

# 国家标准《高质量数据集 质量评测规范》 (征求意见稿) 编制说明

## 一、工作简况

### (一) 任务来源

2025 年 12 月 31 日，根据《国家标准委关于下达 2025 年第十二批推荐性国家标准计划及相关标准外文版计划的通知》（国标委发〔2025〕76 号），国家标准《高质量数据集 质量评测规范》制定计划下达，计划号为 20256913-T-907。该标准由全国数据标准化技术委员会提出并归口，主管部门为国家数据局。

该标准的起草单位为中国电子技术标准化研究院、中国电子信息产业发展研究院、中国科学院计算技术研究所、国家数据发展研究院、国务院国有资产监督管理委员会研究中心、中电数据产业集团有限公司、交通运输部公路科学研究所、北京大学、公安部第三研究所、中国移动通信集团有限公司、中国石油天然气集团有限公司、中国石油化工集团有限公司、中国交通建设集团有限公司、国家能源投资集团有限责任公司信息技术分公司、国家电网有限公司大数据中心、中国南方电网有限责任公司、国家石油天然气管网集团有限公司、中国稀土集团有限公司、华为技术有限公司、科大讯飞股份有限公司、阿里巴巴（中国）有限公司、北京智源人工智能研究院、北京百度网讯科技有限公司、深圳市腾讯计算机系统有限公司、中国联合网络通信集团有限公司、中国电信集团有限公司、工业和信息化部电子第五研究所、

中国信息通信研究院、商业信用中心、石化盈科信息技术有限责任公司、北京智网数科技术有限公司、浦江国家实验室、中国质量认证中心有限公司、煤炭科学研究总院有限公司、中国交通信息科技集团有限公司、中移动信息技术有限公司、国家电投集团数字科技有限公司、中电云计算技术有限公司、中石油（北京）数智研究院有限公司、联通数据智能有限公司、上海库帕思科技有限公司、上海信投智能科技股份有限公司、北京清竞数智科技有限公司、青岛国创智能家电研究院有限公司、成都高新愿景数字科技有限公司、每日互动股份有限公司、航天科工网络信息发展有限公司、中国邮政储蓄银行股份有限公司、中国电子工程设计院股份有限公司、中电金信软件有限公司、江苏省大数据管理中心、内蒙古自治区大数据中心、江西省大数据中心、四川省卫生健康信息中心（四川省健康医疗大数据中心）、北京大学（天津滨海）新一代信息技术研究院、国家开放大学、杭州数美科技有限公司、中天钢铁集团（南通）有限公司、南京南瑞继保工程技术有限公司、南京南瑞瑞中数据股份有限公司、中通服网盈科技有限公司、北京海天瑞声科技股份有限公司、广州数字健康科技有限公司、安徽飞数信息科技有限公司、湖北大数据集团数据开发有限公司、杭州市临安区大数据管理服务中心、软通智慧科技有限公司、同方知网数字科技有限公司、数据堂（北京）科技股份有限公司、睿尔曼智能科技（北京）有限公司、北京银河通用机器人股份有限公司、烽火通信科技股份有限公司、中兴通讯股份有限公司、浪潮电子信息产业股份有限公司、国网山东省电

力公司、蔚来汽车科技（安徽）有限公司、贵州大数据产业集团有限公司、杭州市数据集团有限公司、江苏省数据集团有限公司、四川数据集团有限公司、厦门赛西科技发展有限责任公司、云基华海信息技术股份有限公司、国网江苏省电力有限公司、广东省人民医院、辽宁省电子信息产品监督检验院、数字宁波科技有限公司、杭州景联文科技有限公司、北京腾云天下科技有限公司、北京星河智源科技有限公司、山西集智数据服务有限公司、山东未来集团有限公司、广州维视达数字科技有限公司、厦门身份宝网络科技有限公司、上海森栩医学科技有限公司、北京八月瓜科技有限公司、北京中数睿智科技有限公司、江苏中堃数据技术有限公司。

