

# 国家标准《高质量数据集 数据标注要求》 (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，根据国家标准化管理委员会关于下达 2025 年第十二批推荐性国家标准计划的通知，国家标准《高质量数据集 数据标注要求》制定项目正式立项，该项目计划号为 20256911-T-907，该项目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。

该标准由中国电子信息产业发展研究院、中国电信集团有限公司、中国移动通信集团有限公司、中国电子技术标准化研究院、数据堂（北京）科技股份有限公司、北京海天瑞声科技股份有限公司、安徽飞数信息科技有限公司、数字广东网络建设有限公司、讯飞医疗科技股份有限公司、浙江省数字经济发展中心、天翼云科技有限公司、中国软件评测中心、成都市标准化研究院、中国联合网络通信有限公司软件研究院、国家数据发展研究院、中电信人工智能（北京）科技有限公司、中国信息通信研究院、中国电子科技集团有限公司电子科学研究院、上海数据集团有限公司、四川省大数据中心、中国质量认证中心有限公司、联通数据智能有限公司、中国联合网络通信集团有限公司、中国南方电网有限责任公司、进迭时空（杭州）科技有限公司、数据空间研究院、航天科工网络信息发展有限公司、国家工业信息安全发展研

究中心、工业和信息化部电子第五研究所、中电数据产业集团有限公司、中移互联网有限公司、北京理工大学、科大讯飞股份有限公司、国家电网有限公司大数据中心、国网江苏省电力有限公司、江苏中堃数据技术有限公司、中移（杭州）信息技术有限公司、中电金信软件有限公司、石化盈科信息技术有限责任公司、交通运输部科学研究院、蚂蚁科技集团股份有限公司、北京腾云天下科技有限公司、中国航空发动机研究院、浪潮电子信息产业股份有限公司、宁波博登智能科技有限公司、珠海市政务服务和数据管理局、上海数据交易所有限公司、深圳数据交易所有限公司、中国经济改革研究基金会、下一代互联网关键技术和评测北京市工程研究中心有限公司、上海市数商协会、上海数据研究院有限公司、广州数据集团有限公司、湖南数据产业集团有限公司、广东省珠海市质量技术监督标准与编码所、中国人民大学、北京交通大学、北京化工大学、成都工业学院、国网商用大数据有限公司、中电信数智科技有限公司、中国电信股份有限公司成都分公司、成都数之联科技股份有限公司、成都新津数字科技产业发展集团有限公司、蓝象智联（杭州）科技有限公司、四川省电子信息产业技术研究院有限公司、成都边界元科技有限公司、亚信科技（中国）有限公司、浙江大数据交易中心有限公司、北京国润信达信息科技有限公司、上海零数众合信息科技有限公司、中科斯欧（合肥）科技股份有限公司、重庆回形针信息技术有限公司、北京热热科技有限公司、灵犀科技有限公司、上海新致软件

股份有限公司、北京金山云网络技术有限公司、中冶赛迪信息技术（重庆）有限公司等单位负责起草。

王超、韩冰、邓攀科、张鑫、吴坤、张群、王为中、周渭华、齐红威、赵鹏飞、郝玉峰、谭昶、王波、房建伟、鲍溪河、温晓君、田燕翔、王昆、张连夺、黄莹莹、李宜谦、周鸣爱、李锋、万芊芊、鲁楠、梁小涛、张欢、李成博、裴翔宇、何鸿凌、魏星华、廖晓玲、楚丹丹、王培养、冯汝铨、尤梦祥、周涛、刘莎、白玲玉、李天舒、周润松、时晓光、杨富贵、王尧、程彤彤、毕严先、张瑞、周强、王平、管庆旭、王锋、杜潇霖、王昊、赵兴华、杨瑞、王皓天、蔡卓骏、杨庄媛、范联伟、吴光周、樊威、李荪、周凯、邵元勋、李卫、黄云霞、沈鹏飞、程广明、胡景朝、丁浩、顾延甲、潘小古、唐浩、王树良、阮思捷、刘俊华、郝亚贤、陈振宇、张帆、王鹏飞、夏飞、魏清、陈思源、梅铮、陈倩、李佳、招庚、黄莉莉、张晓蒙、张亚东、吴超、齐园、赵捷、赵鸿辉、刘圣威、古亮、肖云、张旭东、卢勇、刘庆威、刘国庆、蒋乐冰、祁小华、吕文哲、宫大庆、涂群、沈益民、邓春宇、杨经纬、罗莉、傅彦、邵世伟、王超、苗放、陈峥、许小泉、孔俊、陈昊洁、兰春嘉、宣秀芳、龚量亮、鲁楠、姜琳杰、寇祖亮、王文朋、孙小东等起草。

