



T/MIITEC 034-2025

数据标注产业人才岗位能力要求

Industrial talents competency framework of data annotation

2025-09-18 发布

2025-09-18 实施

工业和信息化部人才交流中心

贵州省大数据发展管理局

目 录

目 录..... 1

1 范围..... 3

2 术语和定义..... 3

3 数据标注人才岗位方向及职责..... 4

4 数据标注人才岗位能力要素..... 5

5 数据标注产业人才岗位能力要求..... 5

 5.1 数据处理方向..... 5

 5.2 技术支持方向..... 7

 5.3 模型辅助方向..... 8

 5.4 质量管控方向..... 10

 5.5 安全合规方向..... 11

 5.6 项目管理方向..... 12

 5.7 培训支持方向..... 14

附 录 A （资料性附录） 数据标注人才岗位能力提升..... 15

附 录 B （资料性附录） 数据标注人才岗位能力评价..... 17

参 考 文 献..... 18



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。

本文件由工业和信息化部人才交流中心提出并归口。

本文件起草单位：工业和信息化部人才交流中心、贵州省大数据发展管理局、中国信息通信研究院、信通院河北科技创新研究院、上海翰纬信息科技有限公司、陕西淘丁实业集团有限公司、神龙同城科技（成都）有限责任公司、陕西工业职业技术大学、贵州梦动科技有限公司、贵州迦太利华信息科技有限公司、中国电信股份有限公司成都分公司、工信通(北京)科技有限公司、杭州科技职业技术学院、贵州星长征教育科技有限公司、浙江省经济和信息化教育培训中心有限责任公司、深圳职业技术大学。

本文件主要起草人：李学林、任利华、程宇、赵兴燕、杨熠、冯成、曹峰、李荪、樊威、郭利、夏骆辉、徐倩、王永军、麻杰、余婷婷、苏斌、王欢、李煜、赵永新、宋映徽、宋波、何奇彦、仝丹娜、杨红、盛华、赵杰、陶姝、黄出为、刘庭波、刘青洪、黄珊、罗莉、曾诚、马卫东、杨悦梅、查香云、洪德衍、彭启练、曾祥荣、王才萍、余尚升、周舒扬、王瑞琪、李宁波、李丽。

本文件为首次制定。

数据标注产业人才岗位能力要求

1 范围

本文件规定了数据标注领域主要方向岗位能力要求。

本文件适用于指导各单位开展数据标注人才规划、人才培养、人才评价（人才认证）、人才招聘、人才引进等工作。

2 术语和定义

下列术语和定义适用于本文件。

2.1

数据标注 data labeling

给数据样本制定目标变量和赋值的过程。

[来源：GB/T 42755-2023，定义3.1]

2.2

标注任务 labeling task

按照数据标注说明对数据进行标注的活动。

[来源：GB/T 42755-2023，定义3.2]

2.3

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。

[来源：GB/T 42755-2023，定义3.6]

2.4

数据标注规则 annotation rule

生成式人工智能模型数据标注时所遵循的方法、要求的统称。

[来源：GB/T 45674-2025，定义3.8]

2.5

数据标注人员 annotator

承担数据标注任务的各类人员的统称。

[来源：GB/T 45674-2025，定义3.9]

2.6

数据生存周期 data lifecycle

将原始数据转化为可用于行动的知识的一组过程。

[来源：GB/T 36073-2018，定义3.14]

3 数据标注人才岗位方向及职责

数据标注人才岗位包括数据处理、技术支持、模型辅助、质量管控、安全合规、项目管理和培训支持7个方向，主要涉及以下13个岗位，具体如表1所示。

表 1 数据标注人才主要岗位及职责

序号	方向	岗位名称	岗位职责
1	数据处理	数据采集工程师	负责数据采集方案的设计与开发，实现网络采集、实地拍摄图像、视频和录制的语音等数据采集。
2		数据预处理工程师	负责对原始数据进行清洗、转换、采样等预处理操作，将数据格式转换为适合标注和模型训练的形式，提高数据的质量和可用性。
3		数据标注工程师	负责数据标注方案的执行，确保标注质量、数据质量与数量。
4	技术支持	数据标注规则制定师	负责制定标注规则和指南，跟进标注过程并及时调整完善规则，并为标注员和质检员提供培训与答疑。
5		数据标注工具开发工程师	负责设计和开发标注工具，并不断优化工具的功能和性能，提高标注效率和质量。
6	模型辅助	模型训练数据工程师	负责处理标注数据并输入模型训练，监控训练指标，通过调参、优化数据提升模型性能。
7		模型评估师	负责评估训练完成的大模型，分析其在不同场景的性能，找出优劣并提供改进建议。
8	质量管控	数据标注质量管理工程师	负责数据质检和验收、质检人员培训和考核、数据质量反馈和改善。
9	安全合规	数据安全工程师	负责制定与维护数据安全策略，保障数据全生命周期安全，并承担安全体系设计、策略执行、风险评估与应急响应。
10		数据确权合规工程师	负责明确数据全生命周期权属，保障安全合规，实施差异化访问控制，确保数据及处理可追溯、有法可依，动态追踪并审核访问合规性。
11	项目管理	数据标注项目经理	负责数据采集、标注、质检项目方案制定与执行、质量把控与项目交付管理。