该标准的主要起草人为范科峰、张群、韩冰、郭嘉丰、赵鹏飞、王为中、廖华明、张欢、李成博、李冰、苏越阳、李天舒、时晓光、黄吉海、丁浩、李嘉宁、王超、温晓君、潘宗俊、陈英昊、王亚沙、赵俊峰、邓成龙、汪睿棋、初旭、吴坤、陈文萱、申端明、蒋楠、王奥、杨丽、王宇静、陈振宇、徐欢、宋一纯、李步伟、于海松、储培、林志强、李华杰、赵丽丽、王鑫、方昕、吴峥、李世奇、刘颖、刘广、杨二龙、邱泳钦、刘煜宏、董梁、杨瑞、万芊芊、程广明、樊威、李荪、袁星煜、赵岩、赵艺璇、王锋、程健、骆意、张建中、武光城、金柳、莫洋、梁小涛、王俊杰、薛健、冷家冰、魏如蓝、王昊、谭晓坤、胡力旗、王兴旺、申中一、桂志辉、陈亮、方毅、邵元勋、孙永泉、孟紫、王春涛、杜啸争、吴善鹏、崔连伟、张静、沈明辉、邓韧、李方平、谢魁、

毛子卿、蔡斯博、余其竞、念灿华、解凯、张锦辉、李凡、蒲逸凡、黄宇恒、葛海龙、胡环环、段先明、郑辉、林镇阳、薛德军、齐红威、郑随兵、曾辉、陈刚、吴德亮、陈曦、张闻彬、商泽坤、赵泓青昀、李玮燕、陈康、陈曦、曾良滨、鲁胜强、何金陵、马洲俊、梁会营、王俊吉、毛欢欢、刘云涛、张亚东、严长春、庞俊奇、刘杰、彭荣、陈颖、温冬梅、李长青、韩涵、魏清。

起草单位、起草人及各自完成的工作如下：

范科峰、张群（中国电子技术标准化研究院）、韩冰（中国电子信息产业发展研究院）、郭嘉丰（中国科学院计算技术研究所）、赵鹏飞、王为中（中国电子技术标准化研究院）牵头制定高质量数据集质量评测技术框架，统筹标准主要章节内容、协调处理意见分歧等，负责各阶段的整体进度控制及内容审核。

廖华明（中国科学院计算技术研究所）、张欢、李成博、李冰（中国电子技术标准化研究院）、苏越阳（中国科学院计算技术研究所）牵头编制指标要求章节等相关内容。

李天舒、时晓光（国家数据发展研究院）、黄吉海（国务院国有资产监督管理委员会研究中心）、丁浩、李嘉宁（中电数据产业集团有限公司）牵头编制评测方法章节中评测方式要求等相关内容。

王超、温晓君（中国电子信息产业发展研究院）、潘宗俊、陈英昊（交通运输部公路科学研究所）、王亚沙、赵俊峰（北京大学）牵头编制评测方法章节中说明文档指标评测等相关内容。

邓成龙、汪睿棋（中国电子技术标准化研究院）牵头编写标

准范围、规范性引用文件、术语定义相关章节内容。

初旭（北京大学）、吴坤（中国移动通信集团有限公司）、陈文萱（公安部第三研究所）、申端明（中国石油天然气集团有限公司）、蒋楠（中国石油化工集团有限公司）、王奥（石化盈科信息技术有限责任公司）、杨丽（中国交通建设集团有限公司）、王宇静（国家能源投资集团有限责任公司信息技术分公司）、陈振宇（国家电网有限公司大数据中心）、徐欢（中国南方电网有限责任公司）、宋一纯（国家石油天然气管网集团有限公司）牵头编制评测方法章节中数据质量指标评测等相关内容。

李步伟、于海松（北京智网数科技术有限公司）、储培（浦江国家实验室）、林志强、李华杰（中国稀土集团有限公司）、赵丽丽、王鑫（华为技术有限公司）、方昕、吴峥（科大讯飞股份有限公司）、李世奇（阿里巴巴（中国）有限公司）、刘颖、刘广（北京智源人工智能研究院）、杨二龙、邱泳钦（北京百度网讯科技有限公司）、刘煜宏、董梁（深圳市腾讯计算机系统有限公司）牵头编制评测方法章节中模型应用指标评测等相关内容。

杨瑞（中国联合网络通信集团有限公司）、万芊芊（中国电信集团有限公司）、程广明（工业和信息化部电子第五研究所）、樊威、李荪（中国信息通信研究院）、袁星煜（商业信用中心）参与编制评测方法章节中评测方式要求等相关内容。

赵岩、赵艺璇（石化盈科信息技术有限责任公司）、王锋（中国质量认证中心有限公司）、程健、骆意、张建中、武光城（煤

炭科学研究总院有限公司)、金柳、莫洋(中国交通信息科技集团有限公司)、梁小涛(中国移动通信集团有限公司)、王俊杰(中移动信息技术有限公司)参与编制评测方法章节中数据质量指标评测等相关内容。

