附件1

山东省2025年度重点行业语料库揭榜挂帅

项目申报书

|  |  |
| --- | --- |
| 项 目 名 称： |   |
| 申报单位（盖章）： |   |
| 填 报 日 期： |  年 月 日 |

申报要求

1.请按照申报书的基本格式撰写申报书，用A4幅面编辑。

2.试点项目基本信息表表格字体为4号仿宋体GB\_2312（单倍行距）。试点项目可行性报告正文字体为3号仿宋体GB\_2312（单倍行距），一级标题3号黑体，二级标题3号楷体GB\_2312。

3.按照实际页码更新“目录”。

4.材料顺序打印，试点项目基本信息表试点项目可行性报告、相关附件之间插入分隔纸合并简装。

申报单位声明

根据《关于组织开展2025年度重点行业语料库揭榜挂帅试点工作的通知》要求， （牵头单位）、 （联合体成员 ）、（ 联合体成员） 联合提交了 试点项目申报书参评。

有关情况声明如下：

1.我单位对提供参评的全部资料真实性负责，符合申报条件，不存在安全生产、数据安全、严重失信等限制性条件情形，不存在弄虚作假行为。

2.我单位在参评过程中所涉及的项目内容和程序皆符合国家有关法律法规及相关产业政策要求。

3.该项目为新实施项目，前期未获得过省级财政资金支持。

4.我单位对所提交的项目内容负有保密责任，按照国家相关保密规定，所提交的项目内容未涉及国家秘密、商业秘密和其他敏感信息。

5.我单位申报项目所填写的相关文字和图片已经审核，确认无误。

6.我单位将按照计划要求推动项目实施，并配合开展与该项目有关的监督检查、绩效评价、项目审计、示范推广等工作。

我单位对违反上述声明导致的后果承担全部法律责任。

牵头单位：

法定代表人：（签字）

单位：（单位盖章）

二〇二五年 月 日

联合体成员：

法定代表人：（签字）

单位：（单位盖章）

二〇二五年 月 日

目 录

一、试点项目基本信息表···············X页

二、试点项目可行性报告···············X页

三、相关附件····················X页

1. 试点项目基本信息表

（请如实填写，作为评审、实施和验收依据）

|  |
| --- |
| **牵头**单位基本信息 |
| 单位名称 |  |
| 统一社会信用代码 |  | 注册时间 |  |
| 单位地址 |  |
| 联系人 | 姓名 |  | 座机 |  |
| 职务 |  | 手机 |  |
| 邮箱 |  |
| 是否上市公司 | □否□是（上市时间： ，上市板块： ） |
| 研发资质 | 分条列明，包括省级及以上认定的企业技术中心、重点实验室、工程技术研究中心、数字经济创新平台等资质。 |
| 研发成果 | 近三年新增软件著作权、发明专利（已公开）、标准等，不限于人工智能领域。 |
| 企业主要财务指标 |
| 年度 | 企业总收入 | 主营业务收入 | 营收同比增速 |
| 2024 |  |  |  |
| 2023 |  |  |  |
| 企业研发指标（万元） |
| 研发投入 | 2024 |  | 研发投入占主营业收入比例（%） | 2024 |  |
| 2023 |  | 2023 |  |
| 主要人员情况 |
| 研发人员数 |  | 研发人员占职工总数比例 |  |
| 主要产品（领域）及相应收入 | 主要产品（领域） | 收入（万元） | 占总收入比例 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| 申报主体简介 | （发展历程、主营业务、市场销售、语料建设相关技术成果转化能力、企业荣誉等方面基本情况，不超过600字） |
| 项目基本信息 |
| 项目名称 |  |
| 申报行业 | 围绕31个制造业行业分类（具体行业分类见附件2） |
| 实施周期 |  年 月— 年 月 | 项目投资额（万元） | （总投资金额和实施期内投资金额） |
| 所获与项目直接相关的知识产权情况（近三年，最多写5项） |
| 发明专利名称 |  | 获取时间 | 年 月 日 |
|  | 年 月 日 |
| 软件著作权名称 |  | 获取时间 | 年 月 日 |
|  | 年 月 日 |
