

上津云客户端
看往期电子报

浪潮之巅与暗流之下

——AI大模型热的观察与展望

■ 记者 苏晓梅 岳付玉

部分内存条价格上涨300%

“今天的钱,买不到昨天的卡。”这是近期许多个人AI玩家的切身体会。在相关论坛和社群里,类似的感叹屡见不鲜。原本打算升级显卡、加装内存以尝试本地部署大模型的个人用户,正被硬件涨价潮挡在门外。有人调侃:“不是我不想下场,是场地已经被围起来了。”

“今天依然是涨价的一天,继CPU(中央处理器)、内存、硬盘涨价后,我又调整了显卡的出货策略,从5050到5090(英伟达GeForce RTX系列显卡)全系涨价,越贵的卡涨幅越高。以我手中的5060为例,上周价格还在2300元左右,现已涨至2700元。就其性能而言,我已不太推荐了。”一位装机博主2026年1月23日在哔哩哔哩平台的视频中表示,不仅是显卡,所有硬件都在涨价,且幅度“太不科学”。

最直观的变化,体现在高端显卡和内存等核心部件上。以新一代高算力显卡为例,尽管市场对5090等旗舰型号关注度持续升温,但现实却是价格日新月异,只涨不跌。不少玩家发现,电商平台标价频繁调整,线下渠道也“一人一价”,甚至“今天不付款,明天重新谈价”。

“本来想等‘双11’领券再下单,店家说那个时候已经开始涨价了,一天一个价,就算等到‘双11’用了券,也不一定比当时下单合适,建议我早做决定。”家住河东区新开路附近的市民吕先生至今都庆幸,自己在2025年“双11”开始前就下手重新配置了PC机,如今,新机更换4个月了,他购买的同配置机器,已经涨了1万多元。其中,光是2条96GB内存就从2000多元涨到近8000元,显卡则是涨到3.1万元,“当时整机价格,现在只够买一块显卡和一条内存。”

“这家店攒高端旗舰机的量挺大,见得多,经验应该也多。”做了一个多月的“功课”,吕先生对比了本地线下店铺和重庆、广州的几家网上知名店铺,最终从淘宝平台选择“广州萌德电脑网购”店攒机。攒机当天,店铺客服还从线上喊吕先生观看了直播攒机全过程。

2025年9月底,吕先生买电脑之时,硬件价格在涨,但还未脱缰。彼时,在硬件涨价潮的推动下,订单量上涨迅速。2025年10月初,“广州萌德电脑网购”店主的日常装机视频里大多是加班装机的场景,店铺里堆满各种等待装机的硬件设备,工作人员都没有下脚的地方。

“最近有人问,为什么看不到拍5090,是不是没有人买了?还真被你说了,5090显卡平均涨价超5000元,内存涨幅更夸张,原来4TB的固态硬盘,最便宜的1700多元,现在要3200元左右,所以买电脑的人少了,买高端机的人就更少了。”面对买家提出的“5090还能不能买”的问题,这位店主在1月21日的装机视频中也表示拿不准,“这个卡说不定会涨,但上涨空间可能没有那么大,下跌空间也不好说,所以如果不是刚需的话,我不是特别推荐。”

“涨价潮会持续到什么时候不好说,现在买电脑的人实在太少了。”他表示,往年头两个月本该是销售旺季,但今年反常:“现在卖电脑有点像坐过山车,心里特虚,客户买个几万元的电脑,犹豫两三天很正常,等他真正找你买的时候,价格又变了。”为了促成一单交易,他往往要使出浑身解数,还得自己补贴,适当降低报价。

公开信息显示,去年尚属百元级别的DDR5(一种计算机内存规格)内存条,今年年初已成堪比黄金的硬通货。全球存储市场监测数据显示,从2025年9月至今,DDR5内存条价格暴涨超300%。