薛健(国家电投集团数字科技有限公司)、冷家冰(中电云计算技术有限公司)、魏如蓝(中石油(北京)数智研究院有限公司)、王昊(联通数据智能有限公司)、谭晓坤(上海库帕思科技有限公司)、胡力旗(上海信投智能科技股份有限公司)、王吴越(北京清竞数智科技有限公司)参与编制指标要求章节等相关内容。

王兴旺、申中一(中国电子技术标准化研究院)牵头编写术语定义相关章节内容。

桂志辉(青岛国创智能家电研究院有限公司)、陈亮(成都高新愿景数字科技有限公司)、方毅(每日互动股份有限公司)、邵元勋、孙永泉(航天科工网络信息发展有限公司)、孟紫(中国邮政储蓄银行股份有限公司)参与编制评测方法章节中说明文档指标评测等相关内容。

王春涛(中国电子工程设计院股份有限公司)、杜啸争(中电金信软件有限公司)、吴善鹏(江苏省大数据管理中心)、崔连伟(内蒙古自治区大数据中心)、张静(江西省大数据中心)、沈明辉、邓韧(四川省卫生健康信息中心(四川省健康医疗大数据中心))、李方平、谢魁、毛子卿(北京大学(天津滨海)新一代信息技术研究院)、蔡斯博、余其竞(国家开放大学)参与

编制评测方法章节中数据质量指标评测等相关内容。

念灿华（杭州数美科技有限公司）、符立峰（中天钢铁集团（南通）有限公司）、解凯（南京南瑞继保工程技术有限公司）、张锦辉（南京南瑞瑞中数据股份有限公司）、李凡（中通服网盈科技有限公司）、蒲逸凡（北京海天瑞声科技股份有限公司）、黄宇恒、葛海龙（广州数字健康科技有限公司）、胡环环（安徽飞数信息科技有限公司）、段先明（湖北大数据集团数据开发有限公司）、郑辉（杭州市临安区大数据管理服务中心）、林镇阳（软通智慧科技有限公司）、薛德军（同方知网数字科技有限公司）、齐红威（数据堂（北京）科技股份有限公司）、郑随兵（睿尔曼智能科技（北京）有限公司）、曾辉（北京银河通用机器人股份有限公司）参与编制评测方法章节中模型应用指标评测等相关内容。

陈刚（烽火通信科技股份有限公司）、吴德亮（中兴通讯股份有限公司）、陈曦（浪潮电子信息产业股份有限公司）、张闻彬（国网山东省电力公司）、商泽坤（蔚来汽车科技（安徽）有限公司）、赵泓青昀（贵州大数据产业集团有限公司）、李玮燕（杭州市数据集团有限公司）、陈康（江苏省数据集团有限公司）、陈曦（四川数据集团有限公司）、曾良滨（厦门赛西科技发展有限公司）、鲁胜强（云基华海信息技术股份有限公司）、何金陵、马洲俊（国网江苏省电力有限公司）、梁会营（广东省人民医院）、王俊吉（辽宁省电子信息产品监督检验院）、毛欢欢（数字宁波科技有限公司）、刘云涛（杭州景联文科技有限公司）、

张亚东（北京腾云天下科技有限公司）、严长春（北京星河智源科技有限公司）、庞俊奇（山西集智数据服务有限公司）、刘杰（山东未来集团有限公司）、彭荣（广州维视达数字科技有限公司）、陈颖（厦门身份宝网络科技有限公司）、温冬梅（上海森栩医学科技有限公司）、李长青（北京八月瓜科技有限公司）、韩涵（北京中数睿智科技有限公司）负责标准的试点验证、提供标准修改意见等工作。

魏清（江苏中堃数据技术有限公司）参与标准内容的调研、研讨等工作。

## （二）制定背景及意义

人工智能作为引领新一轮科技革命和产业变革的战略性技术，深刻改变人类生产生活方式。随着人工智能技术快速发展，研发重点正从“重点优化模型架构”转向“模型与数据协同优化”，其中，高质量数据的作用日益凸显。数据作为人工智能发展的三大核心要素之一，已成为人工智能模型开发和训练的核心要素资源。充分发挥标准的支撑和引领作用，加快高质量数据集规范化建设，对于推动人工智能赋能行业发展具有重要意义。

制定该标准的必要性、重要性等主要体现在以下方面。

**一是落实国家政策要求。**国家高度重视高质量数据集建设工作，先后出台《国家数据局等部门关于印发〈“数据要素×”三年行动计划（2024—2026 年）〉的通知》（国数政策〔2023〕11 号）、《国家数据局等部门关于促进企业数据资源开发利用的意见》（国数资源〔2024〕125 号）、《国家发展改革委等部



