

附件 1

广西财经学院 2026 年大数据与人工智能学院数据采集服务采购报价单

技术要求						
品名	规格参数	单位	数量	是否	金额	备注
				响应	(元)	
中国东盟多源数据服务	<p>(一) 东盟国家多模态新闻数据</p> <p>1. 数据范围</p> <p>★覆盖东盟核心国家（至少包括泰国、越南、印度尼西亚、马来西亚、菲律宾、新加坡 6 个国家）的主流媒体新闻信息。</p> <p>★数据采集时间范围为 2025 年 1 月 1 日至今，数据更新截止时间不早于投标截止日前 30 天。</p> <p>2. 数据内容</p> <p>★新闻数据总量不少于 10 万条。</p> <p>★每条新闻数据须包含以下字段：原始标题、原始正文内容、原始语言标识、中文翻译标题、中文翻译正文、中文摘要（不少于 100 字）、发布媒体名称、发布媒体所属国家、新闻发布时间、新闻采集时间、新闻 URL 来源、新闻分类标签。</p> <p>3. 数据质量要求</p> <p>★中文翻译须采用机器翻译与人工智能相结合的方式，翻译准确率不低于 90%。</p> <p>★中文摘要须能够准确概括新闻核心内容，摘要与原文内容一致性不低于 85%。</p> <p>4. 媒体来源要求</p> <p>★每个国家至少覆盖 3 家主流媒体/新闻网站，供应商须在响应文件中提供拟采集媒体清单。</p> <p>6. 数据分类</p> <p>新闻数据须按政治、经济、科技、教育等至少 4 个主题分类标签进行标注，分类准确率不低于 85%。</p> <p>★供应商须提供不少于 5 张新闻数据相关截图供现场查验，包括但不限于：数据库表结构及数据条数截图、新闻全文检索结果截图、中文翻译与摘要结果截图、分类标签标注效果截图等，截图须清晰可辨。</p> <p>(二) 东盟主要国家电商数据</p> <p>1. 数据范围</p> <p>★覆盖 Shopee 平台至少 6 个东盟核心国家站点（泰国、越南、印度尼西亚、马来西亚、菲律宾、新加坡）的商品及评价数据。</p> <p>★数据采集时间范围为 2016 年 1 月 1 日至今。</p> <p>2. 商品数据</p> <p>★商品数量不少于 10 万条。</p> <p>★每条商品数据须包含以下字段：商品 ID、商品名称、商品分类（一至四级分类）、商品价格（原币种及人民币折算）、商品原产地、商品品牌、商品规格参数、商品详情页 URL、</p>	项	1			

	<p>店铺名称、店铺 ID、商品上架时间、数据采集时间、商品评分。</p> <p>3. 评价数据</p> <ul style="list-style-type: none"> <li>★商品评价总量不少于 1 亿条。</li> <li>★每条评价数据须包含以下字段：评价 ID、关联商品 ID、评价内容、评价评分、评价时间、评价者 ID、数据采集时间。</li> </ul> <p>4. 数据质量要求</p> <ul style="list-style-type: none"> <li>★商品分类须采用 Shopee 平台官方分类体系，分类层级至少覆盖一至四级，分类完整率不低于 95%。</li> <li>★评价数据去重率不低于 98%（同一商品同一用户的重复评价须去重处理）。</li> </ul> <p>（三）东盟主要国家统计数据</p> <p>1. 数据范围</p> <ul style="list-style-type: none"> <li>★至少覆盖 2 个东盟核心国家的完整统计数据（新加坡、印度尼西亚）。</li> <li>★统计数据须完整覆盖，与官方提供数据一致。</li> </ul> <p>2. 数据内容</p> <ul style="list-style-type: none"> <li>★每个国家的统计数据须至少涵盖以下主题领域（不少于 8 个）：国民经济核算、人口与就业、价格指数、对外贸易、工业与能源、农业、财政与金融、交通运输与邮电、教育与科技、旅游、环境。</li> <li>★统计数据须包含完整的层级结构：统计主题→统计表格→详细数据项→数据值，每层级均须有中文名称和原语言名称。</li> <li>★每个国家统计数据表格不少于 200 张，数据项不少于 5000 个。</li> </ul> <p>3. 数据来源</p> <ul style="list-style-type: none"> <li>★统计数据须来源于各国官方统计机构，供应商须在响应文件中注明数据来源。</li> </ul> <p>4. 数据结构化要求</p> <ul style="list-style-type: none"> <li>★统计数据须完整结构化，包括指标编码、指标名称、计量单位、时间维度、地区维度、数值、数据来源等字段，确保每条数据可精确定位到原始统计表格。</li> </ul> <p>5. 数据更新</p> <p>供应商须提供数据更新机制说明，包含年度更新计划及更新频率。</p> <ul style="list-style-type: none"> <li>★供应商须提供不少于 5 张电商数据相关截图供现场查验，包括但不限于：数据库商品表结构及记录数截图、四级分类浏览截图、评价数据表结构及记录数截图等，截图须清晰可辨。</li> </ul> <p>（四）中国统计年鉴数据</p> <p>1. 数据范围</p> <ul style="list-style-type: none"> <li>★采集《中国统计年鉴》2010 年至 2025 年共计 15 个年度的完整数据，覆盖年鉴全部章节及统计表格。</li> </ul> <p>2. 数据内容</p> <ul style="list-style-type: none"> <li>★数据须覆盖《中国统计年鉴》以下全部章节（不少于 25 个）：一、行政区划与自然地理；二、人口；三、国民经济核算；四、就业人员和职工工资；五、国家财政；六、资源与环境；七、能源；八、财政；九、价格指数；十、人民生活；十一、城市概况；十二、对外贸易；十三、农业；十四、工业；十五、建筑</li> </ul>					
--	---	--	--	--	--	--