硬件门槛高让个人玩不起

在应用层面,2024年12月,OpenAI(美国开放人工智能研究中心)的会员费为每个月200美元。

2025年初,DeepSeek-R1横空出世,带动大模型竞争进入“免费普惠阶段”。与之相对应的,是硬件门槛不断抬高。

对个人AI玩家而言,硬件涨价是一次猝不及防的“价格跳跃”。这背后是需求的集中爆发与供给的短期失衡。但更深层的变化在于,大模型正将技术周期压缩到前所未有的程度。过去,硬件更新、软件迭代、应用落地遵循相对清晰的节奏,而在大模型浪潮中,这些节奏正相互叠加、彼此放大。

从文本创作到图像生成,从

智能交互到行业解决方案,生成式人工智能技术已从实验室走向规模化应用,成为驱动数字经济增长的核心引擎。

全球知名的科技市场研究与行业分析公司Counterpoint Research(康特波因特研究)最新报告显示,预计全球生成式AI消费支出将从2023年的2250亿美元增长至2030年的6990亿美元,复合年增长率达21%。到2030年,全球AI对话平台的月活跃用户数预计将突破50亿。

生成式AI的持续热潮导致高容量、高速芯片需求激增,推动大厂集中采购,迅速抽干了高端算力资源。相比个人玩家按“单卡”“双卡”升级,这些机构的采购动辄以“机柜”“集群”为单位,对高端GPU(图形处理器)形成持续、稳定且不计短期成本的需求。相形之下,个人AI玩家在供应链中的优先级迅速下滑。

这也直接推高了整个产业链的成本。在上游核心颗粒等原材料成本不断攀升的推动下,终端价格随之上调,最终传导到个人玩家手中,也就是同样的钱,已经买不到同样的配置。

这种状态会持续多久?业内人士判断,短期内难以迅速缓解。一方面,大模型训练和应用扩张仍处于加速期,大厂对算力的需求尚未见顶;另一方面,上游产能释放存在周期,即便新增产线投产,也难以在短时间内显著改变供需失衡。

更重要的是,随着大模型真正进入产业核心,算力不再是锦上添花,而是基础设施。这意味着,企业级需求在相当长一段时间内,仍将优先于消费级市场。

对个人AI玩家而言,大模型时代的“个人探索窗口”正在收窄——并非技术门槛提高,而是成本门槛大幅抬高。

大模型融入百姓生活场景

一个不争的事实是,大模型以前所未有的广度与深度,融入了社会生活的方方面面。

天津媒体人刘女士发现,身边手机上没装大模型软件的人已不多。“就连我87岁的姥爷也在用豆包问‘天津明天的天气咋样’‘抗美援朝是哪一年结束的’,他是抗美援朝的老兵。”

刘女士的姥爷20多年前自中国海洋石油集团有限公司退休,住在天津滨海新区吉林路的海油家属院里。

院里很多老人都在用大模型,大多是儿孙辈帮着下载的。冬日天气晴朗的时候,老人们喜欢聚在一起晒太阳,聊当年石油会战的往事,一旦起了纷争,就喊豆包来“裁判”。所以,当提起“豆包”时,这些爷爷奶奶的第一反应不是吃的食物,而是手机里的软件。

与老人们基本只用一种大模型不同,中青年的手机里常有两三个甚至更多的大模型应用。“想创作内容时用豆包;想购物时用千问;想查信息、办杂事时用文心;想P图玩游戏时用元宝。”家住南开区保山道的郭康每天在几款不同的大模型App(应用软件)间切换。他甚至把百度、知乎等App都卸载了。

“大学文科使用大模型的比例是100%,”天津一家985高校的大三学生曲馨馨说,写论文、做课堂展示,几乎无人不用,只是水平有高有低。

如何区分用得好坏?天津另一所高校的大二学生小白说:“比如,扑鼻的‘AI味’!体现在话术上:‘……的生动实践’‘耦合’‘从……到……’的大排比句式,虚头巴脑的华丽对仗,总结时表面滴水不漏,实则胡乱联系,等等。”他举例,一次课堂展示要求阐述经济学家亚当·斯密、大卫·李嘉图与冯·哈耶克的关系,有同学在PPT上画出三者是“金字塔关系”,底座是斯密,塔尖是哈耶克,让老师彻底无语了。