起草单位、起草人及各自完成的工作如下：

王超、韩冰、邓攀科、温晓君（中国电子信息产业发展研究院）负责标准整体框架设计与统筹协调，牵头制定数据标注框架

总体要求，协调处理意见分歧，负责各阶段的整体进度把控及内容审核。

黄莹莹、李宜谦、周鸣爱、胡景朝（中国电子信息产业发展研究院）参与标准框架制定和研讨，协同推进各阶段的整体进度把控及内容审核。

张鑫、周渭华（中国电信集团有限公司）牵头编写流程要求中的任务实施阶段和交付验收阶段、技术要求中的思维链数据等章节内容。

李锋、万芊芊（中国电信集团有限公司）参与编写流程要求等章节内容。

吴坤、鲁楠、梁小涛、裴翔宇（中国移动通信集团有限公司）牵头编写流程要求中的前期准备阶段、技术要求中的多模态数据、点云数据、时序数据等章节内容。

张群、王为中、赵鹏飞（中国电子技术标准化研究院）牵头编写术语和定义、标注参与方类型与职责等章节内容，参与标准框架制定和研讨。

张欢、李成博（中国电子技术标准化研究院）参与编写流程要求等章节内容。

齐红威、何鸿凌、魏星华（数据堂（北京）科技股份有限公司）牵头编写技术要求中的图像数据、文本数据等章节内容。

郝玉峰、廖晓玲、楚丹丹（北京海天瑞声科技股份有限公司）牵头编写技术要求中的音频数据、视频数据等章节内容。

谭昶、王培养（安徽飞数信息科技有限公司）参与编写流程要求等章节内容。

王波、冯汝铨（数字广东网络建设有限公司）参与编写管理要求等章节内容。

鲍溪河、尤梦祥（讯飞医疗科技股份有限公司）参与编写技术要求等章节内容。

田燕翔（浙江省数字经济发展中心）参与编写管理要求等章节内容。

王昆（天翼云科技有限公司）参与编写管理要求等章节内容。

张连夺、周润松（中国软件评测中心）参与编写管理要求等章节内容。

刘莎、白玲玉（成都市标准化研究院）参与编写管理要求等章节内容。

房建伟、周涛（中国联合网络通信有限公司软件研究院）参与编写管理要求等章节内容。

李天舒、时晓光（国家数据发展研究院）参与标准框架制定和研讨，提出标准主要技术内容相关意见建议。

杨富贵、王尧（中电信人工智能（北京）科技有限公司）参与编写技术要求等章节内容。

程彤彤（中国信息通信研究院）参与编写技术要求等章节内容。

樊威、李荪（中国信息通信研究院）参与编写管理要求等章节内容。

毕严先、张瑞(中国电子科技集团有限公司电子科学研究院)参与编写技术要求等章节内容。

周强(上海数据集团有限公司)参与编写流程要求等章节内容。

王平、管庆旭(四川省大数据中心)参与编写管理要求等章节内容。

王锋、杜潇霖(中国质量认证中心有限公司)参与编写管理要求等章节内容。

王昊、赵兴华(联通数据智能有限公司)参与编写技术要求等章节内容。

杨瑞、王皓天(中国联合网络通信集团有限公司)参与编写流程要求等章节内容。

蔡卓骏(中国南方电网有限责任公司)参与编写技术要求等章节内容。

杨庄媛(进迭时空(杭州)科技有限公司)参与编写技术要求等章节内容。

范联伟、吴光周(数据空间研究院)参与编写管理要求等章节内容。

周凯、邵元勋(航天科工网络信息发展有限公司)参与编写管理要求等章节内容。

李卫、黄云霞(国家工业信息安全发展研究中心)参与编写管理要求等章节内容。

沈鹏飞、程广明（工业和信息化部电子第五研究所）参与编写管理要求等章节内容。

丁浩、顾延甲（中电数据产业集团有限公司）参与编写流程要求等章节内容。

潘小古、唐浩（中移互联网有限公司）参与编写管理要求等章节内容。

王树良、阮思捷（北京理工大学）参与编写技术要求等章节内容。

刘俊华、郝亚贤（科大讯飞股份有限公司）参与编写技术要求等章节内容。

