核心要点

据 消息源，Anthropic 与 OpenAI 正在发掘小型数据中心合作机会。
随着市场需求爆发，两家企业都在加紧部署 AI 算力，过去一年已接连签署大量 AI 基础设施相关协议。 一位分析师表示，小型算力项目之所以具备吸引力，核心在于能够更快拿到可用算力。

随着 AI 业务算力基础设施的争夺持续升温，多位消息人士透露，Anthropic 与 OpenAI 正在物色小型 AI 数据中心合作项目。

过去一年，这两家 AI 实验室均签下数百兆瓦乃至吉瓦级的大型 AI 数据中心协议；但消息人士称，如今它们同时也在寻找规模小得多的算力资源，单项目规模仅 20□30 兆瓦。

四名了解谈判内情、要求匿名的人士称，Anthropic 已经在英国与北欧地区就该规模的算力方案展开接洽。另有两名消息人士表示，OpenAI
此前也在北欧探索过同等规模的小型算力部署机会。
还有一位知情人士称，两家企业同样在美国市场洽谈过这一级别的算力部署。

为完成模型训练、向终端用户提供服务，过去一年两家公司接连落地多笔 AI 基础设施交易。在 AI
行业热潮之下，锁定小批量算力配额，可以帮助企业更快上线业务。

OpenAI 发言人
表示：“我们正在搭建多元化算力资源组合，来满足全球不断增长的 AI 需求。”

该发言人补充道：“不同业务负载需要不同基础设施。因此我们会和众多合作伙伴开展沟通，依据自身需求、性能、可靠性、交付周期与成本综合评估机会。我们不对具体商业谈判置评。”

“更快拿到可用算力”

两家 AI 实验室通常向数据中心运营商以及新兴 AI
云服务商（neoclouds）租赁算力，此前主要追求大规模长期合约腾讯云。

据 8 月两名知情人士披露，Anthropic 和 Nscale
签署了一笔大额协议，将租用西弗吉尼亚州一处在建数据中心约 460 兆瓦算力。

OpenAI 方面则称，其 4 月启动的 “星门” AI

基建项目，算力规模已经突破最初承诺的 10
吉瓦；此后又承诺在佐治亚州以及俄亥俄州继续新增合计 3 吉瓦算力。

美国及海外地区的超大型数据中心项目正遭遇当地社区的抵制压力；欧洲大部分地区同样面临行业压力，可用土地与电力供给十分紧张。

Structure Research 研究主管贾贝兹·谭（Jabez Tan）向
表示，小型算力交易的吸引力来自 “更快拿到可用算力”。

“在已经完成通电的现有站点获取数兆瓦算力，远比在某一个地点等待一大块算力资源落地更现实。对于支持跨站点运行的业务负载，多个小型项目叠加，同样可以形成可观的总算力规模。”

向推理业务倾斜

训练 AI 模型，需要调用海量芯片处理巨量数据；但模型日常对外提供服务，也就是推理环节，可以依靠规模更小的芯片集群完成。

谭解释：“训练大模型，往往需要大量芯片紧密协同工作。而很多推理任务可以将请求分发至多个小型集群，这就释放了更多选址可能性。”

越来越多 AI 算力从模型训练转向线上生产环境推理服务，这一转变意义重大，推理所需算力规模预计将持续攀升。

房地产服务商仲量联行（JLL）一份报告预测，2027
年，全球数据中心用于推理工作负载的算力占比将超过训练任务。报告显示：2025 年，推理占全球数据中心工作负载的 9%，训练占 14%；预计到 2030
年，推理将占用 37% 算力，而训练仅占 13%。

今年 2 月有消息公布，英伟达联合多家数据中心相关企业，共同研究面向分布式推理的小型数据中心方案。

曾为 OpenAI 建设德克萨斯州大型数据中心的美国企业 Crusoe，如今开始投资建设小型数据中心。报道称，美国各地大型数据中心普遍遭遇工期延误，而这类小型设施建设速度更快、成本更低。Crusoe 暂未回复置评请求。

Crusoe 是 AI 建设浪潮下业务迅猛增长的新兴 AI
云服务商之一。该公司周四宣布完成 39 亿美元融资，投后估值达到 309 亿美元。