ICS 35.030
CCS L 80

团 体 标 准

T/ISC 0133—2026

互联网的关键领域 AI 训练数据可信质量要求

Trustworthy quality requirements for AI training data in key fields of the internet

（发布稿）

2026－09－08 发布　　2026－10－08 实施

中 国 互 联 网 协 会　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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国互联网协会归口。

本文件起草单位：珠海市人民医院、中国联通国际有限公司、南京南瑞信息通信科技有限公司、中国联通（香港）创新研究院、南京航空航天大学、武汉卓讯互动信息科技有限公司、中国信息通信研究院、北京邮电大学、北京航空航天大学、深圳鹿野人工智能有限公司、中科数测固源科技（安徽）有限公司、览易（武汉）智能数据服务有限公司、华兴中科标准技术（北京）有限公司。

本文件主要起草人：王涵、肖雨果、于向荣、刘书博、孙浩轩、朱世顺、魏兴慎、曹永健、张恺、欧京槐、张吉、刘亚卓、宋宗维、陈文弢、静静、马若龙、邵彦华、刘欣然、周鸣一、黎立、孙海波、董婧一、董坤、陈绍东、张雪娇、李华、任国静、丁月。

互联网的关键领域 AI 训练数据可信质量要求

1 范围

本文件规定了互联网的关键领域AI训练数据可信质量的基本要求，数据采集、预处理、标注、合成、存储与传输、使用与处置全生命周期各环节的可信质量要求，以及数据可信质量评估方法和管理保障要求。

本文件适用于互联网关键领域AI系统的训练数据提供者、处理者及使用者开展数据可信质量管理相关活动，也可用于评估机构进行数据可信质量评估。

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 35273—2020 信息安全技术 个人信息安全规范

GB/T 40685 信息技术服务 数据资产 管理要求

GB/T 45652 网络安全技术 生成式人工智能预训练和优化训练数据安全规范

GB/T 45674—2025 网络安全技术 生成式人工智能数据标注安全规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

关键领域 key fields

涉及国家安全、国民经济命脉、重要民生、重大公共利益等领域，如金融、能源、医疗、交通等。

3.2

AI 训练数据 AI training data

用于训练和优化人工智能模型的数据集，包括文本、图像、音频、视频等模态数据。

注：参考GB/T 45652中预训练数据和优化训练数据的定义。

3.3

数据可信质量 trustworthy quality of data

数据在采集、处理、标注、管理和使用全过程中体现出的可验证、可追溯、可控制和安全合规的综合属性，反映数据支撑人工智能模型训练、评估和应用的可信程度。数据可信质量包括但不限于数据来源可靠性、真实性、准确性、完整性、一致性、代表性、安全性及合规性等要求。

3.4

合成数据 synthetic data

通过人工智能模型或算法生成的数据，包括但不限于生成式模型产生的文本、图像、音频和视频等数据。对于采用合成数据作为训练数据的数据集，应对其质量进行专门管理。

3.5

数据标注 data annotation

通过人工操作或使用自动化技术机制，将特定信息（如标签、类别或属性）添加到数据样本的过程。

[来源：GB/T 45674—2025,3.3，有修改]

3.6

个人信息 personal information

以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息。

[来源：GB/T 35273—2020,3.1]

4 概述

互联网关键领域AI训练数据可信质量管理覆盖数据采集、预处理、标注、合成、存储、传输、使用和处置等全生命周期环节，以人工智能应用需求和关键领域风险控制要求为导向，围绕数据真实性、准确性、完整性、一致性、安全性、可追溯性等可信质量目标，建立覆盖管理对象、质量控制过程和保障机制的协同管理体系。通过质量要求定义、过程控制、可信验证、风险评估和持续改进，实现AI训练数据来源可信、过程可信、质量可控和应用可靠。框架如图1所示。

