

# 山东省工业和信息化厅 山东省财政厅

---

## 山东省工业和信息化厅 山东省财政厅 关于组织开展 2026 年度重点行业语料库 揭榜挂帅项目申报工作的通知

各市工业和信息化局、财政局：

为深入贯彻全省“人工智能+”创新应用工作会议精神，认真落实《山东省人民政府办公厅印发〈关于支持人工智能全产业链创新发展的若干政策措施〉的通知》要求，加快打造一批高质量、标准化、可流通的行业语料库，赋能一批行业大模型、工业智能体等应用落地，根据《山东省省级“语料券”奖补资金管理办法及实施细则》，现组织开展 2026 年度重点行业语料库揭榜挂帅项目申报工作。有关事项通知如下：

### 一、任务内容

聚焦工业制造重点行业的基础理论研究、产品研发设计、生产管理运行、过程质量检测等关键环节和特定场景的知识语料汇聚，基于结构化数据（如数据库中的表格数据）、非结构化数据（如文本、图像、语音等）和半结构化数据（如日志文件、XML 文件等），通过清洗、去噪和统一格式，用于支持自然语言

---

处理、计算机视觉、机器学习、深度学习等任务，满足行业大模型或场景大模型开发、训练和微调需求的高质量语料库。

## 二、推荐条件

### （一）申报单位条件

1. 具有独立承担民事责任能力的企业，经营状态正常、近三年内无严重失信记录，符合财政涉企资金“绿色门槛”制度要求；
2. 拥有与行业语料库建设相关的核心数据、关键技术、自主知识产权等基础能力；
3. 具备语料收集、整理、标注所需的软硬件设施和专业团队。鼓励项目承担主体联合具有优势学科的高校或科研机构编制建设方案；
4. 进行数据语料采购或治理时，不能违反国家安全、商业秘密、个人隐私、科技伦理等相关政策法规要求，完成必要的安全合规处理。

### （二）申报项目条件

1. 项目为正在实施或计划实施的项目，具有加快实施的必要性、重要性、紧迫性，语料库产品通过持续优化与高质量扩充，能够显著提升大模型训练的效率精度与安全可靠水平，建设周期一般不超过一年；
2. 在技术或产品性能上有较大创新，在特定领域或行业内具有高度的专业化和适用性，有明确的市场需求和较强的市场竞争

争力；

3. 项目应有明确的应用推广方案及生态建设预期，预期目标包括但不限于新增销售收入、服务大模型企业领域及数量，以及组织供需对接、交流推广等活动情况；

4. 项目要有明确的实施计划和资金保障，确保能够按计划完成建设；

5. 项目验收时行业相关语料库数据量不低于 10 万条，具有较高的数据质量、领域覆盖程度、潜在价值和应用成效，项目验收时应通过第三方测评；

6. 鼓励各行业语料库项目加快语料资源整合，积极开放公共语料。

### （三）重点行业

为加快实现对制造业行业门类全覆盖，本次申报重点围绕高端装备、烟草制品业、农副食品加工业、家具制造业、木材加工、皮革毛皮羽毛及其制品和制鞋业、仪器仪表制造业、废弃资源综合利用业等行业，重点推进行业关键数据技术攻关、行业数据语料标准研制、高质量行业语料库打造、语料应用场景落地等。

## 三、申报程序

（一）企业申报。各市工业和信息化局会同市财政局按照企业自愿原则组织申报，省属及中央驻鲁单位按属地申报。申报企业按要求编制《山东省重点行业语料库揭榜挂帅项目申报书》

(见附件1)，申报材料须符合国家保密相关规定。

(二) 初审推荐。各市工业和信息化局牵头对申报项目进行评审论证，于2月27日前将推荐文件、汇总表、申报材料纸质版(一式两份)，以两部门名义正式行文报至省工业和信息化厅、省财政厅，电子版申报材料(盖章PDF版及Word版)同步发至指定邮箱。

(三) 评审公布。省工业和信息化厅组织专家或第三方专业机构对申报项目进行评审，综合采取材料评审、现场答辩等多种方式，确定符合条件的揭榜项目名单，将评审结果商省财政厅同意后予以公示。

(四) 验收评估。省工业和信息化厅将进行动态监督，根据项目实施实际情况，于2026年下半年进行集中验收评估，并对项目实施期内的实际投入资金、建设应用成效等进行核查，公布验收评估结果。