12		数据管理工程师	负责项目数据资源的全生命周期管理，包括制定采集标准、选择存储方式、建立质量管控机制、构建目录标签体系，监控使用情况、评估应用价值，推动二次利用与价值最大化。
13	培训支持	数据标注培训师	负责培训方案的设计与实施，跟踪效果并优化方案，针对技能短板开展进阶培训，提升团队能力。

4 数据标注人才岗位能力要素

数据标注人才岗位能力要素包括专业知识、技术技能、工程实践三个维度。具体如表2所示。

表 2 岗位能力要素列表

维度	要素	说明
专业知识	基础知识	指相应岗位人才应掌握的通用知识，主要包括基本理论、相关标准与规范知识以及有关法律、法规、安全、隐私等。
	专业知识	指相应岗位人才完成工作任务所必备的知识，主要指与具体岗位要求相适应的理论知识、技术要求和操作规程等。
技术技能	基本技能	指相应岗位人才为完成工作任务所应具备的对基础知识应用的水平以及熟练程度。
	专业技能	指相应岗位人才为完成工作任务所应具备的对专业知识应用的水平以及对特殊工具使用的掌握。
工程实践	经验	指相应岗位人才在实际工程与项目推进中应当具备的经验。

5 数据标注产业人才岗位能力要求

5.1 数据处理方向

5.1.1 数据采集工程师

a) 专业知识

- 掌握《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等相关法律法规，明确数据采集的合法边界，确保采集行为符合合规要求；
- 熟悉数据采集过程中的安全防护技术，包括数据传输加密、采集设备安全管理、访问权限控制等，能有效防范数据泄露、篡改等安全风险；
- 熟练掌握数据采集的全流程操作规范与技术要点，涵盖网络爬虫的数据抓取逻辑、图像 / 视频拍摄的构图与光线控制、语音录制的降噪与收音技巧等，能精准把控不同类型数据采集的关键环节；
- 熟悉数据分类与特征提取方法，能根据数据的应用场景对原始数据进行科学分类，准确提取关键特征，为后续的数据整理和筛选奠定基础；
- 了解数据行业相关标准与前沿技术动态，如数据采集的行业最佳实践、新兴的数据获取技术（如API接口调用优化、传感器数据采集等），为工作优化提供方向；
- 了解敏感数据识别标准，在采集环节能精准识别个人信息、商业秘密等敏感数据，严格执行脱敏采集规范，避免违规获取或存储敏感信息。

b) 技术技能

- 能够独立搭建和维护网络爬虫系统，熟练使用Python等编程语言及Scrapy、BeautifulSoup等爬虫框架，可根据需求定制爬虫策略，应对反爬机制；
- 熟练操作各类数据采集设备，包括高清相机、摄像机、录音设备等，能根据采集场景调整设备参数，确保图像、视频、语音数据的采集质量；
- 能够运用数据处理工具对采集数据进行初步清洗与筛选，如使用Excel、SQL或Python 数据处理库（Pandas等），剔除重复、错误、冗余数据等，提升数据可用性。

c) 工程实践

- 具备丰富的实地数据采集项目经验，能制定合理的采集计划，协调团队完成大规模图像、视频或语音数据的采集工作，应对野外、复杂环境等特殊场景的采集挑战；
- 具备数据采集过程中的问题诊断与解决能力，在爬虫运行异常、设备故障、数据质量不达标等情况出现时，能快速定位原因并采取有效措施恢复采集工作；
- 具备数据合规性审查的实践能力，在采集前能对数据源的合法性进行评估，采集过程中做好数据脱敏和权限管理，采集后能出具合规性报告，确保数据全流程符合相关规定。

5.1.2 数据预处理工程师

a) 专业知识

- 掌握数据处理通用理论，如数据生命周期、数据特征基本概念；
- 熟悉数据领域标准规范，像数据质量评估国标、元数据管理规范；
- 精通数据相关法律法规，包括《中华人民共和国数据安全法》保障数据合规处理，《中华人民共和国个人信息保护法》规范隐私数据预处理；
- 具备数据预处理专属知识，涵盖各类数据源（结构化数据库、半结构化、非结构化文本/图像等）特性理论，不同预处理场景（清洗、转换、集成、规约）的技术要求（如缺失值填充算法选型、数据标准化规则），以及预处理操作规程（从数据探查、方案设计到执行验证的全流程），支撑精准、高效的预处理执行；
- 了解数据安全防护（加密、访问控制）与隐私保护（匿名化、去标识化）知识，为预处理全流程筑牢合规与基础认知。

b) 技术技能

- 熟练运用Excel、SQL、LaTeX等基础工具，开展简单数据探查（如统计缺失值、异常值）、基础清洗（去重、格式统一等）；
- 能够快速定位数据基础问题（如格式混乱、重复记录），并执行标准化预处理操作，保障数据可用性底线。

c) 工程实践

- 具备丰富数据预处理项目经验，参与过多个不同行业、不同规模数据预处理项目，熟悉各行业数据特点与预处理痛点，如金融对数据准确性、可追溯性要求高，医疗对隐私与专业标准严格等；
- 经历过项目需求对接、方案落地、问题复盘等全流程，能预判预处理风险（如数据丢失、过度处理），并依据经验制定应对策略（备份机制、试点验证）；
- 具备从需求洞察到流程优化的实战经验，保障预处理工作高效、适配业务，挖掘数据价值。

