

A924一指禅

What 是什么？卖什么

是什么：A924是基于AI处理器，具有强大的计算能力和高校的数据能力，可以满足企业在人工智能、大数据分析、深度学习等领域的需求，A924根据算力的区分，分为训练和推理两类算力产品

卖什么：A924服务器硬件

Why 为什么要用昇腾：痛点

算力消耗大：A924可以应用于大模型、AIGC、自动驾驶、内容审核等场景需要大量高端的GPU算力场景，可以使用昇腾算力部署或替换，算力充足

调优适配难：企业对AI算法的调优能力有限，对于部分算法的迁移适配能力有限

供应风险高：因不可抗因素，非国产受供应限制，长期稳定的供应风险高

Who 卖给谁

- AI企业（商汤、旷视等）：需要强大的算力来训练和优化各种大模型，A924可以满足其对计算能力和效率的高要求
- 互联网企业（字节、腾讯、美团等）：用户产生的大量的数据，需要进行深入的数据分析和智能推荐算法的优化
- 传统行业企业（三一重工、平安医疗等）：传统企业智能化转型（如制造业、医疗、金融、交通等）利用A924提供的算力实现数字化转型，如智能生产化、智能诊断、风险预测、智能交通调度等应用场景
- 高校和教育机构（清华、上海交通大学等）：用于相关专业的教学和科研实验，培养相关领域人才，推动教育领域的应用和发展
- 政府和公共机构（GA、国家气象中心等）：在大数据处理和智能分析领域，如城市管理（智能安防、交通规划等）、气象预测、公共卫生等方面，为公共提供数字化服务
- 云服务提供商：为云服务平台提供人工智能计算资源，以满足不同客户对AI计算的需求，提供更具竞争力的云服务产品

What 主打场景

➤ 313T及以上：

训练：适用于LLM、AIGC模型训练场景，推荐SFT训练场景：如GLMv2-6B、LLaMa-13B、baichuan2-13B、Stable Diffusion 2.1、Stable Diffusion XL等

推理：70B以上模型推理场景：如LLaMa-70B、通义千问1.5 72B等

➤ 280T：

推理：AIGC推理合微调场景：如StableDiffusion2.1、StableDiffusionXL、VideoComposer、MiniGPT4等

中小参数70B以下LLM模型推理：如GLM-6B、GLMv2-6B、LLaMa-7B、LLaMa2-13B、LLaMa2-7B、baichuan2-13B、通义千问 14B等

Story 典型案例

讯飞星火：星火一体机的训练和推理一体化部署，可用于问答系统、对话生成、知识图谱构建、智能推荐等多个领域的应用，具备大模型预训练、多模态理解与生成、多任务学习和迁移等能力