陈振宇、张帆（国家电网有限公司大数据中心）参与编写管理要求等章节内容。

王鹏飞、夏飞（国网江苏省电力有限公司）参与编写技术要求等章节内容。

魏清（江苏中堃数据技术有限公司）参与编写管理要求等章节内容。

陈思源、梅铮（中移（杭州）信息技术有限公司）参与编写技术要求等章节内容。

陈倩、李佳（中电金信软件有限公司）参与编写管理要求等章节内容。

招庚（石化盈科信息技术有限责任公司）参与编写技术要求等章节内容。

黄莉莉（交通运输部科学研究院）参与编写技术要求等章节内容。

张晓蒙（蚂蚁科技集团股份有限公司）参与编写管理要求等章节内容。

张亚东（北京腾云天下科技有限公司）参与编写技术要求等章节内容。

吴超（中国航空发动机研究院）参与编写技术要求等章节内容。

齐园（浪潮电子信息产业股份有限公司）参与编写管理要求等章节内容。

赵捷（宁波博登智能科技有限公司）参与编写技术要求等章节内容。

赵鸿辉（珠海市政务服务和数据管理局）参与编写管理要求等章节内容。

刘圣威（上海数据交易所有限公司）参与编写管理要求等章节内容。

古亮（深圳数据交易所有限公司）参与编写管理要求等章节内容。

肖云（中国经济改革研究基金会）参与编写管理要求等章节内容。

张旭东（下一代互联网关键技术和评测北京市工程研究中心有限公司）参与编写技术要求等章节内容。

卢勇（上海市数商协会）参与编写管理要求等章节内容。



刘庆威（上海数据研究院有限公司）参与编写技术要求等章节内容。

刘国庆（广州数据集团有限公司）参与编写管理要求等章节内容。

蒋乐冰（湖南数据产业集团有限公司）参与编写流程要求等章节内容。

祁小华（广东省珠海市质量技术监督标准与编码所）参与编写流程要求等章节内容。

吕文哲（中国人民大学）参与编写技术要求等章节内容。

官大庆（北京交通大学）参与编写流程要求等章节内容。

涂群（北京化工大学）参与编写技术要求等章节内容。

沈益民（成都工业学院）参与编写流程要求等章节内容。

邓春宇（国网商用大数据有限公司）参与编写技术要求等章节内容。

杨经纬（中电信数智科技有限公司）参与编写技术要求等章节内容。

罗莉（中国电信股份有限公司成都分公司）参与编写流程要求等章节内容。

傅彦（成都数之联科技股份有限公司）参与编写技术要求等章节内容。

邵世伟（成都新津数字科技产业发展集团有限公司）参与编写管理要求等章节内容。

王超（蓝象智联（杭州）科技有限公司）参与编写技术要求等章节内容。

苗放（四川省电子信息产业技术研究院有限公司）参与编写管理要求等章节内容。

陈峥（成都边界元科技有限公司）参与编写技术要求等章节内容。

许小泉（亚信科技（中国）有限公司）参与编写管理要求等章节内容。

孔俊（浙江大数据交易中心有限公司）参与编写管理要求等章节内容。

陈昊洁（北京国润信达信息科技有限公司）参与编写管理要求等章节内容。

兰春嘉（上海零数众合信息科技有限公司）参与编写管理要求等章节内容。

宣秀芳（中科斯欧（合肥）科技股份有限公司）参与编写管理要求等章节内容。

龚量亮（重庆回形针信息技术有限公司）参与编写技术要求等章节内容。

鲁楠（北京热热科技有限公司）参与编写技术要求等章节内容。

姜琳杰（灵犀科技有限公司）参与编写技术要求等章节内容。

寇祖亮（上海新致软件股份有限公司）参与编写管理要求等章节内容。

王文朋（北京金山云网络技术有限公司）参与编写管理要求等章节内容。

孙小东（中冶赛迪信息技术（重庆）有限公司）参与编写管理要求等章节内容。

