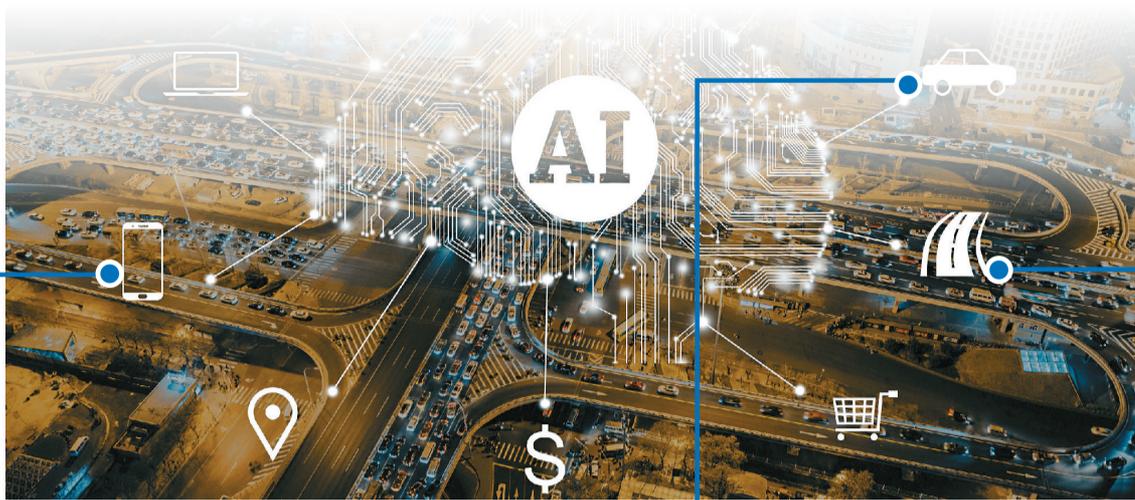


# 群雄并起,国产大模型是“争”还是“乱”



国内国际科技业界忽然就进入了一个仿如“烈火烹油”的阶段,任何点滴相关消息,都能炸起一阵波澜,而这高温烹炸的核心就是“大模型”了。

4月10日,商汤科技发布日日新SenseNova大模型,还宣布推出了自研中文语言大模型应用平台“商量”。昆仑万维同日宣布,其和奇点智源合作自研的国产大语言模型“天工3.5”即将发布。此外,搜狗创始人王小川宣布了自己即将投入创业的大模型产品名叫“百川智能”,该产品拿到了5000万美元的启动资金。

4月11日,商汤在港交所股价高开超8%;A股AIGC、ChatGPT方向领涨,首都在线涨超15%,奥飞娱乐3连板,中文在线涨超12%,昆仑万维、同方股份、万兴科技、三人行、科大讯飞等跟涨。阿里云正式宣布推出大语言模型通义千问,并且表示希望让每家企业都能基于“通义千问”,拥有具备自己行业能力的专属大模型。

这似曾相识的一幕,远的有互联网时代早期的“百团(购)大战”,近的也有元宇宙、NFC。基于这些或远或近的互联网记忆,不少的结局并不美好,难免让人们不得不质疑,这一回是大模型之争,还是大模型之乱呢?

■新快报记者 郑志辉

## 5000万美元就能上车?!

由ChatGPT掀起的这股大语言模型(LLM)浪潮,让人们看到了人工智能的另一种可能,即通过增加模型参数(例如GPT-3的参数1750亿),来模拟人脑。

据国盛证券估算,GPT-3训练一次的成本约为140万美元,对于一些更大的LLM,训练成本介于200万美元至1200万美元之间。以ChatGPT在1月的独立访客平均数1300万计算,其对应芯片需求为3万多片英伟达A100 GPU,初始投入成本约为8亿美元,每日电费在5万美元左右。如此高的入门门槛,只有少数公司能够承担,也是大公司的护城河。所以目前国外大模型基本是谷歌、微软、Meta三家。

不过,这在国内似乎并不成为问题。前有原美团联合创始人王慧文宣布出资5000万美元,称要“组队拥抱新时代,打造中国OpenAI”,后有原搜狗创始人王小川声称已获得5000万美元启动资金,并且其百川智能大模型的训练也已经启动,“争取年内发布国内最好的大模型和颠覆性的产品”。

不少人对这种一窝蜂上大模型之举提出了批评。海通证券研究所科技产业链负责人郑宏达就对王慧文的人局发文质疑,“5000万美元够干什么的?”“互联网的人啥都不懂,就只会营销,一点都不踏实”。

商汤联合创始人徐立在日前发布SenseNova大模型时提出,算力也并非一味追求大,还要在数据准备及质量上下功夫,“高质量数据只有小部分,甚至不够用”,所以“算力的需求实质上没有那么多”,这为中国企业(应对欧美竞争)争取了一些时间。

科技行业观察人士陈根仍对商汤表示怀疑,“对于商汤而言,有建大模型的能力,但能不能训练出真正的类ChatGPT产品是另外一回事。至少从现实来看,商汤依然面临训练数据、参数优化、硬件算力这三大硬伤。”

商汤最新公布的2022年业绩报告显示,公司全年净亏损额虽同比收窄,仍高达60.92亿元,而现金储备已降至166亿元。2018年至2022年五年间,商汤累计亏损超过430亿元。