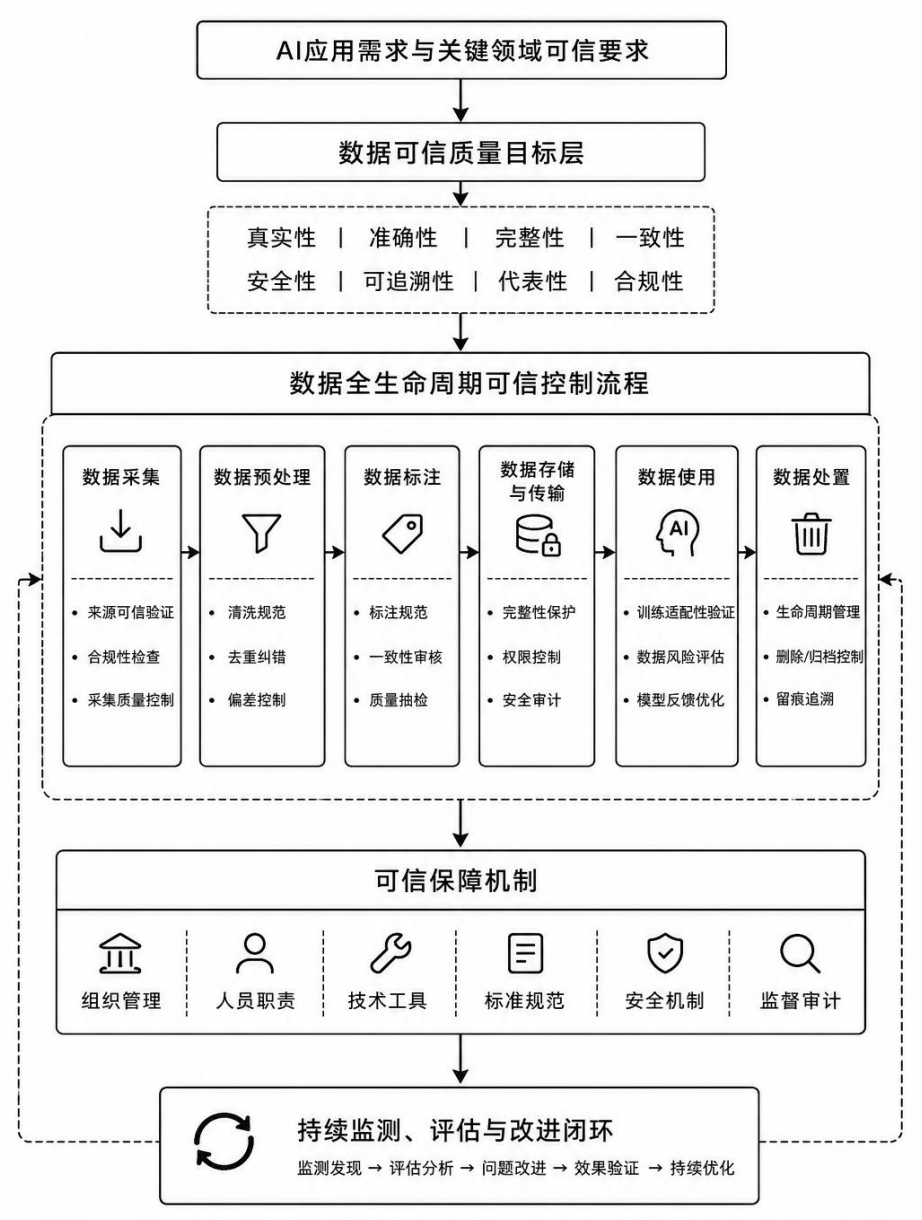

图 1　互联网关键领域 AI 训练数据可信质量控制框架

5 数据可信质量基本要求

5.1 基本原则

开展AI训练数据可信质量管理活动，应遵循以下原则：

a) 合法合规性：数据处理活动应符合国家法律法规和行业监管要求，并符合 GB/T 35273—2020 中关于个人信息保护的规定；
b) 目的明确：数据使用目的应明确、清晰、具体，不得超范围使用，参考 GB/T 35273—2020 中目的明确原则；
c) 质量可控：应建立数据质量评价机制，明确质量指标和阈值，参考 GB/T 40685 中数据质量评估要求；
d) 安全防护：应采取技术和管理双重措施，符合 GB/T 45652 中的通用安全要求。

5.2 数据分类分级

应按照GB/T 40685中数据资产信息要素的要求，对AI训练数据进行分类分级，并采取差异化保护措施。涉及个人敏感信息的，应符合GB/T 35273—2020中个人敏感信息的处理要求。

