

广西如何实现“人工智能+”高质量数据集同频协同发展

在“人工智能+”行动纵深推进的背景下，高质量数据集建设已成为驱动人工智能模型性能跃升、赋能产业智能化转型的关键基石。我区依托面向东盟的区位优势，正加速推动人工智能产业集聚发展，着力打造面向东盟的人工智能合作新高地。让“数据口粮”有结构有质量，是推动我区人工智能高质量发展的关键。

一、我国高质量数据集建设步伐加快

（一）高质量数据集供给持续加强

我国高质量数据集建设取得显著进展，供给能力持续增强。一是数据资源规模加速增长。人工智能场景的深化应用推动数据资源规模加速增长。2024年，全国年度数据生产总量达到 41.04 泽字节（ZB），同比增长 25%，数据存储总量为 2.09 泽字节（ZB），同比增长 20.81%，用于人工智能模型开发、训练和推理的数据量同比增长 40.95%¹。二是数据标注产业快速发展。产业规模方面，2024年我国数据标注企业数量超过 4000 家，近 20 年产业规模增长率达 11.3%；产业创新方面，数据标注相关专利授权增长率达 77%，近 4 年专利授权量翻了 10 倍²。三是高质量数据集供给加速推进。国

¹ 数据来源：全国数据资源调查报告（2024 年）

² 数据来源：上海数据交易所

务院国资委发布人工智能“焕新社区”开源平台，汇聚 158 个高质量数据集和 244 个垂类模型³。地方特色数据集供给取得突破，如山东建成 50PB 海洋数据集⁴、安徽打造首批 55 家行业高质量数据集建设基地⁵。

（二）高质量数据集政策体系不断完善

高质量数据集建设已成为激活数据要素价值的关键支撑，政策体系构建呈现系统性深化、实操性增强的鲜明导向。一是顶层规划持续强化。2023 年 12 月，国家数据局等 17 部门联合印发《“数据要素×”三年行动计划（2024—2026 年）》，明确要强化场景需求牵引，带动数据要素高质量供给、合规高效流通，高质量数据集建设取得积极进展；国家发展改革委等部门联合印发《关于促进数据产业高质量发展的指导意见》，明确到 2029 年数据产业规模年均增速超 15%，推动建设行业共性数据资源库，重点覆盖工业、医疗等 20+ 领域。国家发展改革委等联合印发《国家数据基础设施建设指引》，明确支持农业、工业、交通金融等 10 个行业打造高质量数据集。二是标准规范建设加速成型。2025 年《高质量数据集实践指南（1.0）》提出标准化建设路径，为行业发展提供方法论和操作指引。地方层面，先进省市加快探索步伐，例如，京津冀三地数据管理部门联合实施“京津冀区域携同数据标准化行动”，三地协同共同研制地方标准 8 项、参考执行的

³ 数据来源：央视新闻

⁴ 数据来源：山东省人民政府

⁵ 数据来源：安徽省数据资源管理局

地方标准 8 项。三是地方探索实践持续深化。部分省市初见成效，全国已有如湖北、安徽、福建、江苏等多个省市发布了高质量数据集清单，如湖北累计征集发布 35 个高质量数据集，覆盖多个重点领域，具备模型适配性强、标注准确性高、行业辐射广的特点⁶。多地通过出台政策进行布局，如江苏省提出到 2027 年建成 1000 个高质量数据集，打造 2000 个数据产品，建设 30 个行业数据集支持 AI 大模型应用⁷；上海市明确，到 2025 年形成 1000 个高质量数据集，覆盖金融、医疗、航运等重点领域，并打造 1000 个品牌数据产品，推动数据要素市场规模突破 5000 亿元⁸。

二、广西高质量数据集建设发展空间广阔

（一）制度供给成效初显，但仍需进一步强化

广西在高质量数据集建设的制度供给方面已初步构建了多层次框架体系，其成效主要体现在三方面：一是基础性法规与专项政策协同推进。2024 年以来，广西已密集出台如《广西壮族自治区大数据发展条例》、《广西构建数据基础制度更好发挥数据要素作用总体工作方案》、《广西加快数据要素市场化改革实施方案》等多项加强数据要素发展的政策文件，并设立了规模达 100 亿元的人工智能产业基金，其政策密度与资金支持力度在西部省份中居于前列⁹。二是数据治理模式创新突破。我区目前已形成了“一云承载、一网通达、