| 开发方式（“√”选） | □自主开发 □产学研联合开发□引进技术消化吸收 □集成创新□其它 |
| 语料库权益（“√”选） | □本单位独立开发，拥有语料库全部所有权□本单位与外单位合作开发，联合拥有语料库所有权□本单位与外单位合作开发，本单位拥有语料库全部所有权□由外单位技术转让，本单位拥有语料库全部所有权□其它 |
| 联合单位 | （按照项目参与程度顺序排列） |
| 联合单位项目参与情况 | （资源支持，如资金/人员/技术/设备、量化指标协同、创新/成果贡献等方面项目建设情况，不超过800字） |
| 项目简介 | 阐述行业语料库的整体建设背景、发展定位及核心目标，详尽说明语料的主要来源、采集与更新机制、数据规模及结构特点。介绍语料库的底层技术架构、数据管理与安全保障措施、质量管控体系和持续优化流程。同时，明确语料库的运营管理模式，突出其在行业内的专业性、权威性及独特价值，展现其在行业应用、学术研究、标准制定等方面的典型案例或成果，不超过1000字。 |
| 行业语料库数据种类与体量现状。 | 1．数据种类。需对现有数据的种类及体量进行详细阐述。数据按文本、图像、音视频、多模态四大类别划分；数据进一步细分为预训练类、指令微调类、偏好类、时序类；多模态数据需明确不同模态数据的种类配比，以及模态间的组织方式（如关联逻辑、交互结构等）。2．各类数据均需提供100条数据样例作为参考，其中：标注类数据样例需包含完整标注内容；视频类数据样例需配套提供100张截图，说明对应视频所涵盖的行业知识。3．预训练文本需提供数据总量，其他文本类数据需提供数据条目数和总字数（中文以汉字数为单位，英文以单词数为单位）。图像、音视频类数据，需提供标注情况，及数据条目数，总体数据体积（如GB、TB）。4．数据体量和数据条目数的证明，需配套提供数据库查询截图（建议采用SQL查询方式，需同时附SQL语句及对应的查询结果截图）。5.数据重复性说明。建议采用不低于10个核心关键词检索结果的30条数据，用于专家校验数据重复性（提供SQL语句与查询结果截图），或采用其他能够表明语料库数据重复性的方法。6.数据安全性保障机制建设情况，如是否包含违法违规内容，如暴力、歧视、色情等不良信息。 |
| 项目承担主体基础条件 | 项目承担主体的存储与算力等要素保障能力，说明可用数据存储空间，语料库服务方式；企业数字化程度，数据的有效汇聚、治理和流通的基础。 |
| 技术支撑能力 | 1.具有明确的语料库建设方向和产品发展规划。方向和规划清晰，可行性情况。2.重点实验室、工程技术中心建设情况，拥有人工智能领域研发平台情况。 |
| 试点项目预期投入 | 预期总投入： （万元）其中：牵头单位 （万元）联合单位 （万元）（项目实施期内，联合体成员拟投入的研发和实施费用，包括但不限于与项目相关的设备、材料、产品、软件等费用，以及与项目实施相关的咨询、诊断、设计、检测、评价服务、人员等费用，**在可行性报告中提供预算概要明细**） |
| 近三年省部级以上重大项目或国家标准制定项目参与情况 | （描述重点项目中标情况、项目进度、验收情况，国家标准制定项目参与、制定过程及标准实施情况等，不超过400字） |
| 预期交付物和其他成果 | 预期交付成果包括标准化、模块化行业语料库产品及解决方案等，要有明确的可量化的指标，分条目阐述，不超过1000字 |
| 发明专利数量： 个，目前已获得 个 |
| 发明专利名称 |  | 获取时间 | 年 月 日 |
|  | 年 月 日 |
| 软件著作权数量： 个，目前已获得 个 |
| 软件著作权名称 |  | 获取时间 | 年 月 日 |
|  | 年 月 日 |
| 牵头或参与制定相关标准数量： 个，目前已获得国家标准 个，行业标准 个，团体标准 个；已发布 个，其中国家标准 个，行业标准 个，团体标准 个 |
| 标准名称及类型 | （国标/行标/团标） | 发布时间 | 年 月 日 |
| 预期应用推广成效 | 1.语料库应用的场景。说明数据可覆盖的具体使用场景。例如具体使用的工种、工序、环节等。2.预期新增销售收入、服务大模型企业数量以及组织供需对接、技术交流等活动情况。3.其他应用推广成效。验收时需要场景应用单位提供证明。 |
| 语料库开放形式 | 部分无偿开放 | 例如，公共语料主动开放 |
| 全部无偿开放 |  |
| 有偿开放 |  |

