

正规大小单双快三下载软件

EMCm7DuGMf9IBRLV

正规大小单双快三下载软件不到15万元！清华90后团队发布“褐蚁”一体机，已支持阿里最新Qwen3模型 | 钛媒体AGI

行云集成电路创始人、CEO季宇

4月30日消息，钛媒体AGI获悉，清华90后创立的北京行云集成电路有限公司（简称“行云集成电路”）宣布，推出全新的一体机产品“褐蚁”，仅需最高15万元就可以跑满血版DeepSeek R1/V3大模型，并且对话速度达到了20token/s。

今天下午，行云集成电路创始人、CEO季宇对钛媒体AGI表示，目前“褐蚁”一体机已经支持阿里最新发布的Qwen3系列开源大模型，包括顶配版Qwen3-235B-A22B。

具体来说，“褐蚁”一体机有三款不同的配置：最高性价比的“超大杯”褐蚁HY90，搭载双路AMD EPYC 9355服务器、24条 48G 6400M频率内存和NV 5090D计算卡，支持FP8、INT4两种数据精度，在FP8精度下跑满血版DS能达到21token/s的对话速度，在INT4精度下则能达到28token/s，最高支持128K的上下文，售价14.9万元；此外，行云集成电路还将推出“大杯”褐蚁HY70、“中杯”褐蚁HY50两个配置版本。

季宇团队表示，行云通过对推理引擎的独家优化提升了推理效率，是国内首个在10万价位提供满血FP8精度模型20+tps速率的一体机产品，并且可配置各类调用大模型API的应用。

据悉，行云集成电路成立于2023年8月，其核心团队主要来自清华大学及全球顶尖芯片公司，致力于研发下一代针对大模型推理场景的高效能GPU芯片，创始人包括季宇，以及公司CTO余洪敏。

其中，31岁的季宇是清华大学物理本科、计算机体系结构方向博士，也是“华为天才少年”之一，主攻体系结构、AI芯片方向。而在华为期间，季宇曾是海思昇腾芯片编译器专家，负责多个昇腾编译器项目，同时作为研究科学家，展开AI编译器领域和处理器微架构域诸多挑战性问题攻关；类脑计算机体系结构《自然》论文共同一作，中国计算机学会CCF优博奖获得者。

2024年11月，行云集成电路宣布完成总额数亿元的天使轮及天使+轮融资，投资方包括智谱AI、仁爱集团、中科创星、奇绩创坛、水木清华校友基金、嘉御资本、春华资本、同创伟业、峰瑞资本等机构。

季宇早前表示，当机器的门槛低到消费级可以接受的成本，边际成本承担的商业模型也会发生变化，带来进一步的产业繁荣。如果计算机体系能支撑消费级可以接受的价格享受极致的AI体验，大模型产业会进一步进入一个类似软件行业零边际成本的超高速爆发期。

事实上，目前多家芯片、服务器、云计算、运营商、AI Infra等领域的公司都发布了DeepSeek一体机产品，堪称“百机大战”，但其中可以运行满血版DeepSeek-V3/R1模型的企业并不多，目前国产GPU芯片层面主要有壁仞科技、摩尔线程等。

据第一财经引述券商数据，预计到2027年，国内一体机的市场规模将超过5000亿元，超过30家企业已向市场推出了一体机产品。

一位行业人士早前对钛媒体AGI表示，一体机核心解决的是部署易用性问题。相比复杂的云计算解决方案，项目落地周期更快、私有化安全性更高、价格更有性价比，但DeepSeek一体机比较大的问题在于持

续维护、升级以及大规模数据调度等，以及一体机综合使用成本没有办法复用。

“从当年数据库年代，一体机就兴起，这个大家非常喜欢，但不同的一体机，里面的东西差别是非常大的。DeepSeek出来以后，DeepSeek一体机爆火，我跟一些领导也聊过，数据中心算力这块卖得特别好，但都是短期租赁，因为本身硬件固定资产的年限也就4到5年。所以，我们可以看到一体机比较大的问题是以后持续性的维护、升级以及大规模数据调度等，比如DeepSeek一体机是很新鲜，但是大模型演进非常快，现在DeepSeek已经不是最领先的了，其他公司也能升上去，而客户本身可能根本没有这方面的技术能力。”上述人士表示，推理一体机只是过渡态产品，最终大模型还是要和云计算进行结合，当整个技术成熟以后，跟多云整合，最后会变成云平台的一部分。

对此，在季宇看来，现在一体机依然能够跑得动超大杯的MoE模型和30b当量的dense模型，后续模型迭代其实还好。据悉，相比其他一体机产品，行云集成电路的“褐蚁”在显存、KTransformers等层面有更多的技术研发能力。

据钛媒体AGI了解，下一步，行云集成电路将自研GPU卡，持续带来更低价格，更高tps体验，并且还将实施由30台褐蚁一体机组成的“蚁群”方案，进一步发挥分布式系统优化能力，或将支持500-1000的有效并发数，而价格有望降至300万-400万元——低于英伟达HGX H200集群的整体硬件成本。

钛媒体AGI分析认为，从整个商业市场的视角来看，当前国内有很多人都投身于一体机、AI Infra，以及GPU服务器等类似算力“卖水人”工作。随着推理模型技术迅猛发展，大规模算力投入被质疑，所以国内AI算力行业是否将如同AGI基座模型市场一般，经历大浪淘沙后最终只剩下大厂及头部的几家企业，还是说依然会有更多人“卷”性价比，这一点仍有待进一步观察。

根据规划，季宇的最终目标是：为AI构建x86，目标是让AI PC与互联网底座重回组装机模式（PC及服务器），将组装机打造成AI基座，低成本搭建高端大模型，激活AI服务器组装机体系，让大家能组装出支撑高端AI应用的设备，带动体系运转。

（本文首发于钛媒体App，作者 | 林志佳）

澳洲幸运8官网开奖官方

168澳洲幸运5历史开奖记录查询

168飞艇开奖官网开奖历史

168飞艇开奖官网开奖记录

澳洲幸运5官方开奖直播官方app介绍

澳洲10官网开奖168

澳洲幸运10精准计划3码

飞艇1—7雪球计划网页

澳洲幸运10精准计划走风冠军位

澳洲幸运10计划精准7码计划

澳洲幸运5游戏特色

168澳洲幸运5历史开奖记录查询

飞艇精准冠军计划全天免费

2023澳洲幸运5开奖数据分析

168飞艇全国统一开奖破解方法

澳洲10精准计划系统

必中500万方法

168助赢计划永久免费版

澳洲幸运十开奖结果查询