门关于促进数据标注产业高质量发展的实施意见》（发改数据〔2024〕1822号）、《国家发展改革委等部门关于促进数据产业高质量发展的指导意见》（发改数据〔2024〕1836号）、《国家发展改革委 国家数据局 工业和信息化部关于印发〈国家数据基础设施建设指引〉的通知》（发改数据〔2024〕1853号）等多项政策文件，布局建设行业高质量数据集。标准在高质量数据集建设中可发挥规范和引领作用，《国家发展改革委等部门关于印发〈国家数据标准体系建设指南〉的通知》提出重点推进建设训练数据集采集处理标准，包括训练数据集格式要求、分类分级、采集性能、分析监测、质量要求等标准。

**二是满足行业发展需求。**数据集是开发和训练人工智能模型的基础，开发和训练高质量的模型对数据集的质量要求也越来越高。对数据集进行质量评测是评判其是否“高质量”的基本路径，也是“以评促建”保障高质量数据集建设的重要手段。规范高质量数据集质量评测，对于提升数据集优质供给，促进数据集流通使用，有力支持人工智能模型开发和训练，更好赋能经济社会发展至关重要。

### （三）起草过程

2025年1月：成立标准编制组，开展广泛调研和资料收集，明确工作思路和编制原则，讨论确定标准框架。

2025年2月-5月：编制组内部讨论并编制形成标准草案。

2025年6月底：立项申报。

2025年6月-8月：全国数标委秘书处组织对该标准进行验

证试点，共 36 家单位报名参与。各试点单位结合实际业务场景开展验证工作，编制组结合收集到的意见建议修改完善标准草案，进一步提升标准的科学性、适用性和先进性。

2025 年 9 月-12 月：组织开展研讨，不断完善标准草案。

2025 年 12 月底：国家标准委正式下达标准计划。

2026 年 1 月：持续修改完善标准草案，形成征求意见稿。

## **二、国家标准编制原则、主要内容及其确定依据**

### **（一）编制原则**

该标准的编制原则主要包含两个方面：

1. 该标准涉及相关方众多，鼓励高质量数据集相关建设主体、技术服务厂商、研究机构等广泛参与，以确保标准内容科学合理，具有普适性。

2. 该标准属于《国家数据标准体系建设指南》中的“C 数据资源-CE 训练数据集-CEA 训练数据集采集处理”标准，对人工智能数据产业发展具有重要的支撑作用。该标准应充分借鉴国际、国内相关先进研究成果，与国家相关政策导向相一致。

### **（二）编制依据**

目前，国内外仍主要围绕“数据质量本身”开展数据集质量评测相关工作。高质量数据集的“高质量”不仅是“质量高”，而且是除了“数据质量达标”之外，还具备“模型应用有效”、“说明文档完整”等特点。因此，该标准主要围绕说明文档、数据质量、模型应用等三个维度规范高质量数据集的质量评测。

该标准在编制过程中主要参考了 GB/T 25000.24-2017《系统

与软件工程 系统与软件质量要求和评价(SQuaRE)第24部分：数据质量测量》、GB/T 36344-2018《信息技术 数据质量评价指标》、GB/T 41867-2022《信息技术 人工智能 术语》、YD/T 4522-2023《面向机器学习的电信数据规范 数据质量》、TR-REC-064《数据质量评测方法与指标体系》等标准或文件。

### (三) 主要内容

该标准规定了高质量数据集的质量评测，明确了指标要求、评测方法，可为开展高质量数据集质量评测活动提供指导。

高质量数据集的质量要求应覆盖说明文档、数据质量及模型应用三个维度的指标，其中，说明文档指标要求包括基本信息完整性、内容特征完整性、建设过程完整性、应用说明完整性等；数据质量指标要求包括格式规范性、安全规范性、标注规范性、结构完整性、内容真实性、内容一致性、类型一致性、内容干净性等；模型应用指标要求包括内容多样性、规模完整性、内容时效性、标注准确性、模型适配性等。在评测高质量数据集质量时，需从说明文档、数据质量及模型应用三个维度上，分别根据相应指标计算和评分方法进行评测，数据集同时满足三个维度指标的要求，视为符合高质量数据集的质量要求。

## 三、试验验证的分析

该标准所规定的内容经过公路院、中国石油、中国石化、中国交建、国家管网、国家能源、国投数科、中国联通、中国移动、中电数产、国网江苏电力、国网山东电力、江苏数据集团、四川数据集团、贵州大数据集团、杭州数据集团、辽宁电子信息院、

