一、修改内容

（一）技术需求 2 产品软件（AI 平台软件）、3.2 异构算力调度平台+AI 训推开发平台及招标文件《软硬件产品技术要求比对明显表》第一个▲参数

原条款：▲配置 8 卡；单卡 FP16/BF16 Tensor 稠密算力≥900 TFLOPS；单卡显存≥140GB，总 GPU 显存≥1128GB；GPU 间高速互联带宽≥400GB/s；GPU 品牌与架构不限。

修改为：▲GPU 加速卡配置不少于 8 卡且不超过 24 卡，可由一台或多台 GPU 服务器组成，多台时须通过本项目高带宽 GPU 算力专用交换机互联；加速卡须为本国产品，不接受进口产品；单卡 FP16/BF16 Tensor 稠密算力不低于 400 TFLOPS，单卡显存不低于 96GB；全部加速卡 FP16/BF16 Tensor 稠密算力合计不低于 7200 TFLOPS（按单卡标称稠密算力累加，不得以稀疏算力或低精度算力折算），GPU 显存合计不低于 1128GB；同一 GPU 服务器内各加速卡之间须通过加速卡专用高速片间互联技术（如 HCCS、UB、MetaXLink、NVLink 等）直接互联，不得仅以 PCIe、以太网或 InfiniBand 组网替代；互联带宽由投标人在投标文件中以厂商产品规格书载明，并在验收时核验；GPU 品牌与架构不限。

（二）技术需求 3.1 一体化算力终端及招标文件《软硬件产品技术要求比对明显表》第四个▲参数

原条款：（1）16 台。CPU 总核心≥20 核，CPU 高速缓存≥16MB，CPU 功耗≥140W；GPU 核心≥6144，GPU 位宽≥256Bit；内存≥128GB LPDDR5X，内存带宽≥273GB/s，CPU 与 GPU 共享统一系统内存，无需独立显存；存储≥2TB NVMe M.2 SSD；集成 2×200G QSFP 高速网络接口，板载 1×10G RJ45 网口；支持 SR-IOV 虚拟化功能，芯片内部支持 NVLINK C2C 或等效技术直连互联，实现 CPU 与 GPU 零时差协同；▲支持 FP4 AI 算力数据格式，单台 AI 算力≥1000 TFLOPS。

修改为：（1）不少于 16 台，每台可配置一张或多张加速卡，加速卡须为本国产品，不接受进口产品；可供 AI 计算使用的统一内存或显存不少于 128GB，访存带宽不低于 273GB/s；存储≥2TB NVMe M.2 SSD；不少于 2 个高速网络接口，单口速率不低于 100G，须能接入本项目 200Gb/s 高带宽 GPU 算力专用交换机，板载 1×10G RJ45 网口；支持 SR-IOV 虚拟化功能；CPU 与加速器之间采用高速直连互联（NVLINK C2C、PCIe 5.0 或等效技术）；▲单台 FP16/BF16 Tensor 稠密算力不低于 200 TFLOPS，全部终端合计不

低于 3200 TFLOPS（按稠密算力核算，不得以 FP4 等低精度算力或稀疏算力折算）；须为同一品牌设备统一组网，并纳入异构算力调度平台统一纳管与调度。

（三）技术需求 4.7 项目团队及评分细则第 11 项

原条款：研发人员具备 5 年以上 Java 或 Python 平台研发经验

修改为：研发人员均具备 5 年以上软件平台研发经验

二、补充内容

（一）技术需求 4.7

新增第（7）项：中标人项目组须在合同签订后 15 个工作日内，依据采购人提供的数据库结构、软件架构及接口文档，完成对学校现有业务系统数据库架构、软件架构的调研与熟悉，并据此构建本项目所需的各类 AI 应用，确保 AI 应用与学校现有业务系统顺畅对接。

（二）技术需求 5.1

新增：学校现有业务系统以 Java 技术体系为主，相关数据库结构、软件架构及接口文档中标后提供；中标人应据此完成本项目 AI 应用与现有业务系统的数据对接与集成。

（三）技术需求 3.9

1. “全年提供算力不少于 210240 卡时”是指在技术需求 3.1 所列硬件设备之外，由中标人另行提供的本地化算力服务。中标人须自行配备提供该算力所需的硬件设备，部署在上海大学校内并由学校统一纳管，卡时以上海大学算力调度平台的记录为准。

2. 用于提供上述卡时的加速卡，单卡 FP16/BF16 Tensor 稠密算力不低于 200 TFLOPS，单卡显存不低于 64GB，能够满足模型微调、模型部署、AI 应用开发、智能体开发、强化学习等教学科研实训的需要。低于该要求的加速卡不计入卡时。

3. 中标人在本项目中提供的算力及大模型 TOKEN 服务，均适用技术需求 4.5 第（1）项的本地化要求，相关模型和服务器须部署在上海大学校内，不得以公有云等校外算力替代。