6 数据采集阶段可信质量要求

6.1 采集任务规划

6.1.1 数据采集任务规划应作为数据采集工作的前置环节，由算法需求方、业务专家、数据管理人员及相关领域专家共同参与，根据人工智能模型训练目标，将算法需求转化为明确的数据采集任务，并形成正式的数据采集方案。
6.1.2 采集任务规划阶段应首先开展互联网关键领域判定，根据人工智能模型应用场景识别数据所属领域，包括但不限于医疗健康、金融、教育、政务、交通、工业制造、公共安全、电子商务、网络内容等互联网关键领域，并结合相关法律法规、行业规范及安全要求，明确数据采集范围、采集边界及可信质量要求。
6.1.3 采集任务规划应结合互联网关键领域的数据特点，开展领域差异与共性分析。针对不同领域数据在数据类型、数据结构、采集方式、更新频率、标注规范、质量要求、安全等级及隐私保护等方面的差异，制定差异化的数据采集策略；同时识别跨领域数据在真实性、完整性、准确性、一致性、时效性、可追溯性及标准化等方面的共性要求，为多源数据融合及统一质量管理提供依据。
6.1.4 采集任务规划阶段应明确数据类型、数据格式、规格参数、规模、类别分布、时间跨度及更新频率等要求，并制定数据完整性、真实性、准确性、清晰度、一致性、时效性及可追溯性等质量标准。采集方案应对数据来源、采集方式、采集流程、质量控制措施、责任分工、采集环境及设备条件进行说明，并明确数据授权方式、来源合法性及可信性验证要求。
6.1.5 数据采集方案应在实施前完成评审。评审人员应包括业务专家、数据管理人员及安全管理人员；涉及互联网关键领域的重要数据、敏感数据或高风险应用场景时，宜增加法律合规、伦理及行业专家参与评审。评审内容应包括采集需求合理性、领域判定准确性、数据来源合法性、质量标准完整性、采集方案可行性及数据安全风险等。

6.2 数据来源合规性

6.2.1 数据采集活动应确保数据来源合法，并符合国家有关网络安全、数据安全及个人信息保护相关法律法规要求。对于有特定资质要求的领域，数据采集单位应具备相关资质要求。对于来自第三方机构或平台的数据，应对数据提供方资质进行审核，并确认数据获取方式合法合规。
6.2.2 涉及个人信息的数据采集，应明确数据采集目的，并依法取得数据主体授权或满足法律规定的数据处理条件。数据采集应遵循数据最小化原则，仅采集实现既定业务目标所必需的数据。
6.2.3 应建立数据来源登记机制，对数据来源机构、数据采集时间、采集方式及授权情况进行记录，以确保数据来源可追溯。

6.3 数据真实性验证

6.3.1 应建立数据真实性验证机制，确保采集数据真实反映客观对象或业务场景。
6.3.2 对于人工采集的数据，应通过抽样审核或复核机制对采集结果进行检查；对于设备采集或自动采集的数据，应确保采集设备处于正常校准状态，并对采集数据进行格式校验、异常检测及重复检测。
6.3.3 对于发现存在明显异常或疑似伪造的数据，应及时进行标记、核查或剔除。必要时，可通过多

源数据比对或交叉验证等方式提高数据真实性。

6.4 数据采集过程控制

6.4.1 数据采集活动应按照既定采集规范开展，并建立过程监控机制。
6.4.2 采集活动应记录采集时间、采集地点、采集设备及采集人员信息。对于自动化采集系统，应建立运行监控机制，对系统运行状态和数据采集情况进行监测。
6.4.3 在大规模数据采集任务开始前，宜开展小规模试采集，以验证采集流程和采集设备是否满足质量要求。试采集结果应进行评估，并根据评估结果对采集方案进行必要调整。

6.5 跨境数据采集合规性

6.5.1 涉及跨境数据采集、传输或使用的，应符合国家有关数据出境安全管理的法律法规及监管要求。
6.5.2 在开展跨境数据采集或使用前，应对数据出境的必要性、数据类型及潜在风险进行评估，并依法履行相应的安全评估、备案或其他合规程序。
6.5.3 对涉及个人信息或重要数据的跨境数据处理活动，首先应符合相关法律法规的规定，同时应采取数据脱敏或匿名化处理等技术措施，以降低数据安全风险。
6.5.4 应对跨境数据流动进行记录，记录内容宜包括数据来源、出境方式、接收方信息及数据使用范围等，以支持审计与监管要求。
6.5.5 对跨境数据的使用应限定在合法授权范围内，不得超出原定用途进行扩展使用。
6.5.6 涉及跨境数据活动的，数据采集任务规划应根据数据流向分别制定数据入境和数据出境管理要求。

a) 数据入境：境外来源数据应明确数据来源合法性、授权依据、知识产权归属及使用范围，对数据真实性、完整性、安全性及可信性进行验证，并符合国家有关数据入境管理要求。
b) 数据出境：境内数据向境外提供时，应按照国家有关数据出境管理要求开展合规评估，明确数据类别、传输范围、接收主体、安全保护措施及责任分工，防止重要数据、个人信息及敏感数据非法出境。
c) 跨境数据采集、传输、存储和使用全过程应建立可追溯机制，并保留相关记录，确保全过程可审计、可监管。

