

上海市数据局文件

沪数〔2025〕24号

上海市数据局关于征集 2025 年上海市高质量数据集先行先试项目的通知

各有关单位：

为贯彻落实国务院《关于深入实施“人工智能+”行动的意见》（国发〔2025〕11号），持续加强人工智能高质量数据集建设，结合国家数据局推动高质量数据集先行先试试点建设的工作要求，进一步发掘优质项目，上海市数据局现组织开展 2025 年上海市高质量数据集先行先试项目征集工作。相关事项通知如下：

一、征集时间

即日起至 2025 年 10 月 31 日。

二、征集方向

（一）方向 1：行业领域高质量数据集建设

围绕上海“五个中心”建设，打造相关行业领域的行业通识类或行业专识类高质量数据集，以应用为导向，提炼可复制推广模式，培育高质量数据集产业生态，支撑行业领域智能化水平提升。

（二）方向 2：高质量数据集工具链平台建设

围绕高质量数据集在采集、治理、标注、生产、维护等建设过程中所涉及的作业流程，研制集约化、智能化、高效化的高质量数据集工具链与服务平台，逐步形成体系化服务能力，构建高质量数据集生产支撑工具。

三、工作任务

高质量数据集先行先试项目建设周期为两年，从以下五个维度推进，其中 1-3 为必建内容，4-5 为可选建内容：

1.技术系统维度：聚焦数据采集、处理、管理、标注、合成、质检、服务运营等关键环节，着力推动自动化工具研发应用，加快建设数据集开发运营平台，切实提升数据集生产效能。

2.场景应用维度：聚焦行业领域用数需求与核心痛点，打造可复制应用场景，推动高质量数据集场景化落地应用。

3.标准验证维度：深化对高质量数据集格式要求、分类指南、

质量评测规范、建设指南、建设运营能力成熟度模型、平台建设技术要求等相关标准的验证工作，确保标准科学性与适用性。

4.机制建设维度：积极探索基于贡献度的收益分配、授权使用及版权争议处置等机制，健全数据集安全规范运行制度体系。

5.生态发展维度：注重发挥对相关产业的辐射带动作用，助力构建产业协同发展格局。

四、申报要求

（一）申报范围

1.申报主体须为注册在上海区域内的企事业单位，应具备独立法人资格，财务管理制度健全，信用状况良好。

2.已入选 2025 年度国家数据局行业高质量数据集先行先试工作的试点项目不在本次申报范围内。

3.本次征集不接受联合申报。

（二）申报材料

1.申报材料须重点突出、目标明确、言简意赅，从建设路径、技术方案、场景设计、标准规范、机制建设、生态发展等方面体现创新性和可推广性，预期效果显著，对数据要素市场发展具有较强推动作用和示范价值，同时明确推进计划与阶段性成果。

2.申报材料须对内容真实性、合规性作出承诺且不涉及国家秘密、商业秘密等内容。

（三）申报推荐

获得上海市行业主管部门或区数据局推荐的申报单位享有优先权，推荐单位须在“推荐单位”栏填写单位全称并签章。

五、申报提交

材料规范：申报材料盖章处须加盖公章（复印无效），申报书具体内容和格式须参照《上海市高质量数据集建设先行先试申报书》（见附件）。

提交方式：请将相关材料电子版发送至指定邮箱，邮件主题请按“单位名称+申报名称+上海市高质量数据集先行先试申报书”格式填写。邮箱地址：linyang@shanghai.gov.cn。

咨询电话：021-23012069、021-23012062。

六、评审遴选

申报征集结束后，上海市数据局将组织专家团队开展项目评选，评选采取专家评审、集中答辩或现场考察等方式确定候选项目，根据评审结果划定试点项目名单并对外公布。

七、后续管理

市数据局将对先行先试项目实施动态管理，通过跟踪建设进度、评估实施成效、优化推进举措，确保先行先试工作落地见效。

八、支持措施

本次征集入选的试点项目将在 2025 年全球数商大会上公开发布；结合征集情况，开展供需对接活动和重点推荐等工作，优秀试点项目将优先推荐至国家数据局；推动本市支持政策、资金

