

ICS 35.020  
I651

T/ SIA

中国软件行业协会团体标准

T/SIA 033—2021

## 数据标注工程师能力评估规范

Specification For Capability Assessment Of Data Annotation Engineers

2021-12-16 发布

2021-12-16 实施

中国软件行业协会发布

# 目 录

前言.....	1
1 范围.....	2
2 术语和定义.....	2
2.1 数据 DATA.....	2
2.2 数据标注 DATA ANNOTATION.....	2
2.3 标签 LABEL.....	2
2.4 标注工具 ANNOTATION TOOL.....	2
2.5 标注平台 ANNOTATION PLATFORM.....	2
2.6 标注方法 ANNOTATION METHOD.....	2
3 职业技能等级认证要求.....	2
3.1 面向人群.....	2
3.2 职业技能等级划分.....	2
4 职业技能等级认证要求说明.....	3
4.1 基础知识要求.....	3
4.2 专业技能要求.....	3
4.3 初级数据标注工程师职业技能要求.....	4
4.4 中级数据标注工程师职业技能要求.....	4
4.5 高级数据标注工程师职业技能要求.....	4
5 职业技能等级考核要求.....	5
5.1 报名要求.....	5
5.2 职业技能等级考核标准.....	5
5.3 监考人员、考评人员要求考场要求.....	6
5.4 考核时间要求.....	6
5.5 考核场所设备.....	6

# 前言

人工智能行业处于高速增长期，产业化大潮来临。人工智能人才需求巨大，其中基础数据服务人才是其中一个重要的缺口板块。基础数据服务行业的需求在不断更新迭代中，对于人才的要求不断提升，但整体行业对于人才没有成熟的评估标准，从业人员没有明晰的职业规划和发展路径，导致人才留存时间短、流失严重。为提升数据标注行业从业人员知识技能水平，提升行业人才的留存率，提升数据标注行业影响力，特起草《数据标注工程师能力评估规范》，对数据标注从业人员的专业活动内容进行规范细致描述，明确了各等级从业人员的工作领域、工作内容以及知识水平、专业能力要求。

本标准由中国软件行业协会提出并归口。

本标准起草单位：北京百度网讯科技有限公司，澳鹏数据科技（上海）有限公司，北京安捷智合科技有限公司，北京玛达科技有限公司，北京深度搜索科技有限公司，北京云测信息技术有限公司，北京中软国际科技服务有限公司，软通动力信息技术（集团）股份有限公司，上海商汤智能科技有限公司，数据堂（北京）科技股份有限公司，苏州柏川数据科技有限公司，苏州泛函信息科技有限公司，先导（苏州）数字交通产业投资有限公司，云从科技集团股份有限公司。

本标准主要起草人：施佳樑，李昱霖，李明，陈尚义，蒋晓琳，马婧，季晓霞，陆汀，谭小红，陈茜琪，董丽娅，杜霖，何聪辉，胡希塔，贾宇航，李阳阳，马东升，马艳，秦张波，王斌，王佳利，杨位永，咎智，付晓宇，张然。