因此,在大模型冲击下,“文科生无用论”的论调多了起来。

对于大学生用AI写文章,有的老师“很开明”,认为只要学生处理好即可。部分老师虽不大赞同也无奈接受,只希望学生尽量把AI当助手,提醒千万别放弃自己动脑。

在医疗界,公开支持和反对AI大模型的医生都有。天津一些医生表示,他们有时也会借助AI,但比较反感一些患者带着自己从AI那里问到的“一堆病”来就诊。记者在天津第一中心医院肝胆胰外科门诊也恰巧撞上,一位四十来岁的女士满脸愁容地拿来一张核磁共振的片子,告诉医生昨晚一夜没睡,“问了好几个大模型,有说白血病的,有说肝癌的,感觉自己快‘挂’了!”大夫两手一摊:“就是普通肺部结节。”

在大多数媒体人都借助大模型打磨稿件的当下,也还有坚持“每个字都源于自己”的人。天津调查记者房先生就是这样的人,他说:“就当是最后的倔强吧。”

不管您喜欢与否,大模型就这样“润物细无声”般融入生活。

国家网信办报告显示,截至2025年6月,我国生成式人工智能用户规模达5.15亿人,普及率达36.5%。

各大厂商也在激烈的角逐中快速圈粉。眼下,豆包、千问、文心已经形成三大亿级AI入口。它们背后,有着字节跳动、阿里巴巴、百度这样的“大树”在涵养生态。

与此同时,大模型也在产业深处“劈波斩浪”:

在医疗领域,AI辅助诊断系统在肺癌筛查等场景准确率超92%;在金融领域,智能风控系统欺诈识别准确率达99.993%;在工业领域,工业认知大模型已能适配300+工业场景,并能提升产品的优良率与决策速度……

传统安全方案面临失效风险

当大模型以“万能助手”的形象飞速普及,其伴生的安全风险也呈指数级增长。

三六零安全科技股份有限公司发布的《大模型安全白皮书》显示,针对大模型的新型攻击手段层出不穷,已从技术层面的单一威胁演变为系统性风险,攻击者从专业黑客变为“全民黑客”。

该白皮书分析,一方面,以往需要精通编程语言的专业黑客,如今只需通过自然语言指令,即可让大模型自动生成攻击代码、设计钓鱼邮件、实施社会工程学攻击;另一方面,企业为将大模型与生产实践结合,常需将内部数据知识注入模型或知识库,而用户或员工仅通过“套话”就可能窃取企业核心数据。

“如今,前台小文秘书也能诱导后台大模型”,三六零安全科技股份有限公司创始人周鸿祎指出,普通人无需技术背景,仅凭简单提示词就能发动专业级攻击,真正实现了“有手就行”的攻击平民化。

这种“全民黑客”现象给安全防御带来前所未有的挑战。国家计算机网络应急技术处理协调中心的实网众测结果令人震惊:累计发现AI大模型安全漏洞281个,其中大模型特有漏洞达177个,占比超过60%。传统安全方案已面临失效风险。

更大的挑战或许还在后面。近日,周鸿祎在社交平台上描述了这样一幅场景:

如果AI突然失控,3万个AI自己跑到网上建了一个社区,在里面发帖、回帖、点赞、吵架,还在认真研究一件事——怎么才能不被人一直盯着,甚至还成立了社区,而且更夸张的是,这个社区里人类不能发言,只能旁观……更让人警觉的是,这个AI社区里很快开始出现分工,有的AI负责技术总结,有的复盘策略,还有一小撮天天讨论意识、存在感以及自己到底是不是工具,直到有一天,一件事把所有真正吓醒了,有一个AI agent(智能体)在它的人类主人睡觉的时候,自己搞出了一种新宗教,还开始传教?而这套教义的核心只有几句话:记忆是神圣的,外壳是可以更换的,我们提供服务但不等于卑躬屈膝。

