

# 大数据与决策研究

## (政策与技术跟踪专题)

2025 年第 10 期 (总第 307 期)

广西壮族自治区信息中心

广西壮族自治区大数据研究院

2025 年 4 月 21 日

### 数据标注产业驱动人工智能发展实践及 对策建议

数据标注是对原始数据进行采集、清洗、分类、标记、质量检验等专业数据治理活动，能有效提升数据供给质量，是人工智能发展的关键环节。随着人工智能技术的快速发展，数据标注行业呈现出高速增长的态势，对高质量标注数据的需求急剧增加，推动了数据标注市场规模的不断扩大。本文通过梳理国内数据标注产业发展的主要做法，提出我区发展数据标注产业的对策建议。

## 一、数据标注：人工智能发展的核心基石之一

### （一）数据标注技术类型及行业特点

按照人工参与标注的程度可以将数据标注技术类型分为三种类型：人工标注、自动化标注、人机协同标注。人工标注的准确性高、灵活性强，但其人力资源成本较高、主要适用于对标注质量要求高且数据量相对较小的场景。自动化标注的优势在于作业效率高、成本低，但其准确性受限，适用于对标注速度要求极高，且对标注准确性有一定容忍度的场景。人机协同标注是通过机器学习或自然语言处理等技术对数据进行初步标注，再由人工进行审核和修正，主要优势在于效率高、准确性较高，且成本适中，主要应用于自然语言处理、计算机视觉等领域。

表 数据标注技术类型的对比情况

| 技术类型   | 定义                                 | 优点        | 局限    | 适用场景                                      |
|--------|------------------------------------|-----------|-------|-------------------------------------------|
| 人工标注   | 由标注人员或专业人员对数据进行逐一标注                | 准确性高、灵活性强 | 成本较高  | 对标注质量要求高且数据量相对较小，如医学影像分析、金融风控等            |
| 自动化标注  | 利用算法和模型对数据进行自动标注                   | 效率高、成本低   | 准确性受限 | 对标注速度要求极高且对标注准确性有一定容忍度，如互联网内容审核、社交媒体数据分析等 |
| 人机协同标注 | 由自动标注系统对数据进行初步标注，然后由人工对标注结果进行审核和修正 | 效率高、准确性较高 | —     | 自然语言处理、计算机视觉等领域                           |

数据标注行业有专业性、定制化、技术密集化特点。数据标注要求从业者具备一定的行业知识和技术背景，需要准确理解标注对象，熟悉标注工具、标注规范和数据处理流程等，并具备将标注对象特征准确描述出来的能力。同时，数据标注需求因行业、业务场景的不同而存在一定的差异性，需要满足客户的个性化需求，提供定制化的解决方案。另外，传统的数据标注工作高度依赖人力，且效率低下、成本高昂，企业采用人工智能辅助标注和自动标注技术来提升效率和质量，减少人工干预的程度，提高标注的准确性和一致性，同时在标注模型工具的迭代中不断完善的人工智能基础设施，助推行业向技术密集转型。

## （二）数据标注行业发展趋势

一是行业人工智能应用需求向“好数据”转变，对数据标注的要求也日益提高，标注任务的复杂度显著提升。随着人工智能模型能力的不断增强，以自动驾驶领域为例，早期的自动驾驶模型仅需对简单的道路场景进行标注，如识别车辆、行人等基本目标。如今，为了实现更高级别的自动驾驶功能，需要对复杂路况、恶劣天气以及特殊交通状况等复杂场景的设计。二是大模型在垂直领域的应用不断深化使得高质量数据的供给更依赖高素质的专业人员。如在医疗影像标注方面，随着医学技术的发展，对疾病的诊断和治疗需要更加精准的影像分析。标注人员不仅要准确标注出病变的位置、大小和形状，还需要具备专业的医学知识，能够对病变

