

大模型工作站核心部件--显卡

单一来源情况说明

一、采购大显存显卡的技术必要性

- 当前主流开源大模型（如 LLaMA 70B、Qwen 72B 等）参数量已达 700 亿，单卡至少需要 72GB 显存才能运行。
- 低于 72GB（如 48GB）的显卡，无法在本地完成模型推理和微调，必须依赖云端或集群，既不经济也存在数据安全隐患。
- 大显存还能支持更复杂的 AI 智能体、多模态等前沿科研任务，是开展高水平研究的基础硬件门槛。

二、当前已发布单卡显存 $\geq 72\text{GB}$ 的专业显卡有且仅有

NVIDIA 一个品牌，具体型号如下：

型号	显存	简要说明
NVIDIA RTX PRO 5000 72GB	72GB	性能均衡，功耗 300W，当前主流选择
NVIDIA RTX PRO 6000 96GB	96GB	旗舰级，性能最强，功耗 600W

三、市场行情

- RTX PRO 5000 72GB：市场价约 11 万元/张，价格当前仍在上涨。
- RTX PRO 6000 96GB：官方售价约 32,000 美元（约 21 万元人民币），价格最高且持续上涨。

四、结论

综合考虑性能、功耗、价格和供货，以及实际采购预算，NVIDIA RTX PRO 5000 72GB（约 11 万元）是当前最合理且唯一的选择，能够满足科研需要。

经办人签字：

课题组/部门负责人签字：

2026 年 8 月 24 日

王永睿