他信息

资源记录还可用于确定资源是否可用，若资源不可用，则组织应修订人工智能系统的设计规范或其部署要求。

B.4.3 数据资源

统治

作为资源识别的一部分，该组织应记录用于人工智能系统的数据资源相关信息。

实施指南

数据文档应包括但不限于以下主题：

— 数据来源；

— 数据最后更新或修改的日期（例如元数据中的日期标签）；

——机器学习的数据类别（例如训练、验证、测试和生产数据）；

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

- 数据标签流程；

— 数据的预期用途；

— 数据质量（如ISO/IEC 5259系列2所述）；

适用的数据保留和处置政策；

— 数据中已知或潜在的偏倚问题；

- 数据准备。

B.4.4 工具资源

统治

作为资源识别的一部分，该组织应记录用于人工智能系统（AI系统）的工具资源相关信息。

实施指南

AI系统，特别是机器学习的工具资源，可以包括但不限于：

— 算法类型与机器学习模型；

— 数据处理工具或流程；

—优化方法；

—评价方法；

— 资源调配工具；

- 帮助模型开发的工具；

— 用于人工智能系统设计、开发与部署的软硬件。

其他信息

ISO/IEC 23053标准为机器学习的各种工具资源提供了详细的类型、方法和实施途径指导。

B.4.5 系统与计算资源

统治

作为资源识别的一部分，该组织应记录有关用于人工智能系统的系统和计算资源的信息。

---

2) 正在准备中。发布时的阶段：ISO/IEC DIS 5259-1：2023、ISO/IEC DIS 5259-2：2023、ISO/IEC DIS 5259-3：2023、ISO/IEC DIS 5259-4：2023、ISO/IEC CD 5259-5：2023。

实施指南

关于人工智能系统所涉及的系统与计算资源的信息，可包括但不限于以下内容：

— 人工智能系统的资源需求（即确保系统能在资源受限的设备上运行）；

— 系统与计算资源的部署位置（例如本地部署、云计算或边缘计算）；

— 处理资源（包括网络和存储）；

— 用于运行AI系统工作负载的硬件所产生的影响（例如通过硬件使用或制造对环境造成的影响，或硬件使用成本）。

该组织应考虑，为实现人工智能系统的持续改进可能需要不同的资源。系统的开发、部署和运行可能具有不同的系统需求和要求。

注：ISO/IEC 22989标准阐述了多种系统资源考量因素。

B.4.6 人力资源

统治

在资源识别过程中，组织需详细记录以下人力资源及其能力信息：包括人工智能系统开发、部署、运行、变更管理、维护、迁移与退役， 以及验证与集成等环节所涉及的人员配置。

实施指南

该组织应考虑多元化专业人才的需求，并纳入系统所需的各类岗位。例如，若纳入特定人口统计学群体是系统设计的必要组成部分，则可将与训练机器学习模型所用数据集相关的人群纳入其中。必要的人力资源包括但不限于：

——数据科学家；

— 与人工智能系统的人工监督相关角色；

— 诚信领域的专家，涵盖安全、保密及隐私等主题；

— 人工智能研究人员、专家及相关领域专家。

在人工智能系统生命周期的不同阶段，可能需要不同的资源。

B.5 评估人工智能系统的影响

B.5.1 目标

评估人工智能系统对个体或群体（或两者） 的影响， 以及该系统在整个生命周期中对受影响社会的影响。

B.5.2 人工智能系统影响评估流程

统治

该组织应建立一套流程，用以评估人工智能系统在其整个生命周期中可能对个人或群体（或两者）以及社会造成的潜在影响。

实施指南

鉴于人工智能系统可能对个人、群体或两者以及社会产生重大影响，提供和使用此类系统的组织应根据其预期用途和使用方式，评估这些系统对相关群体的潜在影响。

该组织应考虑人工智能系统是否影响：

— 个人的法律地位或生活机会；

— 个人的身体或心理健康状况；

—普世人权；

— 社会。

该组织的程序应包括但不限于：

a) 应进行人工智能系统影响评估的情形，包括但不限于：

   1) 人工智能系统预期用途及使用情境的关键性，或其中任何重大变更；

   2) 人工智能技术的复杂性、人工智能系统的自动化水平或其任何重大变化；

   3) 人工智能系统处理的数据类型及来源的敏感性，或其发生的任何重大变更；

b) 构成人工智能系统影响评估流程的要素，可能包括：

   1) 识别（例如来源、事件和结局）；

   2) 分析（例如后果与可能性）；

   3) 评估（例如接受决定和优先级排序）；

   4) 治疗（例如缓解措施）；

   5) 文件编制、报告和沟通（参见7.4 、7.5和B.3.3）；

c) 负责执行人工智能系统影响评估；

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

e) 根据系统预期用途、使用方式及特性（例如针对个人、群体或社会的评估）可能受影响的个体与社会。

影响评估应综合考量人工智能系统的多个维度，包括开发该系统所用数据、采用的人工智能技术以及系统整体功能。

这些过程可以根据组织的角色和人工智能应用的领域以及根据评估影响的具体学科（例如安全、隐私和安全）而有所不同。

其他信息

对于某些学科或组织而言，详细评估其对个人、群体或社会的影响，是风险管理的重要组成部分，尤其在信息安全、安全与环境管理等领域。组织应当明确

若作为此类风险管理过程组成部分实施的学科特异性影响评估充分整合了针对特定方面（如隐私）的人工智能考量。

注：ISO/IEC 23894描述了组织如何作为整体风险管理过程的一部分，对组织本身、个人或个人群体或两者以及社会进行影响分析。

B.5.3 人工智能系统影响评估的文档记录

统治

该组织应记录人工智能系统影响评估的结果，并在规定期限内保存评估结果。

实施指南

该文件可为确定应向用户及其他相关利益方传达的信息提供依据。

人工智能系统影响评估应根据需要予以保留和更新，与B.5.2中记录的人工智能系统影响评估要素保持一致。保留期限可遵循组织保留时间表，或根据法律要求或其他要求确定。

组织应考虑记录的项目包括但不限于：

— 人工智能系统的预期用途及任何可预见的合理误用；

— 人工智能系统对相关个人或群体，或两者，以及社会产生的积极与消极影响；

- 可预见的故障、其潜在影响以及为减轻其影响所采取的措施；

— 该系统适用的相关人口统计学群体；

—系统复杂性；

— 人类在系统关系中的作用，包括人类监督能力、流程和工具，可用于避免负面影响；

——就业与员工技能提升。

B.5.4 评估人工智能系统对个人或群体的影响

统治

该组织应评估并记录人工智能系统在整个系统生命周期中对个人或群体可能产生的影响。

实施指南

在评估对个人或群体，或两者以及社会的影响时，组织应考虑其治理原则、人工智能政策和目标。使用人工智能系统或其 PII 被人工智能系统处理的个人，可以对人工智能系统的可信度有预期。应充分考虑儿童、残障人士、老年人及劳动者等特定群体的特殊保护需求。组织机构需评估这些需求，并将满足这些需求的措施纳入系统影响评估体系。

根据人工智能系统用途和使用范围的不同，评估时需考虑的影响领域包括但不限于：

公平；

——问责；

— 透明度和可解释性；

— 安全和隐私；

- 安全和健康；

— 财务后果；

——无障碍；

——人权。

其他信息

必要时，组织应咨询专家（如研究人员、主题专家和用户），以全面了解人工智能系统对个人或群体（或两者）及社会的潜在影响。

B.5.5 评估人工智能系统对社会的影响

统治

该组织应评估并记录其人工智能系统在整个生命周期中的潜在社会影响。

实施指南

社会影响可能因组织背景及人工智能系统类型的不同而存在显著差异。人工智能系统可能产生积极或消极的社会影响。这些潜在社会影响的示例包括：

- 环境可持续性（包括对自然资源和温室气体排放的影响）；

— 经济（包括金融服务、就业机会、税收、贸易和商业）；

— 政府（包括立法程序、为政治利益而散布的虚假信息、国家安全和刑事司法系统）；

- 健康和安全（包括获得医疗保健、医疗诊断和治疗以及潜在的身体和心理伤害）；

— 规范、传统、文化和价值观（包括导致偏见或对个人或群体或两者造成伤害的错误信息，以及对社会造成伤害的错误信息）。

其他信息

人工智能系统的开发与应用往往需要消耗大量计算资源，这可能对环境可持续性产生连锁影响（例如因电力消耗增加导致的温室气体排放，以及对水资源、土地、动植物生态造成的冲击）。与此同时，这些系统也能助力提升其他领域的环境可持续性（比如减少建筑和交通领域产生的温室气体排放）。企业应当将人工智能系统的影响纳入整体环境可持续发展目标和战略框架中进行考量。

该组织应考虑如何滥用其人工智能系统造成社会危害，以及如何利用它们解决历史危害。例如，人工智能系统能否阻止人们获得贷款、补助金、保险和投资等金融服务，人工智能系统能否同样改善这些工具的使用？

人工智能系统已被用于影响选举结果，并制造虚假信息（如数字媒体中的深度伪造内容），这些行为可能引发政治和社会动荡。政府将人工智能系统用于刑事司法领域，已暴露出其对社会、个人或群体存在的偏见风险。

