

国家数据局关于印发《关于推进行业高质量数据集建设行动的实施方案》的通知

国数科基〔2026〕25号

各省、自治区、直辖市及新疆生产建设兵团数据管理部门：

现将《关于推进行业高质量数据集建设行动的实施方案》印发给你们，请认真组织实施，加快推进相关工作。

国家数据局

2026年6月3日

关于推进行业高质量数据集建设行动的实施方案

行业高质量数据集是经过采集、加工等数据处理，可直接用于开发和训练人工智能模型，能有效提升模型性能的行业数据的集合，包含行业通识和行业专识数据集。行业高质量数据集是推动“人工智能+”赋能千行百业、实现产业落地的基础性、关键性资源。为落实国民经济和社会发展规划《纲要》，深入实施“人工智能+”行动，推动行业高质量数据集建设推广与“人工智能+”同频共振、互促共进，强化数据赋能人工智能创新发展，制定本方案。

一、总体要求

以习近平新时代中国特色社会主义思想为指导，深入贯彻党的二十大和二十届历次全会精神，全面落实“人工智能+”行动，主动顺应人工智能发展范式跃迁，按照“需求牵引、急用先行、应用验证、安全保障”原则，聚焦国民经济发展重点行业和战略性新兴产业，围绕行业高质量数据集供给、流通、应用等关键环节，部署强基扩容、标注攻坚、提质增效、应用赋能、管理服务、价值释放六个专项行动，形成“场景牵引数据、数据驱动模型、模型赋能应用、应用创造价值”的“数据飞轮”，加快构建数据要素与人工智能协同演进的共生生态。

到 2028 年底，建成一批覆盖重点领域、经过应用验证的行业高质量数据集，打造一批数据驱动人工智能创新发展的典型应

用场景，培育一批具备领先优势的创新型数据企业 and 专业人才，形成一批行业高质量数据集建设工具和标准。数据从供给到价值释放的良性循环基本形成，数据赋能人工智能创新发展的作用更加凸显，数据产业与人工智能深度融合，持续催生智能经济新增增长点。

二、实施强基扩容行动

顺应人工智能加速向行业渗透，从对话向多模态生成、决策执行、具身智能、物理交互等范式跃迁的趋势，拓宽数据供给渠道，丰富数据供给类型，加快建设行业高质量数据集，为人工智能发展和应用提供充足“燃料”。

（一）聚焦行业领域推进高质量数据集建设。聚焦科学研究、工业制造、农业农村、智慧能源、交通运输、金融服务、医疗卫生、教育教学、电子商务、人力资源、文化旅游、应急管理、气象服务、绿色低碳、公共安全、城市治理、住房建设、自然资源、社会信用等重点领域，以及低空经济、具身智能、智能驾驶、智慧海洋、生物制造等创新领域，加快推进行业高质量数据集建设。

（二）夯实行业高质量数据集建设基础路径。梳理行业数据资源底数和应用场景，建立数据资源清单和数据集需求清单。以应用为牵引，持续推进行业高质量数据集建设先行先试，加快形成一批可复制、可推广的数据驱动型示范场景。强化链主单位牵引带动作用，支持链主单位以联合体等形式推动产业链上下游协同共建和资源整合，持续扩大行业高质量数据集供给规模，鼓励链主单位面向行业开放数据集并提供数据服务，赋能产业链上下

游中小企业。鼓励高等院校、科研院所、行业协会、数据流通服务机构、第三方专业服务机构等各类主体参与行业高质量数据集建设。加大公共数据资源开发利用力度，推动公共数据与行业数据融合利用，建设一批高质量数据集。

（三）面向人工智能应用需求丰富行业高质量数据集建设形态。持续推进文本、代码、图像、音频、视频、点云、时序数据、科学数据等多模态高质量数据集建设，赋能人工智能预训练、指令微调、强化学习、测评等各阶段。加强知识库、知识图谱、本体等数据集建设，加快复杂任务规划、长程推理、人机交互、决策执行等数据集建设，赋能智能体等新型智能应用形态。加快重点场景物理交互、环境感知、运动控制等真机交互数据集建设，积极应用仿真模拟与合成技术扩大数据供给，赋能具身智能发展。积极面向世界模型等前沿方向，推进数据集建设。

（四）强化与数据基础设施建设有机联动。鼓励依托国家数据基础设施，充分运用隐私保护计算、可信数据空间等能力，开展数据集安全存储、可信流通、高效应用，推动数据集从分散持有向集约化、标准化供给转变。鼓励探索建设支撑大规模、多模态数据集的数据基础设施存力中心。

