大数据与决策研究

2025 年第 11 期（总第 308 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 26 日

做强数据标注产业夯实我区人工智能发展基础对策建议研究

数据标注是连接数据资源、算法模型与实际应用场景的关键桥梁，是挖掘数据要素价值的关键环节，是人工智能高质量数据集的核心生产力[1]。数据标注产业是对数据进行筛选、清洗、分类、注释、标记和质量检验等加工处理的新兴产业，对人工智能的发展具有重要支撑作用。当前，我区正在贯彻落实“人工智能+”行动，亟需培育壮大数据标注产业支撑人工智能创新发展。

[1] 魏亮:《繁荣数据标注产业，赋能人工智能高质量发展》

一、当前数据标注产业发展态势

（一）数据标注产业发展成为人工智能创新发展的关键

数据标注为机器提供高质量的数据，赋能机器学习、深度学习等人工智能算法的训练，实现数据价值转化，以医疗领域为例，通过对大量医疗影像数据进行标注，可以训练出精准的疾病诊断模型，提高医疗诊断的准确性和效率，为患者提供更好的医疗服务。当前，全球主流基础大模型，中文语料仅占全部语料的 1%，高质量中文数据成为制约我国基础大模型能力的瓶颈。数据标注是全面提升中文语料质量关键环节，是决定中美两国人工智能科技竞争胜负的关键因素。

（二）人工智能大模型的发展对“数据粮食”的需求高涨

训练一个领先的大模型需要数百万甚至数千万条标注数据，如 OpenAI 在训练 GPT 系列模型，投入数千人力和数亿资金进行数据标注。随着人工智能大模型的发展，我国从 2024 年初日均 Token[2]消耗量 1 千亿，到现在每日消耗量达到 10 万亿级，1 年增长 100 倍[3]。麻省理工大学等研究机构指出，互联网公域高质量文本数据将在 2026 年“耗尽”，人工智能发展将遇到数据壁垒[4]，亟需更多高质量数据集“投喂”人工智能大模型。

（三）国家加大高质量数据集建设力度

我国正积极推动高质量数据集建设，持续增加数据供

[2] Token 作为模型处理文本的最小单元，是指一个单词或者单词的一部分、字符或者其他有意义的文本片段。

[3] 数据来源：国务院新闻办发布会。

[4] 张立：《强化数据标注基地引领作用 带动数据标注产业高质量发展》

给，推动“人工智能+”行动赋能千行百业：**一是**加快推进数据基础制度建设，组织开展数字中国、数字经济、数据要素综合试验区的建设，因地制宜开展先行先试，为数据要素价值释放积累实践经验。**二是**持续推进高质量数据供给。强化公共数据资源登记管理，规范公共数据资源授权运营实施，建立授权运营价格形成机制。**三是**持续推进数据基础设施建设。系统推进全国一体化算力网建设，创新算力电力协同机制，推动算力设施一体化、集约化、绿色化发展。**四是**持续推进数据领域国际合作深化。推进数据领域高水平开放，创造中外数字企业发展良好环境，参与和推动人工智能安全治理，加强国际合作和对话，完善全球治理体系。2024年，国家明确 7 个国家数据标注基地名单，截至 2025 年 3 月，7 个数据标注基地已形成医疗、工业、教育等行业的高质量数据集 335 个，赋能 121 个国产人工智能大模型研发。2023 年我国数据标注产业规模已达 800 亿元[5]。

二、我区数据标注产业发展呈现三个特点

（一）潜在数据标注需求大，但技术能力不高

截至 2024 年底我区人工智能相关企业共计 2708 家[6]，较 2023 年增长了 32.55%，业务主要集中在产业链下游人工智能应用部分。我区实施“人工智能+”产业招商行动，截至 2025 年 4 月，新签约人工智能项目 62 个，涉及投资金额 193 亿元；聚焦“人工智能+制造”，打造 100 个人工智能典型应

[5] 数据来源：央视报道。

[6] 韦泽多，张荃钧：《我区人工智能相关企业发展特点、存在问题和对策建议》。

用场景，形成 100 个标志性智能产品，推动人工智能相关产业产值突破 1000 亿元。我区人工智能应用将释放大量数据标注需求，但技术能力不高。我区数据标注流程尚未形成统一规范，地方性标准体系尚未建立健全，截至 2025 年 2 月，我区暂无数据标注相关的地方标准[7]，同在中西部的山西、贵州分别有 3 个和 1 个地方标准；我区累计获得有效数据标注相关专利[8]共 19 件，占全国（3852 件）的 0.49%，专利数量与发达地区的广东（734 件）、江苏（307 件）、浙江（276 件）存在较大差距，与临近的云南（37 件）、贵州（29 件）也有一定差距；我区累计获得数据标注相关成果登记仅有 2 个[9]。