本标准为首次制定。

本标准按照GB / T 1.1-2009 给出的规则起草。

# 数据标注工程师能力评估规范

## 1 范围

本文件规定了数据标注工程师对应的工作领域、工作内容及需要具备的职业能力。

本文件适用于指导数据标注工程师的培训与考核，与标注相关单位的人员聘用、培训与考核可参考使用。

## 2 术语和定义

下列术语和定义适用于本文件。

### 2.1 数据 data

对事实、概念和指令的一种形式化表示，适于人工或自动方式进行通信、解释或处理。

### 2.2 数据标注 data annotation

对文本、图像、语音、视频等待标注数据进行归类、整理、编辑、纠错、标记和批注等操作，为待标注数据增加标签，生成满足机器学习训练要求的机器可读数据编码。

### 2.3 标签 label

标识数据的特征、类别和属性等，用于建立数据及机器学习训练要求所定义的机器可读数据编码间的联系。

### 2.4 标注工具 annotation tool

数据标注员完成标注任务产生标注结果时所需的工具和软件。

注1:标注工具可生成标签并提供参考模板。

注2:不同的数据类型和标注任务需要不同的标注工具。标注工具按自动化程度可分为手动、半自动、自动三种。

### 2.5 标注平台 annotation platform

开展标注任务的系统化工作台。

注:标注平台在包含标注工具全部功能的基础上将所有标注环节工具化，可有效地对标注任务进行全局管理和跟踪。

### 2.6 标注方法 annotation method

定义数据标注员进行数据标注时的环境和规程，包含标注对象定义、所用标注工具和标注平台、标注格式、标注前的准备工作、标注后的处理工作等。

## 3 职业技能等级认证要求

### 3.1 面向人群

本标准面向职业岗位（群）：主要适用于已经或有志于在人工智能、自然语言处理、模型算法、大数据等行业从业的数据标注人员、数据标注项目经理、数据处理人员、数据分析人员、人工智能训练人员等岗位人员。

### 3.2 职业技能等级划分

#### 3.2.1 等级划分

数据标注员职业技能分为三个等级：初级、中级、高级，三个级别依次递进，高级别涵盖低级别职业技能要求。

### 3.2.2 数据标注工程师（初级）

对数据标注有一定的认知和理解，掌握数据标注基础理论知识、掌握数据标注工作流程；熟练使用数据标注工作平台，掌握文本、图像、语音、视频四大标注类型的基本标注规则和方法，能够通过学习给定的标注规则，完成特定的数据标注任务。

### 3.2.3 数据标注工程师（中级）

对数据标注具有较深入的理解，熟练掌握常见标注任务的目标及基本标注方法，能够高质量完成数据标注任务；能够总结提炼标注规则中的易错点、重难点，并及时沉淀；能够对标注规则提出改进建议，并对标注规则做简单的修正；同时在项目任务实施过程中，可参与一定的标注业务培训及标注规则辅导、答疑，以提升整个项目的质量与效率。

### 3.2.4 数据标注工程师（高级）

深度理解标注行业和标注业务需求，精通常见标注任务的目标及基本标注方法，能够高质量完成数据标注任务；精通各类项目标注规则，且能够独立根据项目需要完善和细化规则；或能够独立主持标注规则、标注体系的制定和完善；同时掌握标注项目运营的全流程，承担标注项目管理职责，保证整个项目的质量与效率。

## 4 职业技能等级认证要求说明

### 4.1 基础知识要求

#### 4.1.1 标注基础知识

包括人工智能基础知识，如发展历史、相关领域、应用场景等；数据标注基础知识，如基本概念、专业术语、应用场景等相关知识。

#### 4.1.2 项目管理知识

包括数据标注项目运营流程环节，数据标注项目接入与筹备，数据标注项目实施、管理、交付，数据标注质量检验等相关知识。

#### 4.1.3 标注平台知识

数据标注平台基础功能介绍、国内外领先平台介绍等相关知识。

#### 4.1.4 文本标注知识

文本标注基本概念、发展与现状、应用场景、标注内容、标注案例等相关知识。

#### 4.1.5 图像标注知识

图像标注基本概念、发展与现状、应用场景、标注内容、标注案例等相关知识。

#### 4.1.6 语音标注知识

语音标注基本概念、发展与现状、应用场景、标注内容、标注案例等相关知识。

#### 4.1.7 视频标注知识

视频标注基本概念、发展与现状、应用场景、标注内容、标注案例等相关知识。

### 4.2 专业技能要求

#### 4.2.1 专业技能要求涵盖范围

数据标注工程师的专业技能要求主要涵盖在标注任务实施、标注规则理解与制定、标注项目管理等工作方向上的要求。

#### 4.2.2 专业技能要求说明

a) 本文件对数据标注员初级、中级、高级的相关知识要求和专业技能要求依次递进，高级别涵盖低级别的要求。

b) 本文件中提到的各标注项目，包括但不限于文本、图像、语音、视频等类型；且各标注项目要求会根据实际业务需求进行变化调整。

### 4.3 初级数据标注工程师职业技能要求

#### 4.3.1 标注任务实施

标注任务实施，应满足以下要求：

- a) 熟知各标注项目运营流程，并依据项目流程执行标注任务；
- b) 能够根据各标注项目要求，通过相关标注项目的练习要求与标注任务能力准入；
- c) 能够严格按照项目安排，实施数据标注作业并如期提交；
- d) 能够按照各标注项目中的要求和指导，利用标注平台完成相应标注项目，并满足标注项目对初级数据标注工程师正确率的要求。