	<p>业；十六、运输邮政业；十七、批发零售业；十八、住宿餐饮业；十九、旅游业；二十、金融业；二十一、教育；二十二、科技；二十三、卫生健康；二十四、体育；二十五、公共管理、社会保障和社会组织。</p> <p>★每个年度统计表格不少于 300 张，数据项不少于 1 万个，5 个年度合计数据项不少于 5 万个。</p> <p>3. 数据来源</p> <p>★数据须来源于国家统计局官方发布的《中国统计年鉴》（电子版），供应商须在响应文件中注明数据获取渠道。</p> <p>4. 数据结构化要求</p> <p>★统计年鉴数据须提供完整 excel 表格，包括：年鉴年度、章节编号、章节名称、表格编号、表格名称、指标编码、指标名称、计量单位、时间维度（年份）、地区维度（省/自治区/直辖市）、数值、备注。确保每条数据可精确定位到年鉴原始表格。</p> <p>5. 数据质量要求</p> <p>★统计年鉴数据与国家统计局官方发布数据一致性不低于 99%。</p> <p>（五）马来西亚人工智能产业主题知识图谱</p> <p>基于本需求书第（一）部分采集的马来西亚新闻数据，进行分类清洗后，抽取人工智能产业主题相关的实体、关系和属性，形成完整的知识图谱，以 Neo4j 图数据库方式交付，同步交付知识图谱抽取、清洗、构建的程序源码，并提供部署支持服务。</p> <p>1. 数据来源</p> <p>★知识图谱须基于本需求书第（一）部分采集的马来西亚新闻数据进行构建，供应商须说明从原始新闻数据到知识图谱的完整数据处理流程。</p> <p>2. 实体抽取要求</p> <p>★实体总数不少于 1,000 个，须进行本体约束（控制在 8 类）：人工智能企业/机构、AI 产品/服务、AI 技术/算法、AI 领域人物、AI 相关政策/法规、AI 相关事件/活动、AI 应用领域/场景、AI 学术研究/论文。</p> <p>★每个实体须包含以下属性：实体 ID、实体名称（中文翻译）、实体类型、实体描述（中文，不少于 50 字）、首次出现时间、最后出现时间、出现频次、数据来源标识。</p> <p>3. 关系抽取要求</p> <p>★关系总数不少于 3,000 条，须进行关系约束（控制在 15 类）：企业开发产品、企业投资企业、企业合作企业、人物任职企业、政策监管领域、技术应用于领域、事件涉及企业、事件涉及技术、产品使用技术、学术研究关联技术、企业参与活动、政府支持政策、技术衍生技术、产品服务于领域、机构发布研究。</p> <p>★每条关系须包含以下属性：关系 ID、起始实体 ID、关系类型、目标实体 ID、关系描述（中文）、关系强度/权重、数据来源标识。</p> <p>4. 主题分类与清洗</p> <p>★须对马来西亚新闻数据按人工智能产业主题进行分类标注，AI 相关新闻筛选覆盖率不低</p>				
--	---	--	--	--	--