5.1.3 数据标注工程师

a) 专业知识

- 掌握数据标注领域通用理论，如数据标注流程、分类体系的基础原理；
- 掌握与数据标注岗位适配的理论知识和技术要求，比如图像标注中计算机视觉基础概念（图像特征、语义分割原理）、文本标注里的自然语言处理基础（词性、实体识别规则）、标注工具的算法逻辑对标注结果的影响等；
- 掌握操作规程，如标注质量核查流程、不同类型数据（图像、文本、音频等）标注的具体步骤；
- 熟悉数据标注相关标准规范，如智能辅助驾驶数据标注要符合交通行业标准等的标注格式、精度要求；
- 了解数据安全、隐私保护法规，如处理用户信息数据标注时，要遵循《中华人民共和国个人信息保护法》，保障数据主体权益。

b) 技术技能

- 熟练运用基础知识，如依据数据安全知识，规范处理含敏感信息的数据标注；精准应用通用理论，按照分类体系原理，高效完成数据分类标注，保证标注结果符合基础逻辑；
- 深度应用专业知识，如利用计算机视觉知识，高质量完成复杂图像（如医学影像多器官标注）标注；
- 熟练掌握特殊标注工具，如专业标注平台的高级功能（批量标注、智能辅助驾驶标注工具的运用），提升标注效率与精度。

c) 工程实践

- 具备多类型标注项目经验，如参与过大规模图像、文本数据标注项目等；
- 熟悉项目流程，具备从需求分析、任务分配，到质量管控、交付验收全流程经验；
- 能够面对项目中数据复杂、工期紧张等问题，有应对解决经验，如通过优化标注流程、引入辅助工具，保障项目按时高质量完成；
- 具备多种数据标注服务方式的经验：集中式（大型企业/机构自建团队集中完成）、分布式（通过众包、外包分配给多团队或个人）、混合式（结合前两种模式，按需求灵活组建团队优化效果）。

5.2 技术支持方向

5.2.1 数据标注规则制定师

a) 专业知识

- 掌握数据标注领域通用知识，包括数据标注基本理论，如标注流程框架、不同数据类型标注逻辑等，熟悉数据标注相关标准与规范，如图像标注的VOC标准、文本标注的实体识别标注规范等；
- 熟悉数据安全、隐私保护法规，能保障标注数据中涉及用户信息、商业机密等安全；
- 了解标注项目安全管理方法，如数据访问权限设置、标注环境安全防护等，为规则制定筑牢合规与基础理论根基；
- 具备数据标注规则制定专属知识，涵盖不同标注类型的理论知识，如图像语义理解、文本语法结构等；
- 掌握各类标注场景技术要求，如医疗影像标注的医学术语适配、智能辅助驾驶数据标注的精度标准等；
- 熟知标注规则制定的操作规程，涵盖需求调研、初稿撰写，到多轮验证、迭代优化全流程，以及如何平衡标注效率与质量、协调不同业务方（标注团队、需求方、审核方）对规则的诉求，确保规则科学合理、可落地执行。

b) 技术技能

- 熟练依据通用标准与规范，快速梳理标注需求核心要点；
- 熟练对简单标注场景，制定清晰、可执行的初步规则，识别规则制定中基础逻辑冲突并修正，保障规则基本可用性；
- 能够对复杂标注场景，运用专业工具进行深度需求拆解与规则预演；
- 熟练掌握特殊工具，对不同业务线、不同数据类型的标注规则进行协同设计，能快速解决规则制定中的疑难问题；
- 能够通过专业技能实现规则的精细化、体系化构建，支持大规模、高质量标注作业，推动标注流程高效运转。

c) 工程实践

- 具备丰富的数据标注项目规则制定经验，参与过多个不同行业、不同类型数据标注项目规则构建；
- 熟悉各行业标注业务痛点，能依据过往经验预判规则制定风险，提前规划应对策略；
- 具备复杂规则迭代场景的经历，沉淀出从规则初建、验证到持续优化的全流程实战经验，保障规则始终适配业务需求，为数据标注质量与效率提供长效支撑。

5.2.2 数据标注工具开发工程师

a) 专业知识

- 掌握数据标注领域通用知识，包括数据标注基本理论，熟悉数据标注相关标准与规范，熟悉数据安全、隐私保护法规，能保障工具开发中数据处理合规，了解软件开发基础安全知识，为工具开发筑牢合规与业务认知基础；
- 掌握数据标注工具开发的基础理论（含前后端架构、数据库原理等），熟知不同标注类型（如图像、文本）的工具功能适配要点，遵循需求分析、架构设计、编码测试到部署迭代的开发流程规范，理解数据标注业务需求与工具功能的融合逻辑，能支撑工具从需求到落地的全流程开发工作。