## （二）制定背景及意义

当前，数据标注已成为人工智能产业链的关键环节。随着深度学习、大语言模型等技术快速发展，数据标注产业规模持续扩大，形成了一批专业化数据标注企业和平台。数据标注在图像识别、语音识别、自然语言处理等领域的应用不断深化，有效支撑了人工智能技术的创新发展。然而，数据标注行业在快速发展的同时也面临标注规范不统一、流程不规范、技术要求不明确、管理机制不完善等问题，制约了数据标注产业的健康发展，也影响了人工智能模型的训练效果。

2025年1月13日，国家数据局等4部门联合印发《关于促进数据标注产业高质量发展的实施意见》，明确提出“制定数据标注技术、质量、能力等国家标准”。随着生成式人工智能、大模型技术的快速发展，数据标注工作出现了新的变化和要求，如大模型对高质量、大规模训练数据提出了更高需求，多模态数据的联合标注成为新的发展方向，思维链条数据、点云数据、时序数据等新型数据类型的标注需求快速增长，人机协同标注、智能辅助标注等新技术不断涌现，这些变化都对数据标注的标准化、规范化提出了迫切要求。本标准着眼于当前人工智能发展需求和数

据要素市场建设要求，规范数据标注全流程、全要素，为高质量数据集建设提供技术支撑。

数据标注是数据加工处理的重要环节，是高质量数据集建设的基础工作。数据标注标准是数据标准体系的重要组成部分，系统化地规定了数据标注的流程要求、技术要求和管理要求，有利于为数据标注项目的规划、实施与管理提供技术指导，同时有利于统筹推进高质量数据集标准体系的建设工作，促进数据要素市场化配置和人工智能产业高质量发展。

### （三）起草过程

2025 年 2 月：成立标准编制组，开展广泛调研和资料收集。明确标准定位为面向人工智能模型训练的高质量数据集数据标注要求，确定工作思路和编制原则，讨论确定标准总体框架。

2025 年 3 月-5 月：编制组分工协作开展草案编写工作。重点研究数据标注全流程规范、不同模态数据的标注技术要求以及标注管理要求。编制组内部多轮讨论修改，形成标准草案初稿。

2025 年 6 月底：完成立项申报材料准备，正式向国家标准化管理委员会提交立项申请。

2025 年 9 月-10 月：组织召开多次标准研讨会，邀请数据标注企业、人工智能应用单位、第三方评测机构等各方代表参与讨论，根据意见不断完善标准草案。

2025 年 11 月：全国数标委 2025 年第二次标准周活动期间，本标准面向数据技术工作组成员单位进行了介绍，并充分听取各方意见和建议。

2025 年 11 月-12 月：结合标准周期间各方建议持续修改完善标准草案。

2025 年 12 月底：国家标准化管理委员会正式下达标准计划。

2026 年 1 月：结合前期各方意见反馈，对标准草案进行全面修改完善，形成标准征求意见稿及编制说明。

## 二、国家标准编制原则、主要内容及其确定依据

### （一）编制原则

本标准的编制原则包含以下方面：

（1）科学性与实用性相结合。本标准基于数据标注产业实践和人工智能技术发展需求，充分调研国内外数据标注方法和流程，提出科学合理的标注要求，确保标准内容既具有理论支撑，又能够有效指导实际工作。

（2）系统性与完整性。本标准覆盖数据标注全流程，从前期准备、任务实施到交付验收，形成完整的流程链条；涵盖图像、文本、音频、视频、多模态、思维链条、点云、时序等主流数据类型，构建全面的技术体系；明确流程要求、技术要求、管理要求三个维度，建立系统的标准框架。

（3）通用性与扩展性。本标准规定的标注要求应具有普适性，能够适用于不同行业、不同场景的数据标注工作，同时应具有良好的扩展性，能够适应数据标注技术和方法的持续演进。