⁶ 数据来源：湖北日报

⁷ 数据来源：江苏省人民政府

⁸ 数据来源：上海市人民政府

⁹ 数据来源：广西人工智能产业发展白皮书（2025 年）

一池共享”的治理架构，逐步推动非涉密系统上云，有效整合了公共数据资源，目前已依托自治区公共数据开放平台完成数据汇聚总量达 93.79 亿条，较 2023 年增长 32%¹⁰，初步破解“数据孤岛”难题。在此基础上，我区通过建设广西数字政府大脑，实施数字资源“一本账”机制，加强各类数据的全面治理和应用。三是逐步完善资产化管理制度。通过发布《广西壮族自治区财政厅关于加强行政事业单位数据资产管理的通知》，强化数据资产登记、核算与授权运营规范，数据交易所等市场化平台形成监管协同。

但我区在制度层面仍存在三方面不足：一是垂直领域制度供给不足。尽管通用数据开放框架基本建立，但工业、医疗、金融等高价值领域缺乏专门数据集建设标准与授权细则，导致企业数据流通动力不足，核心业务数据开放比例偏低。以行业高质量数据集为例，苏州已发布 30 个高质量数据集，覆盖工业制造、交通运输、金融服务等领域¹¹，但我区暂未形成行业高质量数据集，相对较为落后。二是数据质量管控机制薄弱。现有制度对数据完整性、鲜活度约束有限，部分公共数据存在更新滞后、规范性欠缺问题，非结构化数据转化效率不足，制约数据集可用性。三是跨境与权益分配制度尚待突破。虽然已经提出探索面向东盟的数据跨境流动国际标准，但实操层面仍缺乏安全评估、收益分配等配套细则，企业数据授权模式与公共数据有偿使用边界亟待明晰。

¹⁰ 数据来源：广西壮族自治区公共数据开放平台公告

¹¹ 数据来源：高质量数据集实践指南 1.0

（二）数据供给质量提升，结构性矛盾与效能短板亟待破解

一是公共数据供给方面。我区公共数据供给体系目前已初步形成统一平台支撑与分级管理机制，实现了跨部门、跨层级的公共数据归集共享，覆盖自治区、市、县三级单位，接入 716 个单位，发布公共数据资源目录 1.02 万条，累计接口调用超 327.14 万次¹²，数据资源目录与库表资源规模显著提升。通过建设公共数据开放平台，支撑了政务服务、民生应用等场景创新，如“一网通办”精准服务等。但数据供给质量与效能仍存短板：一是数据开放主动性不足，部分部门因安全顾虑或权责不清存在“不敢开、不愿开”现象，导致高价值数据开放有限；二是数据鲜活度与规范性待提升，部分数据更新滞后、格式非标准化，影响开发利用效率；三是供需匹配度不高，部分数据存在“形式开放”，核心业务数据与社会需求脱节，制约了数据要素价值释放。

二是产业数据供给方面。我区产业数据供给在特色农业、制造业等领域取得局部进展，例如农业产业数据方面，以一个中心（广西农业农村大数据中心）和三大平台（广西农业云平台、广西农业农村大数据管理平台、广西农业全产业链公共服务平台）为基础，构建了 4 大类 39 小类农业专项数据资源库，范围涉及农业生产、农产品市场流通等领域¹³。

¹² 数据来源：广西壮族自治区公共数据开放平台公告

¹³ 数据来源：数字广西发展报告（2024）

但产业数据整体供给仍面临结构性矛盾：一是农业数据碎片化突出，当前农业经济发展的模式导致生产数据分散、标准化程度低，难以支撑规模化分析；二是制造业数据融合深度不足，企业数据孤岛现象普遍，产业链上下游数据未能有效贯通；三是数据驱动产业升级能力较弱，数据在优化生产决策、供应链管理等核心环节的应用仍显薄弱，且缺乏行业性数据交易与融合创新机制。

（三）数据基础支撑形成，要素短板亟待补齐

我区当前依托“桂林—柳州—南宁—钦州”中轴构建的“一轴双核多点”的算力网络布局，为数据集开发提供基础支撑环境，建成覆盖 14 个设区市“一城一池”及 30 余个边缘节点的的数据中心集群，形成“同城 1 毫秒、全区 5 毫秒”的超低时延算力圈。通过加强基础设施建设，保障数据有效流通。2024 年全区累计建成 5G 基站超 14 万个，互联网出口带宽达 6175 万兆¹⁴。