三、实施标注攻坚行动

数据标注是将知识和经验注入到训练数据的过程，是行业高质量数据集建设不可或缺的关键环节。引导数据标注从“以人为主”向“人机协同、专家深度参与”的多层次标注模式转变，推动数据标注向专业化、智能化跃升。

（五）推动数据标注转型升级。加强数据标注领域科技创新，强化自动化工具和平台的研发与应用，发展“模型预标注+人工校准”“人工标注+模型检验”“模型预标注+模型检验”等智能化标注服务，全面提升数据标注水平。发展专家型数据标注服务，建立行业专家认证机制，推动专家深度参与指令微调、强化学习等阶段所需的专业知识标注，生产领域知识、逻辑推理等高质量数据集，提高数据集的知识密度与专业价值。

（六）持续推动数据标注先行先试。指导首批七个承担数据标注先行先试任务的城市，持续做强做深数据标注产业。面向创新能力强、发展基础好、产业特色优的地区，梯次布局一批数据标注创新试验区。有序引导具备条件的地区因地制宜开展试验区建设，促进数据标注产业链上下游紧密协同，形成产业集聚效应。培育一批数据标注龙头企业、独角兽企业、瞪羚企业，壮大数据标注产业。

（七）扩大数据标注人才供给。支持有条件的院校增设数据标注相关课程，依托产教融合、校企协同等方式，培育具备专业知识的数据标注人才。鼓励开展数据标注职业技能等级认定，强化继续教育与在职培训，畅通人才发展通道。完善分层分类人才评价体系，建设专职与兼职相结合的专业标注人才队伍。鼓励高校毕业生等群体参与标注工作，通过灵活就业、项目协作、多元岗位供给等方式，扩展就业渠道。

四、实施提质增效行动

推动构建符合结构完整性、内容多样性、标注准确性、模型

适配性等质量标准、满足人工智能就绪（AI-Ready）的高质量数据集，降低训练推理成本，有效提升模型性能。

（八）提升行业高质量数据集建设质效。加强数据清洗、增强、标注、对齐、质检等关键技术攻关和全过程自动化工具研发应用，支撑高效率、高标准构建行业高质量数据集。针对行业特定应用场景，鼓励运用数据智能过滤与配比等技术，构建更精、更强的高知识密度数据集，降低训练推理成本。发挥数据合成在数据集建设中的积极作用，利用模型、仿真系统等生成的数据，解决稀缺场景数据集构造难、真实场景数据采集成本高等问题。

（九）推动高质量数据集标准体系建设和应用落地。加快推进高质量数据集格式、类型、标注、质量测评等相关国家标准研制、应用验证和贯标，健全高质量数据集标准体系。鼓励各行业、各地方与国家标准联动，推动重点行业领域高质量数据集标准研制。加强在产业政策、政府采购、招投标中引用相关标准，引导产业执行高水平标准，用好检验检测等手段推动标准实施，促进高质量数据集规范化建设。

（十）强化高质量数据集质量测评和结果互认。持续完善“数据质量验证+模型应用反馈”的测评方法，加快建设覆盖多行业、多场景、多模态的测评数据集，有效评估高质量数据集应用效果。发起高质量数据集测评联合行动倡议，推动相关单位按照标准，采用统一测评方案和工具开展测评和封装工作，实现“一次测评、全国互认”。

五、实施应用赋能行动

坚持行业高质量数据集建设与实际应用深度融合，以模引数、用数赋模，促进高质量数据集建设与“数据要素×”“人工智能+”同频共振，全面赋能产业数智化转型。

（十一）打造“数据飞轮”应用闭环。以模型应用牵引数据供给、以数据赋能模型迭代，推动形成“场景—数据—模型”协同发展的良性循环。发挥“人工智能+”场景牵引作用，推动数据供给和场景的精准匹配，以用促建，以实际需求吸引更多数据资源汇聚，推动行业高质量数据集有效供给和持续优化。结合“数据要素×”行动，深化行业高质量数据集建设，以建促用，充分利用模型应用产生的动态交互数据等，驱动行业模型能力持续提升。

（十二）打造行业应用标杆和典型案例。着力打造集“数据集生产加工和流通利用、支撑模型训练应用”于一体的数据赋能工场，打造一批行业标杆，加速人工智能应用落地。打造一批数据赋能智能体解决实际问题的典型案例，推动高质量数据集规模化应用。

（十三）繁荣数据集协同发展生态。加强统筹协调，强化部门联动，分行业分领域有序推进高质量数据集建设和应用。搭建“政产学研用金”多方交流平台，形成共建共享、互利共赢的产业生态，破解“数据孤岛”“数据烟囱”。常态化举办供需对接活动，提升供需匹配效率，促进实质性合作落地。组织遴选行业高质量数据集建设和应用典型案例，发挥行业示范引领作用。支持举办行业高质量数据集创新赛事活动，以赛促建、以赛促用。深化国际交流合作，建立健全数据集跨境安全有序流动机制，推动跨境