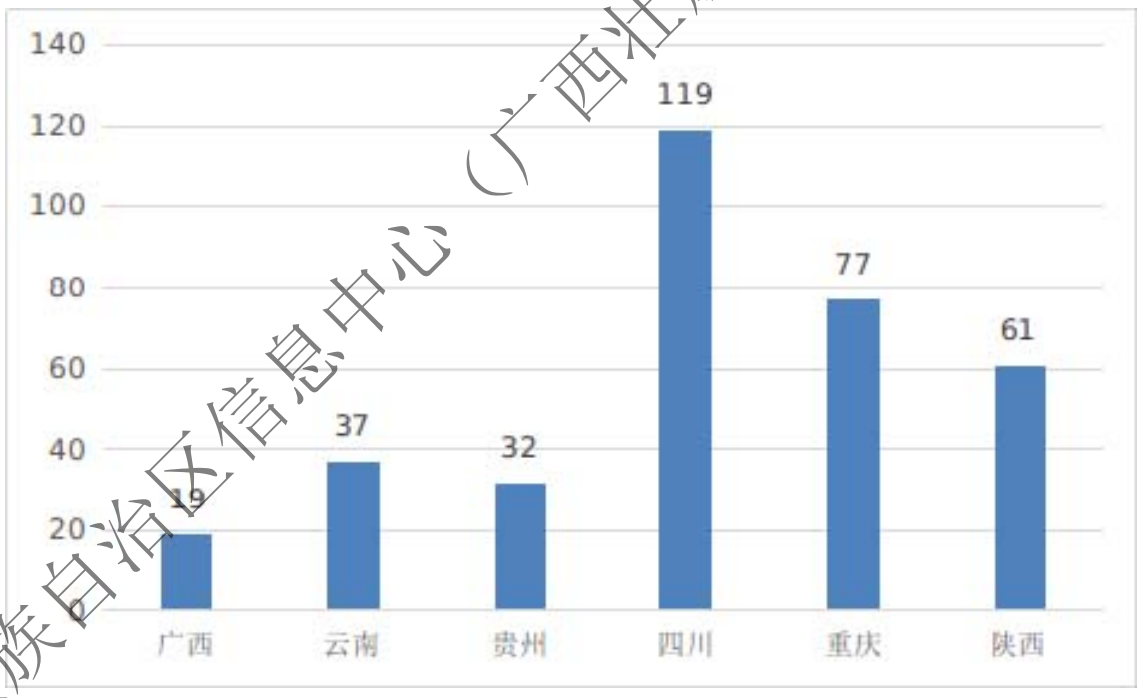

图 1　部分西部省份数据标注相关专利情况

[7] 根据地方标准信息服务平台数据整理，统计口径：地方标准名称中含有“数据标注 数据清洗 数据筛选 数据注释 数据标记 数据质量检测 文本标注 图像标注 视频标注 语音标注 4D 标注 3D 标注 2D 标注 大模型标注 ”等关键字。

[8] 根据国家知识产权局专利业务办理系统数据整理，专利统计口径：申请人所在省为广西，且发明名称中含有“数据标注 数据清洗 数据筛选 数据注释 数据标记 数据质量检测 文本标注 图像标注 视频标注 语音标注 4D 标注 3D 标注 2D 标注 大模型标注 ”等关键字。

[9] 数据来源：广西大数据分析应用平台。

（二）我区数据标注相关企业增长较快，但总体竞争力不强

截至 2025 年 3 月中旬，全区开展数据标注业务的企业共有 37 家[10]，较 2023 年底增长 32.14%，整体呈增长趋势。企业规模以小微型企业为主，占比高达 94.59%；近三年成立企业的占比达 54.05%；45.94%企业分布在南宁；超七成企业分布在科技推广和应用服务业、软件和信息技术服务业等行业。数据标注产业链上游为数据资源提供和应用，中游为高质量数据集开发和治理，下游为能力支持与生态发展[11]。我区上游汇聚数字广西集团、各行业龙头企业及相关互联网企业等数据资源供给体系；下游通过人才培养、生态培育、数据安全与数据标准等赋能的产业发展，产业上、下游有一定基础，但在数据标注产业的中游核心环节缺乏数据标注专业型服务商、标注工具开发者、标注质量评估机构等关键的链主企业，数据标注完整的产业链条尚未形成。我区开展数据标注企业位于产业链中游，呈现“小散弱”特点，规模普遍较小，最大的企业参保人数仅有 39 人，企业零散分布在全区各地，业务单一，以劳动密集型企业为主，数据加工处理能力弱，限制了数据标注的专业化和规模化发展。