向高质量数据集先行先试入选项目倾斜。

附件：上海市高质量数据集建设先行先试申报书

上海市数据局

2025 年 10 月 13 日

（此件主动公开）

上海市高质量数据集建设 先行先试申报书

项目名称：

申报单位：

推荐单位（如有）：

申报日期：

上海市数据局编制

2025 年 10 月

申报须知

- 1.申报材料要求盖章处，须加盖公章，复印无效。
- 2.高质量数据集建设先行先试的申报主体为注册在上海区域内的企事业单位。
- 3.如有推荐单位，需在印有“推荐单位”处填写单位全称并签章。
- 4.电子版以邮件形式提交，当页面不够时可另加页。

一、项目基本信息情况表

项目名称			
申报单位名称			
申报方向	☐ 方向 1：行业领域高质量数据集建设 ☐ 方向 2：高质量数据集工具链平台建设		
项目周期	2025 年 12 月至 2027 年 11 月		
拟投入资金（万元）			
已投入资金（万元）			
项目联系人		职务	
电话			
邮箱			
单位地址			
所在区（注册区县）			
法定代表人			
统一社会信用代码			
单位性质	☐ 国有企业 ☐ 事业单位 ☐ 民营企业 ☐ 外资企业 ☐ 合资企业 ☐ 其他（请注明）		
近三年企业发展情况	2022 年	2023 年	2024 年
资产总额（万元）			

资产负债率（%）			
主营业务收入（万元）			
利润率（%）			

二、发展基础

（一）申报单位基本情况（不超过 1000 字）

介绍申报单位主营业务等基本概况，以及高质量数据集相关工作基础和能力等情况。

（二）已开展工作（不超过 1000 字）

请从建设方法、应用场景、技术方案、运营管理等方面描述相关工作基础，如已开发的相关场景，已应用的大模型情况，已形成的大模型与数据集协同机制等。

三、总体思路

请简要说明项目总体思路和主要任务。

四、试点方案

试点方案除包含基本情况、背景与必要性、工作基础等必须部分之外，还需包含：

1.建设路径：针对行业领域大模型需求，明确从原始数据资源获取，到高质量数据集建成的关键步骤和路径，形成符合行业特点、可落地验证的数据集质量评价依据。

2.技术方案：描述数据采集、处理、管理、标注、合成、质检、服务运营等关键加工环节的技术应用，以及拟构建数据集技术方案、实施计划的合理性与可行性论证。

3.场景建设：描述重点领域中相关场景数据集规模与质量情况，应用场景复用性、行业共性需求解决价值及量化成效，还有与跨行业主体的合作及需求收集机制。

4.标准规范：描述数据采集加工、标注处理、发布运营等全流程管理规范覆盖情况，以及数据集质量管理规则、安全风险审查机制与防控预案的落地情况。

5.机制建设（选填）：描述数据集流通交易方式、可持续商业模式及人才培养方案，版权保护与利益分配运营机制，收益测算及与不同生态主体的深度合作。

6.生态建设（选填）：描述数据集生产过程与不同生态主体间形成的协同生态网络，数据集生产对行业创新、中小企业发展的赋能效果等。

7.进度计划：描述数据集构建、场景落地、生态合作等各阶段时间节点与阶段性成果，以及预期达成的行业数据质量提升、数据价值释放等社会效益与经济效益。

8.投资明细：请列明投资明细。

9.项目财务分析、经济分析及主要指标，如收入预测分析、

经济与社会效益分析，考核指标明细等。

五、申报主体承诺书

本单位近三年未出现违规违法行为，未发生重大信息、网络安全事件，未发生重大、特大生产和环境安全事故，本单位不属于失信被执行人。

所开展的高质量数据集建设和运营行为合法合规，申报材料内容真实有效。

承诺单位（签章）：

承诺日期：

六、推荐单位意见（如有）

推荐意见：

推荐单位（签章）：

日期：

七、项目单位相关证明材料

1. 须提供基本材料，含营业执照或法人证书、近 3 年财务状况等证明材料；

2. 须提供技术基础材料，能证明高质量数据集已具备的技术基础条件，亦可提供系统关键页面截图以直观展示用户界面和操作界面；

3. 须提供应用情况材料，用于证明单位提供数据资源、数据产品及数据服务的相关情况；

4. 可提供运营情况材料，能体现运营团队成立情况、与数据提供方、使用方、服务方等生态主体的合作情况，以及内部编制或发布的高质量数据集运营与流通利用规则、流程、标准等。

（上述证明材料包括但不限于客户服务合同、客户出具的场景应用成效证明、已完成的规则机制扫描件、产品专利和知识产权证书、申报通知发布日期前的系统和软件运行日志、系统访问地址与账号等相关支撑材料）