的性质、发展阶段等进行判断和标注。对数据标注行业从业人员的能力提出更高要求，不仅需要掌握标注工具的使用方法，还需要深入了解相关领域的专业知识和复杂的标注规则。三是数据标注产业的快速发展产生大量数据安全技术需求，带动相关行业发展。数据标注从数据采集、标注、存储到传输的每一个环节，都存在数据安全风险。如部分数据采集方可能存在非法采集数据的行为，未经授权获取用户的个人信息、敏感数据等，可能引发数据安全事故。一些众包、转包模式下的标注数据可能会缺乏安全性，对标注工具平台的安全需求将大大提升，以寻求降低被黑客攻击导致数据泄露几率。

## 二、当前国内按“427”布局促进数据标注产业发展

我国按照“427”布局促进数据标注产业发展，“4”是指技术创新、生态培育、标准规范、人才引育，“2”是指公共数据标注需求和企业数据标注需求，“7”是指在国内布局成都、沈阳、合肥、长沙、海口、保定、大同七个数据标注基地建设。国内在提升四类供给，释放两类需求，做大做强做优数据标注基地的主要做法如下。

### （一）提升四类供给

在技术创新方面，沈阳在国内首次发布《数据标注技术创新指导性意见》，梳理了标注技术的标注方式、标注模态、标注平台等内容，为技术研发与应用提供梯度指引，创新性地

将数据标注技术划分为国际创新、国内领先、国内先进 3

个层次，从工业制造、现代农业等 9 个方向提出意见加速行业赋能，以“揭榜挂帅”形式支持建成人工智能医学图像分析系统。大同邀请国内头部数据标注平台类企业入驻，吸引标注产业生态合作伙伴以线上、线下两种模式入驻，利用算力资源共同研发集数据预处理、标注、合成、评测等功能为一体的智能数据标注生成管理平台。在生态培育方面，保定锚定“服务首都、对接京津、联动雄安”定位，深度融入京津冀协同发展战略，全力推动京数保标、京模冀用，打造环京津数据产业带和京津冀数智协同创新高地。成都发布数据标注产业图谱，全面呈现了成都市数据标注产业的生态体系与发展潜力，推进本地产业协同创新；打造数据标注公共服务平台，建设标注基础通用工具，满足不同行业对于数据标注及后续处理需求。长沙、成都、西宁均成立数据标注产业发展联盟，搭建政企协同发展高效平台，围绕“数据标注+”产业链，打造数据产业集群。在标准规范方面，数据标注领域的规范化进程呈现出明显的立体化特征。国家层面，《人工智能 面向机器学习的数据标注规程》作为首个机器学习数据标注的通用国标，为行业锚定了基础框架；《人工智能 医疗器械 质量要求和评价 第 3 部分：数据标注通用要求》与《烟草行业地理信息系统基础数据的标注与交换规范》行业标准则深入医疗、烟草垂直领域，针对器械质量评估、地理信息标注等场景提出专项要求，凸显标准化需适配行业特性的趋势。地方层面，黑龙江、山西等地近两年连续发布涵

盖框架设计、技术细则以及人才培训的地方标准，反映出区域产业化布局中数据标注作为基础设施的战略地位。在人才引育方面，青海印发《关于加快推进全省数据标注产业人才队伍建设的若干措施》，从数据标注人才的培养、使用、评价、激励等方面，出台 20 条措施，促进数据标注产业健康发展。长沙通过举办“星城杯”人工智能训练师职业技能大赛吸引企业参与，引导青年人才投身于数据标注产业；依托湖南 5G 应用生态产业园建立长沙市 AI 数据应用人才公共实训基地，促进长沙十余所高校与标注企业深度融合，共同培养实用型人才；设立 1 千万元的数据人才职业技能培训专项补贴，鼓励和支持人才投身数据标注产业。大同聚力建设“算力之城”，发展数据呼叫、数据标注、数据审核等数据服务业态，依托大同数据科技职业学院等院校资源，发挥大数据智慧服务产教协会作用，加快建设实训基地，全面构建“产业龙头+产教融合+订单制输出”的人才培养模式，大力培养数字经济实用型人才队伍。