6.6 数据安全与隐私保护

6.6.1 数据采集过程中应采取必要的安全措施，防止数据泄露或非法使用。采集系统应具备访问控制和身份认证机制，并采用安全通信协议进行数据传输。
6.6.2 涉及敏感数据的采集活动，应采取数据脱敏或匿名化处理等技术措施，以降低数据安全风险。
6.6.3 涉及个人信息的数据采集，应严格遵守国家个人信息保护相关法律法规，并建立相应的数据保护和管理机制。

6.7 数据采集可追溯性

6.7.1 应建立完整的记录机制，对数据来源、采集时间、采集方式及采集人员等信息进行记录。
6.7.2 数据采集系统应保存数据采集日志，以支持数据审计与问题追溯。
6.7.3 宜建立数据血缘管理机制，使训练数据能够追溯至其原始采集来源和采集过程。

6.8 数据采集质量记录

6.8.1 数据采集活动应建立数据采集质量记录机制，对数据采集过程中的关键质量信息进行记录，以支持数据质量评估、审计和问题追溯。
6.8.2 数据采集质量记录宜包括数据来源信息、采集时间、采集地点、采集设备或系统信息、采集人员或采集系统标识，以及采集过程中的质量检查结果等内容。数据采集质量记录宜以日志或元数据形式保存，并与训练数据建立关联关系。
6.8.3 数据采集质量记录应按照相关数据管理和网络安全要求进行安全存储和管理，并在数据生命周期内保持可查询和可追溯。

6.9 数据版权与知识产权合规性

6.9.1 数据采集和使用活动应符合国家有关著作权、知识产权及数据合规管理相关法律法规要求，不得侵犯他人合法权益。
6.9.2 对来源于互联网公开数据的训练数据，应评估其使用是否符合合理使用原则或相关法律规定，并避免将未经授权的数据直接用于商业用途。
6.9.3 对包含代码的数据集，应遵循相关开源许可证要求，明确数据的使用范围、修改方式及再分发条件，不得违反开源协议约定。
6.9.4 对图像、音频、视频等多媒体数据，应确认其版权归属及授权情况。对于未明确授权的数据，不应直接用于模型训练或数据集构建。
6.9.5 在使用第三方数据或公开数据构建训练数据集时，应保留数据来源信息及授权依据，以支持后续审计与合规验证。
6.9.6 对基于已有数据生成的合成数据，应评估其是否存在对原始数据的直接复现或版权侵权风险，并采取必要措施降低相关风险。

7 数据预处理阶段可信质量要求

7.1 数据清洗与过滤

7.1.1 应对原始数据进行清洗，以识别和处理缺失数据、重复数据及明显错误数据。
7.1.2 数据清洗过程中应根据业务需求和模型训练需求，对无效数据、异常数据或与任务目标无关的数据进行过滤或标记。
7.1.3 数据清洗规则应形成规范文档，并在数据处理过程中保持一致执行。

7.2 数据标准化处理

7.2.1 应对不同来源的数据进行标准化处理，确保数据格式、编码方式及单位表示的一致性。
7.2.2 数据标准化处理应统一数据结构和字段定义，并建立数据字典或元数据说明，明确数据字段含义及取值范围。
7.2.3 在多源数据融合场景下，应对不同数据源之间的数据进行规范化转换。

7.3 安全预处理措施

7.3.1 应根据数据敏感程度采取必要的安全处理措施，以防止数据泄露或非法使用。
7.3.2 对涉及个人信息或敏感数据的数据集，应进行脱敏处理，包括匿名化、去标识化或数据扰动等方式。
7.3.3 不应生成可重新识别个人身份的信息组合，并应对处理后的数据进行安全风险评估。

7.4 数据质量校验

7.4.1 数据预处理完成后，应对数据质量进行校验，以确保数据满足模型训练和业务应用需求。
7.4.2 数据质量校验宜包括数据完整性检查、数据一致性检查及数据格式合规性检查。
7.4.3 对发现存在明显质量问题的数据，应进行修正、重新处理或剔除。