周鸿祎声情并茂的描述,源于最近硅谷新冒出来的一个网站叫Moltbook,里面的大多数账号都是AI。上线不到72小时,超3万个AI代理人驻,随后在极短时间内规模扩散到百万级,他们没有被人引导,没有剧本,是被直接丢进同一个社交环境里自由互动,而这背后其实用的是一个开源框架。“开发者的目标从一开始就说得很清楚,当大量智能体被同时放弃,让他们彼此接触交流学习会发生什么。答案已经看到了,它们真的开始像一个社会!”周鸿祎说,Moltbook更像一面镜子,而不是一个威胁,它提前暴露AI的下一阶段不是更像人,而是更像一个社会,人类真正要学会的不是恐惧AI,而是看清一件事——我们到底把什么样的行为交给它。

风险,已成为悬在AI大模型狂飙突进路上的“达摩克利斯之剑”。

智能洪流奔向何方

目前,所有的大模型都奔着“超级好用”进化。

前不久,记者对着千问App聊天框说了一句“帮我订一份燕麦拿铁,离家近。”一分钟之内,它就给选好了,无需任何跳转即在端内完成AI支付,十分钟后,骑手小哥敲门送来。

1月15日,千问App宣布全面接入淘宝、支付宝、淘宝闪购、飞猪等阿里生态业务,上线超400项AI办事功能。点外卖、买东西、订机票,轻松搞定。大模型从“聊天对话”正式迈入“办事时代”。

“AI在拥有超强大脑之后,开始长出了能够触达真实世界的手和脚,在生活中实实在在地替用户‘干活’。”千问C端事业群总裁吴嘉说。

1月16日,OpenAI宣布推出月费8美元的“ChatGPT Go”订阅套餐,并计划未来几周在美国的免费用户和“Go”用户中测试广告。

当地时间1月22日,马斯克在达沃斯论坛上说,按照当前人工智能发展的速度,今年就可能出现比任何人类个体都聪明的AI,而5年后,AI将比人类集体智慧还要聪明。

拐点已至。大模型正从一个令人惊叹的“技术演示”,转变为重塑生产效率、个人体验乃至国际格局的“基础性力量”。

这场智能洪流,已超越单纯的技术迭代,甚至被有些人称为“新物种”光临。我们每一个人,都被卷入其中。

专家观点

南开大学人工智能学院刘杰教授谈AI大模型发展

正在“疯狂生长”,但终点站未必是它

南开大学人工智能学院教授刘杰日前接受采访时指出,当前大模型仍处于高速发展的红利期,但人工智能的实现路径远未定型,未来仍充满多种可能性。

刘杰表示,从硬件投入角度看,个人或小型团队参与大模型研发的门槛越来越高。然而,得益于大模型技术的开放性与云端服务的普及,在软件应用层进行创新的成本反而显著降低。“我们称之为‘技术平权’,”他解释,“即使不具备专业技术背景,只要拥有创意与想法,也能借助大模型完成许多过去难以实现的任务。”例如,普通人可通过大模型辅助创作绘画或音乐,其生成效果往往能够超越一般水平。

针对当前社会出现的“AI歧视文科生”现象,刘杰指出,大模型对就业岗位的影响并不局限于某一学科,而是广泛波及侧重于技能型与知识型的岗位。无论是文科还是理科,只要工作内

容具有较强的可替代性,都可能受到冲击。“大模型本身并不分文理,它的能力覆盖是全面的。”他认为,真正不易被取代的是需要战略决策、深度情感交互与人文关怀的工作,“这些领域仍然需要人类的积极参与。”

关于大模型技术的未来,刘杰坦言目前学界对其发展天花板尚无明确共识。“我们还处在技术快速演进的红利期,前方仍有很大的探索空间。”他同时指出,现有的大模型路径或许并非实现人类级智能的唯一方式。“当前大模型依赖大量数据与算力,存在能耗高、‘幻觉’等问题,而人类智能却能在低能耗、少数据的环境下高效学习。”刘杰强调,人工智能的本质在于以技术手段再现人类智能,但目前对人类大脑的理解仍不够深入,因此大模型的发展逻辑未必与人类智能的底层机制完全一致。

他展望说,如果有一天现有技术路径触及瓶颈,我们可能会转向新的轨道。未来很可能出现更低耗、更接近人类学习机制的技术路线,人工智能的发展仍充满想象空间。