组织应分析行为者如何滥用人工智能系统， 以及这些系统如何强化历史遗留的社会偏见。

人工智能系统可用于疾病诊断与治疗， 以及健康福利资格判定。在自动驾驶汽车、人机协同等可能导致人员伤亡的场景中，该系统同样被部署使用。组织机构在应用人工智能系统时（如涉及健康安全的场景），应综合考量其正反两面影响。

注：ISO/IEC TR 24368标准对人工智能系统及应用相关的伦理与社会问题提供了高层次概述。

B.6 人工智能系统生命周期

B.6.1 人工智能系统开发的管理指南

B.6.1.1 目标

确保组织识别并记录目标，并实施流程以负责人工智能系统的设计与开发。

B.6.1.2 人工智能系统负责任开发的目标

统治

该组织应明确并记录指导人工智能系统负责任开发的目标，并在开发生命周期中考虑这些目标，整合实现目标的措施。

实施指南

组织应当明确那些影响人工智能系统设计与开发流程的目标（参见6.2），并将这些目标贯穿于整个开发过程。例如，若将“公平性”作为核心目标之一，就必须将其纳入需求规范、数据采集、数据预处理、模型训练、验证与确认等环节。组织需要根据实际需求制定相应要求和指导方针，确保各阶段都能落实相关措施（比如要求使用特定测试工具或方法来消除不公平现象或不良偏差），从而有效达成这些目标。

其他信息

人工智能技术正被用于增强安全措施，例如威胁预测检测和安全攻击预防。这是人工智能技术的一个应用，可用于加强安全措施， 以保护人工智能系统和传统的非人工智能软件系统。附件C提供了管理风险的组织目标示例，这些示例可用于确定人工智能系统开发的目标。

B.6.1.3 人工智能系统负责任设计与开发的流程

统治

该组织应明确并记录人工智能系统责任设计与开发的具体流程。

实施指南

对人工智能系统流程负责的开发应包括但不限于以下内容：

— 生命周期阶段（ISO/IEC 22989标准提供了通用的人工智能系统生命周期模型，但组织可自行定义其生命周期阶段）；

— 测试要求及计划的测试方法；

— 人类监督要求，包括流程和工具，特别是当人工智能系统可能影响自然人时；

— 应在哪些阶段进行人工智能系统影响评估；

— 培训数据的预期与规则（例如可使用哪些数据、经批准的数据供应商及标签）；

— 需要专业知识（特定领域或其他）或对人工智能系统开发人员的培训，或两者兼有；

- 释放标准；

— 各阶段需进行审批与签字确认；

— 变更控制；

— 可用性和可控性；

— 涉及各方的参与。

具体的设计与开发流程取决于拟用于该AI系统的功能及AI技术。

B.6.2 人工智能系统生命周期

B.6.2.1 目标

明确人工智能系统生命周期各阶段的标准与要求。

B.6.2.2 人工智能系统的需求与规范

统治

该组织应明确并记录对新人工智能系统或现有系统重大改进的要求。

实施指南

该组织应记录开发人工智能系统（AI）的理论依据及其目标。需考虑、记录并理解的因素包括：

a) 例如，开发人工智能系统的原因，是基于商业案例、客户需求还是政府政策？

b) 如何训练该模型以及如何满足数据需求。

需明确人工智能系统的需求规范，并覆盖其整个生命周期。当开发的AI系统无法按预期运行，或出现可用于调整和优化需求的新信息时，应重新评估这些需求。例如，从财务角度考虑，开发该AI系统可能变得不可行。

其他信息

描述人工智能系统生命周期的流程由ISO/IEC 5338标准提供。关于交互系统中以人为中心的设计的更多信息，请参阅ISO 9241-210标准。

B.6.2.3 人工智能系统设计与开发的文档记录

统治

该组织应基于组织目标、已记录的需求及规范标准，对人工智能系统的设计与开发进行文档化记录。

实施指南

人工智能系统需要进行多种设计选择，包括但不限于：

— 机器学习方法（如监督式与无监督式）；

— 采用的学习算法及机器学习模型类型；

-如何对模型进行训练，数据质量如何（参见B.7）；

— 模型的评估与优化；

— 硬件和软件组件；

— 考虑人工智能系统全生命周期的安全威胁；人工智能系统特有的安全威胁包括数据投毒、模型窃取或模型反演攻击；

— 输出界面与呈现方式；

— 人类如何与系统互动；

— 互操作性和可移植性考量。

设计与开发之间可能存在多次迭代，但应保持阶段文档的完整性，并确保最终系统架构文档的可用性。

其他信息

有关交互式系统以人为中心设计的更多信息，请参阅ISO 9241-210标准。

B.6.2.4 人工智能系统验证与确认

统治

该组织应制定并记录人工智能系统的验证与确认措施，并明确其使用标准。

实施指南

验证与确认措施可包括但不限于：

- 测试方法和工具；

— 测试数据的选择及其对预期使用领域的表征；

- 释放标准要求。

该组织应制定并记录评估标准，包括但不限于：

评估人工智能系统组件和整个人工智能系统对个人或群体或两者以及社会的影响所涉及的风险的计划；

— 评估计划可基于以下因素制定：

  — 人工智能系统的可靠性与安全性要求，包括该系统性能的可接受错误率；

  —负责B.6.1.2和B.9.3中所述的AI系统开发和使用目标；

  — 操作因素，如数据质量、预期用途，包括各操作因素的可接受范围；

  — 需要定义更严格操作因素的预期用途，包括操作因素的不同可接受范围或更低的错误率；

— 用于评估相关利益方（即基于AI系统输出做出决策或受其决策影响的主体）能否充分解读AI系统输出的方法、指导原则或评估指标。评估频率应予以确定，可依据AI系统影响评估结果进行。

— 任何可接受的因素，这些因素可能导致无法达到最低绩效目标水平，特别是在评估人工智能系统对个人或群体（或两者）以及社会的影响时。计算机视觉系统图像分辨率不足或背景噪声影响语音识别系统）。针对这些因素导致的AI系统性能不佳，应记录相应的应对机制。

应根据已记录的评估标准对AI系统进行评估。

如果AI系统不能满足记录的评价标准，特别是针对负责任的AI系统开发和使用目标（见B.6.1.2和B.9.3），组织应重新考虑或管理AI系统预期用途的缺陷、其性能要求以及组织如何有效解决对个人或群体或两者的影响。

备注　关于如何处理神经网络稳健性的更多信息可参见ISO/IECTR 24029-1.

B.6.2.5 人工智能系统的部署

统治

该组织应制定部署计划，并在部署前确保满足相关要求。

实施指南

人工智能系统可在不同环境中开发部署（例如本地开发后通过云计算部署），企业在制定部署方案时需充分考虑这些差异。同时，企业还需评估各组件是否可独立部署（例如软件与模型能否分别部署）。此外，企业应在发布前制定一套必须满足的要求清单（业内称为“发布标准”），包括需通过的验证与确认措施、需达成的性能指标、需完成的用户测试，以及需获得的管理层审批和签字确认。部署方案的制定应充分考虑相关利益方的立场及其可能产生的影响。

B.6.2.6 人工智能系统操作与监控

统治

该组织应明确并记录人工智能系统持续运行所需的必要要素。至少应包括系统与性能监控、维修、更新及技术支持。

实施指南

每次操作与监测的最低活动量可综合考虑多种因素。例如：

— 系统与性能监控可涵盖常规错误与故障监测，以及生产数据运行是否符合预期的评估。技术性能指标包括问题解决成功率、任务达成率或置信度等。其他评估标准则涉及满足相关方的承诺、预期及需求，例如持续监控以确保符合客户要求或适用的法律规范。

— 部分部署的AI系统通过机器学习（ML）实现性能优化，其生产数据与输出数据将用于持续训练模型。当采用持续学习时，企业需实时监控AI系统性能，确保其始终符合设计目标，并按预期处理生产数据。

— 一些人工智能系统的性能即使不采用持续学习，也可能发生变化，通常由生产数据中的概念漂移或数据漂移引起。在此类情况下，监测可识别是否需要重新培训，以确保人工智能系统持续满足其设计目标，并按预期在生产数据上运行。更多信息可参见ISO/IEC 23053标准。

— 系统维护可能涉及对错误和故障的响应。企业应建立完善的应对和修复流程。此外，当系统持续演进、发现关键问题或因外部因素（如未满足客户期望或法律要求）引发问题时，可能需要进行系统更新。企业需制定系统更新流程，包括受影响组件、更新时间表以及向用户说明更新内容等具体措施。

— 系统更新还可能涉及系统操作的变更、预期用途的新增或修改，或其他系统功能的调整。组织应制定相应程序以应对操作变更，包括与用户进行沟通。

— 对系统的支持可以是内部的、外部的或两者兼有，具体取决于组织的需求以及系统是如何获得的。支持流程应考虑用户如何联系适当的帮助、如何报告问题和突发事件、支持服务水平协议和指标。

— 如果人工智能系统被用于其设计目的以外的用途或以未预料到的方式使用，应考虑此类用途的适当性。

— 应识别与组织应用和开发的AI系统相关的人工智能特定信息安全威胁。人工智能特定信息安全威胁包括但不限于数据投毒、模型窃取和模型反演攻击。

其他信息

该组织应考虑可能影响利益相关方的运营绩效，并在设计和确定绩效标准时予以考量。

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

性能指标可包含统计学方面，如错误率和处理时长。针对每个评估标准，组织需明确所有相关指标及其相互依存关系。针对每个指标，组织应基于领域专家建议及利益相关方对现有非人工智能实践的期望分析，确定可接受的数值范围。

