《互联网的关键领域AI训练数据可信质量要求》

编制说明

2026 年 4 月

一、工作简况

（一）任务来源

本项目根据中国互联网协会2025年团体标准制定计划（标准计划号：142-T/ISC-25），项目名称为“《网络安全与关键领域AI训练数据可信质量要求》”的任务而进行制订。

（二）起草单位及主要起草人

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、中国信息通信研究院、武汉卓讯互动信息科技有限公司、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、中科数测科技有限公司、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、朱世顺、魏兴慎、曹永健、张恺、张吉、陈文弢、静静、马若龙、邵彦华、刘亚卓、宋宗维、刘欣然、周鸣一、黎立、孙海波、董坤、董婧一、成瑾、李华、任国静、丁月。

（三）标准制定目的和意义

AI技术是驱动数字经济创新发展的核心动力，网络安全领域AI应用的有效性直接依赖于高质量训练数据。训练数据的可信质量作为AI安全技术落地的基础前提，其保障能力关系到关键信息基础设施防护、数据安全等核心领域的风险防控水平。《新一代人工智能发展规划》等政策文件明确要求“提升AI训练数据质量，加强数据安全与可信管理”，为网络安全领域AI数据标准建设提供了重要遵循。

网络安全与关键领域AI训练数据可信质量，是指面向网络攻击检测、漏洞挖掘、态势感知等核心场景，确保训练数据具备真实性、完整性、保密性、可用性的质量特性集合；可信质量要求则是围绕数据采集、标注、清洗、存储等全生命周期，制定的技术规范与评估准则。关键领域AI训练数据管理体系是通过标准化流程与技术手段，将数据质量管控与AI模型开发、安全应用深度融合，实现网络安全AI系统的可信、可控运行。

随着AI在网络安全关键领域应用的加速渗透，攻击检测模型误报漏报、漏洞预测准确率不足等问题频发，其核心症结在于训练数据存在来源不可靠、标注精度低、隐私信息泄露等质量隐患。行业当前依赖分散化的数据治理方式，缺乏统一的可信质量评估框架，导致AI安全产品性能参差不齐，难以满足关键领域的高安全需求。亟需通过标准制定填补技术空白，构建全流程数据质量管控体系。本标准的制定可完善网络

安全 AI 技术标准体系，推动训练数据管理规范化发展，同时有助于提升 AI 安全产品的可靠性与安全性，为关键领域网络安全防护提供坚实支撑。

（四）主要工作过程

2025 年 12 月本团体标准由中国互联网协会正式立项，立项名称为：《互联网的关键领域 AI 训练数据可信质量要求》。

2025 年 12 月 5 日通过腾讯会议召开了本团体标准的第一次讨论会。本次会议确定了本标准的框架结构，收集了与会各单位对标准草案（初稿）的修改意见或建议，明确工作组各单位的工作分工，以及会后的具体工作安排。

2026 年 3 月 24 日通过腾讯会议线上召开了本团体标准的第二次讨论会。编制组就本团体标准第一次讨论后的修改情况、意见采纳情况进行了综合阐述。与会专家及代表对该项团体标准的定义内容、技术内容等进行了进一步的深入讨论和修改，编制组认真听取和记录了与会代表提出的问题和调整建议。

二、标准编制原则和依据

（一）编制原则

标准起草小组在编制标准过程中，以国家、行业现有的标准为制订基础，结合我国电子油泵产品的现状，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》的规定及相关要求编制。

（二）标准范围

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、合成数据、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，也可用于第三方评估机构进行可信数据质量评估。

（三）标准主要内容与确定依据

1.概述

本章节参考 GB/T 40685 中数据资产管理框架列出了本文件的基本框架。

2.数据可信质量基本要求

本章节对互联网的关键领域 AI 训练数据可信质量的基本做出了规定。包含基本原

则、数据分类分级等内容

依据：参考 GB/T 35273—2020、GB/T 40685、GB/T 45652 等相关标准的要求，结合行业需求和企业生产实际。

3.数据采集阶段可信质量要求

本章节对数据采集阶段可信质量要求的采集任务规划、数据来源合规性、数据真实性验证、数据采集过程控制、跨境数据采集合规性、数据安全与隐私保护、数据采集可追溯性、数据采集质量记录、数据版权与知识产权合规性等做出了规定。

依据：结合行业需求和企业生产实际。

4.数据预处理阶段可信质量要求

本章节对数据预处理阶段可信质量要求的数据清洗与过滤、数据标准化处理、安全预处理措施、数据质量校验、数据处理可追溯性、数据偏差控制、数据预处理质量记录做出了规定。

依据：结合行业需求和企业生产实际。

5.数据标注阶段可信质量要求

本章节对数据标注阶段可信质量要求的任务规划、任务实施、任务评审、交付验收、后期维护做出了规定。

依据：遵循《中华人民共和国网络安全法》、《中华人民共和国数据安全法》及《个人信息保护法》，结合行业需求和企业生产实际。

6.合成数据质量要求

本章节对合成数据质量要求的基本要求、偏差与公平性控制、版权与合规性要求、数据标识与可追溯性、使用控制做出了规定。

依据：结合行业需求和企业生产实际。

7.数据存储与传输可信质量要求

本章节对数据存储与传输可信质量要求的数据存储安全、数据访问控制、数据传输安全、数据完整性保护、数据可追溯性、数据存储与传输质量记录、做出了规定。

依据：结合行业需求和企业生产实际。

8.数据使用与处置可信质量要求

本章节对数据使用与处置可信质量要求的数据使用管理、数据共享与对外提供、数据处置管理、数据销毁、数据使用与处置可追溯性、数据使用与处置质量记录做出了规定。

依据：结合行业需求和企业生产实际。

9.数据可信质量评估方法

本章节对数据可信质量评估方法的评估原则、评估指标、评估方法、评估流程、抽样方法、第三方评估、评估记录与持续改进、多主体责任划分做出了规定。

依据：结合行业需求和企业生产实际。

10.管理保障要求

本章节对管理保障要求的组织管理、制度建设、技术保障、人员管理与培训、监督与审计、管理记录与持续改进做出了规定。

依据：结合行业需求和企业生产实际。

三、国内外情况简要说明

经查，暂无相同类型的国际标准与国外标准，故没有相应的国际标准、国外标准可采用。本标准达到国内先进水平。

四、与有关的现行法律、法规和强制性国家标准的关系

本标准的制定过程、技术要求的选定、试验方法的确定、检验项目设置等符合现行法律法规和强制性国家标准的规定。

五、重大分歧意见的处理经过和依据

本文件在制定过程中未出现重大分歧意见。

六、废止现行有关标准的建议

本标准不涉及对现行标准的废止。

七、其他应予说明的事项

无。

团体标准编制组

2026 年 4 月