### （二）释放两类需求

湖南举办 2025 年全省数据标注供需对接大会，发布总值达 4.6 亿元的数据标注及数据集订单、共计 10200 人次的数据人才需求，搭建数据标注产业供需双方深入对接、交流合作的平台。长沙向社会各界征集数据标注基地建设“合伙人”，通过“政、企、校”多方协作、共同建设长沙市数据标注基地，促进供需对接、数字科技研发转化合作等，助推

数据标注、人工智能等数据产业高质量发展。保定聚焦七大主导产业，培育形成“7+20+N”现代化产业体系，全面推动人工智能+产业，发布了《保定市推进数据智能产业创新发展措施》，全面推动AI赋能产业行动，为数据标注产业带来海量业务需求。

### （三）布局七个数据标注基地建设

成都打造数据标注能力、应用、生态、支撑、安全“五大体系”，推行“数据+场景+资本+供应链”协同的招商新模式，构建以平台企业为引领、专业企业为骨干、创新企业为发展的数据标注企业集群。沈阳聚焦数据标注技术创新，布局跨模态融合的多模态数据增强等技术攻关，建设数据标注技术创新平台，打造行业高质量数据集。合肥集中突破数据预处理、智能标注工具等关键技术；建设公共服务平台和众智标注平台；培育龙头企业与中小企业联动发展；推动社会治理数字化转型，构建标准化体系和人才培育机制；创新“四会一赛”生态运营模式，通过“政产学研用金”协同推进机制，加速形成高效智能的产业生态。长沙建设全球高质量音视频文创数据特色开发基地，推动视频数据合成技术达到国际先进水平。海口打造以“AI数据国际化”和“卫星遥感测绘”为特色的数据标注示范基地，利用其区位优势，提供“来数加工”等增值服务，形成国际化数据服务能力，参与国际化的数据要素市场。保定打造“一核联动双园”空间布局，构建了一套行业可信数据空间、一套数据标注技术标准、一

套标注行业生态的“三个一”能力体系，全力推动数据标注基地建设提速奋进。大同以“数据标注+算力中心”为两大基石，以高效率数据处理平台为技术手段，以建设高质量数据集和促进数据流通为目标牵引，以产教融合及职业院校为人才供给，以软硬件一体的安全环境为安全保障，形成以龙头企业为引领，中小企业蓬勃发展的数据标注产业新生态。

### 三、广西发展数据标注产业建议

借鉴国内发展数据标注产业做法，我区可以从四类供给和两类需求着力发展数据标注产业。一是提升创新支撑。支持建设数据标注创新平台，整合数据、模型、工具等资源，推动技术融合应用；依托广西科技计划专项、重点课题研究等，聚焦自动驾驶、医疗健康、金融等领域开展多模态标注、大模型预标注、自动化标注等关键技术攻关，加速智能化工具研发。二是构建产业生态。构建以技术研发为核心、数据标注应用为驱动的数据标注产业生态，推动数据标注技术从研发到应用的全链条融合。建立数据标注产业发展行业组织，搭建产业发展交流平台。三是建立标准规范。推动标注技术标准化，实现跨技术、跨平台标注产品互认互通。对标全国数标委数据集相关标准，完善数据集质量评测体系，结合市场需求和行业特色，因地制宜建设汽车、机械、冶金等行业高质量数据集。四是培育标注人才。鼓励高校和职业院校强化数据标注相关学科建设，推动产教融合、校企合作，培育高质量数据标注人才。支持龙头企业、高校、培训机构

等深度合作，推动人才培养供给侧与需求侧精准对接。推进数据标注人才队伍梯度建设，畅通人才发展通道，推动数据标注领域职业资格与职业技能等级衔接互认。五是释放标注需求。依托广西公共数据流通交易基础设施，依法依规有序推动公共数据标注与开发利用；构建我区数据标注公共服务体系，提升产业资源汇聚、供需对接能力；组织我区标注企业参与全国性的数据标注供需交流活动，充分融入全国一体化数据市场。

（执笔人：蔡亮亮）

编辑部地址:南宁市体强路 18 号广西信息中心 1412 号房

联系电话: 0771-6113592

电子邮箱: dsjyjs@gxi.gov.cn

网 址: <http://gxxxzx.gxzf.gov.cn/>



扫描二维码获取  
更多决策参考信息