ICS 35.240
CCS L 70

DB11

北　京　市　地　方　标　准

DB11/T 2251—2024

信息安全 人工智能数据安全通用要求

Information security—General requirements for data security of artificial intelligence

2024 - 06 - 28 发布　　2024 - 10 - 01 实施

北京市市场监督管理局　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由北京市公安局提出并归口。

本文件由北京市公安局组织实施。

本文件起草单位：北京市公安局、北京市公安局人工智能安全研究中心、北京百度网讯科技有限公司、北京瑞莱智慧科技有限公司、中国科学院计算技术研究所、科大讯飞股份有限公司、公安部第三研究所、北京市标准化研究院。

本文件主要起草人：蔡瑜坤、曹奇、王崇鹏、张晓飞、孙毅、刘鸣、常峰博、吴尚、孔凡真、杨彬、李晓波、王东明、辛铮、王海棠、郭建岭、周巧霖、樊子风、韦云霞、张旭东、曹娟、唐胜、朱莉莉、孔凡胜、程鸣、孙文琦。

信息安全 人工智能数据安全通用要求

1 范围

本文件规定了人工智能数据安全的基本原则、管理要求和处理活动安全要求。

本文件适用于指导人工智能服务提供者开展人工智能数据安全管理及数据处理活动的安全。

2 规范性引用文件

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

GB/T 25069 信息安全技术 术语

GB/T 35273 信息安全技术 个人信息安全规范

GB/T 41479 信息安全技术 网络数据处理安全要求

GB/T 41867 信息技术 人工智能 术语

GB/T 43697 信息安全技术 数据分类分级规则

3 术语和定义

GB/T 25069、GB/T 35273、GB/T 41867和GB/T 43697界定的以及下列术语和定义适用于本文件。

3.1

数据安全 data security

通过采取必要措施，确保数据处于有效保护和合法利用的状态，以及具备保障持续安全状态的能力。

3.2

重要数据 key data

特定领域、特定群体、特定区域或达到一定精度和规模的数据，一旦被泄露或篡改、损毁，可能直接危害国家安全、经济运行、社会稳定、公共健康和安全。

3.3

数据脱敏 data masking

通过一系列数据处理方法对原始数据进行处理，以屏蔽敏感数据的一种数据保护方法。

3.4

敏感个人信息 sensitive personal information

一旦泄露或者非法使用，容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息，包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息，以及不满十四周岁未成年人的个人信息。

4 基本原则

人工智能数据安全在管理、数据处理活动等应遵循以下原则：

a) 保密性：采用技术和管理手段保证人工智能数据的保密性，并记录数据处理日志及确保日志的保密性；
b) 完整性：采用完整性校验的技术和管理手段，并对人工智能数据的完整性进行监测；
c) 隐私性：采用技术和流程以保护个人信息主体的个人信息，涉及个人信息的应符合 GB/T 35273 的要求；
d) 合规性：数据处理活动应符合法规、政策文件、标准规范相关要求；
e) 分类分级：对人工智能数据进行分类和分级，并对不同类别不同级别的数据建立相应的全流程数据安全保护措施。涉及敏感个人信息的，宜采取更强的保护措施；
f) 伦理安全：人工智能数据全生命周期保障伦理安全，促进公平、公正、和谐，避免偏见、歧视、隐私和信息泄露等问题。

5 管理要求

5.1 组织和人员管理

组织和人员管理应满足以下要求：

a) 宜建立人工智能数据安全管理团队及监督职能部门，明确数据安全岗位和用户角色职责；
b) 制定人员管理安全策略，明确不同岗位人员在数据生存周期各阶段数据服务和系统服务相关的工作范畴和安全管控措施；
c) 应确保人工智能数据安全人员符合相关岗位要求，与所有涉及人工智能数据岗位人员签订安全责任协议，并定期开展安全教育；
d) 建立第三方人员安全管理制度，对接触人工智能数据的人员进行审批和记录，定期开展操作安全审查；
e) 在重要岗位人员调离或终止劳动合同时，应核查其在该岗位期间数据操作行为。

