一图读懂

关于推进行业高质量数据集建设行动的实施方案

为落实国民经济和社会发展“十五五”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。

总体要求

基本原则

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。

核心任务

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。

总体目标

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和专业人才，形成一批行业高质量数据集建设工具和标准。

实施强基扩容行动

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。

夯实行业高质量数据集建设基础路径

建立**数据资源清单**和**数据集需求清单**，持续推进**行业高质量数据集建设先行先试**，强化**链主单位**牵引带动作用，鼓励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。

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**多模态**高质量数据集建设。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。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。积极面向世界模型等前沿方向，推进数据集建设。

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**。

实施标注攻坚行动

推动数据标注转型升级

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**。发展**专家型数据标注服务**，建立**行业专家认证机制**，推动专家深度参与专业知识标注。

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。布局一批**数据标注创新试验区**。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。

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**，鼓励开展**数据标注职业技能等级认定**，完善**分层分类人才评价体系**，鼓励高校毕业生等群体参与标注工作。

实施提质增效行动

提升行业高质量数据集建设质效

加强**关键技术攻关**和**全过程自动化工具**研发应用。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。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。

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**，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**。

强化高质量数据集质量评测和结果互认

持续完善**“数据质量验证+模型应用反馈”**的测评方法，加快建设评测数据集。发起**高质量数据集测评联合行动倡议**，推动相关单位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”**。

实施应用赋能行动

打造“数据飞轮”应用闭环

推动**“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**。结合**“数据要素×”**行动，**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。

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**。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。

繁荣数据集协同发展生态

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。常态化举办**供需对接活动**，组织遴选**行业高质量数据集建设和应用典型案例**，支持举办**行业高质量数据集创新赛事活动**，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**。

实施管理服务行动

构建数据集全生命周期的管理体系

加强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。

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。完善人工智能**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。

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，**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。

实施价值释放行动

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**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，激发用数活力。

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**以词元为基础的数据价值体系**。

探索行业高质量数据集资产化创新路径

探索开展**数据集资产盘点、登记、评估**等试点工作，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**等多元化资产化创新模式。

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，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**，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。

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。

监制单位：国家数据局
制图单位：国家数据发展研究院