1. 试点项目可行性报告

（一）试点项目内容

1.行业语料库需求分析（不超过500字）

（阐述行业语料库建设背景，分析面临的痛点问题及共性需求，说明建设行业语料库的必要性、迫切性。）

2.建设实施方案（不超过5000字）

（介绍试点项目实施方案的主要内容，包括但不限于实施方案总体架构、关键技术攻关、产品研发、系统及设备应用、调试优化、推广应用等。）

3.行业语料库现状（不超过500字）

（对照项目验收预期指标，说明项目目前实际情况。）

4.预期目标（不超过1000字）

（阐述**目标成果**和**示范应用预期成效**。**目标成果**包括交付物、研发成果、知识产权等；**示范应用预期成效**包括语料库应用的场景、预期新增销售收入、服务大模型企业数量以及组织供需对接、技术交流等活动情况等**至少3项，**确保可量化、可评估、可验证。）

5.投资概算（不超过500字）

（按照实施方案，综合测算项目实施期内投资额度，包括牵头单位项目投资额、带动场景落地单位投资额以及联合体成员投资额，按用途列明主要费用概算，需要列明资金投入方）

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 用途 | 金额（万元） | 投入方 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

6.进度安排（不超过300字）

（试点项目实施任务进度安排）

7.创新性（不超过500字）

（介绍项目实施过程中采用的创新技术或应用）

8.典型性及可推广性（不超过500字）

（阐述揭榜任务目标成果在同行业、同体量企业中可复制推广性。）

9.语料库安全性与合规性：对敏感信息进行脱敏处理、遵循相关隐私法规；语料来源的版权情况，避免侵权风险；数据语料安全制度建设情况，数据安全风险处置情况。

10.未来发展规划：语料库项目建设推广规划、迭代升级方向。

（二）联合单位基础条件（不超过1000字）

（从研发条件、技术能力、产品水平、实施能力、应用案例等方面对揭榜基础条件进行描述。）

（三）项目负责人和团队基本情况（不超过500字）

（项目负责人及项目实施团队的技术水平、主要成果和实施业绩等情况介绍，同时提供团队人员名单、所属单位和任务分工）

（四）项目组织实施（不超过1000字）

（组织管理方式、协调机制、保障措施、应用推广计划等，联合体实施机制等。）

1. 相关附件

联合体成员申报必要的证明文件。

1.联合体各成员法人营业执照或事业单位法人证书副本复印件。

2.联合体协议，需明确责任分工和权利义务，细化项目目标任务、合作内容、工作计划、出资方式和知识产权归属等内容。

3.联合体各成员近三年财务状况证明材料。

4.牵头单位省级以上荣誉复印件。

5.联合体各成员在信用中国（山东）官网：https://credit.shandong.gov.cn/snipet/1028.html，查询的《经营主体公共信用报告（无违法违规记录证明普通版）》。

6.联合体各单位法定代表人在全国法院失信被执行人名单信息公布与查询平台：http://zxgk.court.gov.cn/shixin/，的查询截图。

7.中央驻鲁企业、省属国企的子公司申报的需提供省级公司推荐函。