5.2 数据安全策略

数据安全策略应满足以下要求：

a) 制定符合组织数据安全要求且能覆盖全部数据处理活动数据安全策略，明确数据安全方针、目标和原则，内容应包括目的、范围、岗位、责任、管理层承诺、内外部协调及合规性要求等；
b) 制定并实施与安全策略和规程相适应的人工智能数据安全实施细则，并分发至组织内数据安全部门、岗位和人员；
c) 定期审核和更新人工智能数据安全策略和规程，并随之审核和更新人工智能数据安全实施细则。

5.3 合规管理

合规管理应满足以下要求：

a) 明确组织所适用的外部合规要求并形成清单，定期跟进监管机构合规要求以更新该清单，同时将其进行分发宣贯；
b) 参考 GB/T 35273 制定组织统一的个人信息保护制度，建立符合国家法律法规和相关标准的个人信息保护能力；
c) 参考 GB/T 41479 建立组织统一的重要数据全生存周期保护的制度和管控措施；

d) 明确组织统一的数据跨境安全制度和管控措施，定期开展人工智能数据安全风险评估。

5.4 数据资产管理

数据资产管理应满足以下要求：

a) 在组织层面建立数据资产安全管理制度，定义数据资产相关的角色定位和职责，宜按照GB/T 43697对数据进行分类分级管理；
b) 明确数据资产登记机制，明确数据资产管理范围和属性；
c) 根据数据资产安全管理制度识别数据并编制数据资产清单/数据保护目录，包括数据类别、数据量、数据级别、数据资产责任部门和所处位置等内容；
d) 应建立数据资产使用制度及审批流程，并保留数据处理记录。

5.5 数据安全审计

数据安全审计应满足以下要求：

a) 在组织层面建立数据安全审计制度，定义数据安全审计的相关角色定位和职责；
b) 对人工智能数据生存周期的处理活动进行记录，确保数据可审计、可追溯。

5.6 数据供应链安全管理

数据供应链安全管理应满足以下要求：

a) 确保数据供应链上下游各方对数据的收集、交换、使用符合相关法律法规要求；
b) 制定数据供应链安全管理规范，明确数据供应链安全目标、原则和范围、数据供应商选择和管理等；
c) 要求数据提供者说明数据来源，并对数据提供者的身份进行审核，留存审核记录；
d) 签署合作协议明确数据供应链上下游各方责任和义务，包括数据供应链中数据的使用目的、供应方式、保密约定、有效期等内容，确保数据供应链相关数据服务真实可用。
e) 建立数据供应链目录和相关数据源数据字典，用于及时查看供应链的整体情况和事后追踪分析数据供应链合规情况；
f) 对数据供应链上下游的数据提供者和数据使用者的行为进行合规性审核和分析；
g) 定期对数据供应链上下游数据处理活动安全风险和数据安全管理能力进行评估。

5.7 数据安全事件应急管理

数据安全事件应急管理应满足以下要求：

a) 制定数据安全应急管理工作指南，定义数据安全事件类型、划分数据安全事件等级明确不同类别、不同级别事件的处置流程和方法；
b) 建立数据安全事件应急预案，定期开展应急演练活动。

6 数据处理活动安全要求

6.1 数据收集

6.1.1 数据获取

数据获取应满足以下要求：

a) 明确数据的收集目的、用途以及数据的获取源、范围、频度，确保数据收集和获取的合法性和正当性，对数据源进行评估；
b) 获取个人信息时，应征得个人信息主体的同意，并确保获取数据过程中数据不被泄露；
c) 对数据收集和获取环境中的设施和技术进行安全管控，确保数据的完整性、一致性和真实

性；

d) 对于所需公开标准数据集，应确保数据未遭受污染；
e) 记录数据收集和获取过程，包括但不限于收集活动的时间、用户、收集目的、收集方式等,对数据收集和获取操作过程可追溯；
f) 在委托数据收集时，受委托方应同样遵照要求执行；
g) 数据收集过程中应提高数据的多样性，对每种语言、类型的语料应来自多个数据源。