流动规则互认，鼓励行业高质量数据集建设主体积极参与全球数据生态建设。

六、实施管理服务行动

加强数据集管理，完善数据伦理和治理机制，推动落实数据权益相关制度，推进数据集建设体系更加规范有序。

（十四）构建数据集全生命周期的管理体系。加强覆盖数据采集、清洗、加工、标注、质检、测评、迭代、审计等全生命周期的数据集管理服务能力建设，依托数据基础设施，强化隐私保护计算、区块链等技术应用，确保数据可管、可控、可追溯。建设“物理分散、逻辑集中”的国家数据集管理服务系统，实现数据集目录、供需等信息互联互通。支持各地方、行业依托国家系统设置专区，支持已有系统与国家系统对接。

（十五）探索面向人工智能发展的数据相关制度。落实数据持有权、使用权、经营权三权分置制度。研究合成数据等新情况新问题。兼顾产权保护与创新发展需求，完善人工智能训练阶段数据使用规则，推动版权作品数据等有序用于模型训练，完善数据授权使用机制和收益分配规则，打造权责清晰、合规包容的制度环境。

（十六）坚持伦理先行与公平普惠。研究探索高质量数据集伦理道德规范，坚持有益社会的价值导向，严禁非法收集或使用敏感数据。防范数据集建设全流程产生数据偏见与歧视，充分考虑社会各界多元需求，最大限度确保数据集建设成果惠及全民。

七、实施价值释放行动

发挥数据集的应用价值，以行业高质量数据集赋能人工智能发展。释放数据要素价值，推动数据集商业化、资产化，培育为数据付费的市场共识，探索以词元（Token）为基础的价值体系。

（十七）发挥行业高质量数据集的应用价值。建立数据集和模型需求对接机制，推动行业高质量数据集和模型精准适配，提高模型质量和效率，深度赋能行业发展。推动行业高质量数据集跨行业、跨领域、跨场景融合利用，鼓励“以数换数”“数模互换”“数据托管”“数算一体”等多种应用模式。积极参与开源社区建设，鼓励中介机构、公益机构将基础性、公益性数据集作为公共产品向社会公开，激发用数活力。

（十八）创新行业高质量数据集商业模式。完善数据集长效运营机制，鼓励数据集在数据交易所（中心）等数据流通服务机构挂牌交易，发展“订阅模式”“商场模式”“定制模式”等多元服务形态，推动商业模式从基础数据包销售向 API 调用、模型化解决方案及全栈服务梯次跃升。探索词元交易等新型交易模式，构建以词元为基础，可量化、可定价的数据价值体系。

（十九）探索行业高质量数据集资产化创新路径。鼓励有条件的单位率先探索开展数据集资产盘点、登记、评估等试点工作，为数据资产化积累可复制、可推广的经验。鼓励探索数据集质押融资、作价入股、资产证券化、数据信托、数据保险等多元资产化创新模式，拓宽数据价值转化渠道。

（二十）培育为高质量数据付费的市场共识。建立健全市场化利益分配机制，确保数据供给、加工、流通、应用等各环节主

体均能获得市场化价值回报，共享数据红利。鼓励数据需求方通过购买、合作等形式加大投入，充分发挥政府部门、国有企业、模型企业等单位的示范引领作用，推动数据采买纳入预算编制，率先开展数据采购实践，带动形成数据有偿使用市场共识，构建健康可持续的数据市场生态。

八、保障措施

国家数据局发挥统筹协调作用，会同有关部门协同推进高质量数据集建设与应用，构建“部门协同、领域联动”的工作格局。各地要落实属地管理责任，加大组织实施力度，避免一哄而上，防止同质化、低水平重复建设，结合实际制定配套措施，统筹安排数据产品和服务采购经费，用于支持行业高质量数据集建设。支持各类主体以应用为牵引，积极开展和参与行业高质量数据集建设。引导金融机构、耐心资本、产业基金等，加大对行业高质量数据集建设的投资力度，鼓励地方设立专项资金，探索多元化、多渠道投入机制。鼓励在依法依规、风险可控前提下开展创新探索。持续跟踪行业高质量数据集建设工作，完善监测指标，评估建设和应用成效，阶段性总结经验。强化安全保障，落实数据安全相关法律法规要求，建立全流程安全治理机制，防范数据投毒与污染、数据泄露等安全风险，守牢数据安全底线。