例如，一个组织可以确定F_1分数是适当的性能度量，这是基于其对假阳性和假阴性影响的评估，如所述

ISO/IEC TS 4213 。组织可据此确定人工智能系统应达到的F_1值。需评估现有措施是否足以应对这些问题。若无法满足，则应考虑调整现有措施或制定新措施以检测和处理这些问题。

该组织应考虑非人工智能系统或流程在运行中的表现，并将其作为制定绩效标准时潜在的相关背景。

该组织还应确保用于评估人工智能系统的手段和流程（包括在适用情况下对评估数据的选择和管理）能够提升其性能评估的完整性和可靠性，以符合既定标准。

绩效评估方法的开发可基于标准、指标和价值观。这些要素应指导评估中所用数据量及流程类型，以及执行评估人员的角色与专业能力。

绩效评估方法应尽可能准确地反映操作与使用过程的属性及特征， 以确保评估结果具有实用性和相关性。在绩效评估的某些环节，可能需要通过受控方式引入错误或虚假数据或流程，以评估其对绩效的影响。

ISO/IEC 25059中的质量模型可用于定义性能标准。

B.6.2.7 人工智能系统技术文档

统治

该组织应明确界定各相关利益方（包括用户、合作伙伴及监管机构）所需的人工智能系统技术文档类型，并以适当形式向其提供相关技术文档。

实施指南

人工智能系统技术文档可包含但不限于以下要素：

— 人工智能系统的总体描述，包括其预期用途；

— 使用说明；

— 关于其部署与运行的技术假设（运行时环境、相关软硬件能力、数据处理假设等）；

— 技术限制（如可接受的错误率、准确性、可靠性、稳健性）；

— 监控功能与操作权限，使用户或操作员能够影响系统运行。

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

— 设计与系统架构规范；

— 系统开发过程中所作的设计选择及采取的质量控制措施；

— 系统开发过程中所用数据的相关信息；

— 对数据质量所作的假设及采取的质量控制措施（例如：假设的统计分布）；

— 在人工智能系统开发或运行期间采取的管理活动（如风险管理）；

— 认证和验证记录；

— AI系统运行时所做的更改；

—B.5中所述的影响评估文件。

该组织应记录与人工智能系统责任操作相关的技术信息。这包括但不限于：

— 编制故障管理方案。具体包括：制定AI系统的回滚方案、关闭AI系统功能、更新流程，以及向客户和用户通报AI系统变更、系统故障信息更新及缓解措施的方案。

—记录用于监测人工智能系统运行状态（即人工智能系统按预期运行且处于正常操作范围内，亦称可观察性）的流程，以及用于处理人工智能系统故障的流程；

— 记录人工智能系统的标准操作规程，包括应监测哪些事件以及如何对事件日志进行优先级排序和审查。还可包括如何调查故障及预防故障；

— 记录负责人工智能系统运行的人员以及负责系统使用问责的人员的职责，特别是与处理人工智能系统故障的影响或管理人工智能系统更新相关的内容；

— 记录系统更新，如系统操作变更、预期用途新增或修改，或其他系统功能变更。

该组织应制定相应程序以应对运营变更，包括向用户传达信息及对变更类型进行内部评估。

文件资料应保持最新且准确无误，并需经组织内相关管理层批准。

当作为用户文档的一部分提供时，应考虑表A.1中提供的控制措施。

B.6.2.8 人工智能系统对事件日志的记录

统治

该组织应确定在人工智能系统生命周期的哪些阶段启用事件日志记录，但至少应在人工智能系统使用期间启用。

实施指南

组织应确保对部署的AI系统进行日志记录， 以自动收集和记录与运行期间发生的特定事件相关的事件日志。此类日志包括但不限于：

— 人工智能系统功能的可追溯性，以确保该系统按预期运行；

— 通过监测人工智能系统的运行状态，检测其在预期操作条件之外的表现，该表现可能导致生产数据出现异常或对相关利益方造成影响。

AI系统事件日志可记录以下信息：每次使用时间与日期、系统运行的生产数据、超出预期操作范围的输出结果等。

事件日志应根据人工智能系统的预期用途及机构数据保留政策保存至所需时长。相关数据保留的法律要求可能适用。

其他信息

某些人工智能系统（如生物识别系统）可能因所在司法管辖区不同而产生额外的记录要求。相关机构应充分了解这些规定。

B.7 AI系统数据

B.7.1 目标

确保组织理解数据在人工智能系统应用与开发、提供或使用其整个生命周期中的作用与影响。

B.7.2 用于开发和增强人工智能系统的数据

统治

该组织应定义、记录并实施与人工智能系统开发相关的数据管理流程。

实施指南

数据管理可涵盖以下主题，但不限于：

— 由于使用数据（其中部分数据可能具有敏感性）而产生的隐私和安全影响；

— 数据依赖型人工智能系统开发可能引发的安全与安全威胁；

— 透明度与可解释性要求，包括数据溯源机制，以及当系统需要具备透明度和可解释性时，需能说明数据如何用于确定人工智能系统的输出；

— 训练数据与实际应用领域的代表性对比；

— 数据的准确性和完整性。

注：关于人工智能系统生命周期及数据管理概念的详细信息，可参阅ISO/IEC 22989标准。

B.7.3 数据的获取

统治

该组织应确定并记录人工智能系统所用数据的获取与筛选细节。

实施指南

该组织可能需要根据其人工智能系统的应用范围和用途，从不同来源获取不同类别的数据。数据采集的具体细节可包括：

— 人工智能系统所需的数据类别；

—所需数据量；

— 数据来源（如内部数据、购买数据、共享数据、公开数据、合成数据）；

— 数据源的特性（如静态、流式、采集式、机器生成）；

— 数据主体的人口统计学特征与属性（如已知或潜在的偏倚或其他系统性误差）；

— 数据的预先处理（例如既往使用情况、是否符合隐私与安全要求）；

——数据权利（如 PII 、版权）；

— 相关元数据（例如数据标注与增强的详细信息）；

— 数据来源。

其他信息

ISO/IEC 19944-1标准中规定的数据类别及数据使用结构，可用于详细记录数据采集与使用的相关细节。

B.7.4 人工智能系统数据的质量

统治

该组织应明确并记录数据质量要求，确保用于开发和运行人工智能系统的数据符合这些要求。

实施指南

人工智能系统在开发和运行过程中所使用的数据质量，可能对系统输出结果的有效性产生重大影响。根据ISO/IEC 25024标准，数据质量是指在特定使用条件下，数据特性满足明示和隐含需求的程度。对于采用监督式或半监督式机器学习的人工智能系统而言，必须对训练数据、验证数据、测试数据和生产数据的质量进行明确定义、量化评估和持续优化，同时企业应确保这些数据完全符合其预期应用场景。该组织应评估偏差对系统性能及公平性的影响，并根据实际需求对所用模型和数据进行必要调整，以提升性能与公平性，确保其符合使用场景要求。

其他信息

有关数据质量的更多信息，可参见[2)]关于分析和机器学习的数据质量的ISO/IEC 5259系列。有关人工智能系统所用数据中不同形式偏倚的更多信息，可参见ISO/IEC TR 24027。

B.7.5 数据溯源

统治

该组织应制定并记录一套流程，用于在其人工智能系统数据及系统生命周期内追踪所用数据的来源。

实施指南

根据ISO 8000-2标准，数据溯源记录应包含数据的创建、更新、转录、抽象、验证及控制权转移等信息。此外，数据共享（不涉及控制权转移）和数据转换也可纳入数据溯源范畴。根据数据来源、内容及使用背景等因素，组织应评估是否需要采取措施验证数据的来源。

B.7.6 数据准备

统治

该组织应明确并记录其数据准备选择标准及所采用的数据制备方法。

实施指南

人工智能系统中使用的数据通常需要经过预处理，才能满足特定任务的需求。例如，机器学习算法有时无法容忍数据缺失或错误输入，非

数据呈现正态分布且尺度差异显著。通过数据预处理方法和转换技术可提升数据质量，若预处理不当可能导致人工智能系统出现错误。人工智能系统常用的数据预处理方法和转换技术包括：

- 数据的统计探索（例如分布、平均值、中位数、标准差、范围、分层、抽样）和统计元数据（例如数据文档化计划（DDI）规范[28]）；

— 清理（即更正条目、处理缺失条目）；

— 填补法（即用于填补缺失条目的方法）；

—标准化；

- 缩小；

— 目标变量的标注；

— 编码（例如将分类变量转换为数值）。

针对特定的人工智能任务，组织应记录其选择特定数据准备方法及转换的标准，以及该任务中实际采用的具体方法和转换。

注：有关机器学习数据准备的更多信息，请参阅ISO/IEC 5259系列2和ISO/IEC 23053。

B.8 向有关各方提供的信息

B.8.1 目的

确保相关利益方掌握必要信息，以充分理解并评估风险及其影响（包括积极与消极两方面）。

B.8.2 系统文档及用户信息

统治

该组织应确定并向系统用户提供必要信息。

实施指南

关于人工智能系统的信息可包含技术细节与操作说明，以及根据具体情境向用户发出的交互提示。该信息既涵盖系统本身，也包含其可能产生的输出结果（例如提示用户某张图片由AI生成）。