[10] 通过广西大数据应用分析平台，结合招聘信息统计。
[11] 数据标注产业链上中下游划分来源于中国信息通信院。

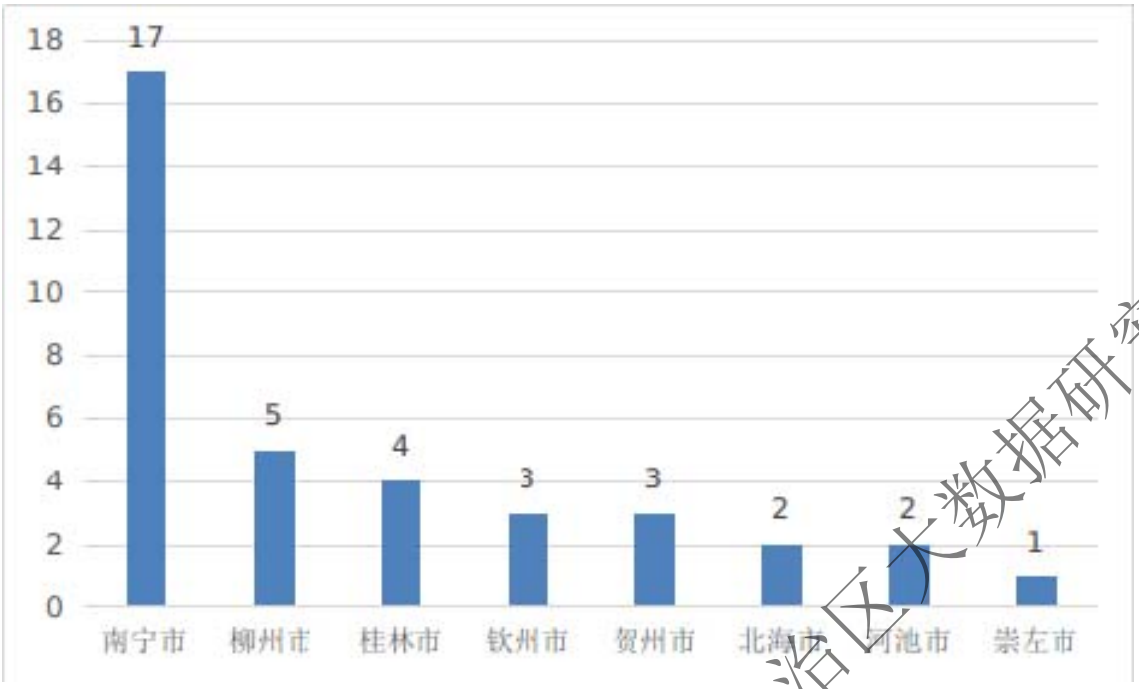

图 2　我区开展数据标注相关企业区域分布情况

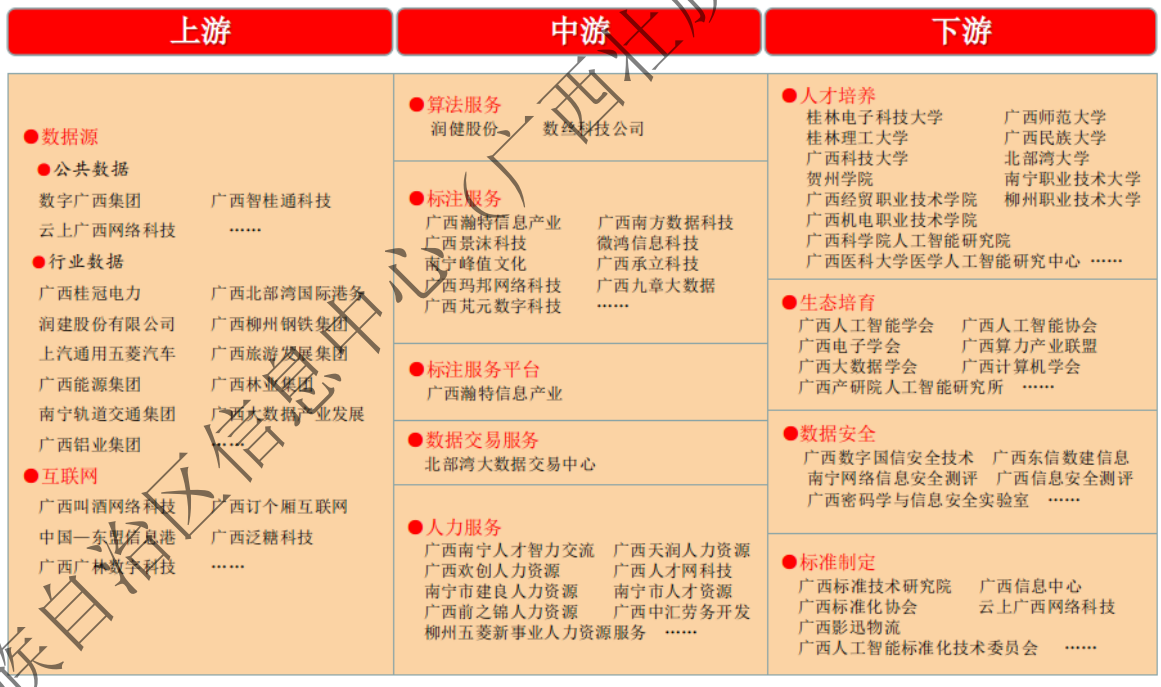

图 3　我区数据标注产业链生态图谱

（三）我区发展数据标注产业具备一定基础，但配套支撑存在短板

在行业数据方面，我区传统优势产业如制糖业、有色金

属产业、林业及林产品加工、汽车制造等的发展，积累了大量行业数据。**在公共数据方面**，截至 2025 年 4 月，广西公共数据开放平台已汇聚 87 个部门的 93.58 亿条，建设广西公共数据授权运营平台等基础设施，为公共数据开发利用打下了基础。**在人才方面**，2024 年广西高校毕业生达 45.25 万人，较 2023 年增加 5.62 万人，规模的持续攀升[12]，为产业发展带来大量人才支撑。但我区发展数据标注配套支撑仍存在短板，如我区数据标注相关的政策零散分布在数字经济发展和数据中心规划建设中，支持力度不强。产业人才吸引力不足，如 2024 年我区数据标注相关岗位平均招聘月薪资仅为 2810 元[13]，远低于 2023 年广西城镇私营单位就业人员的月平均工资 4294 元[14]，而北京专业领域数据标注员月薪接近 2 万元。我区各市发展数据标注产业协同性不足，未融入全国一体化的数据标注市场。