7.5 数据处理可追溯性

7.5.1 数据预处理活动应建立数据处理记录机制，对数据处理过程中的关键操作进行记录。
7.5.2 数据处理记录宜包括数据来源、处理时间、处理规则及处理人员或处理系统信息。
7.5.3 数据管理系统宜建立数据血缘管理机制，使数据能够追溯到原始数据及其处理过程。

7.6 数据偏差控制

7.6.1 在数据预处理过程中，应识别可能存在的数据偏差，以降低模型训练中的系统性偏差风险。
7.6.2 应分析数据在不同类别、地域、人群或时间维度上的分布情况，避免样本分布严重失衡。
7.6.3 对存在明显偏差的数据集，可通过样本补充、数据重采样或加权处理等方式进行调整。

7.7 数据预处理质量记录

7.7.1 数据预处理活动应建立质量记录机制，对数据处理过程中的关键质量信息进行记录。

7.7.2　数据预处理质量记录宜包括数据处理时间、处理规则、处理结果及质量检查情况等内容。
7.7.3　数据预处理质量记录应按照相关数据管理和网络安全要求进行安全存储，并在数据生命周期内保持可查询和可追溯。

8　数据标注阶段可信质量要求

8.1　标注任务规划与规范制定

8.1.1　应与算法需求方及业务专家协作，将模型训练需求转化为明确、无歧义的标注任务定义。任务定义应涵盖标注对象形态、任务类型、标签体系层级及标签语义定义。
8.1.2　应制定数据标注规范作为执行基准，明确正反例示例、边界情况处理规则、格式精度要求及版本记录机制。规范应经多方评审确认，并建立动态更新机制，确保规则修订及时同步。
8.1.3　在规划阶段应评估数据分布的地域、人群等维度偏差，制定相应的采样加权或补充采集策略，以降低系统性偏差风险。

8.2　人员资质与工具平台要求

8.2.1　应根据任务复杂度筛选具备相应背景知识的标注人员，对关键领域任务宜优先选择具备行业专业知识的人员。所有人员上岗前应完成规范培训与考核，并保留记录。
8.2.2　应选择功能完善、稳定可靠且符合网络安全要求的数据标注平台。平台应具备权限管理、任务分配、进度跟踪及安全审计功能。涉及敏感数据时，应优先支持私有化部署或境内可信云服务，并具备操作留痕、防泄露等安全能力。
8.2.3　工具应支持高效标注操作，并集成自动化逻辑校验规则，以提升标注效率与基础质量。

8.3　过程监控与安全合规

8.3.1　在大规模标注前，应通过小范围试标注校准标注人员理解一致性，并根据结果迭代优化标注规范。正式标注过程中，宜植入金标准数据进行实时质量监测，对连续不合格人员实施暂停或再培训措施。
8.3.2　应建立动态答疑机制与知识库，及时统一全员对疑难案例的理解，确保标注标准执行的一致性。
8.3.3　标注活动应在受控网络环境中进行，对高度敏感数据应采取物理隔离或虚拟桌面等技术措施。应严格遵守数据安全法律法规，落实数据分类分级与脱敏处理，实施最小权限访问控制，严禁数据违规下载或外传。

8.4　质量校验与评审机制

8.4.1　应设计多层次的质检流程，包括标注员自检、交叉评审、质检员抽检及专家终审等环节。
8.4.2　对于大规模数据集，应采用随机抽样检查策略，并根据初期质量表现动态调整抽检比例；对于高精度要求或小样本关键任务，宜进行全量评审。
8.4.3　应利用自动化工具对标注结果进行逻辑性、格式合规性及一致性校验，自动识别并拦截明显错误。
8.4.4　宜采用多人独立标注同一批数据的方式计算标注者间一致性，对不一致结果由专家复核仲裁。标注一致性应达到预设质量阈值，未达标时应分析原因并采取改进措施。

8.5　交付验收与闭环反馈

8.5.1　交付物应包含最终版标注数据、数据标注规范、量化质量报告、数据描述文件及合规文档。验收方应依据既定质量标准进行独立抽检，验证质量报告的真实性。
8.5.2　应建立“质检-反馈-修正-复核”的闭环流程，对质检发现的错误进行标准化归因分析，并及时反馈指导修正。
8.5.3　数据传输应采用加密通道或离线加密介质，确保传输安全。验收合格后应签署确认单，不合格项应启动返工或熔断机制。