b) 技术技能

- 掌握从前端到后端的全栈技术（具体包括Vue、Python、MySQL等基础技术栈），并具备将AI算法与业务场景（如智能辅助驾驶、医疗）融合的能力；
- 能够依据通用理论与标准，快速规划工具基础技术框架，运用基础开发工具搭建工具雏形，通过识别工具开发中基础问题并修正，保障工具基本可用，满足简单标注场景需求；
- 熟练使用专业开发工具对复杂数据标注场景进行深度功能开发；
- 掌握特殊技术，如AI辅助标注算法集成、标注流程自动化脚本编写等，实现工具智能功能，如自动预标注、标注质量实时校验；
- 能够对工具性能进行优化，解决工具开发中的疑难问题，能够打造适配多场景、高性能的标注工具。

c) 工程实践

- 具备丰富的数据标注工具开发与项目实践经验，参与过数据标注工具从需求到上线的完整项目，覆盖不同标注类型；
- 熟悉多类数据标注业务场景（医疗影像标注、智能辅助驾驶数据标注），熟悉各场景对工具的特殊需求。

5.3 模型辅助方向

5.3.1 模型训练数据工程师

a) 专业知识

- 掌握机器学习与深度学习的基本理论，如监督学习、无监督学习、神经网络的基础概念；
- 熟悉数据相关的标准与规范，包括数据格式标准（CSV、JSON等）、数据标注规范；
- 掌握数据清洗、转换和预处理的技术，包括处理缺失值、异常值，进行数据标准化、归一化等操作；
- 熟悉特征工程相关知识，如特征提取、特征选择、特征编码等，能够根据模型需求构建有效的特征集；
- 了解数据安全与隐私保护的法律法规，如《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等，确保在模型训练数据处理过程中，数据的存储、使用、传输等环节合规，避免数据泄露和违规使用；
- 了解模型训练过程中数据的作用机制，清楚不同类型的模型（分类、回归、聚类等）对数据的要求，以及如何评估数据对模型性能的影响。

b) 技术技能

- 能够熟练使用SQL语言进行数据库操作，如查询、插入、更新、删除数据，以获取所需的训练数据；
- 能够运用基础的数据分析工具进行数据的初步探索性分析，了解数据的基本统计特征（均值、中位数、标准差等）；
- 能够使用Python等编程语言进行简单的数据处理脚本编写，例如读取和写入数据文件、数据格式转换等；
- 熟练掌握Python中的数据处理库，如Pandas、Numpy，进行大规模数据的高效处理；
- 能够运用Scikit-learn等机器学习库进行数据预处理、特征工程以及简单模型的训练和评估；对于深度学习，熟悉TensorFlow、PyTorch等框架，能够构建和训练神经网络模型，并根据模型训练结果对数据进行针对性优化；
- 具备使用数据可视化工具的能力，如Matplotlib、Seaborn等，将数据的分布、特征关系等进行可视化展示，辅助分析数据质量和模型训练效果。

c) 工程实践

- 参与过多个实际的模型训练项目，涵盖不同领域（如金融风控、医疗诊断、图像识别等），了解不同领域数据的特点和处理难点；
- 在项目中负责过数据采集、清洗、预处理以及特征工程等工作，积累了从原始数据到可用训练数据的全流程经验；
- 遇到并解决过数据质量差、数据不平衡、特征冗余等实际问题，能够根据项目需求和数据情况，制定合理的数据处理策略和应急方案；
- 具备与算法工程师、数据科学家等团队成员协作的经验，能够准确理解模型对数据的需求，并提供高质量的训练数据，保障模型训练的顺利进行。

5.3.2 模型评估师

a) 专业知识

- 掌握机器学习与深度学习基础理论知识，如随机森林、支持向量机、卷积神经网络、循环神经网络等算法；熟悉如大语言模型、多模态模型等模型的原理与内部机制；

- 熟悉模型评估通用指标体系，可遵循<场景-能力-任务-指标>四层结构。如在通用场景下对于其理解能力的阅读理解任务，则可以采用而阅读理解可以利用准确率、F1 Scores、BLUE、ROUGE等指标进行评估和考察；
 - 熟悉模型评估相关行业规范（如金融领域模型风险评估标准、医疗AI模型验证指南）及法律法规（《人工智能伦理治理规范》《生成式人工智能服务管理暂行办法》），确保评估过程合规且符合伦理要求；
 - 熟悉不同类型数据集对模型评估时的影响，如对于大语言模型的通用能力可采用MMLU、BBH、GSM8K等数据集评测，对于多模态模型的通用能力MMBench、LLaVA-Bench、POPEO、CRBench等评测数据集更加适用；
 - 熟悉大语言模型的基础能力与推理能力，以及多模态模型在代码能力、理解能力、生成能力等相关维度的专业知识。
- b) 技术技能
- 熟练掌握至少一种编程语言，如Python，并能够使用Python等编程语言进行简单的数据处理脚本编写，例如读取和写入数据文件、数据格式转换等；
 - 熟练掌握并使用核心科学计算库（如NumPy、Pandas）和深度学习框架（如PyTorch、TensorFlow）等；
 - 熟练掌握并使用主流模型评估库（如Hugging Face、Evaluate、Scikit-learn、OpenCompass、Robustness Metrics）以及可视化工具；
 - 掌握数据处理和建模相关的工具，如Hadoop、Spark（用于大数据处理），Jupyter Notebook（用于代码编写和文档记录）等。此外，还需熟悉办公软件，用于数据整理和报告撰写。
- c) 工程实践
- 在项目中主导或深度参与过从模型评估需求分析、评估环境准备、评估数据集构建、评估流程执行、评估结果显示等全流程的评估环节，熟悉各流程要点；
 - 具备参与一个或多个实际的模型评估项目，涵盖不同领域（如金融风控、医疗诊断、图像识别等），了解不同领域数据的特点和处理难点；
 - 在模型评估过程中，能基于评估结果深度诊断模型在特定维度的缺陷，制定有效优化建议，并通过模型重训与复评估验证方案效果，切实提升模型在该维度的性能及整体鲁棒性。