尽管人工智能系统可能较为复杂，但用户必须清楚自己与AI系统的交互方式及其运作原理。用户还需理解系统的预期用途、潜在危害或益处。部分系统文档可能需要针对特定技术需求（如系统管理员）进行优化，组织应当充分考虑不同利益相关方的需求，明确可理解性对他们而言的具体含义。这些信息还应便于获取，既便于用户查找，也应满足需要额外无障碍功能的用户需求。

可向用户提供以下信息，但不限于：

— 系统的目的；

— 用户正在与人工智能系统进行交互；

— 如何与系统交互；

— 如何及何时覆盖系统；

— 系统运行的技术要求，包括所需计算资源、系统局限性及其预期使用寿命；

—需要人类监督；

- 关于准确性和性能的信息；

- 影响评估的相关信息，包括潜在的益处和危害，特别是如果它们适用于特定背景或某些人口群体（见B.5.2和B.5.4）；

— 修订关于该系统效益的声明；

— 系统运行方式的更新与变更，以及必要的维护措施，包括其频率；

— 联系信息；

- 系统使用的教育材料。

该组织用于确定是否提供及提供何种信息的标准应予以记录。相关标准包括但不限于：人工智能系统的预期用途及可合理预见的误用、使用者的专业水平以及人工智能系统的具体影响。

信息可通过多种方式向用户提供，包括书面使用说明、系统内置的警报及其他通知、网页信息等。根据组织采用的信息提供方式，应验证用户是否具备访问权限，并确保所提供信息完整、最新且准确。

B.8.3 外部报告

统治

该组织应为相关方提供报告系统不良影响的能力。

实施指南

在监测系统运行过程中报告的问题和故障的同时，组织还应为用户提供或第三方报告不良影响（如不公平性）的能力。

B.8.4 事件通报

统治

该组织应制定并记录向系统用户通报事件的计划。

实施指南

与人工智能系统相关的事件可能特指该系统本身，或涉及信息安全或隐私问题（如数据泄露）。该组织应当明确其在系统运行情境下向用户及其他利益相关方通报事件的义务。例如，涉及产品中影响安全的人工智能组件的事件，其通知要求可能与其他类型系统不同。法律要求（如合同）和监管活动可能适用，这些要求可规定：

必须通报的事件类型；

— 通知时限；

— 是否及应通知哪些主管部门；

— 需要传达的细节。

该组织可以将AI的事件响应和报告活动整合到其更广泛的组织事件管理活动中，但应意识到与AI系统或AI系统单个组件相关的特殊要求（例如，系统培训数据的 PII 数据泄露可能具有与隐私相关的不同报告要求）。

其他信息

ISO/IEC 27001和ISO/IEC 27701分别提供了关于安全和隐私事件管理的更多细节。

B.8.5 供相关方参考的信息

统治

该组织应确定并记录其向利益相关方报告人工智能系统信息的义务。

实施指南

在某些情况下，司法管辖区可能要求将系统信息与监管机构等主管部门共享。相关信息可在规定时限内向客户或监管机构等利益相关方报告。共享的信息可包括但不限于：

— 技术系统文档，包括但不限于：训练集、验证集和测试集数据集，以及算法选择依据、验证与确认记录；

- 与系统相关的风险；

— 评估结果；

— 日志和其他系统记录。

该组织应明确其在此方面的义务，并确保向相关主管部门提供准确信息。此外，该组织还应了解与执法部门共享信息相关的管辖权要求。

B.9 使用AI系统

B.9.1 目标

确保组织负责任地使用人工智能系统，并遵守组织政策。

B.9.2 人工智能系统的负责任使用流程

统治

该组织应明确并记录人工智能系统负责任使用的流程。

实施指南

根据具体情境，该组织在决定是否采用特定人工智能系统时需综合考量多项因素。无论人工智能系统是由组织自行开发还是从第三方获取，组织都应明确这些考量因素，并制定相关政策予以应对。例如：

— 需要批准；

— cost (including for ongoing monitoring and maintenance);

- 批准的采购要求；

— 适用于该组织的法律要求。

若组织已接受其他系统、资产等使用的相关政策，可根据需要将其纳入。

B.9.3 人工智能系统负责任使用的相关目标

统治

该组织应明确并记录指导人工智能系统负责任使用的相关目标。

实施指南

在不同情境下运作的组织对人工智能系统负责任开发的构成要素可能持有不同的期望与目标。根据具体情境，组织应明确其与负责任使用相关的目标。部分目标包括：

公平；

——问责；

透明度；

— 可解释性；

—可靠性；

- 安全；

—稳健性和冗余性；

——隐私与安全；

— 可及性。

组织在明确目标后，应建立相应机制以实现其内部目标。这包括评估第三方解决方案是否符合组织目标，或内部开发的解决方案是否适用于预期用途。该组织应确定在人工智能系统生命周期的哪些阶段应纳入有意义的人工监督目标。这可能包括：

— 由人工评审员对人工智能系统的输出结果进行核查，包括有权推翻人工智能系统作出的决策；

— 确保在需要时纳入人工监督，以确保人工智能系统按照相关说明或其他与预期部署相关文件的要求被合理使用；

— 监测人工智能系统的性能，包括其输出结果的准确性；

— 报告与人工智能系统输出相关的问题及其对相关利益方的影响；

— 报告对人工智能系统性能或能力变化的担忧，即该系统无法对生产数据作出正确输出；

— 考虑自动化决策是否适用于负责任地使用人工智能系统及其预期用途。

人工智能系统影响评估可为是否需要人工监督提供依据（见B.5）。应告知并培训参与人工智能系统相关人类监督活动的人员

理解向人工智能系统传达的指令及其他文档，并理解其为满足人类监督目标而执行的职责。在报告性能问题时，人类监督可增强自动化监测。

其他信息

附录C提供了风险管理的组织目标示例，可用于确定人工智能系统使用的目标。

B.9.4 人工智能系统的预期用途

统治

该组织应确保人工智能系统按照其预期用途及随附文件进行使用。

实施指南

AI系统应根据AI系统相关说明和其他文件（参见B.8.2）进行部署。部署可能需要特定资源来支持部署，包括确保按要求实施人工监督（参见B.9.3）。为了可接受地使用AI系统，可能需要确保AI系统所用数据与AI系统相关文件一致，以确保AI系统性能准确。

应监测人工智能系统的运行（参见B.6.2.6）。如果按照相关说明正确部署人工智能系统，但对相关利益方或组织的法律要求造成影响，组织应向组织内的相关人员以及人工智能系统的任何第三方供应商传达其关切。

组织应当保存与人工智能系统部署和运行相关的事件日志或其他文档，这些记录可用于证明人工智能系统正按预期使用，或帮助传达与系统预期用途相关的关切。事件日志及其他文档的保存期限取决于人工智能系统的预期用途、组织的数据保留政策以及相关数据保留的法律要求。

B.10 第三方和客户关系

B.10.1 目标

为确保组织明确自身职责并保持问责制，同时在人工智能系统生命周期的任何阶段涉及第三方时，风险得到适当分摊。

B.10.2 责任分配

统治

该组织应确保其人工智能系统生命周期内的职责在组织、合作伙伴、供应商、客户及第三方之间进行分配。

实施指南

在人工智能系统生命周期中，责任可划分为数据提供方、算法与模型提供方、AI系统开发或使用方，并需对部分或全部相关方负责。组织应记录所有参与AI系统生命周期的各方及其职责，并明确其责任范围。

如果组织向第三方提供AI系统，组织应确保以负责任的方式开发AI系统。请参阅B.6中的控制和指导。该组织应能够提供为AI系统所需的文件（请参阅B.6.2.7和B.8.2）。

向相关利益方及组织所供人工智能系统第三方进行系统说明。

当处理的数据包含 PII 时，责任通常由PII 处理者和控制者共同承担。ISO/IEC 29100提供了关于 PII 控制者和 PII 处理者的更多信息。若要保护 PII 的隐私，应考虑采用ISO/IEC 27701中描述的控制措施。根据组织和人工智能系统在 PII 上的数据处理活动，以及组织在整个生命周期中对人工智能系统应用和开发的角色，组织可以担任 PII 控制者（或联合 PII 控制者）、 PII 处理者或两者兼有。

B.10.3 供应商

统治

该组织应建立相应流程，确保其对供应商所提供服务、产品或材料的使用符合组织在人工智能系统负责任开发与应用方面的方针。

实施指南

开发或使用人工智能系统的组织可通过多种方式利用供应商，包括采购数据集、机器学习算法或模型， 以及系统组件（如软件库），甚至整个AI系统本身，既可独立使用，也可作为其他产品（如车辆）的组成部分。

各组织在确定供应商的选择、对供应商的要求以及对供应商需要进行的持续监测和评价的水平时，应考虑不同类型的供应商、它们提供的服务以及这可能对系统和整个组织造成的不同程度的风险。

各组织应记录人工智能系统及其组件如何整合至本组织开发或使用的AI系统中。

若组织认为供应商提供的AI系统或其组件未能按预期运行，或可能对个人、群体或社会造成影响，且这些影响与组织所采取的AI系统责任管理方针不符，则应要求供应商采取纠正措施。该组织可决定与供应商合作以实现这一目标。