6.1.2 数据清洗和转换

数据清洗和转换应满足以下要求：

a) 建立清洗和转换前后数据间映射关系，确保数据清洗和转换过程可溯源；对于不存在映射关系的清洗规则，应记录数据清洗的方法及影响范围；
b) 在数据清洗和数据转换完成后对数据清洗和转换过程中的中间数据或临时数据进行删除；
c) 采取必要的技术手段和管理措施，在数据有还原需求时能有效地还原数据。

6.2 数据存储

6.2.1 数据备份

数据备份应满足以下要求：

a) 对人工智能应用开发中的训练数据、特征数据、模型数据等建立备份或副本；
b) 建立数据复制、备份与恢复操作规范，包括复制、备份和恢复的日志记录规范；
c) 对备份数据执行和数据源同样的安全管控措施，包括访问控制、加密管理、完整性校验等，对数据存储状态进行监控，主动发现违规行为并进行处理；
d) 建立数据复制、备份与恢复的定期检查和更新工作程序，包括数据副本更新频率、保存期限等，确保数据副本或备份数据的有效性。

6.2.2 数据留存

数据留存应满足以下要求：

a) 对人工智能应用开发中的训练数据、验证数据、模型数据等分别明确访问、禁止使用和删除的有效期；
b) 具备对过期存储数据延长有效期或彻底删除的能力；
c) 对于不同敏感类型的数据，分别选用针对敏感度差异的存储加密技术和工具；
d) 涉及敏感个人信息的，用于恢复识别的数据和去标识化后的数据应分开存储。

6.2.3 权限管理

权限管理应满足以下要求：

a) 对存储系统用户使用身份标识与鉴别、权限分配等安全控制措施；
b) 建立审计信息和使用日志的存储机制和管控措施；
c) 建立用户数据访问的追溯机制，对数据访问进行监控，主动识别异常访问并进行阻断和告警。

6.3 数据加工

6.3.1 数据加工环境

数据加工环境应满足以下要求：

a) 在数据加工系统的设计、开发和运维阶段制定相应的安全控制措施，确保数据不被泄

露；

b) 建立数据加工日志管理工具，记录用户在数据加工系统上的操作，建立数据加工前后数据间映射关系，确保数据加工过程可溯源；

c) 应定期对数据加工环境进行安全评估，发现潜在的安全风险，及时进行修复，并详细记录漏洞发现及处置情况。

6.3.2 数据脱敏

数据脱敏应满足以下要求：

a) 基于数据脱敏的安全审计要求，对数据脱敏过程中的操作进行记录，保留其原始数据格式和特定属性；

b) 根据人工智能应用场景和数据属性，明确数据脱敏规则、脱敏方法、使用限制和处理流程。

6.3.3 数据标注

数据标注应满足以下要求：

a) 明确数据标注过程中的安全操作规范,建立与数据级别匹配的安全标注环境；

b) 建立健全标注规则机制，标注规则应至少覆盖数据标注以及数据审核等环节；

c) 采取机器验证或人工检测的方式对数据标注的结果进行质量检查和控制，必要时建立仲裁机制对标注结果进行最终核验，形成核验报告，核验报告包括核验时间、核验方式、核验内容等；

