

# 人工智能 数据标注总体框架

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 总则.....	1
5 生产流程.....	1
6 生产内容.....	2
参考文献.....	4

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

本文件主要起草人：郁鸣钢、高果荣、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 数据标注总体框架

## 1 范围

本文件提供了人工智能机器学习中数据标注生产的生产流程和生产内容的建议,并给出了相关信息。  
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

本文件没有规范性引用文件。

## 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 annotation**

对文本、图像、语音、视频等数据添加标签的过程。

### 3.2

**数据标注工具 data annotation tool**

数据标注中具有管理、标注、质量控制、验收等部分功能或全部功能的软件统称。

### 3.3

**数据标注规则 data annotation rules**

数据标注中对数据进行人工处理所遵循的方法、质量控制等要求的统称。

### 3.4

**数据标注员 data annotator**

按照数据标注规则对数据进行标注的人员。

### 3.5

**数据标注培训师 data annotation trainer**

根据数据标注规则,对数据标注员和数据标注质量控制员进行培训的人员。

## 4 总则

4.1 数据标注应遵循规范的原则。

4.2 数据标注生产应遵循安全、有效和准确的原则。

## 5 生产流程

数据标注生产是按照需求方提出的数据标注规则，组织具有相应技能的数据标注员利用数据标注工具进行生产的过程。包括规划、实施、交付以及全生产过程的监控。具体生产过程可根据生产规模简化或裁剪，生产流程见图 1。

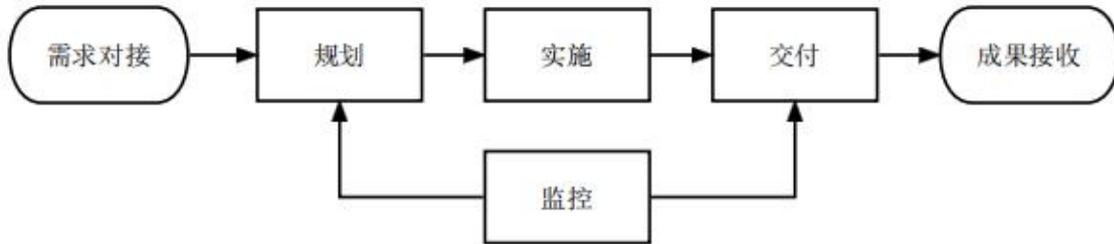


图 1 数据标注生产流程

## 6 生产内容

### 6.1 组成

数据标注生产内容主要包括标注实现和基础保障两部分，具体生产内容见图 2。

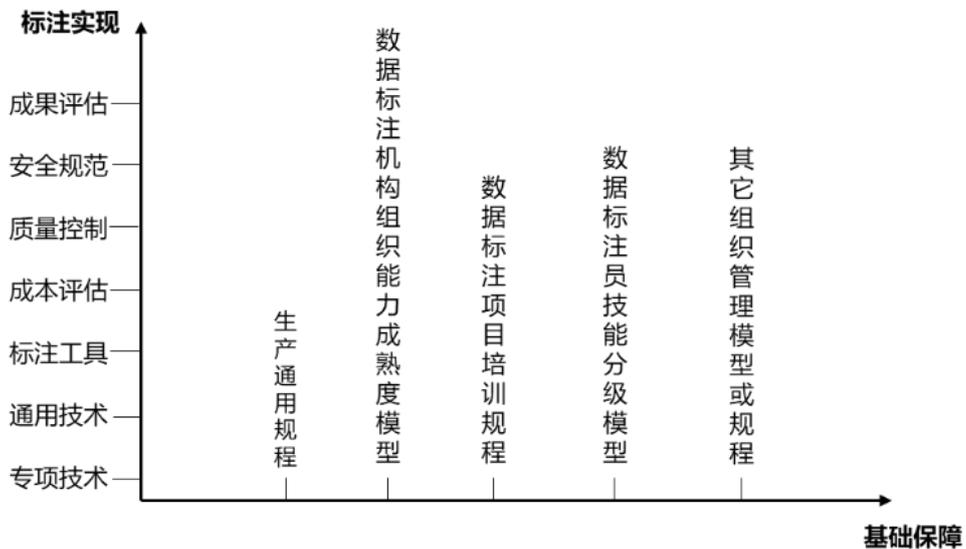


图 2 数据标注生产内容

### 6.2 标注实现

#### 6.2.1 概述

数据标注实现包括数据标注工具、安全规范、质量控制、成本评估、成果评估、通用技术、专项技术要求。

#### 6.2.2 数据标注工具

数据标注软件工具规定了该工具的兼容性、齐全度、易用性和管理性。

### 6.2.3 数据标注安全规范

数据标注安全规范规定了数据标注在安全管控上所遵循的要求。

### 6.2.4 数据标注质量控制

数据标注质量控制规定了数据标注在保证质量上所采取的控制手段。

### 6.2.5 数据标注成本评估

数据标注成本评估规定了数据标注成本评估方法和模型要素。

### 6.2.6 数据标注成果评估

数据标注成果评估规定了数据标注验收评估方法和模型要素。

### 6.2.7 数据标注通用技术

数据标注通用技术规定了人工智能行业通用领域的数据进行标注的基本技术。

### 6.2.8 数据标注专项技术

数据标注专项技术规定了人工智能行业专业领域的数据进行标注的特定技术。

## 6.3 基础保障

### 6.3.1 概述

数据标注基础保障包括生产通用规程、数据标注机构组织能力成熟度模型、数据标注员技能分级模型、数据标注项目培训规程、以及其他组织管理模型或规程。

### 6.3.2 数据标注生产通用规程

数据标注生产通用规程规定了数据标注生产中应当共同遵循的要求。

### 6.3.3 数据标注机构组织能力成熟度模型

数据标注机构组织能力成熟度模型规定了数据标注机构组织能力的衡量模型。

### 6.3.4 数据标注员技能分级模型

数据标注员技能分级模型规定了数据标注从业人员技能等级划分的模型。

### 6.3.5 数据标注项目培训规程

数据标注项目培训规程规定了在数据标注项目中，培训工作应当遵循的方法、步骤及需要达到的要求。

### 参 考 文 献

- [1] GB/T 38548.5 内容资源数字化加工 第5部分：质量控制
  - [2] 《项目管理知识体系指南》第5版
  - [3] 《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
-