组织应确保AI系统的供应商提供与AI系统相关的适当和充分的文件（参见B.6.2.7和B.8.2）。

B.10.4 客户

统治

该组织应确保其对人工智能系统开发与使用的负责任方法，充分考虑客户的期望与需求。

实施指南

企业在提供与人工智能系统相关的产品或服务时（即作为供应商时），必须充分理解客户的期望与需求。这些需求可能体现在产品或服务的设计/工程阶段的具体要求，也可能通过合同条款或通用使用协议的形式呈现。同一企业可能与不同类型的客户建立合作关系，而这些客户群体往往具有差异化的具体需求与期望。

该组织需特别理解供应商与客户关系的复杂性，明确在满足需求与期望的同时，人工智能系统提供商与客户各自应承担的责任。

例如，该组织可识别客户使用其人工智能产品和服务时产生的风险，并通过向客户提供适当信息来处理这些风险，从而帮助客户自行应对相关风险。

作为适当信息的示例，当AI系统适用于特定使用领域时，应向客户传达该领域的限制。参见B.6.2.7和B.8.2。

附件C
提供信息的

人工智能相关组织潜在目标与风险源

C.1 概要

本附录概述了企业在进行风险管理时可考虑的潜在组织目标、风险源及描述。需要说明的是，本附录并非详尽无遗的指南，也不适用于所有组织。企业应自行确定相关的目标和风险源。ISO/IEC 23894标准对这些目标、风险源及其与风险管理的关系提供了更详细的信息。对人工智能系统的评估——无论是初始评估、定期评估还是必要时的评估——都为验证该系统是否符合组织目标提供了依据。

C.2 目标

C.2.1 责任追究

人工智能的应用可以改变现有的问责框架。

C.2.2 人工智能专业知识

需要选拔一批具备跨学科技能、在评估、开发和部署人工智能系统方面具有专业知识的专职专家。

C.2.3 训练数据与测试数据的可用性及质量

基于机器学习（ML）的AI系统需要训练数据、验证数据和测试数据，以实现系统预期行为的训练与验证。

C.2.4 环境影响

人工智能的应用可能对环境产生积极和消极的影响。

C.2.5 公平性

人工智能系统在自动化决策中的不当应用可能对特定个人或群体不公平。

C.2.6 可维护性

可维护性是指组织处理人工智能系统修改的能力，以纠正缺陷或适应新需求。

C.2.7 隐私

个人及敏感数据（如健康记录）的滥用或泄露可能对数据主体造成有害影响。

C.2.8 耐受性

在人工智能领域，鲁棒性特性体现系统在新数据上保持与训练数据或典型操作数据相当性能的能力（或无能）。

C.2.9 安全

安全性是指在特定条件下，系统不会导致人类生命、健康、财产或环境处于危险状态的预期。

C.2.10 安全

在人工智能（AI）领域，尤其是基于机器学习（ML）方法的AI系统中，需要考虑超越传统信息与系统安全关注的新安全问题。

C.2.11 透明度与可解释性

透明度既涉及运营人工智能系统的组织特征，也涉及系统本身。可解释性则指以人类可理解的方式，向相关方提供影响人工智能系统结果的重要因素的说明。

C.3 风险源

C.3.1 环境的复杂性

当人工智能系统在复杂环境中运行时， 由于情境范围广泛，其性能可能存在不确定性，从而成为风险来源（例如复杂自主驾驶环境)。

C.3.2 缺乏透明度与可解释性

无法向利益相关方提供适当信息可能构成风险来源（即涉及组织的可信度与问责性）。

C.3.3 自动化水平

自动化水平可能对安全、公平性或保密性等多个关注领域产生影响。

C.3.4 与机器学习相关的风险源

用于机器学习（ML）的数据质量及数据收集过程可能成为风险来源，因其可能影响安全性与稳健性等目标（例如因数据质量问题或数据污染所致）。

C.3.5 系统硬件问题

硬件相关风险源包括基于缺陷组件的硬件错误或在不同系统间转移经过训练的机器学习模型。

C.3.6 系统生命周期相关问题

风险来源可能出现在人工智能系统整个生命周期中（例如设计缺陷、部署不足、维护缺失、退役问题）。

C.3.7 技术准备度

风险源既可能源于未知因素（如系统局限性、边界条件及性能漂移）导致的技术不成熟，也可能源于技术自满导致的成熟技术问题。

附件D

跨领域或跨部门（提供信息）使用人工智能管理系统

D.1 概要

该管理体系适用于所有开发、提供或使用人工智能系统产品及服务的组织。因此，其适用范围广泛，涵盖不同行业领域中需对利益相关方履行义务、遵循良好实践、满足预期或承担合同承诺的各类产品与服务。具体行业示例包括：

健康；

——防御；

——运输；

——金融；

——就业；

- 能量。

在制定和使用人工智能系统时，可考虑多种组织目标（具体目标示例参见附录C）。本文件从人工智能技术特性角度出发，提供相关要求与指导建议。针对部分潜在目标，目前已有通用或行业专用的管理体系标准可供参考。这些标准通常从技术中立的角度考量目标设定，而人工智能管理体系则专门针对该技术特性提出具体要求。

人工智能系统不仅包含运用AI技术的组件，还可能整合多种技术与组件。因此，负责任地开发和使用AI系统时，不仅要考虑AI特有的考量因素，还需将整个系统及其所采用的所有技术组件纳入考量。即便针对AI技术本身，除AI特有的考量因素外，还需考虑其他方面。例如， 由于AI属于信息处理技术，信息安全原则普遍适用。诸如安全性、保密性、隐私保护及环境影响等目标，应当进行整体管理而非分别针对AI系统与其他组件单独处理。因此，将AI管理系统与通用或行业特定的管理标准进行整合，对于负责任地开发和使用AI系统至关重要。

D.2 AI管理系统与其他管理系统的标准整合

在提供或使用人工智能系统时，组织可能涉及其他管理体系标准所涵盖的方面，从而产生相关目标或义务。例如，这些可包括ISO/IEC 27001、ISO/IEC 27701和ISO 9001中分别涵盖的安全、隐私和质量主题。

在提供、使用或开发人工智能系统时，潜在的相关通用管理系统标准（但不限于以下标准）包括：

—ISO/IEC 27001：在大多数情况下，安全是实现组织与人工智能系统目标的关键。组织如何追求安全目标取决于其具体情境和自身政策。如果组织认为有必要实施人工智能管理系统

为以同样全面系统的方式实现安全目标，企业可依据ISO/IEC 27001标准实施信息安全管理体系。 由于ISO/IEC 27001与人工智能管理体系均采用统一的高层架构，其整合应用不仅便于操作，更能为企业带来显著效益。在此情况下，本文件中涉及信息安全的管控措施（详见B.6.1.2节）可与企业实施ISO/IEC 27001的流程实现无缝衔接。

—ISO/IEC 27701：在众多应用场景中，个人身份信息（PII）常由人工智能系统处理。企业通过采用该标准，既能满足隐私保护的法定要求，又能实现自身政策目标。与ISO/IEC 27001标准类似，将ISO/IEC 27701标准融入人工智能管理体系能为企业带来双重效益。人工智能管理体系中涉及隐私保护的目标与管控措施（详见B.2.3和B.5.4节），可与企业实施的ISO/IEC 27701标准形成协同效应。

— ISO 9001：对众多企业而言，符合ISO 9001标准是其以客户为中心、切实关注内部效能的关键标志。通过独立开展ISO 9001合规评估，不仅能促进跨组织业务发展，更能增强客户对产品或服务的信任。当涉及人工智能技术时，若将AI管理系统与ISO 9001标准同步实施，可显著提升客户对组织或AI系统的信任度。该管理系统能与ISO 9001要求（如风险管理、软件开发、供应链协调等）形成互补，助力企业实现其战略目标。

除上述通用管理体系标准外，人工智能管理系统还可与特定行业专用管理系统协同使用。例如，ISO 22000与人工智能管理系统均适用于食品生产、加工及物流领域的人工智能系统。另一个典型案例是ISO 13485标准。实施AI管理系统可支持ISO 13485中与医疗器械软件相关的要求或来自医疗部门的其他国际标准的要求，如IEC 62304。

参考文献

[1] ISO 8000-2，*数据质量 第2部分：词汇*

[2] ISO9001，*质量管理体系——要求*

[3] ISO 9241-210，*人机交互的人体工程学——第210部分：交互式系统的人本设计*

[4] ISO 13485，*医疗器械——质量管理体系——法规要求*

[5] ISO 22000，*食品安全管理体系——食品链中任何组织的要求*

[6] IEC 62304，*医疗器械软件——软件生命周期过程*

[7] ISO/IEC指南51，*安全方面——标准中所含安全方面的准则*

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

[9] ISO/IEC 5259（所有部分[2)]），*分析和机器学习（ML）的数据质量*

[10] ISO/IEC 5338，*信息技术——人工智能——人工智能系统生命周期过程*

[11] ISO/IEC 17065，*符合性评估——对产品、过程和服务认证机构的要求*

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

[13] ISO/IEC 23053，*人工智能（AI）系统使用机器学习（ML）的框架*

[14] ISO/IEC 23894，*信息技术——人工智能——风险管理指南*

[15] ISO/IEC TR 24027，*信息技术——人工智能（AI）——人工智能系统和人工智能辅助决策中的偏倚*

[16] ISO/IEC TR 24029-1，*人工智能（AI）——神经网络稳健性评估——第1部分：概述*

[17] ISO/IEC TR 24368，*信息技术——人工智能——道德和社会问题概述*

[18] ISO/IEC 25024，*系统和软件工程——系统和软件质量要求和评价（SQuaRE）——数据质量的测量*

[19] ISO/IEC 25059，*软件工程——系统和软件质量要求和评价（SQuaRE）——人工智能系统质量模型*

[20] ISO/IEC 27000：2018，*信息技术——安全技术——信息安全管理系统——概述和词汇*

[21] ISO/IEC 27701，*安全技术-隐私信息管理的ISO/IEC 27001和ISO/IEC 27002扩展-要求和指南*

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

[23] ISO/IEC 29100，*信息技术——安全技术——隐私框架*

[24] ISO 31000：2018，*风险管理——指南*

[25] ISO 37002，*举报管理—指南*

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

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

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

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

本页故意留白

INTERNATIONAL STANDARD

ISO/IEC 42001

First edition
2023-12

Information technology — Artificial intelligence — Management system

Reference number
ISO/IEC 42001:2023(E)