d) 开展标注人员培训及考核，对标注过程中的行为进行规范和检测，暂停或取消不合格者的标注上岗资格。

6.4 数据使用

6.4.1 数据使用环境

数据使用环境应满足以下要求：

a) 在服务器和移动端的处理器中为数据和代码执行提供安全区域，建立与数据级别匹配的安全处理环境；

b) 在合作开发情况下，宜采用联邦学习等技术构建联合开发环境，避免数据的交换和泄露；

c) 采用访问控制模块实施数据使用用户身份标识与鉴别策略、数据访问权限分配策略等，并采取安全控制措施。

6.4.2 数据检测

数据检测应满足以下要求：

a) 通过人工核查或自动化检测技术对投毒数据、后门数据等恶意数据做筛查和剔除，确保参与模型训练的数据集没有受到人为污染；

b) 依据模型训练的任务要求，明确数据质量评估方法，对数据进行质量评估；

c) 对存在特征值缺失、规模不达标、分布不均衡、偏见等质量问题的数据，采取必要的技术手段进行调整或删除。

6.4.3 数据利用

数据利用应满足以下要求：

a) 确保人工智能模型训练、推理等数据利用的目的、范围符合国家相关法律法规要求和数据收集目的以及使用范围；

b) 对于通过人工智能模型训练、推理等数据利用获得的中间数据和结果数据，保证对其中的个人信息、重要数据合法使用；
c) 记录并管理完整的人工智能模型训练、推理等数据利用操作，以备违约责任的识别和追责；
d) 建立数据使用和数据访问的安全审计机制和管控措施；
e) 涉及数据展示的，应对展示的必要性和安全性进行评估，采取数据脱敏、数据水印等技术手段对个人信息、重要数据等进行保护，不应展示违法不良信息。

6.5 数据传输

6.5.1 传输安全

传输安全应满足以下要求：
a) 区分不同数据形态和安全域内、安全域间等不同数据传输场景，建立相应的数据传输安全策略和操作规程；
b) 建立数据传输安全策略相应的安全控制措施，以保证数据在传输过程中的保密性；
c) 依照数据传输场景，对数据传输相关方进行身份鉴别，并对传输数据加密传输；
d) 采用基于密码技术的完整性和机密性保护机制，保护数据传输过程中的完整性和机密性，防止数据篡改和泄露。

6.5.2 传输监控

传输监控应满足以下要求：
a) 使用监控技术与机制对数据传输接口进行审核及监控，发现数据传输失败/错误等安全事件，并自动应急处置；
b) 如发现数据泄露事件，及时告警并应急处置。

6.6 数据提供

6.6.1 提供安全

提供安全应满足以下要求：
a) 向他人提供数据前，应按照 GB/T 41479 安全要求进行评估；
b) 所提供数据需经过提供者授权，数据提供的与方应保证数据安全防护能力处于同等级别；
c) 采用数据加密、安全通道等措施保护数据提供过程中的个人信息、重要数据等；
d) 数据提供完成后应清除数据提供通道的缓存数据并确保数据不能被恢复；
e) 采用自动和人工审计相结合的方法或手段对数据提供操作进行监控，记录操作事件，并制定数据提供风险行为识别和评估规则；
f) 在跨组织数据交换共享活动中部署必要的数据防泄漏实时监控工具，采用技术手段确保数据提供的可追溯性；
g) 涉及数据跨境提供的，应按照国家相关要求开展安全评估，并符合参与方所在地的数据安全要求。

6.6.2 第三方使用

数据处理中涉及第三方的，还应满足以下：
a) 人工智能数据处理者应与第三方签订合同、协议等有效约定，明确双方权责及相关安全义务；

b) 涉及个人信息的，应参考 GB/T 35273 中关于第三方的相关要求；
c) 宜要求第三方提供透明和可解释的数据处理方式，以便理解和验证第三方处理的过程和结果；
d) 宜要求第三方建立与提供方相应的安全能力，建立应急预案并定期演练。

6.7 数据删除

6.7.1 数据删除

数据删除应满足以下要求：
a) 建立数据删除策略和操作规范，明确删除数据对象；
b) 数据删除前应按策略和规范完成审批流程，并留存审批及删除日志；
c) 在数据主体主张删除时，在约定时间内对含有数据主体个人信息的数据进行删除，当数据在人工智能模型中已用于模型训练且无法删除时，应采用模型输出结果屏蔽等技术手段降低对个人信息主体的影响。

6.7.2 介质销毁

人工智能数据处理者应满足以下要求：
a) 建立存储介质销毁处理策略、管理制度和机制，明确销毁介质对象和操作规范；
b) 制定对存储介质进行销毁的监管措施，确保对销毁介质登记、审批、交接等介质销毁过程；
c) 对载有重要数据的存储介质，应进行消磁或物理粉碎等不可恢复性销毁处理，并做好相应的销毁记录。