	<p>于 95%。</p> <p>★须对抽取结果进行清洗去重，实体去重率不低于 95%，关系去重率不低于 95%，杜绝同一实体/关系的重复记录。</p> <p>5. 图谱交付格式</p> <p>★知识图谱须以 Neo4j 图数据库（5.x 及以上版本）方式交付，包含完整的节点（实体）、边（关系）及属性数据，图谱数据库须可直接启动并查询。</p> <p>★须提供 Neo4j 数据库完整备份文件（.dump 格式）及 Cypher 脚本导入方式双重交付，确保采购方可在自有服务器上独立部署和恢复全部图谱数据。</p> <p>6. 源码交付</p> <p>★须同步交付对新闻数据进行知识图谱抽取、清洗、构建的完整程序源码，源码须包含以下模块：数据预处理模块、实体抽取模块、关系抽取模块、属性填充模块、去重清洗模块、图谱构建模块。</p> <p>★源码须附带完整的技术文档，包括：各模块功能说明、运行环境要求（Python 版本、依赖包清单）、配置参数说明、运行步骤说明、输入输出格式说明。</p> <p>7. 部署支持服务</p> <p>★供应商须提供知识图谱系统部署支持服务，包括：Neo4j 数据库环境搭建、图谱数据导入、源码环境配置及运行验证，部署支持服务不少于 3 次现场或远程支持，每次不少于 4 小时。</p> <p>8. 数据质量要求</p> <p>★供应商须提供不少于 5 张知识图谱相关截图供现场查验，包括但不限于：Neo4j 数据库节点/关系全景截图、Cypher 查询语句及结果截图（至少 2 种不同类型查询）、源码运行过程截图（至少 2 个模块）、知识图谱可视化展示（至少 2 类实体/关系的子图）、图谱 schema 结构截图等，截图须清晰可辨。</p> <p>（六）结构化数据库建设及数据字典</p> <p>★除特别说明的 Excel 格式外，所有交付数据必须以结构化形式保存至关系型数据库（MySQL 8.0 及以上版本或 PostgreSQL 12 及以上版本），不得以 PDF、Word、Excel 等非结构化文件形式交付。</p> <p>★供应商须提供完整的数据字典，数据字典须包含：数据库名称、数据表清单、每张表的表名及中文名称、每张表的字段清单（字段名、中文名称、数据类型、字段长度、是否允许为空、默认值、字段说明、约束条件）、表间关系说明、索引设计说明。</p> <p>★数据字典须以结构化文档（数据库 DDL 脚本+Excel/CSV 格式数据字典文档）双重形式交付，DDL 脚本须可直接在指定数据库版本上执行创建数据库及数据表。</p> <p>数据库字符集须统一采用 UTF-8（utf8mb4），以支持多语言文字存储。</p> <p>★供应商须提供数据库初始化脚本、数据导入脚本及完整的部署文档，确保采购方可在自有服务器上独立部署和恢复全部数据。</p>					
--	--	--	--	--	--	--

<b>商务要求</b>	
参数要求	<p>1、本采购需求书所列带“★”号条款为实质性响应条款，供应商须完全满足，若有一项带“★”号条款未响应或不满足，将按无效投标处理。</p> <p>2、本采购需求书所列带“▲”号条款为重要技术指标，供应商无法满足或响应时将影响其技术评分。</p> <p>3、供应商应对本需求书所有条款逐条响应，并在响应文件中提供详细的技术方案、实施方案及相关证明材料。</p> <p>4、本项目预算金额为人民币壹拾万元整（¥98,000.00），投标报价超过预算金额的按无效投标处理。</p> <p>5、供应商须保证所提供数据的合法性、真实性、准确性和完整性，并承担由此产生的一切法律责任。</p>
交付要求	<p>交付时间 ★合同签订之日起5个日历日内完成全部数据交付及部署。</p> <p>交付方式 ★数据须部署至采购方指定的服务器环境（Linux操作系统，MySQL 8.0/PostgreSQL 12及以上版本），确保数据可正常访问和查询。 同时提供完整数据库备份文件（.sql格式），便于采购方自主备份和迁移。</p>
验收要求	<p>★验收分为数据完整性验收、数据质量验收、数据库部署验收三个环节，三个环节全部合格方为验收通过。</p> <p>（一）数据完整性验收</p> <ol style="list-style-type: none"> <li>1. 各类数据量须达到本需求书规定的最低要求；</li> <li>2. 各类数据字段须完整，关键字段缺失率不超过2%；</li> <li>3. 数据时间范围须覆盖本需求书规定的时间区间。</li> </ol> <p>（二）数据质量验收</p> <ol style="list-style-type: none"> <li>1. 中文翻译准确率不低于90%（随机抽取5%样本评估）；</li> <li>2. 中文摘要与原文一致性不低于85%；</li> <li>3. 电商评价数据去重率不低于98%；</li> <li>4. 新闻分类准确率不低于85%。</li> </ol> <p>（三）数据库部署验收</p> <p>★1. 数据库可正常启动，全部数据表可正常访问；</p> <p>★2. 数据字典与实际数据库结构完全一致；</p> <p>★3. DDL脚本可在全新数据库环境中正确执行，数据导入脚本可正确导入全部数据；</p>
知识产权与保密要求	<ol style="list-style-type: none"> <li>1. 供应商如使用第三方数据源，须确保已获得合法授权或许可，并提供相关证明文件。</li> <li>2. 合同终止后，供应商须销毁其保留的全部交付数据及相关衍生数据，并出具销毁证明。</li> </ol>

报价公司：

法定代表人签字：

联系人及电话：