## A股有人在裸泳

昆仑万维4月10日发布的公告称,自研“天工3.5”将于4月17日启动邀请测试,并称这将是“中国第一个真正实现智能涌现的国产大语言模型”,“已经非常接近ChatGPT的智能水平”。

让昆仑万维没想到的是,ChatGPT概念却在当天迎来集体下跌,其中云从科技、三六零、科大讯飞等集体跌停,昆仑万维、海天瑞声、蓝色光标等跌逾10%。

就在前不久,包括昆仑万维、云从科技、同花顺等公司发布股价异动或股票交易异动公告,纷纷表示目前公司AI等业务领域并未产生业务收入或收入较少,且公司AI相关技术和世界领先公司存在着较大差距,呼吁投资者注意投资风险。

和昆仑万维同享了大模型概念首波红利的三六零,4月9日才宣布基于360GPT大模型开发的人工智能产品矩阵“360智脑”率先落地搜索场景,将面向企业用户开放内测。在此之前,360股价已经从年初的6元出头,狂涨到了4月4日最高点20.85元。

与之对应的是360连续下挫的业绩。2020年360净利润下滑至28.38亿元,跌幅52.3%;2021年,360全年净利润只有8.404亿元,暴跌70.39%;2022年前三季度,360总营收为69.35亿元,同比下滑18.79%,净亏损19.93亿元。

中信证券4月9日发布研报为AI概念“降温”,“AI相关板块纯主题炒作热度或降温,机构进一步调仓带来的增量对于AI主题的边际影响非常有限”,“从高热度的数字经济板块轮动到低热度板块,建议围绕全球流动性拐点和财报季业绩两条主线布局”。

## 大厂MaaS才是终局?

互联网分析师易方寒认为,中国需要有自己的AI大模型,这是彰显科技创新实力的需要,也是经济和社会发展的需要。但中国不需要那么多企业盲从做AI大模型,那是人才和资源的浪费。有实力的企业加入大模型混战,让更具创新能力的企业营运而生,会推动人工智能时代的BATJ大企业的诞生。

而中信证券的研报则指出,当未来模型大小与日活用户数接近阈值后,成本端的优化将会收束模型整体成本的膨胀,也提供给一些中小公司切入垂直领域的机会,最终形成:AI巨头提供泛用性模型以及服务主要ToC需求,中小公司切入特定垂直行业并根据需求做微调优化模型的格局。

据记者了解,随着OpenAI、微软、百度等大厂竞相发布大模型产品,一种新的商业模式MaaS(模型即服务)应运而生并成为新趋势。在百度“文心一言”发布会上,李彦宏提到,大模型时代的主要商业机会之一就是未来云计算业务,主流商业模式将变为MaaS。

阿里云智能集团CEO张勇昨日表示,阿里云致力于做基础大模型,未来每一个企业都可以调用通义千问的全部能力,也可以结合企业自己的行业知识和应用场景,训练自己的企业大模型。比如,每个企业都可以有自己的智能客服、智能导购、智能语音助手、文案助手、AI设计师、自动驾驶模型等。

中信建投最新研报指出,未来阿里大模型有望在办公、应用开发、智能家居、电商等领域落地。其中,在办公领域,有望和钉钉深度融合,帮助快速完成文件工作,提供新一代智能办公工具。

## 链接

说起大模型,其实也不算什么新鲜概念。

国外超大规模预训练模型始于2018年谷歌提出BERT大规模预训练语言模型,并在2021年进入“军备竞赛”阶段。

2020年,OpenAI推出GPT-3超大规模语言训练模型,参数达到1750亿,用了大约两年的时间,实现了模型规模从1亿到上千亿级的突破,并能实现作诗、聊天、生成代码等功能。

再到2021年,谷歌推出的Switch Transformer模型成为历史上首个万亿级语言模型多达1.6万亿个参数,并在

12月提出了具有1.2万亿参数的GLaM通用稀疏语言模型,在7项小样本学习领域的性能优于GPT-3。

同期,国内超大规模模型的研发发展也很迅速,2021年成为中国AI大模型爆发的一年。

商汤科技的计算机视觉模型、华为云联合发布的盘古NLP超大规模预训练语言模型和盘古α超大规模预训练模型、阿里达摩院PLUG中文预训练模型以及联合发布的M6中文多模态预训练模型还有百度的ERNIE 3.0 Titan模型陆续面世,其中达摩院M6模型的参数达到10万亿,直接将大模型的参数提升

了一个量级。

虽然当时国内院校跟大厂都在做大模型,但这更多是科研上的效果复现,对国外较为前沿的AI模型的跟进。

直到2022年年底StableDiffusion、ChatGPT的出现,才开始让更多业内人士从观望走向下场。从技术和战略层面上看,ChatGPT让更多人意识到这不单是一个应用工具,甚至会成为一个新的PaaS层基础设施。从技术和战略层面上,群雄争霸,都想试图成为一个新时代的基础设施供应商。

再到2023年,上市公司和大厂陆续放出做大模型的信号,AIGC概念股股

价翻番,外加“AI的iPhone时刻”无不在这股热度从二级市场涌向一级,引发对语言大模型的重点关注。

然而赛道虽香,语言大模型的火爆表象下,技术差距并不止一点。以当前国内大厂语言基础大模型十几亿、几百亿的参数规模,以及大多数创业公司还在做一两百亿的参数模型来看,跟ChatGPT相比,还有很长的路要追赶。



一起了解更多  
3C前沿资讯