广东医科院、腾讯、中兴通讯、浪潮电子、烽火通信、同方知网数科、蔚来汽车、云基华海、数字宁波、中数睿智、杭州景联文、腾云天下、星河智源、厦门身份宝、集智数据、山东未来、维视达数科、森栩医学、八月瓜等 36 家企事业单位验证，已被证明确实可行，对数据集相关建设方、技术服务方等开展质量评测活动具有实际指导价值。

#### **四、与国际、国外同类标准技术内容的对比情况**

国际方面，ISO/IEC JTC 1/SC 42/WG 2 已推动制定 ISO/IEC 5259《人工智能 面向分析和机器学习的数据质量》系列标准，包括概述、术语及示例，数据质量度量，数据质量管理要求和指南，数据质量过程框架，数据质量治理框架，数据质量可视化框架等 6 部分，其中，前 5 部分已发布，第 3、4 部分与数据集质量评估相关。总体上，国际方面仍主要聚焦“数据质量”，较少涉及“模型应用”、“说明文档”等方面。

#### **五、产业化情况、推广应用论证和预期达到的经济效益、社会效益和生态效益**

开发和训练高质量的模型对数据集的质量要求越来越高，其质量要求不明确会不利于供需双方在质量要求进行匹配，进而阻碍其流通，同时，其质量要求不达标会对其应用造成负面影响，即相关模型开发和训练效果不佳。对于数据集，业界当前已经在其质量方面开展了一些评测工作，但是尚未明确其在质量方面的“高质量”要求及评测规范，这不利于促进其质量提升，进而阻碍其流通、应用等。

当前，随着新一代信息技术持续快速发展，人工智能正加速融入各行业领域，赋能实体经济高质量发展。高质量数据集是开发和训练人工智能模型的基础，本标准的发布必定有助于提升数据集优质供给，促进数据集流通应用，推动人工智能高效赋能行业发展。在推广应用方面，在本标准发布后，有关建设主体可依据标准开展高质量数据集质量评测，“以评促建”提升所建设数据集质量水平；有关技术服务方可基于标准，为建设主体提供在质量方面符合要求的数据集建设相关服务；有关第三方评测机构可基于标准，为建设主体提供符合规范的数据集质量评测服务。

## **六、是否合规引用或者采用国际国外标准**

该标准未引用或者采用国际国外标准。

## **七、与现行相关法律、法规、规章及相关标准的协调性**

该标准与现行相关法律、法规、规章及相关标准协调一致。

## **八、重大分歧意见的处理经过和依据**

该标准研制过程中未涉及重大分歧意见。

## **九、涉及知识产权或专利的情况说明**

该标准不涉及知识产权或专利。

## **十、实施国家标准的要求**

建议作为推荐性国家标准，在标准报批阶段及正式发布后，同步开展标准宣贯培训与应用示范工作。建议标准发布 6 个月后正式实施。

## **十一、贯彻标准的要求和措施建议**

1. 加强政府引导与宣传推广。标准发布后，在国家数据局

指导下，由全国数据标准化技术委员会组织开展标准宣贯活动，在高质量数据集相关产业链和应用领域加强宣传，提升标准的宣传权威性和受众针对性；

2. 完善配套政策与激励措施。建议相关部门结合本标准评测维度及要求，在数据集建设、运营、流通等方面研究出台配套政策措施。鼓励各类企事业单位依据本标准开展质量评测，对通过质量评测的“高质量数据集”建设项目给予资金等方面激励。推动数据集质量评测规范与行业监管、产业扶持政策精准衔接，将评测结果作为衡量数据集价值的重要参考，切实增强企业应用标准提升数据集质量的积极性；

3. 推动第三方评测与生态协同。鼓励第三方机构依据本标准，开展涵盖说明文档、数据质量及模型应用的数据集质量评测工作，建立健全高质量数据集质量评测能力。同时，推动建立政府与市场协同的多方采信机制，促进质量评测结果在数据集流通与价值实现中发挥作用，提升标准的实施效力与行业认可度。

## **十二、替代或废止现行相关标准的建议**

无。

## **十三、公平竞争审查结论**

该标准已完成公平竞争审查，并填写了《公平竞争审查表》。该标准起草过程中无限制或变相限制市场准入和退出、商品要素自由流动等情况，未对经营者生产经营成本、生产经营行为造成不利影响，不存在违反《公平竞争审查条例》规定的情况，符合公平竞争审查标准。

#### 十四、其它应予说明的事项

无。

国家标准《高质量数据集 质量评测规范》

编制工作组

2026-01-30