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

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

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

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

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

Introduction

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

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

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

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

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

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

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

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

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

— management of risks and opportunities;

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

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

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

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

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

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

Compatibility with other management system standards

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

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

INTERNATIONAL STANDARD ISO/IEC 42001:2023(E)

Information technology — Artificial intelligence — Management system

1 Scope

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

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

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

2 Normative references

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

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

3 Terms and definitions

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.6
objective
result to be achieved

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

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

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

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

3.7
risk
effect of uncertainty

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

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

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

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

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

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

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

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

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

Note 2 to entry: Documented information can refer to:

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

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

— evidence of results achieved (records).

3.11
performance
measurable result

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

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

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

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

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

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

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

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

3.15
conformity
fulfilment of a *requirement* (3.14)

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

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

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

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

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

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

3.19
measurement
process (3.8) to determine a value

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[SOURCE: ISO/IEC 5259-1:—[1)], 3.4]

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

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

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

4 Context of the organization

4.1 Understanding the organization and its context

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

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

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

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

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

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

— AI customers, including AI users;

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

— AI subjects, including data subjects and other subjects;

— relevant authorities, including policymakers and regulators.

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

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

a) external context related considerations such as:

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

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

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

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

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

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

b) internal context related considerations such as:

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

2) contractual obligations;

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

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

4.2 Understanding the needs and expectations of interested parties

The organization shall determine:

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

— the relevant requirements of these interested parties;

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

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

4.3 Determining the scope of the AI management system

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

When determining this scope, the organization shall consider:

— the external and internal issues referred to in 4.1;

— the requirements referred to in 4.2.

The scope shall be available as documented information.

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

4.4 AI management system

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

5 Leadership

5.1 Leadership and commitment

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

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

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

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

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

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

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

— promoting continual improvement;

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

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

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

5.2 AI policy

Top management shall establish an AI policy that:

a) is appropriate to the purpose of the organization;

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

c) includes a commitment to meet applicable requirements;

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

The AI policy shall:

— be available as documented information;

— refer as relevant to other organizational policies;

— be communicated within the organization;

— be available to interested parties, as appropriate.

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

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

5.3 Roles, responsibilities and authorities

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

Top management shall assign the responsibility and authority for:

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

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

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

6 Planning

6.1 Actions to address risks and opportunities

6.1.1 General

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

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

— prevent or reduce undesired effects;

— achieve continual improvement.

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

— distinguishing acceptable from non-acceptable risks;

— performing AI risk assessments;

— conducting AI risk treatment;

— assessing AI risk impacts.

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

The organization shall determine the risks and opportunities according to:

— the domain and application context of an AI system;

— the intended use;

— the external and internal context described in 4.1.

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

The organization shall plan:

a) actions to address these risks and opportunities;

b) how to:

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

2) evaluate the effectiveness of these actions.

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

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

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

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

6.1.2 AI risk assessment

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

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

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

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

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

d) analyses the AI risks to:

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

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

3) determine the levels of risk;

e) evaluates the AI risks to:

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

2) prioritize the assessed risks for risk treatment.

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

6.1.3 AI risk treatment

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

a) select appropriate AI risk treatment options;

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

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

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

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

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

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

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

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

g) formulate an AI risk treatment plan.

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

— aligned to the objectives in 6.2;

— available as documented information;

— communicated within the organization;

— available to interested parties, as appropriate.

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

6.1.4 AI system impact assessment

The organization shall define a process for assessing the potential consequences for individuals or groups of individuals, or both, and societies that can result from the development, provision or use of AI systems.

The AI system impact assessment shall determine the potential consequences an AI system's deployment, intended use and foreseeable misuse has on individuals or groups of individuals, or both, and societies.

The AI system impact assessment shall take into account the specific technical and societal context where the AI system is deployed and applicable jurisdictions.

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

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

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

6.2 AI objectives and planning to achieve them

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

The AI objectives shall:

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

b) be measurable (if practicable);

c) take into account applicable requirements;

d) be monitored;

e) be communicated;

f) be updated as appropriate;

g) be available as documented information.

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

— what will be done;

— what resources will be required;

— who will be responsible;

— when it will be completed;

— how the results will be evaluated.

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

6.3 Planning of changes

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

7 Support

7.1 Resources

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

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

7.2 Competence

The organization shall:

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

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

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

Appropriate documented information shall be available as evidence of competence.

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

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

7.3 Awareness

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

— the AI policy (see 5.2);

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

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

7.4 Communication

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

— what it will communicate;

— when to communicate;

— with whom to communicate;

— how to communicate.

7.5 Documented information

7.5.1 General

The organization's AI management system shall include:

a) documented information required by this document;

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

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

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

— the complexity of processes and their interactions;

— the competence of persons.

7.5.2 Creating and updating documented information

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

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

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

— review and approval for suitability and adequacy.

7.5.3 Control of documented information

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

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

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

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

— distribution, access, retrieval and use;

— storage and preservation, including preservation of legibility;

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

— retention and disposition.

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

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

8 Operation

8.1 Operational planning and control

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

— establishing criteria for the processes;

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

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

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

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

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

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

8.2 AI risk assessment

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

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

8.3 AI risk treatment

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

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

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

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

8.4 AI system impact assessment

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

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

9 Performance evaluation

9.1 Monitoring, measurement, analysis and evaluation

The organization shall determine:

— what needs to be monitored and measured;

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

— when the monitoring and measuring shall be performed;

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

Documented information shall be available as evidence of the results.

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

9.2 Internal audit

9.2.1 General

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

a) conforms to:

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

   2) the requirements of this document;

b) is effectively implemented and maintained.

9.2.2 Internal audit programme

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

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

The organization shall:

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

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

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

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

9.3 Management review

9.3.1 General

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

9.3.2 Management review inputs

The management review shall include:

a) the status of actions from previous management reviews;

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

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

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

   1) nonconformities and corrective actions;

   2) monitoring and measurement results;

   3) audit results;

e) opportunities for continual improvement.

9.3.3 Management review results

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

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

10 Improvement

10.1 Continual improvement

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

10.2 Nonconformity and corrective action

When a nonconformity occurs, the organization shall:

a) react to the nonconformity and as applicable:
    1) take action to control and correct it;
    2) deal with the consequences;
b) evaluate the need for action to eliminate the cause(s) of the nonconformity, so that it does not recur or occur elsewhere, by:
    1) reviewing the nonconformity;
    2) determining the causes of the nonconformity;
    3) determining if similar nonconformities exist or can potentially occur;
c) implement any action needed;
d) review the effectiveness of any corrective action taken;
e) make changes to the AI management system, if necessary.

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

Documented information shall be available as evidence of:

— the nature of the nonconformities and any subsequent actions taken;
— the results of any corrective action.

Annex A
(normative)

Reference control objectives and controls

A.1 General

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

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

Table A.1 — Control objectives and controls

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

Table A.1 *(continued)*

A.4.5	System and computing resources	As part of resource identification, the organization shall document information about the system and computing resources utilized for the AI system.
A.4.6	Human resources	As part of resource identification, the organization shall document information about the human resources and their competences utilized for the development, deployment, operation, change management, maintenance, transfer and decommissioning, as well as verification and integration of the AI system.
A.5 Assessing impacts of AI systems		
Objective: To assess AI system impacts to individuals or groups of individuals, or both, and societies affected by the AI system throughout its life cycle.		
	Topic	Control
A.5.2	AI system impact assessment process	The organization shall establish a process to assess the potential consequences for individuals or groups of individuals, or both, and societies that can result from the AI system throughout its life cycle.
A.5.3	Documentation of AI system impact assessments	The organization shall document the results of AI system impact assessments and retain results for a defined period.
A.5.4	Assessing AI system impact on individuals or groups of individuals	The organization shall assess and document the potential impacts of AI systems to individuals or groups of individuals throughout the system's life cycle.
A.5.5	Assessing societal impacts of AI systems	The organization shall assess and document the potential societal impacts oftheir AI systems throughout their life cycle.
A.6 AI system life cycle		
A.6.1 Management guidance for AI system development		
Objective: To ensure that the organization identifies and documents objectives and implements processes for the responsible design and development of AI systems.		
	Topic	Control
A.6.1.2	Objectives for responsible development ofAI system	The organization shall identify and document objectives to guide the responsible development AI systems, and take those objectives into account and integrate measures to achieve them in the development life cycle.
A.6.1.3	Processes for responsible AI system design and development	The organization shall define and document the specific processes for the responsible design and development of the AI system.
A.6.2 AI system life cycle		
Objective: To define the criteria and requirements for each stage of the AI system life cycle.		
	Topic	Control
A.6.2.2	AI system requirements and specification	The organization shall specify and document requirements for new AI systems or material enhancements to existing systems.
A.6.2.3	Documentation of AI system design and development	The organization shall document the AI system design and development based on organizational objectives, documented requirements and specification criteria.
A.6.2.4	AI system verification and validation	The organization shall define and document verification and validation measures for the AI system and specify criteria for their use.
A.6.2.5	AI system deployment	The organization shall document a deployment plan and ensure that appropriate requirements are met prior to deployment.