#### 4.3.2 标注规则理解与制定

标注规则理解与制定，应满足以下要求：

- a) 能够读懂并理解标注项目的目标及规则；
- b) 能够在阅读学习标注规则时，发现项目规则中存在的问题，并及时向上反馈；
- c) 能够在实际标注工作中，发现项目规则有待改进或更新的信息，并及时向上反馈；
- d) 初步了解标注项目中，标注规则的编写思路与注意事项。

### 4.4 中级数据标注工程师职业技能要求

#### 4.4.1 标注任务实施

标注任务实施，应满足以下要求：

- a) 掌握各标注项目运营流程，清晰明确标注项目运营实施阶段；
- b) 能够根据各标注项目要求，通过相关标注项目的练习要求与标注任务能力准入；
- c) 能够严格按照项目安排，实施数据标注作业并如期提交；
- d) 能够按照标注项目的要求和指导，利用标注平台高质量完成各标注项目，并满足标注项目对中级数据标注工程师正确率的要求。

#### 4.4.2 标注项目管理

标注项目管理，应满足以下要求：

- a) 能够参与一定的标注业务培训及标注规则辅导、答疑；
- b) 能够承担各标注项目质检工作，且质检正确率不低于项目要求；
- c) 能够通过办公软件或相关系统，完成标注项目测试题筛选或评审工作；
- d) 能够初步制定提升项目质量或提升标注效率的执行方案。

#### 4.4.3 标注规则理解与制定

标注规则理解与制定，应满足以下要求：

- a) 能够快速理解各标注项目中的目标及规则；
- b) 能够总结提炼标注规则中的易错点、重难点，并及时沉淀；
- c) 能够在理解项目规则的基础上，发现项目规则中的问题，并及时向上反馈；
- d) 能够在实际标注工作中，发现项目规则有待改进或更新的信息，并及时向上反馈；
- e) 深度理解各标注项目需求，能对标注规则做简单的修正。

### 4.5 高级数据标注工程师职业技能要求

#### 4.5.1 标注任务实施

标注任务实施，应满足以下要求：

- a) 能够根据各标注项目要求，通过相关标注项目的练习要求与标注任务能力准入；
- b) 熟练掌握各标注项目运营流程，能即时参与到各标注项目运营实施的各个阶段；

c) 精通常见标注任务的目标及基本标注方法，能够高质量完成数据标注任务，并满足标注项目对高级数据标注工程师正确率的要求；

d) 能够严格按照标注项目安排，实施数据标注作业并如期提交。

#### 4.5.2 标注项目管理

标注项目管理，应满足以下要求：

- a) 掌握标注项目运营的全流程，可承担标注项目相关管理职责；
- b) 能够独立制定项目实施和管理方案，相关方案必需可落地执行；
- c) 能够独立运用相关软件录制培训讲解视频，并可在线上线下组织培训活动；
- d) 能够独立通过办公软件或相关系统，完成标注项目测试题筛选或评审工作；
- e) 能够独立制定提升项目质量或提升标注效率的执行方案；
- f) 能够有效把控标注项目的运营实施，保证项目如期、高质交付。

#### 4.5.3 标注规则理解与制定

标注规则理解与制定，应满足以下要求：

- a) 精通各标注规则，且能够根据项目需要完善和细化规则；
- b) 能够提前预判各标注规则中的易错点、重难点，并提前辅导、培训标注团队；
- c) 或能够独立主持标注规则、标注体系的制定和完善。