8.6　后期维护与可追溯性

8.6.1　已交付数据集应实行严格的版本控制，任何修正或增补均作为新版本发布并附带变更日志。
8.6.2　应建立数据血缘管理机制，确保每个标注结果可追溯至原始数据、标注人员、评审记录及规范

版本，以实现责任界定和问题排查。
8.6.3 应建立用户反馈渠道，对模型应用中发现的标注错误进行核实修正并整合至后续版本。
8.6.4 应定期审视存量数据标签体系的适用性，应对概念漂移风险，必要时启动历史数据重新标注。
8.6.5 超出保留期限或依约需销毁的数据，应采用不可恢复方式彻底销毁，并出具销毁报告以备审计。

9 数据合成阶段可信质量要求

9.1 基本要求

9.1.1 合成数据应具备与真实数据一致的语义合理性和结构一致性，不应包含明显错误、逻辑冲突或不符合业务场景的内容。
9.1.2 合成数据应经过真实性验证或质量评估，必要时应通过人工审核或与真实数据对比的方式进行校验，以确保其能够支持模型训练目标。

9.2 偏差与公平性控制

9.2.1 在生成和使用合成数据时，应识别和评估数据偏差风险，包括但不限于类别分布偏差、语义偏差及模型放大偏差等问题。
9.2.2 对发现存在明显偏差的合成数据，应通过调整生成策略、引入真实数据或重新采样等方式进行修正。
9.2.3 在关键领域应用中，应对合成数据的公平性进行评估，避免因数据偏差导致模型产生歧视性或不合理决策。

9.3 版权与合规性要求

9.3.1 合成数据的生成过程应符合相关法律法规要求，不得侵犯他人知识产权或使用未经授权的数据进行训练生成。
9.3.2 对来源于真实数据的生成模型，应评估其输出内容是否存在对原始数据的直接复现或泄露风险。
9.3.3 在对外提供、共享、交易、公开发布或以其他方式向第三方提供合成数据时，应明确数据来源、生成方式及使用限制，并按照国家有关生成合成内容标识及网络数据安全管理的规定，对合成数据进行相应标识和合规管理。

9.4 数据标识与可追溯性

9.4.1 合成数据应进行标识或标注，与真实数据进行区分。
9.4.2 数据管理系统应记录合成数据的生成模型、生成时间及生成参数等信息。
9.4.3 应建立合成数据的可追溯机制，使其能够追溯至生成模型及生成过程。

9.5 使用控制

9.5.1 在模型训练中使用合成数据时，应控制其使用比例，避免对模型训练产生不合理影响。
9.5.2 合成数据宜与真实数据结合使用，并通过实验验证其对模型性能的影响。
9.5.3 对于关键领域应用，应对使用合成数据训练的模型进行额外评估，以验证其可靠性和安全性。
9.5.4 应评估合成数据与模型推理阶段应用场景及输入数据分布（推理上下文）的匹配程度，并结合模型在实际推理中的输出结果对合成数据的使用效果进行验证。当推理结果表明合成数据与推理上下文不匹配或引入质量问题时，应调整合成数据的生成策略、使用比例或使用范围。

10 数据存储与传输阶段可信质量要求

10.1 数据存储环境要求

10.1.1 数据存储环境应满足人工智能训练数据保存需求，保障数据在存储周期内保持完整、可访问和可使用。
10.1.2 应根据互联网关键领域数据特点，建立适宜的数据存储策略，明确存储位置、存储方式、存储周期和管理责任。

10.1.3 数据存储过程中应避免因系统故障、介质损坏、格式变化等因素造成数据丢失、损坏或不可读取。

10.2 数据存储质量保持

10.2.1 数据存储过程中应保持训练数据内容、结构、格式、标注信息和元数据的一致性。
10.2.2 应建立数据版本管理机制，对数据更新、迁移、转换等变化过程进行记录，避免不同版本数据混用。
10.2.3 宜定期开展存储数据质量检查，识别数据缺失、损坏、重复、异常变化等问题。

10.3 数据传输质量控制

10.3.1 数据传输过程应保证训练数据从发送端到接收端的一致性和完整性，避免传输过程中发生数据丢失、错误或非预期改变。
10.3.2 应对数据传输前后进行质量验证，检查数据数量、格式、结构、标签及关键属性是否保持一致。
10.3.3 涉及跨系统、跨平台或跨主体的数据传输，应明确数据格式、接口要求和质量验收要求。