5.4 质量管控方向

5.4.1 数据标注质量管理工程师

- a) 专业知识
- 掌握数据标注领域通用理论，如数据标注流程、分类体系的基础原理；
 - 熟悉数据标注相关标准与规范，如图像标注的COCO标准、文本标注的BIESO标注规范等；
 - 熟悉数据安全、隐私保护法规，能保障标注数据中涉及个人信息、敏感信息等不泄露；
 - 掌握不同标注类型的技术要求，如图像语义分割标注的精度标准、文本情感分类标注的标签定义规则等；
 - 熟悉标注操作规程及质量管控流程，从数据标注人员培训、任务分配，到标注过程抽检、最终验收全流程质量把控要点；
 - 具备数据标注质量管理专属知识，涵盖标注质量评估理论，如准确率、召回率、F1 值等质量指标计算与分析；
 - 了解标注项目中的安全管理，如数据存储安全、传输安全防范等。
- b) 技术技能

- 熟练运用基础知识开展工作，如依据标注标准与规范，快速识别标注数据中的基础问题，如格式错误、标签明显不合理等；
- 借助通用办公软件对标注质量基础数据进行统计分析，通过基础技能及时发现标注流程中的漏洞并提出初步优化建议；
- 能够精准应用专业知识，针对复杂标注场景，如多模态数据标注质量校验，运用专业工具，如质量评估系统、标注审核平台等，进行深度质量检测；
- 熟练掌握特殊工具，如 AI 辅助质量核查算法、标注差异对比软件，对大规模标注数据进行自动化+人工复核的质量管控，能快速定位标注争议点，并制定合理的质量仲裁与修正方案。

c) 工程实践

- 具备丰富的数据标注项目质量管理经验，参与过不同类型数据标注项目全流程质量管控；
- 熟悉项目推进各阶段质量风险，如项目初期数据标注人员技能不达标、项目中期标注任务量激增导致质量下滑，能依据过往经验提前规划质量保障措施，如制定针对性培训计划、动态调整抽检比例；
- 经历过复杂质量问题解决，如标注质量纠纷处理、因业务需求变更导致的标注标准迭代后的质量适配，沉淀出高效的质量问题复盘与改进经验，保障标注项目高质量交付。

5.5 安全合规方向

5.5.1 数据安全工程师

a) 专业知识

- 熟悉计算机网络知识、计算机基础操作知识、计算机常用应用软件的安装及使用方法、数据库知识、Linux基本操作知识；
- 熟悉计算机网络架构、基础操作体系，掌握常用应用软件（如办公、安全检测类）的安装与应用；
- 精通数据库知识，涵盖Linux基本操作，熟悉关系型数据库概念、SQL基础，以及Oracle、MySQL、NoSql等数据库基础原理；
- 掌握数据安全管理体系构建逻辑，熟练运用数据管理宏观体系及关键职能操作方法（如数据资产梳理、安全策略制定）；
- 具备独立开展数据防护设计、安全运维执行、风险主动排查、安全运营管控的知识储备，能将理论转化为实际安全管理动作；
- 了解国家数据安全法规（如《中华人民共和国数据安全法》《网络安全法》《关键信息基础设施安全保护条例》）、信息安全防护体系、安全保密规范，以及云计算、大数据基础概念。

b) 技术技能

- 熟练操作并高效使用云产品（虚拟主机、云数据库、云存储、云网络等），熟练操作使用典型大数据平台；
- 精通主流操作系统，实现安全配置、故障排查与性能优化。

c) 工程实践

- 能够通过实施数据分类分级、访问控制、加密等技术手段，确保数据在存储、传输过程中的保密性、完整性和可用性，降低数据泄露、篡改等安全风险，保障核心数据资产安全；
- 能够通过安全审计和风险评估，提供数据安全保障。