### 5 职业技能等级考核要求

#### 5.1 报名要求

- a) 具备以下条件之一，可报名参与“初级”数据标注工程师认证：
  - ① 欲从事数据标注行业人员，高中、中专同等学历及以上学历者；
  - ② 或距报名为止，连续参与标注项目 $\geq 0.5$ 年。
- b) 具备以下条件之一，可报名参与“中级”数据标注工程师认证：
  - ① 取得初级数据标注工程师认证后，连续参与标注项目 $\geq 0.5$ 年；
  - ② 或距报名为止，连续参与标注项目 $\geq 1$ 年。
- c) 具备以下条件之一，可报名参与“高级”数据标注工程师认证：
  - ① 取得中级数据标注工程师认证后，连续参与标注项目 $\geq 1$ 年；
  - ② 或距报名为止，连续参与标注项目 $\geq 2$ 年。

#### 5.2 职业技能等级考核标准

##### 5.2.1 考核内容

- a) 数据标注工程师认证，考核方式分为理论知识考核、标注实操考核（技能考核），且均以机考方式进行。
- b) 理论知识考核，主要对数据标注工程师应掌握的行业知识和专业理论知识进行考核；
- c) 标注实操考核（技能考核），主要对数据标注工程师应具备的职业技能进行考核。

##### 5.2.2 达标要求

- a) “初级”数据标注工程师合格标准：理论知识考核，总分100分，合格分数为 $\geq 60$ 分；标注实操考核（技能考核），总分100分，合格分数为 $\geq 60$ 分。
- b) “中级”数据标注工程师合格标准：理论知识考核，总分100分，合格分数为 $\geq 60$ 分；标注实操考核（技能考核），总分100分，合格分数为 $\geq 60$ 分。
- c) “高级”数据标注工程师合格标准：理论知识考核，总分100分，合格分数为 $\geq 60$ 分；标注实操考核（技能考核），总分100分，合格分数为 $\geq 60$ 分。
- d) 在取得以上成绩后，学员需申请【高级】数据标注工程师专业答辩，答辩通过后，方可获得“高级”数据标注工程师合格证书。

## 5.2.3 理论知识权重表

表 1 数据标注工程师职业技能等级认证理论知识权重表

项目		技能等级	初级数据标注 工程师 (%)	中级数据标注 工程师 (%)	高级数据标注 工程师 (%)
基础理论知识	标注基础知识		25	15	10
	项目管理知识		--	20	30
	标注平台知识		10	10	5
专业理论知识	文本标注		15	任选两项, 各 占 27.5	任选两项, 各 占 27.5
	图像标注		20		
	音频标注		15		
	视频标注		15		
合计			100	100	100

## 5.2.4 专业技能权重表

表 2 数据标注工程师职业技能等级认证专业技能要求权重表

项目		技能等级	初级数据标注 工程师 (%)	中级数据标注 工程师 (%)	高级数据标注 工程师 (%)
技能要求	文本标注		25	任选两项, 各 占 50	任选两项, 各 占 50
	图像标注		35		
	音频标注		20		
	视频标注		20		
合计			100	100	100

## 5.3 监考人员、考评人员要求考场要求

监考人员与考生配比不低于 1:30, 且每个考场不少于 2 名监考人员。

## 5.4 考核时间要求

理论知识考核时间不少于 30 min; 技能考核时间初级、中级、高级均不少于 90 min。

## 5.5 考核场所设备

理论知识考核、技能考核均需配备有必要设施(电脑), 通风条件良好, 光线充足的场所进行。

台式计算机(笔记本)	
CPU	≥双核处理器
内存	≥8G
显卡	G405/512M/独立显卡
接口	≥6 个 USB 接口, ≥1 个串口, ≥2 个 PS2 接口
硬盘	≥120G
网卡	≥10/100/1000M 以太网卡
显示器	≥14 寸宽屏液晶显示器, 分辨率不低于 1440x900
键盘	PS2 防水抗菌键盘
鼠标	USB 光电鼠标
操作系统	Win7 系统以上