

人工智能 数据标注通用工作规程

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 原则.....	1
5 工作流程.....	1
6 规划要求.....	2
7 实施要求.....	3
8 监控要求.....	3
9 交付要求.....	3
10 归档.....	4
参考文献.....	5

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、山西灵波微步科技有限公司、山西物联网行业技术中心（有限公司）、山西集智数据服务有限公司、山西省科技评估学会。

本文件主要起草人：司文、施佳樑、王洋、闵楠、胡驰、尉赤、张辰妹、吴泽銜、查晓丹、郭瑞鹏、李先军、赵波、李鹏、李鹏飞。

人工智能 数据标注通用工作规程

1 范围

本文件规定了人工智能机器学习中数据标注生产所需要的通用工作要求，包括原则、流程、规划、实施、监控与交付。

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

数据标注合格率 *qualified rate of data annotation*

一个标注任务中，符合标注规则要求的已标注数据总量与已标注数据总量的比值。

3.2

数据标注质量控制员 *data annotation quality controller*

数据标注工作中，按照标注规则从事数据审核等质量控制工作的人员。

3.3

数据标注任务 *data annotation task*

数据标注机构按照需方要求或合同约定，在特定的时间、成本、质量要求下完成指定数据标注的工作。

4 原则

4.1 数据标注机构对待标注数据应遵循数据安全、保密的原则。

4.2 制定数据标注计划制定时应遵循合理性、可量化性、及时性和可调整性的原则。

4.3 数据标注生产过程应遵循全过程质量控制原则，确保生产过程质量可控。

5 工作流程

数据标注工作流程见图 1。

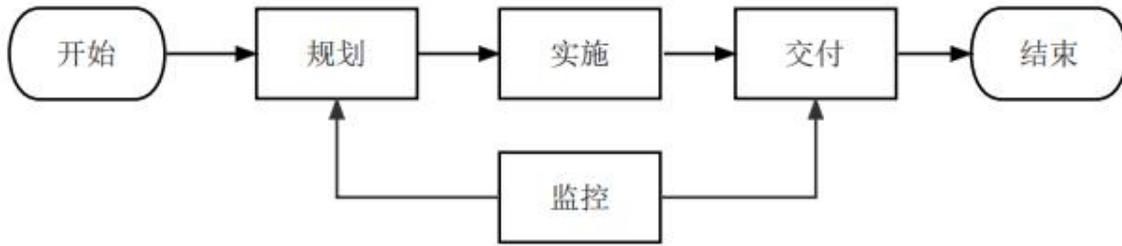


图 1 数据标注工作流程图

6 规划要求

6.1 明确需求

数据标注生产开始前，数据标注项目负责人应及时获取数据需求方数据标注的需求说明书，并应包括以下内容：

- a) 明确数据标注需求完成的时间节点；
- b) 明确数据标注的内容；
- c) 明确数据标注的方法；
- d) 明确数据验收规则；
- e) 其他要求。

6.2 获取数据

数据标注项目负责人应及时从数据需求方处获取待标注数据，数据获取方式包括：

- a) 平台接口对接的方式；
- b) 存储介质复制的方式。

6.3 规则验证与细化

数据标注项目负责人应按照标注规划进行标注规则验证与细化，并完成以下内容：

- a) 确定数据试标数量：应遵循数据标注内容覆盖全面性及数据标注成本最小化要求，抽取需求方指定数量作为试标数量；
- b) 进行试标：应选择中等以上技能的数据标注员，按照标注规则对抽样数据进行标注，以验证标注规则的合理性；
- c) 细化数据标注规则：应遵循易理解、易操作的原则，且与数据需求方要求一致的原则，对数据试标中发现的不合理规则进行修改细化。

6.4 明确计划

数据标注项目负责人应根据标注需求、数据试标效率确定标注计划。计划内容包括：进度计划、人员计划、资金计划、工具计划、质量控制计划、验收计划。

6.5 专项培训

数据标注培训师应按照标注计划和标注规则,在数据标注生产过程中对数据标注员进行有针对性的培训,确保标注质量。

7 实施要求

7.1 任务创建

数据标注任务负责人应利用标注工具创建数据标注项目相关内容。

7.2 任务分发

数据标注任务负责人应利用标注工具将数据标注任务分派给数据标注员或数据标注团队。

7.3 任务实施

数据标注员或数据标注团队应使用数据标注工具完成指派的数据标注任务。

8 监控要求

8.1 任务检验

8.1.1 任务质量检验

数据标注质量控制员在任务进行过程中按照标注规则进行质量检验和质量纠偏。

8.1.2 任务进度检验

数据标注项目负责人在任务进行中应使用数据标注工具,对数据标注日产能和数据标注计划的数值偏差情况进行检验,并及时调整资源和日工作量要求。

8.2 任务数据回收

数据标注项目负责人应对标注不合格数据进行收回,并重新分派进行标注。

9 交付要求

9.1 验收

数据标注项目负责人应将审核结果达到数据标注合格率要求的数据,交付需求方验收。

9.2 总结

数据标注项目负责人应组织各工作角色参与项目的总结工作,编制总结报告,内容包括:

- a) 数据标注项目基本信息:描述本次数据标注项目的合作背景、数据标注内容及类型;
- b) 数据标注过程控制措施:描述本次数据标注项目在执行过程中采取的执行方法,还原标注过程;
- c) 数据标注方法总结:描述本次数据标注项目遇到的困难及采取的有效措施,总结可行的方法做经验复用;

- d) 数据标注措施改进：描述本次数据标注项目控制过程的不足，总结可以改进的方法和措施，应对下次同类项目；
- e) 其他需要总结的内容。

9.3 售后

数据需求方确认数据标注结果验收通过后，数据标注机构应跟进整个项目的售后工作。包括项目终期审计、项目综合评价、数据检查改进和服务监督处理。

10 归档

数据标注机构的项目管理部门应在项目完成后或定期对生产过程文档进行分类、整理、归档，并根据生产文档对生产过程进行核验，验证生产过程与合同要求符合性。

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
-