**附件一：**

山东师范大学招标采购项目需求公示表

**编号:SDSL-2024-175**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 基本情况 | | | | | | |
| 项目名称 | | 山东师范大学人工智能训练服务器系统平台采购项目 | | | | |
| 项目主管单位 | | 山东师范大学信息科学与工程学院 | | | | |
| 第三方咨询服务单位 | | 山东善立招标有限公司 | | | | |
| 项目信息 | | | | | | |
| 包号 | 标的名称 | 数量 | 是否允许进口 | 是否  创新产品 | 是否专门面向中小微或预留份额 | 预算金额  （最高限价） |
| A包 | 人工智能大模型开发训练平台 | 6 | 否 | 否 | 否 | 470万元 |
| 信创算力服务终端 | 2 | 否 | 否 | 否 |
| 信创边缘计算推理服务器 | 1 | 否 | 否 | 否 |
| 万兆业务交换机 | 2 | 否 | 否 | 否 |
| 需求描述 | | | | | | |
| 技术要求 | | 一、项目概况  本项目为山东师范大学人工智能训练服务器系统平台采购项目，共一个包，总预算470万元。  二、技术要求  详见附件详细技术参数。 | | | | |
| 商务要求 | | **1、交货期**：国产设备自合同签订之日起30天内供货、安装调试完毕。  **2、交付地点**：设备全部到达山东师范大学校内指定地点。  **3、付款条件：**  国产设备：招标人在签订合同后7个工作日支付合同金额的30%；中标人供货并安装调试完成，经招标人验收合格后，招标人支付至合同价款的100%。  4、验收：  采购人根据《山东省政府采购履约验收管理办法》及采购人要求进行验收。  4.1货物运抵现场后，采购人将对货物数量、质量、规格等进行检验。如发现货物和规格或者两者都与招标文件、投标文件、合同不符，采购人有权限根据检验结果要求中标人立即更换或者提出索赔要求。  4.2货物由中标人进行安装，完毕后，采购人应对货物的数量、质量、规格、性能等进行详细而全面的检验。安装完毕7日后，证明货物以及安装质量无任何问题，由采购人组成的验收小组签署验收报告，作为付款凭据之一。  5、质量保证期：  5.1质保期：国产设备原厂质保3年。（清单技术规格要求中有要求的按照清单要求执行）。国家主管部门或者行业标准对货物本身有更高要求的，从其规定并在合同中约定，投标人亦可提报更长的质保期，质保及售后须注明质保主体（原厂或投标人）。  5.2质量保证期内，如果证实货物是有缺陷的，包括潜在的缺陷或者使用不符合要求的材料等，中标人应立即免费维修或者更换有缺陷的货物或者部件，保证达到合同规定的技术以及性能要求。如果中标人在收到通知后5天内没有弥补缺陷，采购人可自行采取必要的补救措施，但风险和费用由中标人承担，采购人同时保留通过法律途径进行索赔的权利。  6、售后服务：  6.1投标人应提供及时周到的售后服务，应保证每季度至少一次上门回访、检修。  6.2投标人自报响应、维修时间以及备品备件情况。  6.3投标人应提供7×24小时的技术支持服务，确保故障能及时处理，且需具备远程和现场支持能力。   1. 培训   投标人提供现场技术培训，包括设备的日常保养和维护，操作的技术要领，常见故障处理的技术培训等，直至其掌握操作技能为止。 | | | | |
| 政策要求 | | 严格落实财政部、工业和信息化部《政府采购促进中小企业发展管理办法》（财库〔2020〕46号）、《财政部 司法部关于政府采购支持监狱企业发展有关问题的通知》（财库〔2014〕68号）、《关于促进残疾人就业政府采购政策的通知》（财库[2017]141号）、《关于进一步发挥政府采购政策功能支持中小企业发展的通知》、《关于调整优化节能产品、环境标志产品政府采购执行机制的通知》、《关于印发节能产品政府采购品目清单的通知》、《关于印发环境标志产品政府采购品目清单的通知》等国家最新政府采购政策。 | | | | |
| 绩效目标 | | 1、高效计算能力提升：通过采购高性能服务器，显著提升实验室在深度学习和高性能计算任务中的计算速度和处理能力。  2、科研成果产出：支持人工智能研究的项目实施，促进国家自然科学基金项目的高质量完成，推动科研成果的产出和应用。  3、系统稳定性和安全性：确保设备长时间稳定运行，减少故障率，提升系统的安全性，保障数据安全。  4、节能环保：实现设备在高效运作的同时符合国家节能环保标准，降低能耗并减少环境影响。  5、技术人员能力提升：通过培训，提升实验室技术人员的设备操作和维护能力。 | | | | |