#### 四、有关要求

(一) 项目申报单位须对申报材料中涉及的指标、数据和相关证明、附件材料的真实性负责，承担因提供虚假材料造成的后果。

(二) 省工业和信息化厅不委托任何单位或个人从事项目申报代理服务工作，申报单位申报书和数据等信息资料仅用于项目评审和过程管理使用。

## 五、联系方式

省工业和信息化厅：杨洪于 0531 - 51782640, 18888302292

省财政厅：马子腾 0531 - 51769593

邮 箱：cytjc@shandong.cn

地 址：济南市历下区省府前街1号省工信厅数据产业推进处（邮编250011）

- 附件：
1. 山东省2026年度重点行业语料库揭榜挂帅项目申报书
  2. 制造业行业目录
  3. 山东省重点行业语料库揭榜挂帅项目推荐汇总表



附件 1

# 山东省 2026 年度重点行业语料库揭榜挂帅

## 项目申报书

项 目 名 称: \*\*\*行业 AI 语料库

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 2026 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）

二〇二六年 月 日

联合体成员：

法定代表人：（签字）

单位：（单位盖章）

二〇二六年 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

# 一、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	<input type="checkbox"/> 否 <input type="checkbox"/> 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。			
牵头单位近三年发展情况				
年度	企业总收入	主营业务收入	资产负债率	营收同比增速
2025				
2024				
2023				

企业研发指标（万元）					
研发投入	2025		研发投入占主营业务收入比例（%）	2025	
	2024			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)	<input type="checkbox"/> 自主开发 <input type="checkbox"/> 引进技术消化吸收 <input type="checkbox"/> 其它	<input type="checkbox"/> 产学研联合开发 <input type="checkbox"/> 集成创新	
语料库权益 ("√" 选)	<input type="checkbox"/> 本单位独立开发，拥有语料库全部所有权 <input type="checkbox"/> 本单位与外单位合作开发，联合拥有语料库所有权 <input type="checkbox"/> 本单位与外单位合作开发，本单位拥有语料库全部所有权 <input type="checkbox"/> 由外单位技术转让，本单位拥有语料库全部所有权 <input type="checkbox"/> 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)		

领域与行业情况概述	<p>简要介绍语料库所属行业有关数字化转型、语料库开发利用等基本情况、痛点问题、发展定位、核心目标及场景探索等，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、标准制定等方面典型案例或成果，不超过 1000 字。</p>
已有工作基础	<p>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）。</p> <p>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。</p> <p>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）。</p> <p>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）。</p> <p>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。</p> <p>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。</p>
项目承担主体基础条件	<p>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。</p>

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	<p>预期总投入：_____（万元）      其中：牵头单位 _____（万元）      联合单位 _____（万元）</p> <p>（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）</p>			
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			
预期应用推广成效	标准名称及类型	(国标/行标/团标)	发布时间	年月日
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			

## 二、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。）

### 三、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/snippet/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。

## 附件 2

# 制造业行业目录

1	制造业	农副食品加工业
2		食品制造业
3		酒、饮料和精制茶制造业
4		烟草制品业
5		纺织业
6		纺织服装、服饰业
7		皮革、毛皮、羽毛及其制品和制鞋业
8		木材加工和木、竹、藤、棕、草制品业
9		家具制造业
10		造纸和纸制品业
11		印刷和记录媒介复制业
12		文教、工美、体育和娱乐用品制造业
13		石油、煤炭及其他燃料加工业
14		化学原料和化学制品制造业
15		医药制造业
16		化学纤维制造业
17		橡胶和塑料制品业
18		非金属矿物制品业
19		黑色金属冶炼和压延加工业
20		有色金属冶炼和压延加工业
21		金属制品业
22		通用设备制造业
23		专用设备制造业
24		汽车制造业
25		铁路、船舶、航空航天和其他运输设备制造业
26		电气机械和器材制造业
27		计算机、通信和其他电子设备制造业
28		仪器仪表制造业
29		其他制造业
30		废弃资源综合利用业
31		金属制品、机械和设备修理业

附件 3

## 山东省重点行业语料库揭榜挂帅项目 推荐汇总表

推荐单位(盖章):

填表人:

联系电话:

时间: 年 月 日

序号	市	语料库名称	牵头单位名称	行业领域	语料库数据总量(TB)	建设性质 (在建/规划)	联系人	联系电话
1	XX 市							
2								
3								
.....	.....	.....	.....	.....	.....	.....	.....	.....

注: 各市按照推荐优先级进行排序。