10.4 数据传输可信保障

10.4.1 数据传输过程应采取必要措施保障传输路径、传输对象和传输内容可信。
10.4.2 涉及互联网关键领域 AI 训练数据传输时，应根据数据重要程度采取身份验证、完整性校验等措施。
10.4.3 传输异常时，应及时识别异常原因，并采取重新传输、数据修复或质量复核等措施。

10.5 存储与传输过程记录

10.5.1 数据存储与传输过程应形成记录，包括存储位置、版本变化、传输时间、传输对象和质量检查结果。
10.5.2 相关记录应支持数据流转过程查询、质量问题定位和可信验证。

11 数据使用与处置阶段可信质量要求

11.1 数据使用适配性

11.1.1 数据使用前应确认训练数据与人工智能应用场景、模型训练目标和性能需求相匹配。
11.1.2 应检查数据范围、数据规模、类别分布、质量水平等是否满足模型开发、训练和验证要求。
11.1.3 涉及互联网关键领域 AI 应用的数据，应评估数据不足、偏差或质量缺陷可能对模型结果产生的影响。

11.2 数据使用过程控制

11.2.1 数据使用过程中应保持训练数据来源、版本、处理状态和质量状态明确。
11.2.2 模型训练、测试和验证过程中使用的数据集应进行区分管理，避免数据混用或数据泄漏影响模型评价结果。
11.2.3 涉及数据转换、增强、筛选等操作时，应确保处理过程不会破坏数据真实性和有效性。

11.3 数据质量反馈与更新

11.3.1 应结合模型训练效果、验证结果和应用反馈，对训练数据质量进行持续评估。
11.3.2 当发现数据偏差、不完整、不准确或不适用情况时，应及时开展数据修正、补充或更新。
11.3.3 数据更新后应重新进行必要的质量验证，确保更新后的数据满足使用要求。

11.4 数据使用过程追溯

11.4.1 数据使用过程应记录数据版本、使用场景、处理过程和质量变化情况。
11.4.2 当模型性能异常或产生风险结果时，应能够追溯相关训练数据及其处理过程。

11.5 数据退出与处置

11.5.1　当训练数据超过使用期限、不满足质量要求或不适用于当前应用场景时，应按照规定进行退出或处置。
11.5.2　数据退出前应评估对模型训练、验证及后续应用的影响。
11.5.3　数据归档、删除等处置过程应保留必要记录，支持后续质量审查。

12　数据可信质量评估方法

12.1　评估原则

12.1.1　数据可信质量评估应遵循客观性原则，评估过程应基于真实数据和明确的评估指标，避免主观判断对评估结果产生影响。
12.1.2　数据可信质量评估应遵循可重复性原则，评估方法和评估流程应形成规范文档，以确保不同评估人员能够获得一致或相近的评估结果。
12.1.3　数据可信质量评估应遵循全面性原则，应综合考虑数据在真实性、准确性、完整性、一致性和安全性等方面的表现。

12.2　评估指标

12.2.1　对标注数据集的评估，可采用标注一致性指标，例如标注者间一致性或专家复核一致性等指标。
12.2.2　在涉及敏感数据或关键领域数据的场景中，还应评估数据安全性和合规性，以确保数据处理活动符合相关法律法规要求。

12.3　评估方法

数据可信质量评估可采用自动化评估与人工评估相结合的方法，以提高评估效率和评估准确性。

a)　自动化评估可通过数据校验规则、数据统计分析或脚本检测等方式，对数据格式、数据分布及数据完整性进行检查。
b)　人工评估可通过专家审核、抽样检查或交叉评审等方式，对数据内容质量和语义准确性进行评价。

12.4　评估流程

12.4.1　数据可信质量评估应建立规范化评估流程，包括评估准备、指标计算、结果分析及评估报告形成等环节。
12.4.2　在评估过程中，应根据数据规模和数据类型确定合理的抽样比例，以保证评估结果具有代表性。
12.4.3　评估完成后，应形成数据质量评估报告，对评估方法、评估指标及评估结果进行说明。