| 其他要求 | | 1、投标人资格要求：  （1）符合《中华人民共和国政府采购法》第二十二条的规定。  （2）在“信用中国”（www.creditchina.gov.cn）、中国政府采购网（www.ccgp.gov.cn）、“信用山东”（credit.shandong.gov.cn）等网站中被列入失信被执行人、重大税收违法失信主体、政府采购严重违法失信行为记录名单的投标人，不得参加本次政府采购活动；  （3）单位负责人为同一人或者存在直接控股、管理关系的不同投标人，不得参加同一合同项下（同一包号）的政府采购活动；  （4）本项目不接受联合体投标。 | | | | |

**附件、详细技术参数**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 设备名称 | 数量 | 单位 | 是否强制节能 | 是否核心产品 | 是否接受进口 | 重要程度 | 指标需求 |
| 1 | 人工智能大模型开发训练平台 | 6 | 套 | 否 | 是 | 否 | 一般 | 1、形态：机架式服务器，高度≤4U。 |
| 一般 | 2、处理器：配置2颗非ARM架构处理器，单颗≥24个核心、≥48个线程，主频≥2.1GHz。 |
| 一般 | 3、内存：配置不低于1024GB 3200MHz DDR4内存。 |
| 一般 | 4、硬盘：配置≥2块960G SSD硬盘，≥4块7.68TB NVMe。 |
| 一般 | 5、RAID支持：配置≥2G缓存独立raid卡，支持RAID 0, 1, 5, 6, 10, 50, 60。 |
| ★ | 6、GPU：配置≥8块CUDA生态GPU加速卡，单卡提供≥48G显存、FP32算力≥58TFLOPS，FP8算力≥230 TFLOPS，显存带宽≥850GB/s。 |
| △ | 7、GPU拓扑：支持在不同场景下CPU-GPU间连接拓扑切换。 |
| 一般 | 8、网络：配置≥2个10G网络端口（含光模块），≥2个千兆网口。 |
| △ | 9、PCIe扩展：配置8个GPU后仍支持≥4个PCIe槽位。 |
| 一般 | 10、管理功能：提供服务器统一管理功能，实现不同品牌、不同平台的服务器统一管理，监控告警、性能分析、批量配置更新、OS部署等功能。 |
| 一般 | 11、模型微调功能  11.1.图形化界面：支持大模型微调，提供SFT任务的图形化界面；支持开源大模型包括DeepSeek、GLM、Qwen、Baichuan、Llama、YI等。 |
| 一般 | 11.2.参数在线配置：支持在线配置微调参数，包括训练轮数、batchsize、学习率、加速方式；支持配置参数，包括梯度累积、学习率调节器、验证集比例、优化器等。 |
| 一般 | 11.3.断点续训：支持自动从上一次任务基础上继续训练的能力。 |
| 一般 | 11.4.模型评估：支持自动模型评估。 |
| 一般 | 11.5.模型仓库：支持平台本身模型和第三方大模型的导入、发布、导出等统一管理；支持模型发布，可以生成服务配置。 |
| 一般 | 12、服务部署功能  12.1.部署方式：提供模型部署和镜像部署两种方式。 |
| 一般 | 12.2.服务监控：支持CPU利用率、内存使用率、内存已用、内存总量、部署网络传输速度、部署网络接收速率、请求数、请求成功率、请求延迟等服务的监控。 |
| 一般 | 12.3.弹性缩扩容：支持服务可基于实时并发请求数或者CPU自动触发服务实例扩缩容。 |
| 一般 | 12.4.服务测试：支持服务的在线测试、服务评估和周期性服务部署。 |
| 一般 | 13、应用开发功能  13.1.知识检索：支持知识库创建、编辑和删除；支持多种格式的知识库文档上传和解析；支持检索类型、引用数量、关键词相似度权重、混合相似度权重等配置检索策略；支持配置Prompt模板。 |
| 一般 | 13.2.插件管理：支持敏感词检测、知识库检索、网络检索、文档阅读插件。 |
| 一般 | 13.3.支持在线提供对话应用。 |
| 一般 | 14、系统管理功能  14.1.租户管理：支持业务模块的使用权限控制，支持用户的资源配额设置，支持用户的批量管理。 |
| 一般 | 14.2.资源组管理：支持设置资源的共享、专享策略。 |
| 一般 | 14.3.集群监控：提供多层级监控信息查看，提供GPU、CPU、内存、IB/RoCE网络、存储等关键设施的性能曲线、图表展示。 |