5.5.2 数据确权合规工程师

a) 专业知识

- 熟悉相关法律法规，包含《中华人民共和国民法典》《中华人民共和国数据安全法》《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》及商法体系（含公司法、合伙企业法、破产法等商事单行法），在处理数据确权及合规问题时，能精准依据法律条文作出专业判断并实施规范操作；
- 掌握法律专业知识，中级职称以上需有律师执业资格证书，高级职称则需要透彻理解国际规则和国内法规，能独立编写高质量确权合规报告；
- 熟练掌握数字经济产业逻辑、数据管理全流程知识（含数据采集、加工、流通），理解数据资产化路径中确权合规关键节点；
- 了解信息技术、数据传输技术、数据存储技术、数据加密技术及信息管理体系。

b) 技术技能

- 具备快速学习能力，适配法规、技术迭代；
- 拥有较强语言沟通及文字表达能力，精准传递合规要求；
- 熟练用Office等工具，高效输出合规文档；
- 构建“法律框架认知+技术实现理解+管理协调推进”三维能力，精准界定不同业务场景（如数据交易、跨境传输）确权合规边界；
- 熟悉数据确权流程性工具（如合规评审模板、确权协议框架）；
- 具备数据资产法律属性梳理能力，快速识别确权环节风险点；
- 掌握合规方案落地路径，协同技术、业务团队推进确权措施执行。

c) 工程实践

- 具备识别并评估组织在数据管理和技术措施方面的法律风险的能力，对资源采取法律保护措施；
- 能够为数据资产的数据资源持有权、加工使用权、产品经营权等环节中出具法律意见书，支撑数据资产入表的合规评估；
- 能够从数据全生命周期出发，规避法律风险，制定控制措施，预防潜在法律责任。

5.6 项目管理方向

5.6.1 数据标注项目经理

a) 专业知识

- 掌握数据标注行业基本理论，如数据标注流程、分类体系等；
- 熟悉数据标注相关标准与规范，像数据质量标准、标注精度要求等；
- 了解数据安全、隐私保护法规，如《中华人民共和国个人信息保护法》，以及项目管理通用法规、安全生产等知识；
- 精通数据标注项目管理理论，包括项目规划、进度控制、成本管理、人员管理等；
- 熟悉数据标注技术要求，如不同标注类型的标注规则、算法适配知识等；
- 掌握数据标注项目操作规程，从项目启动、任务分配到验收全流程规范。

b) 技术技能

- 熟练运用项目管理基础知识，如用甘特图等工具规划项目进度、用WBS分解工作任务等；
- 熟练操作基础办公软件及项目管理工具基础功能，实现项目文档管理、沟通协作；
- 精准应用数据标注专业知识，把控标注质量，识别标注偏差并指导修正；

- 能够制定清晰的标注规则（含边界、逻辑及示例），并能根据项目中发现的漏洞快速组织评审修订，避免批量偏差；
- 熟练掌握数据标注项目管理工具，进行任务分发、进度追踪、质量抽检等操作；
- 能够基于客户需求和资源情况，设计包含需求拆解、资源匹配、进度排期、风险预判的可执行方案。

c) 工程实践

- 具备多类型数据标注项目全生命周期管理经验；
- 了解不同行业数据标注需求差异；
- 具备应对项目突发状况，如数据标注人员流失、数据质量争议等的处理经验，能快速调整保障项目推进。

5.6.2 数据管理工程师

a) 专业知识

- 掌握数据管理通用知识，包括数据生存周期管理基本理论，如数据采集、存储、使用、销毁等阶段逻辑；
- 熟悉数据管理相关标准与规范，如数据分类分级国标、存储安全规范等，精通数据领域法律法规，保障数据资产合规管理；
- 了解数据安全与隐私保护知识，确保数据管理全流程合法合规、安全可控；
- 具备数据生存周期管理专属知识，涵盖数据采集标准制定理论，如明确不同数据源采集规则、格式要求等；
- 掌握存储方式选型知识，如关系型与非关系型数据库适用场景、分布式存储架构等；
- 熟悉质量管控机制构建方法，如质量校验规则设计、质量问题追溯流程等；
- 精通目录标签体系搭建逻辑（元数据管理、标签分类与关联），以及数据使用监控（使用权限、行为审计）、价值评估（ROI分析、业务价值映射）和流程优化（PDCA循环在数据管理中的应用）等专业理论、技术要求与操作规程，支撑全流程数据管理。

b) 技术技能

- 熟练运用基础知识开展工作，如依据通用理论与标准，快速规划数据采集标准框架，运用基础工具（Excel、SQL基础语句）对数据存储、质量基础信息进行梳理分析；
- 能够完成数据管理流程初步搭建，识别数据采集、存储阶段基础问题（如格式不统一、存储冗余）并提出简易优化方案，保障数据基础管理有序；
- 熟练使用专业工具，如数据治理平台、元数据管理系统等，进行数据生存周期管理；
- 能够构建目录标签体系，利用工具完成元数据关联与标签自动化打标；
- 能够依据《中华人民共和国数据安全法》《数据管理能力成熟度评估模型》（GB/T 36073-2018）等法规与标准，遵循科学、系统、规范、明确、时效、动态原则，按数据敏感程度、影响范围、业务属性及风险类型等完成分级分类标签标注，采用ABC等管理方法定期审核数据级别，并结合应用场景、法规政策及数据有效性变化动态调整更新，确保数据管理合规适配；
- 能够对数据使用情况进行深度审计（如分析高频使用场景、异常访问），开展价值评估建模（如构建数据价值评估指标体系），推动数据二次利用（数据挖掘、模型训练）；
- 能够优化管理流程，提升数据运营效率。

c) 工程实践

- 具备数据生存周期管理项目经验，参与过多个完整数据管理项目；

- 熟悉各阶段项目推进风险（如采集数据源冲突、存储扩容延迟等），能依据经验提前制定应对预案（备用数据源、存储弹性扩容方案）；
- 经历过复杂数据管理场景（多源异构数据整合、高并发数据访问），沉淀出质量管控机制落地、目录标签体系迭代、价值评估优化等实战经验，保障数据生存周期管理高效、优质，持续推动数据价值最大化与管理流程升级。