三、加强我区数据标注产业发展的对策建议

综上所述，针对我区数据标注产业处在培育阶段特征，提出以下加快我区数据标注产业发展对策建议。**一是**建立数据标注地方标准规范。规范我区数据标注流程，建立全区统一的数据标注质量评估体系；发挥广西与东盟地缘相近、文缘相通以及广西东盟小语种人才等优势，持续完善中国—东盟国家语言资料库，探索制定汉—东盟语言互译、智能语音识别等关键环节数据标注标准。**二是**培育市场主体强化引领

[12] 数据来源：《广西壮族自治区 2024 届普通高校毕业生就业质量年度报告》。
[13] 数据来源：基于广西大数据分析应用公共服务平台中公开招聘数据。
[14] 数据来源：广西统计局。

带动作用。重点引进产业链中游的数据标注专业服务商和数据标注质量评估机构在广西落地；依托南宁国际通信出入口局建设，构建面向东盟的数据跨境流动试验区，支持本土数据标注企业开展面向东盟国家数据标注业务；参照国家数据局的做法，结合南宁、柳州、桂林等地产业基础建设特色数据标注基地。**三是**优化数据标注产业发展配套支撑。出台我区数据标注产业发展政策，明确数据标注产业发展的目标和路径；加强企业与高校的产教融合，打造“教育链—人才链—产业链”闭环，促进大学生就业，举办形式多样的行业技术技能大赛竞赛活动，激励从业者不断提升自身素质；建设全区统一的数据标注公共服务体系，汇聚行业发展资源，促进供需对接。

（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