| 一般 | 14.4.告警：支持节点CPU、内存、网络、磁盘、加速卡健康状态及性能情况的报警设置。 |
| 一般 | 15、满配冗余电源模块和风扇模组，配置导轨。 |
|  |  |  |  |  |  |  | 一般 | 16、支持用户对预置的大语言模型进行开发、训练和微调。  •提供Jupyter、Webshell等在线开发环境，支持对接第三方开发工具（如VSCode，PyCharm）；支持0.x张卡(GPU显存隔离/MIG)、单卡/多卡，单机/多机等不同算力环境创建。  提供图形化大模型微调界面，支持微调任务的创建、编辑、删除等操作。支持查看微调日志，模型收敛过程可视化。 |
| 一般 | 17、支持知识库的创建、编辑、合并、拆分和删除。  •支持配置检索策略；支持知识编码；支持在线测试。 |
| 一般 | 18、支持通过可视化方式创建特定功能的大模型助手，支持知识库助手、对话助手和任务助手等多类型助手开发配置。  •支持大模型以挂载知识库、配置prompt的形式进行问答。  •支持用户通过配置prompt设置开发对话助手。  •支持用户配置平台内置或自定义的插件，开发任务助手。 |
| 一般 | 19、  支持与已有服务器平台、存储设备的集成适配。 |
| 一般 | 20、保修：3年原厂质保。 |
| 2 | 信创算力服务终端 | 2 | 套 | 否 | 否 | 否 | 一般 | 1、形态：≥4U。 |
| ★ | 2、处理器：配置2颗非ARM架构处理器，单颗≥32个核心、≥64个线程，主频≥2.2GHz，支持国密算法。 |
| 一般 | 3、内存：配置≥1024GB 3200MHz DDR4。 |
| 一般 | 4、硬盘：配置≥6块1.92T SSD。 |
| 一般 | 5、RAID支持：配置≥4G缓存独立raid卡，支持RAID 0, 1, 5, 6, 10, 50, 60。 |
| 一般 | 6、网络：配置≥2个10G万兆光口（含光模块），配置≥4个千兆网口。 |
| ★ | 7、GPU：配置≥8块CUDA生态GPU加速卡，单卡提供≥48G显存、FP32算力≥58TFLOPS，FP8算力≥230 TFLOPS，显存带宽≥850GB/s。 |
| 一般 | 8、PCIe扩展：配置8块GPU后仍支持≥4个PCIe插槽。 |
| 一般 | 9、管理：支持IPMI2.0、Redfish、SNMP等标准接口，提供故障诊断、自动化运维、性能监控和远程管理，提供对系统/内存/电源/风扇/温度/网络/硬盘等关键部件的故障诊断报警功能等。 |
| 一般 | 10、电源：满配冗余电源，配置导轨。 |
| 一般 | 11、安全：支持TPM 2.0和TCM。 |
| 一般 | 12、保修：3年原厂质保。 |
| 3 | 信创边缘计算推理服务器 | 1 | 套 | 否 | 否 | 否 | 一般 | 1、ARM架构处理器，总核心数≥48，频率≥2.6GHZ。 |
| 一般 | 2、1024G DDR4 3200内存。 |
| 一般 | 3、支持8个硬盘槽位。 |
| 一般 | 4、4千兆口网卡。 |
| 一般 | 5、硬盘：固态硬盘≥3.8T，机械硬盘≥64T。 |
| ★ | 6、2张推理卡，单卡显存≥64GB，INT8≥256 TOPS，FP16≥128 TFLOPS。 |
| 一般 | 7、4G缓存RAID卡一张。 |
| 一般 | 8、满配冗余电源。 |
| 4 | 万兆业务交换机 | 2 | 台 | 否 | 否 | 否 | 一般 | 1、24个1G/10Gbps SFP+端口2个40G/100G QSFP28端口（支持拆分为10G/25G端口使用）。 |
| 一般 | 2、MAC地址表容量：≥80K，路由表容量：≥48K，提供检测报告 |
| 一般 | 3、静态路由、RIPv1/v2、OSPF、BGP。 |
| △ | 4、支持M-LAG技术，支持VRRP技术。 |
| △ | 5、支持VxLAN二层/三层网关，支持VXLAN Tunnel及BGP EVPN |
| 一般 | 6、支持N:M镜像，N与M数量≥32。 |
| 一般 | 7、BGP协议支持MD5加密和SM3国密加密。 |
| 一般 | 8、满配冗余电源，24个万兆多模模块，2个40G多模模块，原厂3年质保。 |

**备注：重要程度中标注“★”为实质性响应指标，如存在不满足或未响应投标无效；重要程度中标注“△”为重要指标，如存在不满足或未响应扣分多；标注“一般”为次要指标，如存在不满足或未响应扣分少。**