5.7 培训支持方向

5.7.1 数据标注培训师

a) 专业知识

- 掌握数据标注领域通用知识，包括数据标注基本理论，如标注流程、各类标注任务原理等；
- 熟悉数据标注相关标准与规范，如图像标注的COCO、文本标注的BIESO等规范；
- 熟悉数据安全、隐私保护法规，保障培训中数据合规使用；
- 了解数据安全（加密、访问控制）与隐私保护知识（数据脱敏、匿名化），为培训内容奠定合规与基础认知根基；
- 具备数据标注培训专属知识，涵盖不同标注类型的深度理论知识，如图像分割算法基础、文本语义理解模型等；
- 掌握各类标注任务技术要求，熟知标注操作规程及质量管控要点、培训教学理论、培训课程设计逻辑，确保培训内容专业且适配教学规律。

b) 技术技能

- 能够依据通用标准与规范，快速梳理培训核心知识点，制作培训课件，对简单标注任务，能够设计清晰教学流程，解答学员基础疑问，识别培训内容基础漏洞并优化，保障基础培训有效开展；
- 能够针对复杂标注场景，如多模态数据标注、高精度工业标注等，运用专业工具，如标注模拟训练平台、AI辅助教学系统等，设计沉浸式培训环节；
- 熟练掌握特殊工具，如标注错误案例分析软件、技能水平评估系统等，对学员标注技能进行精准测评，依据测评结果定制个性化提升方案；
- 能够通过专业技能，将复杂标注理论，如深度学习辅助标注原理，转化为通俗教学内容，解决培训中重难点知识传递难题，助力学员快速掌握高阶标注技能，支撑高质量标注项目。

c) 工程实践

- 具备丰富的数据标注项目培训经验，参与过多个不同行业标注项目，熟悉各行业标注业务痛点，如医疗对专业医学知识依赖、智能辅助驾驶对安全精度要求等；
- 开展过多场数据标注培训，覆盖不同层级学员，能依据项目实践经验，将真实标注案例融入培训，预判培训中学员常见问题，提前设计解决方案；
- 经历过培训体系从0到1搭建、培训效果持续优化等过程，沉淀出培训需求洞察、课程迭代、效果评估全流程经验，保障培训切实提升团队标注能力，为项目输出高质量标注成果。

附 录 A
(资料性附录)
数据标注人才岗位能力提升

A.1 数据标注人才岗位能力提升内容

岗位能力提升内容应包括：

- a) 基础知识、专业知识等相关知识提升；
- b) 基本技能、专业技能等相关技术技能提升；
- c) 基于项目经验的工程实践能力提升。

A.2 数据标注人才岗位能力提升阶段和方式

数据标注人才岗位能力提升分为岗前提升和在岗提升两个阶段，构成数据标注相关岗位从业人员不同阶段和能力水平的终身教育体系。

- a) 岗前提升方式，包括：
 - 1) 理论教学；
 - 2) 理论与实践一体化教学；
 - 3) 项目实训、企业实习等方式。
- b) 在岗提升方式，包括：
 - 1) 内部在岗培训；
 - 2) 外部脱岗培训；
 - 3) 项目实践或导师辅导等。

