



中华人民共和国国家标准

GB/T 42755—2023

人工智能 面向机器学习的数据标注规程

Artificial intelligence—Code of practice for data labeling of machine learning

2023-05-23 发布

2023-12-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 数据标注流程	2
5 标注任务前期准备	3
5.1 标注任务	3
5.2 标注人员	4
5.3 标注环境	4
6 标注任务执行	4
6.1 过程控制	4
6.2 质量保证	5
6.3 管理机制	6
7 标注结果输出	7
7.1 内部质检	7
7.2 数据交付	8
7.3 后期维护	8
图 1 数据标注流程框架	2

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。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：北京航空航天大学、中国电子技术标准化研究院、北京百度网讯科技有限公司、浪潮软件科技有限公司、山东省人工智能研究院、美的集团(上海)有限公司、北京智谱华章科技有限公司、北京爱数智慧科技有限公司、腾讯云计算(北京)有限责任公司、北京航天自动控制研究所、郑州中业科技股份有限公司、东软集团股份有限公司、北京海天瑞声科技股份有限公司、云从科技集团股份有限公司、深圳云天励飞技术股份有限公司、中国科学院软件研究所、上海依图网络科技有限公司、中国医学科学院生物医学工程研究所、平安科技(深圳)有限公司、上海商汤智能科技有限公司、上海人工智能实验室、上海计算机软件技术开发中心、中国航空综合技术研究所、中国科学院新疆理化技术研究所、中国质量认证中心、中汽数据(天津)有限公司、北京眼神科技有限公司、上海人工智能研究院有限公司、浙江大华技术股份有限公司、杭州趣链科技有限公司、常州微亿智造科技有限公司、长春博立电子科技有限公司、罗克佳华科技集团股份有限公司、上海交通大学、上海计算机软件技术开发中心。

本文件主要起草人：吴文峻、董建、马珊珊、刘祥龙、徐洋、贾一君、孟令中、任健、陈斌、赵豪杰、刘海涛、陈尚义、脱立恒、左家平、王丽娜、徐颂、王健宗、张楠、蔡亚森、王功明、陈敏刚、赵赫、金铸、郝玉峰、刘永辉、李玮、赵春昊、黄志龙、杨春林、王潇蔓、施佳樑、舒明雷、王英龙、匡立中、陈晓丰、吴庚、蒋慧、蒲江波、马元巍、邢警、乔宇、何聪辉、杨雅婷、马博、陶剑、胡进伟、楚思思、李军、宋海涛、沈灏、程淼、郑忠斌、李爽。

人工智能 面向机器学习的数据标注规程

1 范围

本文件规定了人工智能领域面向机器学习的数据标注框架流程。

本文件适用于指导人工智能领域面向机器学习的数据标注以及与之相关的研究、开发和应用等。

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 35274—2017 信息安全技术 大数据服务安全能力要求

GB/T 37973—2019 信息安全技术 大数据安全管理指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据标注 data labeling

给数据样本指定目标变量和赋值的过程。

3.2

标注任务 labeling task

按照数据标注说明对数据进行标注的活动。

3.3

数据标注方 data labeler

承担数据标注任务的人员或机构。

3.4

数据需求方 data user

提出数据标注需求的人员或机构。

3.5

标注管理方 data labeling administrator

管理数据标注任务评估、分发、交付、验收以及质量把控的人员或机构。

3.6

标注工具 labeling tool

数据标注方执行数据标注时使用的工具,标注管理方管理数据标注时使用的工具,数据需求方验收数据标注时使用的工具等所有流程相关的工具。

3.7

标注任务说明 labeling task description

数据需求方用于向标注管理方以及数据标注方明确标注任务的书面表达。

注:标注任务说明通常包含对要执行的标注任务的描述、标注方法、正反示例、验收方法与验收指标等内容。