问题集中体现在两方面：一是技术支撑体系薄弱，数据治理工具链成熟度不足，尤其在隐私计算、多模态数据融合等关键技术领域存在瓶颈，对比全国首批 19 个数据基础设施试点城市，武汉已搭建隐私计算公共服务平台，南京、苏州试点数联网技术并形成 35 个应用场景¹⁵，而广西尚未入围国家试点。二是产业生态尚未健全，2024 年广西存续人工智能相关企业 2708 家，同比增长 32.55%，但规模以上企业仅

¹⁴ 数据来源：广西人工智能产业发展白皮书（2025 年）

¹⁵ 数据来源：江苏经济报

267 家¹⁶，仅占总体的 9%；南宁虽集聚全区 57.31% 的企业，仍缺乏具备全域数据整合能力的龙头企业，跨主体数据协作机制与价值分配模式尚未成熟，导致数据集建设与产业需求脱节。

（四）中国—东盟数据联通体系建成，规则壁垒阻碍合作深化

我区着力构建以中国—东盟信息港为核心的数据枢纽，建成南宁国际通信出入口局、国家级互联网骨干直联点及跨境数据专用通道等基础设施，以及联通越南、老挝、缅甸等东盟国家的跨境陆路光缆 11 条¹⁷，初步实现与东盟国家的高效数据联通，同时建成中国—东盟跨境征信服务平台、北部湾大数据交易中心等载体，推动区域数据资源整合与应用创新。

然而，广西与东盟的合作仍面临多重挑战：一是东盟成员国数据治理规则碎片化，新加坡、泰国等 6 国采用 GDPR 兼容框架，越南《2024 年数据法》强制金融、医疗数据本地化，而缅甸、柬埔寨尚无基本数据分类标准¹⁸，这导致跨境数据流动存在制度壁垒。二是受欧盟 GDPR、美国 CBPR 等域外规则渗透，企业面临多标准合规冲突与成本攀升问题，如中国—东盟跨境电商因 GDPR 的“长臂管辖”额外承担了营收 5—7% 的合规成本¹⁹。三是数据分级分类标准互认不足，各国对敏感数据界定差异制约高质量数据集的共建共享。

¹⁶ 数据来源：广西人工智能产业发展白皮书（2025 年）

¹⁷ 数据来源：广西壮族自治区发展和改革委员会

¹⁸ 数据来源：中国社会科学网

¹⁹ 数据来源：中国社会科学网

三、对策建议

（一）强化垂直领域制度供给与权益分配机制

一是加快出台工业、医疗、金融等高价值领域数据集建设专项标准与授权细则，建立行业数据分类分级开放目录，配套企业数据流通激励机制，破解核心业务数据开放难题。二是构建数据全生命周期质量管理体系，制定覆盖数据采集、清洗、更新的强制性规范，强化非结构化数据智能转化能力建设，设立公共数据质量问责机制。三是深化跨境数据流动规则实践创新，依托中国—东盟信息港开展安全评估试点，探索跨境数据“监管沙盒”与收益分配模型，明晰公共数据有偿使用边界，推动数据要素市场化配置改革走深走实。

（二）构建分层激励机制打通数据供给堵点

一是在公共数据领域实施“负面清单+授权运营”双轨机制，对非敏感数据实行“开放为原则、不开放为例外”的负面清单管理，同步建立数据授权运营制度，通过收益分成激发部门开放积极性。二是针对产业数据实施“链主企业牵引计划”由糖业、汽车制造等特色产业龙头企业牵头搭建行业数据空间，通过“数据高铁”技术贯通产业链上下游节点，推动农业数据标准化采集与小农户数据合作社建设。

（三）以智能融合基建赋能数据集开发应用

加速推进算力网络与先进技术融合应用，在“一轴双核多点”布局中嵌入隐私计算、联邦学习等安全技术组件，建设跨域异构算力智能调度平台。打造垂直行业高质量数据集

开发工场，聚焦工业互联网、中国—东盟跨境贸易等场景，建设行业专属的多模态数据集训练平台，配套数据治理工具链开源社区培育计划。

（四）创新区域协作机制破解东盟合作壁垒

率先构建中国—东盟数据规则“软联通”示范体系，联合新加坡、泰国等规则成熟国家共同制定跨境数据流动互认白名单。建立区域分级分类标准互认联盟，依托中国—东盟信息港发布跨境敏感数据识别指引，建设区域跨境征信数据池。实施数字鸿沟弥合专项行动，通过建设中国—东盟人工智能算力共享平台，为柬埔寨、老挝等国提供普惠型数据开发工具链和技术培训，构建区域数据要素共同体合作框架。

（执笔人：黄子川、周曼）