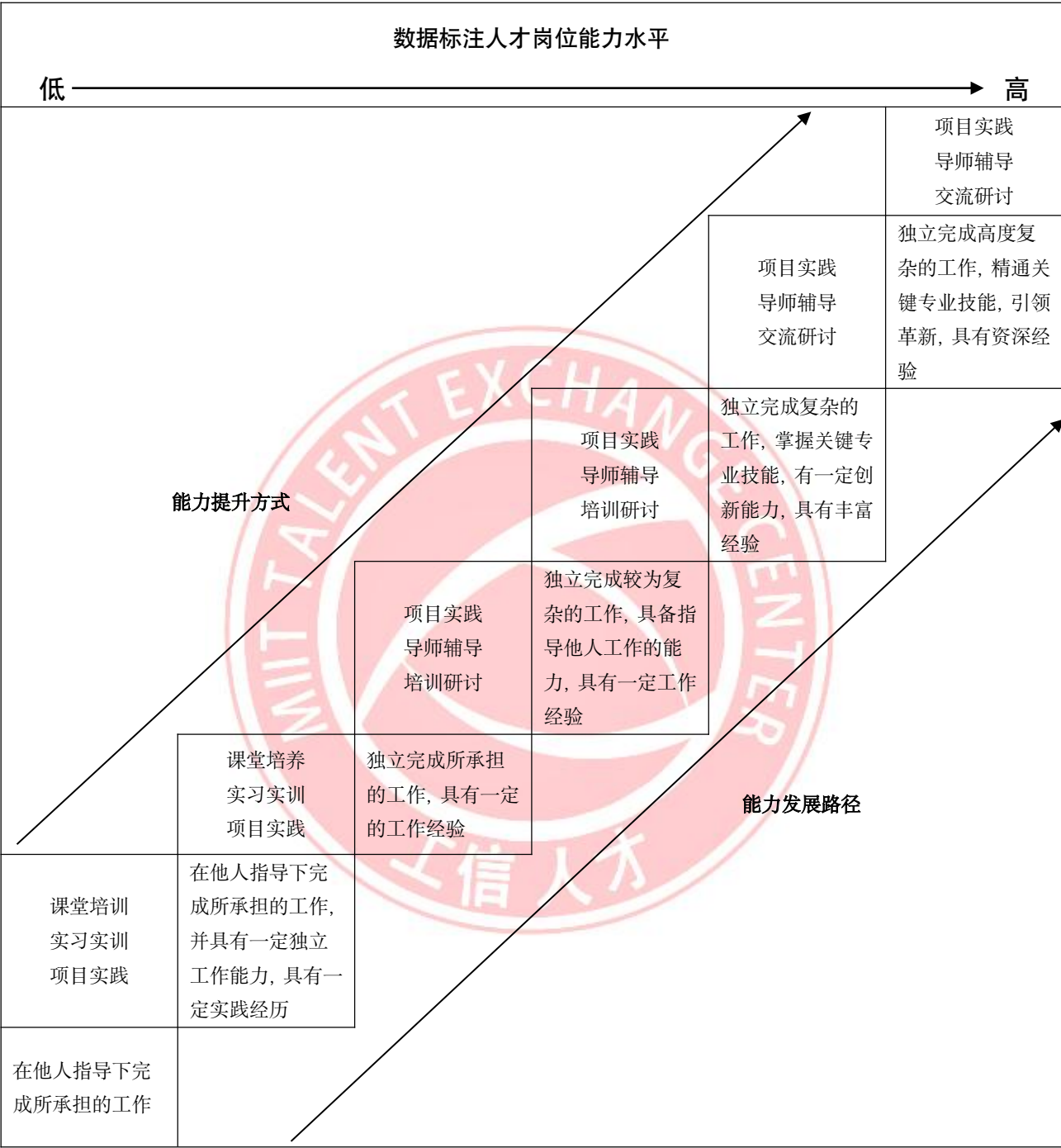
A.3 数据标注人才岗位能力提升活动供给类别

数据标注人才岗位能力提升活动供给包括：

- a) 教育、培训机构培养：符合要求的各级教育机构（普通高校、中等和高等职业院校等）及培训机构应根据数据标注领域各岗位能力要求，制定人才能力提升方案，为数据标注领域及企业培养合格的从业人员，满足个人发展需要；
- b) 企业培养：企业结合业务发展需要，应根据数据标注领域各岗位能力要求有针对性、有计划的实施岗位能力提升计划，满足个人发展需要，增强企业竞争力；
- c) 个人培养：从业人员根据个人发展计划，做好职业规划与岗位定位，对标数据标注产业人才岗位能力要求，不断提升专业知识、技术技能水平，丰富工程实践经验。

数据标注人才岗位能力提升路径见图A.1。

图 A.1 数据标注人才岗位能力提升路径



附 录 B
(资料性附录)
数据标注人才岗位能力评价

B.1 数据标注人才岗位能力评价方法

对从业人员进行评价和定级，评价结果可以作为数据标注人才能力胜任、职业发展等活动的依据。评价方式包括：

- a) 专业知识主要通过笔试考核的方式进行评价；
- b) 技术技能主要通过实验考核方式进行评价；
- c) 工程实践主要通过成果评价方式进行评价。

B.2 数据标注人才岗位能力评价等级

数据标注人才岗位能力评价等级可以分为初、中、高级三级，能力分为9等。

- a) 初级（1—3级）：在他人指导下完成所承担的工作，并具有一定独立工作能力，具有一定实践经历；
- b) 中级（4—6级）：独立完成较为复杂的工作，具备指导他人工作的能力，具有3年及以上工作经验；
- c) 高级（7—9级）：独立完成高度复杂的工作，精通关键专业技能，引领革新，具有5年及以上工作经验。

B.3 数据标注人才岗位能力等级评价权重

数据标注人才岗位能力等级评价权重表如下：

B.1 数据标注人才岗位能力等级评价权重表

评价维度		专业知识	技术技能	工程实践
岗位等级		评价分值权重		
高级	9 级	20%	30%	50%
	8 级			
	7 级			
中级	6 级	50%	25%	25%
	5 级			
	4 级			
初级	3 级	70%	25%	5%
	2 级			
	1 级			
备注		评价总分满分为 100 分，由专业知识、技术技能、工程实践三项评价维度的权重总分所得。		

参 考 文 献

- [1] GB/T 42129-2022 数据管理能力成熟度评估方法
- [2] GB/T 36073-2018 数据管理能力成熟度评估模型
- [3] GB/T 40685-2021 信息技术服务 数据资产 管理要求
- [4] GB/T 37550-2019 电子商务数据资产评价指标体系
- [5] GB/T 35295-2017 信息技术 大数据 术语
- [6] GB/T 42755-2023 人工智能 面向机器学习的数据标注规程
- [7] GB/T 45674-2025 网络安全技术 生成式人工智能数据标注安全规范

