

2025年2月12日 星期三 编辑 张海 版式 罗梅 校对 汪智博

中国造人工智能 DeepSeek 有多强?



短短30天,中国初创企业深度求索(DeepSeek)发布两款性能比肩GPT-4o的大模型。据数据显示,DeepSeek活跃用户数2月1日突破3000万,成为史上最快达成这一里程碑的应用。



价格“感人”是让DeepSeek快速出圈的一个标签。DeepSeek-R1的API服务定价为每百万输入tokens(指文本被分割成的基本单位)1元或4元,每百万输出tokens 16元。



相较谷歌 Gemini, DeepSeek-R1 更便宜、上下文更长、推理性能更佳。低成本比肩 OpenAI 的 o1 模型,令硅谷的“烧钱模式”一时间遭到猛烈质疑。



DeepSeek-V3模型的训练成本为557.6万美元。相比之下,同样是开源模型的Meta旗下Llama-3.1模型训练成本超过6000万美元,而OpenAI的GPT-4o模型的训练成本为1亿美元。



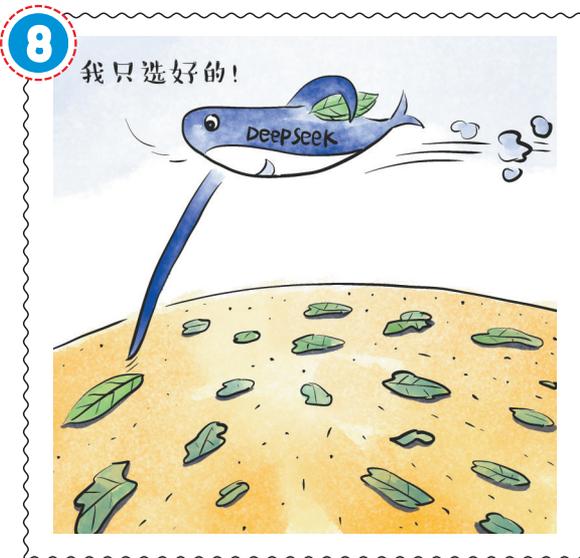
DeepSeek问世之前,全球人工智能巨头不约而同走上一条参数越“炼”越大的路线。另辟蹊径的DeepSeek并不盲目追求参数之大,而是选择了一条更高效的训练方法提升性能。



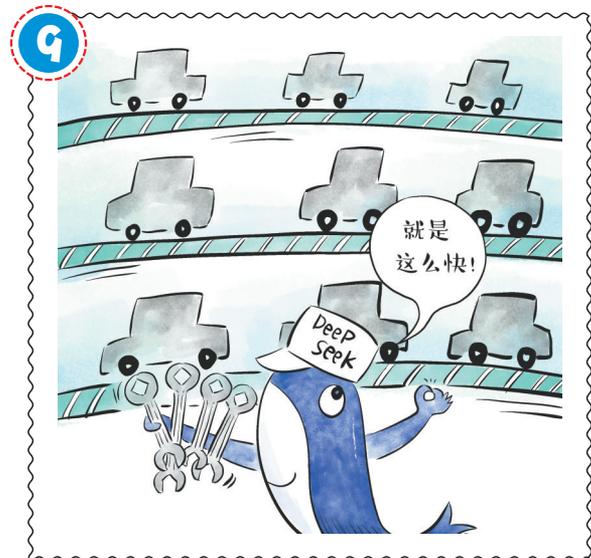
“DeepSeek通过有限资源的极致高效利用,实现了以少胜多。”面壁智能首席科学家刘知远说。算力封锁下的破局,得益于DeepSeek技术架构、数据策略、工程实践三方面的关键突破。



传统大模型的架构好比高速公路,当车辆(数据)多时会导致堵车(计算慢、能耗高)。而DeepSeek架构则把一条串行的高速路,变成了辐射状的快递分拣中心,因此既能提高速度又能节约能耗。



传统的数据策略好比去农场随便采捡,常有价值不高的烂菜叶(低质量数据)。而DeepSeek创新的数据蒸馏技术,有针对性地筛选掉质量不高的烂菜叶,相比随机采样训练效率提升3.2倍。



传统大模型的训练方式好比手工造车,一次只能装配一台,效率低下。而DeepSeek的3D并行相当于通过流水线并行把造车流程拆分为10个步骤,同时组装10辆车(数据分块处理)。

文/新华社客户端 漫画/杨仕成

◎ 华西都市报副刊“少年派”定位于亲子共读刊物,设有作文版、少儿新闻版、科普读物版、漫画版,每周星期一至星期五,都有精彩好看的内容呈现。
◎ 让有温度的纸质阅读,助力孩子养成良好阅读习惯,提升核心素养。

订阅热线:028-86969110
大家也可以通过微信小程序中国邮政微商城搜索《华西都市报》,即可订阅。
欢迎小朋友向我们投稿! 投稿邮箱:shaonianpai@thecover.cn
你投来的每一篇文章,都有机会被大家看到! 快来投稿吧!