（4）标准协调性。本标准与《人工智能 面向机器学习的数据标注规程》《网络安全技术 生成式人工智能数据标注安全规

范》等相关标准相互配套、协调一致，同时注重与高质量数据集相关标准的衔接。

（5）本标准编制过程中广泛征集行业专家、企业代表、科研机构的意见建议，确保标准内容能够反映产业共识，满足各方实际需求。

## （二）编制依据

在广泛参考各标准化组织、数据标注行业领先企业、人工智能技术研究机构相关研究成果的基础上，分析提出数据标注规范的研究思路和方法，并结合我国数据标注产业实践和人工智能应用需求，构建了适用于我国高质量数据集建设的数据标注标准体系。由于数据标注是一个涉及多数据类型、多任务类型的复杂工作，涉及流程管理、技术实施、质量控制等多个环节，结合数据管理、质量管理、项目管理等理论与方法，本标准从流程要求、技术要求、管理要求三个维度建立标准框架，系统规范数据标注全流程、全要素。

## （三）主要内容

本标准给出了高质量数据集数据标注的总体框架，规定了数据标注参与方类型与职责、数据标注流程要求、技术要求和管理要求。

本标准适用于面向人工智能应用的高质量数据集的数据标注工作，可为数据标注项目的规划、实施、管理和验收提供指导。

本标准主要技术内容包括：

（1）标注参与方类型与职责：明确了数据标注工作涉及的需求方、标注方、审核方等参与方的类型及其职责分工。

（2）数据标注框架：给出了数据标注的总体框架，包括流程要求、技术要求、管理要求三个维度。

（3）流程要求：规定了数据标注的总体流程，包括前期准备阶段、任务实施阶段、交付验收阶段。

（4）技术要求：针对不同数据类型和标注任务，规定了具体的技术要求。

（5）管理要求：规定了数据标注工作的管理要求，包括标注人员要求、标注工具要求、标注文档要求、标注质量要求等。

### **三、试验验证的分析**

选取若干数据标注企业和机构进行应用示范和试点验证。一是选择若干数据标注基地，基于典型标注场景，以标准内容与实际项目相结合，开展数据标注应用示范项目，为推广本标准落地应用开展验证。二是选取数据标注行业龙头企业推广使用本标准，以数据标注基地为样板，推动本标准在不同数据类型、不同任务类型、不同应用领域的推广应用。三是在主要数据标注平台开展试用，验证标准在实际标注流程中的可操作性和适用性，推动标准在数据标注产业的全面应用。

### **四、与国际、国外同类标准技术内容的对比情况**

国际上有数据管理、数据质量相关的标准，以及人工智能、机器学习等领域的标注标准，暂无专门针对高质量数据集的数据标注规范标准。本标准在数据标注流程规范、保障数据准确性、

完整性等方面与国际通用原则、做法保持一致，制定完成后优先在国内推广应用，验证其实用性和可操作性，若应用效果显著，可考虑提升为国际标准。

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

本标准能够指导数据标注项目的规划、实施与管理，通过规范数据标注流程和技术要求，全面提升数据标注质量和效率，充分释放数据要素价值，为人工智能技术创新和应用提供高质量数据支撑。同时强化数据标注质量管理体系，实现标注过程的精准控制和质量的有效保障，充分提升数据集可用性。

本标准赋能数据标注产业发展，为数据标注企业提供统一的技术规范和管理框架，提升标注效率、数据质量与服务能力，增强产业竞争力，同时能激发产业创新活力，推动标注技术和方法创新，催生智能标注、人机协同等新技术应用。通过采用本标准，有助于优化数据标注资源配置，降低标注成本，提高数据流通效率，进而促进数据要素市场化配置和人工智能产业高质量发展。

## **六、是否合规引用或者采用国际国外标准**

该标准未采用国际国外标准。

## **七、与现行相关法律、法规、规章及相关标准的协调性**

与现行法律法规以及相关标准协调一致。

## **八、重大分歧意见的处理经过和依据**

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和专利。

## **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**

为更好地发挥标准实施效力，建议标准发布后，开展标准宣贯会，在确保标准的宣传权威性和受众针对性的基础上，鼓励相关单位根据标准开展数据标注工作。

## **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**

本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 数据标注要求》

编制工作组

2026 年 1 月 27 日