Table A.1 *(continued)*

A.6.2.6	AI system operation and monitoring	The organization shall define and document the necessary elements for the ongoing operation of the AI system. At the minimum, this should include system and performance monitoring, repairs, updates and support.
A.6.2.7	AI system technical documentation	The organization shall determine what AI system technical documentation is needed for each relevant category of interested parties, such as users, partners, supervisory authorities, and provide the technical documentation to them in the appropriate form.
A.6.2.8	AI system recording of event logs	The organization shall determine at which phases ofthe AI system life cycle, record keeping of event logs should be enabled, but at the minimum when the AI system is in use.
A.7 Data for AI systems		
Objective: To ensure that the organization understands the role and impacts of data in AI systems in the application and development, provision or use of AI systems throughout their life cycles.		
	Topic	Control
A.7.2	Data for development and enhancement ofAI system	The organization shall define, document and implement data management processes related to the development of AI systems.
A.7.3	Acquisition of data	The organization shall determine and document details about the acquisition and selection of the data used in AI systems.
A.7.4	Quality of data for AI systems	The organization shall define and document requirements for data quality and ensure that data used to develop and operate the AI system meet those requirements.
A.7.5	Data provenance	The organization shall define and document a process for recording the provenance of data used in its AI systems over the life cycles ofthe data and the AI system.
A.7.6	Data preparation	The organization shall define and document its criteria for selecting data preparations and the data preparation methods to be used.
A.8 Information for interested parties of AI systems		
Objective: To ensure that relevant interested parties have the necessary information to understand and assess the risks and their impacts (both positive and negative).		
	Topic	Control
A.8.2	System documentation and information for users	The organization shall determine and provide the necessary information to users of the AI system.
A.8.3	External reporting	The organization shall provide capabilities for interested parties to report adverse impacts ofthe AI system.
A.8.4	Communication of incidents	The organization shall determine and document a plan for communicating incidents to users of the AI system.
A.8.5	Information for interested parties	The organization shall determine and document their obligations to reporting information about the AI system to interested parties.
A.9 Use of AI systems		
Objective: To ensure that the organization uses AI systems responsibly and per organizational policies.		
	Topic	Control
A.9.2	Processes for responsible use of AI systems	The organization shall define and document the processes for the responsible use of AI systems.
A.9.3	Objectives for responsible use of AI system	The organization shall identify and document objectives to guide the responsible use of AI systems.

Table A.1 *(continued)*

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

Annex B
(normative)

Implementation guidance for AI controls

B.1 General

The implementation guidance documented in this annex relates to the controls listed in Table A.1. It provides information to support the implementation of the controls listed in Table A.1 and to meet the control objective, but organizations do not have to document or justify inclusion or exclusion of implementation guidance in the statement of applicability (see 6.1.3).

The implementation guidance is not always suitable or sufficient in all situations and does not always fulfil the organization's specific control requirements. The organization can extend or modify the implementation guidance or define their own implementation of a control according to their specific requirements and risk treatment needs.

This annex is to be used as guidance for determining and implementing controls for AI risk treatment in the AI management system defined in this document. Additional organizational and technical controls other than those included in this annex can be determined (see AI system management risk treatment in 6.1.3). This annex can be regarded as a starting point for developing organization-specific implementation of controls.

B.2 Policies related to AI

B.2.1 Objective

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

B.2.2 AI policy

Control

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

Implementation guidance

The AI policy should be informed by:

— business strategy;

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

— the level of risk posed by the AI systems;

— legal requirements, including contracts;

— the risk environment of the organization;

— impact to relevant interested parties (see 6.1.4).

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

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

— processes for handling deviations and exceptions to policy.

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

— AI resources and assets;

— AI system impact assessments (see 6.1.4);

— AI system development.

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

B.2.3 Alignment with other organizational policies

Control

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

Implementation guidance

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

Other information

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

B.2.4 Review of the AI policy

Control

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

Implementation guidance

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

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

B.3 Internal organization

B.3.1 Objective

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

B.3.2 AI roles and responsibilities

Control

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

Implementation guidance

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

— risk management;

— AI system impact assessments;

— asset and resource management;

— security;

— safety;

— privacy;

— development;

— performance;

— human oversight;

— supplier relationships;

— demonstrate its ability to consistently fulfil legal requirements;

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

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

B.3.3 Reporting of concerns

Control

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

Implementation guidance

The reporting mechanism should fulfil the following functions:

a) options for confidentiality or anonymity or both;

b) available and promoted to employed and contracted persons;

c) staffed with qualified persons;

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

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

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

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

h) provides response mechanisms within an appropriate time frame.

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

Other information

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

B.4 Resources for AI systems

B.4.1 Objective

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

B.4.2 Resource documentation

Control

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

Implementation guidance

Documentation of resources of the AI system is critical for understanding risks, as well as potential AI system impacts (both positive and negative) to individuals or groups of individuals, or both, and societies. The documentation of such resources (which can utilize, for instance, data flow diagrams or system architecture diagrams) can inform the AI system impact assessments (see B.5).

Resources can include, but are not limited to:

— AI system components;

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

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

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

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

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

Other information

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

B.4.3 Data resources

Control

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

Implementation guidance

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

— the provenance of the data;

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

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

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

— process for labelling data;

— intended use of the data;

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

— applicable data retention and disposal policies;

— known or potential bias issues in the data;

— data preparation.

B.4.4 Tooling resources

Control

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

Implementation guidance

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

— algorithm types and machine learning models;

— data conditioning tools or processes;

— optimization methods;

— evaluation methods;

— provisioning tools for resources;

— tools to aid model development;

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

Other information

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

B.4.5 System and computing resources

Control

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

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

Implementation guidance

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

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

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

— processing resources (including network and storage);

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

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

NOTE ISO/IEC 22989 describes various system resource considerations.

B.4.6 Human resources

Control

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

Implementation guidance

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

— data scientists;

— roles related to human oversight of AI systems;

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

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

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

B.5 Assessing impacts of AI systems

B.5.1 Objective

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

B.5.2 AI system impact assessment process

Control

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

Implementation guidance

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

The organization should consider whether an AI system affects:

— the legal position or life opportunities of individuals;

— the physical or psychological well-being of individuals;

— universal human rights;

— societies.

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

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

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

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

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

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

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

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

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

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

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

c) who performs the AI system impact assessment;

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

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

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

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

Other information

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

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

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

B.5.3 Documentation of AI system impact assessments

Control

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

Implementation guidance

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

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

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

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

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

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

— relevant demographic groups the system is applicable to;

— complexity of the system;

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

— employment and staff skilling.

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

Control

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

Implementation guidance

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

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

— fairness;

— accountability;

— transparency and explainability;

— security and privacy;

— safety and health;

— financial consequences;

— accessibility;

— human rights.

Other information

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

B.5.5 Assessing societal impacts of AI systems

Control

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

Implementation guidance

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

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

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

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

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

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

Other information

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

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

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

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

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

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

B.6 AI system life cycle

B.6.1 Management guidance for AI system development

B.6.1.1 Objective

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

B.6.1.2 Objectives for responsible development of AI system

Control

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

Implementation guidance

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

Other information

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

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

Control

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

Implementation guidance

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

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

— testing requirements and planned means for testing;

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

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

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

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

— release criteria;

— approvals and sign-offs necessary at various stages;

— change control;

— usability and controllability;

— engagement of interested parties.

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

B.6.2 AI system life cycle

B.6.2.1 Objective

To define the criteria and requirements for each stage of the AI system life cycle.

B.6.2.2 AI system requirements and specification

Control

The organization should specify and document requirements for new AI systems or material enhancements to existing systems.

Implementation guidance

The organization should document the rationale for developing an AI system and its goals. Some of the factors that should be considered, documented and understood can include:

a) why the AI system is to be developed, for example, is this driven by a business case, customer request or by government policy;

b) how the model can be trained and how data requirements can be achieved.

AI system requirements should be specified and should span the entire AI system life cycle. Such requirements should be revisited in cases where the developed AI system is unable to operate as intended or new information arises that can be used to change and to improve the requirements. For instance, it can become unfeasible from a financial perspective to develop the AI system.

Other information

The processes for describing the AI system life cycle are provided by ISO/IEC 5338. For more information about human-centred design for interactive systems, see ISO 9241-210.

B.6.2.3 Documentation of AI system design and development

Control

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

Implementation guidance

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

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

— learning algorithm and type of machine learning model utilized;

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

— evaluation and refinement of models;

— hardware and software components;

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

— interface and presentation of outputs;

— how humans can interact with the system;

— interoperability and portability considerations.

