技术参数要求确认单

|  |
| --- |
| 项目名称：中医药大模型算力租用和运维服务 |
| 联系人 | 郎许锋 | 联系电话 | 18251955296 |
| 项目预算 | 15万 |
| 主要用途描述：用于学校灵山中医大模型、素山中药大模型部署和运维，确保其具备良好的运行稳定性与响应速度；大模型将应用于支撑中药学、中医学、中西医结合主体学科的“助学、助教、助研”，以及中药学“教育部首批生成式人工智能教育专用大模型建设项目”申报，并面向全国提供具有南中医特色的中医药大模型服务，助力中医药传承创新高质量发展。 |
| 参数要求：一、硬件配置参数（数量：2套）计算卡：具备高算力、大显存的GPU（如NVIDIA A100、H100、A800等）8卡，支持NVLink全互联。内存：不低于DDR5 4800MHz RDIMM 1.5TB。CPU：不低于AMD EPYC 9654 (96核/192线程)或同等核心及算力。存储：不低于PCIe 5.0 SSD U.2 8TB。网络：不低于100GbE RoCEv2网卡。互联拓扑：支持NVIDIA Quantum-2 InfiniBand，服务器间通信不低于200Gb/s。1. 软件配置参数（数量：2套）

支持pytorch、cuda。1. 服务要求

（1）验收之日起，提供61天的服务。（2）服务期间提供7×24的算力运维服务，出现问题及时解决；因服务方原因导致的大模型服务中断，根据中断时长（小时），折算相应费用双倍扣款。（3）提供中医药复杂数据处理和大模型训练和微调服务，持续提升大模型性能。（4）提供大模型网络安全相关服务支持。**四、公司资质**国家高新技术企业。 |