人工智能 数据标注一般技术要求

2022 - 06 - 14 发布

2022 - 09 - 15 实施

目 次

前言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 技术分类.....	1
5 技术要求.....	1
参考文献.....	3

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。

本文件由山西省工业和信息化厅提出、组织实施和监督检查。

山西省市场监督管理局对标准的组织实施情况进行监督检查。

本文件由山西省物联网和人工智能标准化技术委员会归口。

本文件起草单位：百度云计算技术（山西）有限公司、山西省大众科技评估中心、山西集智数据服务有限公司、山西物联网行业技术中心（有限公司）、山西灵波微步科技有限公司、山西省科技评估学会。

本文件主要起草人：高海峰、李昱霖、李兴华、闵楠、胡驰、尉赤、张辰妹、吴泽衔、查晓丹、李先军、郭瑞鹏、赵波、李鹏、李鹏飞。

人工智能 数据标注一般技术要求

1 范围

本文件规定了人工智能机器学习中数据标注通用技术要求的术语和定义、通用技术分类和通用技术要求。

本文件适用于人工智能机器学习中数据标注的企业、高校、研究机构、政府机关。其他行业和机构可参照执行。

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。

DB14/T 2463 人工智能 数据标注总体框架

3 术语和定义

DB14/T 2463 界定的以及下列术语和定义适用于本文件。

3.1

数据标注预处理技术 data annotation preprocessing technology

在正式开始标注作业前，通过人工或技术的方式对待标注数据做筛选或预备标注，以实现提高后续标注效率的技术手段。

4 技术分类

4.1 数据标注通用技术

包括数据标注预处理技术和数据标注技术。

4.2 数据标注预处理技术

包括对数据清洗、预标注。

4.3 数据标注技术

包括对数据的转写、切割、评分、提取、选择、富集标注、点标注、线标注、框标注、区域标注。

5 技术要求

5.1 预处理技术要求

5.1.1 清洗

清洗是对数据进行结构化标注前做的预处理技术。应通过人工判断方式剔除不符合标注标准的无效数据，筛选出可标注数据。

5.1.2 预标注

预标注技术要求应用机器学习算法对待标注数据做识别和预先标注处理。

5.2 数据标注技术要求

5.2.1 转写

应按照标注规则要求将数据转换为相应文本。

5.2.2 切割

应按照标注规则将数据分割并标注为要求的片段。

5.2.3 评分

应按照标注规则对数据进行主观判断并打分。

5.2.4 提取

应按照标注规则将主题内容提炼，描述为相应文本。

5.2.5 选择

应按照标注规则，对数据在给定选项中选择标签及属性。

5.2.6 富集标注

应根据标注规则给定的场景、主题及描述，编写出与主题相关联文本。

5.2.7 点标注

应按照标注规则，对图像中的指定位置进行标点。

5.2.8 线标注

应按照标注规则，对图像中的指定区域用线条进行标注。

5.2.9 框标注

应按照标注规则，对图像中的指定对象，用物体的最小外接框进行框选（框选范围允许包含指定对象外的部分区域）。

注：如二维框选、三维框选。

5.2.10 区域标注

应按照标注规则，对图像中的指定对象进行多边形贴合标注。

参 考 文 献

- [1] 《项目管理知识体系指南》第5版
 - [2] 《信息技术 人工智能 面向机器学习的数据标注规范》中国电子工业标准化技术协会
-