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

Other information

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

B.6.2.4 AI system verification and validation

Control

The organization should define and document verification and validation measures for the AI system and specify criteria for their use.

Implementation guidance

The verification and validation measures can include, but are not limited to:

— testing methodologies and tools;

— selection of test data and their representation of the intended domain of use;

— release criteria requirements.

The organization should define and document evaluation criteria such as, but not limited to:

— a plan to evaluate the AI system components and the whole AI system for risks related to impacts on individuals or groups of individuals, or both, and societies;

— the evaluation plan can be based on, for example:

  — reliability and safety requirements of the AI system, including acceptable error rates for the AI system performance;

  — responsible AI system development and use objectives such as those in B.6.1.2 and B.9.3;

  — operational factors such as quality of data, intended use, including acceptable ranges of each operational factor;

  — any intended uses which can require more rigorous operational factors to be defined, including different acceptable ranges for operational factors or lower error rates;

— the methods, guidance or metrics to be used to evaluate whether relevant interested parties who make decisions or are subject to decisions based on the AI system outputs can adequately interpret the AI system outputs. The frequency of evaluation should be determined and can be based upon results from an AI system impact assessment;

— any acceptable factors that can account for an inability to meet a target minimum performance level, especially when the AI system is evaluated for impacts on individuals or groups of individuals, or both, and societies (e.g. poor image resolution for computer vision systems or background noise affecting speech recognition systems). Mechanisms to deal with poor AI system performance as a result of these factors should also be documented.

The AI system should be evaluated against the documented criteria for evaluation.

Where the AI system cannot meet the documented criteria for evaluation, especially against responsible AI system development and use objectives (see B.6.1.2 and B.9.3), the organization should reconsider or manage the deficiencies of the intended use of the AI system, its performance requirements and how the organization can effectively address the impacts to individuals or groups of individuals, or both, and societies.

NOTE Further information on how to deal with robustness of neural networks can be found in ISO/IEC TR 24029-1.

B.6.2.5 AI system deployment

Control

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

Implementation guidance

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

B.6.2.6 AI system operation and monitoring

Control

The organization should define and document the necessary elements for the ongoing operation of the AI system. At the minimum this should include system and performance monitoring, repairs, updates and support.

Implementation guidance

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

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

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

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

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

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

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

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

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

Other information

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

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

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

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

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

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

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

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

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

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

B.6.2.7 AI system technical documentation

Control

The organization should determine what AI system technical documentation is needed for each relevant category of interested parties, such as users, partners, supervisory authorities, and provide the technical documentation to them in the appropriate form.

Implementation guidance

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

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

— usage instructions;

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

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

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

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

— design and system architecture specification;

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

— information about the data used during system development;

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

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

— verification and validation records;

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

— impact assessment documentation as described in B.5.

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

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

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

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

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

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

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

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

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

B.6.2.8 AI system recording of event logs

Control

The organization should determine at which phases of the AI system life cycle, record keeping of event logs should be enabled, but at the minimum when the AI system is in use.

Implementation guidance

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

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

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

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

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

Other information

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

B.7 Data for AI systems

B.7.1 Objective

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

B.7.2 Data for development and enhancement of AI system

Control

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

Implementation guidance

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

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

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

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

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

— accuracy and integrity of the data.

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

B.7.3 Acquisition of data

Control

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

Implementation guidance

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

— categories of data needed for the AI system;

— quantity of data needed;

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

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

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

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

— data rights (e.g. PII, copyright);

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

— provenance of the data.

Other information

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

B.7.4 Quality of data for AI systems

Control

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

Implementation guidance

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

Other information

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

B.7.5 Data provenance

Control

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

Implementation guidance

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

B.7.6 Data preparation

Control

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

Implementation guidance

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

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

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

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

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

— normalization;

— scaling;

— labelling of the target variables;

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

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

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

B.8 Information for interested parties

B.8.1 Objective

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

B.8.2 System documentation and information for users

Control

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

Implementation guidance

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

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

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

— purpose of the system;

— that the user is interacting with an AI system;

— how to interact with the system;

— how and when to override the system;

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

— needs for human oversight;

— information about accuracy and performance;

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

— revisions to claims about the system's benefits;

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

— contact information;

— educational materials for system use.

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

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

B.8.3 External reporting

Control

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

Implementation guidance

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

B.8.4 Communication of incidents

Control

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

Implementation guidance

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

— types of incidents that must be communicated;

— the timeline for notification;

— whether and which authorities must be notified;

— the details required to be communicated.

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

Other information

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

B.8.5 Information for interested parties

Control

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

Implementation guidance

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

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

— risks related to the system;

— results of impact assessments;

— logs and other system records.

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

B.9 Use of AI systems

B.9.1 Objective

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

B.9.2 Processes for responsible use of AI systems

Control

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

Implementation guidance

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

— required approvals;

— cost (including for ongoing monitoring and maintenance);

— approved sourcing requirements;

— legal requirements applicable to the organization.

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

B.9.3 Objectives for responsible use of AI system

Control

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

Implementation guidance

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

— fairness;

— accountability;

— transparency;

— explainability;

— reliability;

— safety;

— robustness and redundancy;

— privacy and security;

— accessibility.

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

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

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

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

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

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

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

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

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

Other information

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

B.9.4 Intended use of the AI system

Control

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

Implementation guidance

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

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

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

B.10 Third-party and customer relationships

B.10.1 Objective

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

B.10.2 Allocating responsibilities

Control

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

Implementation guidance

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

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

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

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

B.10.3 Suppliers

Control

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

Implementation guidance

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

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

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

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

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

B.10.4 Customers

Control

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

Implementation guidance

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

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

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

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

Annex C
(informative)

Potential AI-related organizational objectives and risk sources

C.1 General

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

C.2 Objectives

C.2.1 Accountability

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

C.2.2 AI expertise

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

C.2.3 Availability and quality of training and test data

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

C.2.4 Environmental impact

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

C.2.5 Fairness

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

C.2.6 Maintainability

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

C.2.7 Privacy

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

C.2.8 Robustness

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

C.2.9 Safety

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

C.2.10 Security

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

C.2.11 Transparency and explainability

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

C.3 Risk sources

C.3.1 Complexity of environment

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

C.3.2 Lack of transparency and explainability

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

C.3.3 Level of automation

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

C.3.4 Risk sources related to machine learning

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

C.3.5 System hardware issues

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

C.3.6 System life cycle issues

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

C.3.7 Technology readiness

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

Annex D
(informative)

Use of the AI management system across domains or sectors

D.1 General

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

— health;

— defence;

— transport;

— finance;

— employment;

— energy.

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

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

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

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

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

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

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

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

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

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

Bibliography

[1] ISO 8000-2, *Data quality — Part 2: Vocabulary*

[2] ISO 9001, *Quality management systems — Requirements*

[3] ISO 9241-210, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*

[4] ISO 13485, *Medical devices — Quality management systems — Requirements for regulatory purposes*

[5] ISO 22000, *Food safety management systems — Requirements for any organization in the food chain*

[6] IEC 62304, *Medical device software — Software life cycle processes*

[7] ISO/IEC Guide 51, *Safety aspects — Guidelines for their inclusion in standards*

[8] ISO/IEC TS 4213, *Information technology — Artificial intelligence — Assessment of machine learning classification performance*

[9] ISO/IEC 5259 (all parts[2]), *Data quality for analytics and machine learning (ML)*

[10] ISO/IEC 5338, *Information technology — Artificial intelligence — AI system life cycle process*

[11] ISO/IEC 17065, *Conformity assessment — Requirements for bodies certifying products, processes and services*

[12] ISO/IEC 19944-1, *Cloud computing and distributed platforms — Dataflow, data categories and data use — Part 1: Fundamentals*

[13] ISO/IEC 23053, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

[14] ISO/IEC 23894, *Information technology — Artificial intelligence — Guidance on risk management*

[15] ISO/IEC TR 24027, *Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*

[16] ISO/IEC TR 24029-1, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*

[17] ISO/IEC TR 24368, *Information technology — Artificial intelligence — Overview of ethical and societal concerns*

[18] ISO/IEC 25024, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality*

[19] ISO/IEC 25059, *Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems*

[20] ISO/IEC 27000:2018, *Information technology — Security techniques — Information security management systems — Overview and vocabulary*

[21] ISO/IEC 27701, *Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines*

[22] ISO/IEC 27001, *Information security, cybersecurity and privacy protection — Information security management systems — Requirements*

[23] ISO/IEC 29100, *Information technology — Security techniques — Privacy framework*

[24] ISO 31000:2018, *Risk management — Guidelines*

[25] ISO 37002, *Whistleblowing management systems — Guidelines*

[26] ISO/IEC 38500:2015, *Information technology — Governance ofITfor the organization*

[27] ISO/IEC 38507, *Information technology — Governance ofIT — Governance implications of the use ofartificial intelligence by organizations*

[28] Lifecycle D.D.I. 3.3, 2020-04-15. Data Documentation Initiative (DDI) Alliance. [viewed on 2022-02-19]. Available at: https://ddialliance.org/Specification/DDI-Lifecycle/3.3/

[29] Risk Framework N.I.S.T.-A.I. 1.0, 2023-01-26. National Institute of Technology (NIST) [viewed on 2023-04-17] https://www.nist.gov/itl/ai-risk-management-framework

This page deliberately left blank