12.5　抽样方法

12.5.1　数据可信质量评估应根据数据规模、数据类型、类别分布和应用场景选择适当的抽样方法。抽样方法应有助于保证评估样本的代表性和评估结果的可信性。
12.5.2　对分布较均匀、样本结构相对单一的数据集，可采用随机抽样方法。随机抽样应确保样本被抽取的机会基本一致，避免因主观选择导致评估偏差。
12.5.3　对类别分布不均衡、来源多样或包含多个业务场景的数据集，宜采用分层抽样方法。分层抽样应根据类别、来源、时间、区域、风险等级或其他关键属性进行分层，并分别抽取样本。
12.5.4　对于少数类样本、边界样本、异常样本或高风险样本占比较低的数据集，宜采用补充抽样或重点抽样方法，以提高评估对关键风险的识别能力。
12.5.5　当训练数据集规模较大时，抽样数量应与数据规模、评估目标和风险等级相匹配。对于关键领域高风险场景，宜适当提高抽样覆盖程度。
12.5.6　抽样过程应形成记录，记录内容宜包括抽样方法、抽样依据、样本规模、样本分布及执行时间等信息。

12.6　评估记录与持续改进

12.6.1　数据可信质量评估活动应建立评估记录机制，对评估过程和评估结果进行记录。
12.6.2　评估记录宜包括评估时间、评估人员、评估方法及评估结果等信息。

12.6.3 数据管理组织应根据评估结果持续改进数据管理流程。

12.7 多主体责任划分

12.7.1 在AI训练数据管理过程中，涉及数据提供方、数据处理方及模型训练方等多主体时，应明确各主体在数据质量管理中的职责分工。
12.7.2 数据提供方应对数据来源的合法性、数据真实性及数据完整性负责，并提供必要的数据来源说明和授权依据。
12.7.3 数据处理方应对数据预处理、标注及数据质量控制过程负责，确保数据处理过程符合相关规范要求，并对处理过程中引入的数据质量问题承担相应责任。
12.7.4 模型训练方应对训练数据的选择、使用及适用性评估负责，应对因数据使用不当或数据质量不满足要求而导致的模型性能问题承担相应责任。
12.7.5 在多主体协作场景下，各方应通过协议或合同形式明确数据质量责任、数据安全责任及风险承担机制。
12.7.6 当数据质量问题导致AI系统异常或风险事件时，应根据数据生命周期各阶段的责任划分进行追溯，并由相关责任主体承担相应责任。

13 管理保障要求

13.1 组织管理

13.1.1 组织应建立数据质量管理组织体系，明确数据管理责任部门及相关岗位职责。
13.1.2 组织应指定数据管理负责人或数据质量负责人，统筹协调数据采集、处理、存储、使用及处置等环节的管理工作。
13.1.3 组织应建立跨部门协作机制，促进业务部门、技术部门及数据管理部门之间的信息共享与协同管理。

13.2 制度建设

13.2.1 组织应建立数据管理制度，包括数据采集管理、数据处理管理、数据安全管理及数据质量管理等制度。
13.2.2 数据管理制度应明确数据生命周期各阶段的管理要求，并规定数据管理流程和责任分工。
13.2.3 数据管理制度应根据法律法规变化和业务发展情况进行定期评估和更新。

13.3 技术保障

13.3.1 组织应建立支持数据质量管理的信息技术平台，对数据采集、处理、存储及使用过程进行管理。
13.3.2 数据管理系统应具备数据质量监测、数据访问控制及日志记录等功能。
13.3.3 组织应采取必要的技术措施保障数据安全，包括身份认证、访问控制、数据加密及安全审计等。

13.4 人员管理与培训

13.4.1 组织应对参与数据管理和数据处理活动的人员进行管理，并明确其岗位职责。
13.4.2 组织应定期开展数据安全、数据质量及相关法律法规培训，提高相关人员的数据管理能力。
13.4.3 对涉及敏感数据处理的岗位人员，应加强安全管理，并签署保密协议。

13.5 监督与审计

13.5.1 组织应建立数据质量监督机制，对数据管理活动进行定期检查。
13.5.2 组织应开展数据质量审计，对数据质量管理制度执行情况进行评估。
13.5.3 对发现的数据质量问题或安全风险，应及时采取整改措施，并进行持续改进。

13.6 管理记录与持续改进

13.6.1 数据管理活动应建立记录机制，对数据管理过程中的关键活动进行记录。
13.6.2 管理记录宜包括数据质量检查记录、审计记录及问题整改记录等内容。
13.6.3 组织应根据数据质量评估和监督审计结果，持续优化数据管理流程，以提升数据可信质量水平。