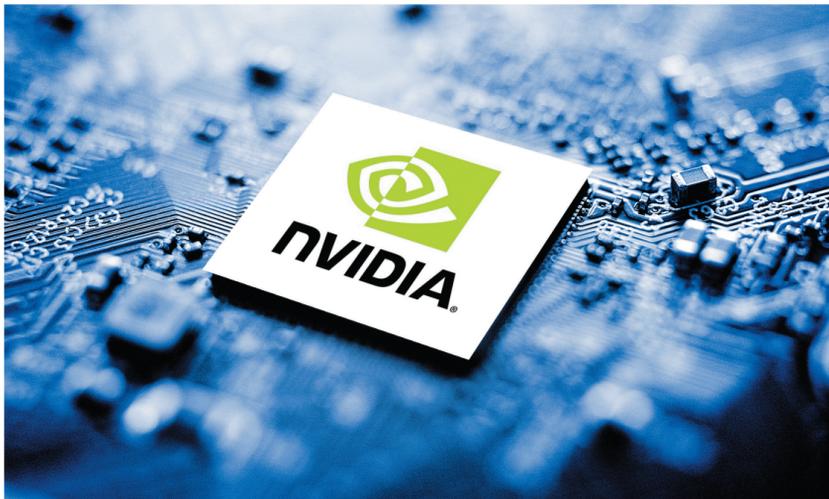


# 英伟达下场做AI大模型

如果有人问AI时代最大的赢家是谁,答案几乎毋庸置疑——英伟达。凭借着一个个供不应求的H100,它就像淘金热里卖铲子的人,看着全世界的AI公司打得头破血流,自己却闷声发大财,市值一路冲破天际。最新财务文件显示,英伟达将在未来5年累计投入260亿美元巨资,全力推进开源AI大模型的研发。这意味着英伟达不再满足于卖铲子,而是要亲自下场挖金子了。



## 大手笔投资

3月12日,根据英伟达公司向美国证券交易委员会(SEC)提交的财务文件,英伟达将在未来5年累计投入260亿美元(约合1788亿元人民币)巨资,全力推进开源AI大模型的研发。英伟达公司也正式开启了从“芯片制造商”向“全栈式AI顶尖实验室”的战略转型。

根据规划,英伟达此次260亿美元的投入并非单纯聚焦于单一模型研发,而是覆盖开源AI大模型全产业链,资金将在未来18至24个月内逐步落地,首批自研开源AI模型最快将于2026年底至2027年初正式问世。

作为对比,这一投资规模远超OpenAI训练GPT-4时所花费的30亿美元。而在技术路线上,英伟达选择了一条“开放权重”(Open-weight)的“中间道路”。这一模式介于OpenAI的完全闭源与Meta旗下Llama系列的完

全开源之间。

具体而言,英伟达将公开模型的关键参数(权重),允许企业和开发者免费下载,并在自己的设备或私有云上运行、微调,从而满足企业对数据隐私、定制化和成本控制的需求。但模型的训练数据和代码可能不会完全公开。

专注推动AI开放性的非营利机构Laude Institute创始人、计算机科学家Andy Konwinski则将英伟达的这笔投资定性为里程碑式信号。“他们处于众多开放与封闭AI项目的前沿交汇点,”Konwinski表示,“这是他们对开放性信念的一次史无前例的表态。”

此外,行业分析指出,开源策略对英伟达还具有更长远的商业意义。英伟达的模型发布时会公开权重及技术细节,方便初创公司和研究人员在其技术基础上进行修改和创新,这有助于形成围绕英伟达硬件生态的开发者网络,进一步强化其芯片的市场黏性。

## 比肩OpenAI

英伟达自2023年11月发布首个Nemotron模型以来,已陆续推出面向机器人、气候建模及蛋白质折叠等垂直领域的专用模型。英伟达应用深度学习研究副总裁Bryan Catanzaro还透露,英伟达近期已完成一个5500亿参数模型的预训练工作。在核心模型研发上,英伟达将重点研发多模态、多领域前沿大模型,覆盖语言、代码、科学计算、智能体等多个方向。

近日,英伟达还推出专为企业级多智能体系统设计的新一代开源大语言模型Nemotron 3 Super,模型总参数量达1280亿(推理仅激活120亿),原生支持100万token超长上下文窗口。与主流API访问模式不同,英伟达此次开放了模型权重、预训练/后训练数据集及完整训练方案。

1280亿参数,规模与OpenAI GPT-OSS的最大版本大体相当。英伟达声称,在人工智能指数综合评分中,Nemotron 3 Super获得37分,而GPT-OSS仅为33分。

值得一提的是,英伟达同时承认,部分中国模型的得分高于这一水平。此外,英伟达表示Nemotron 3 Super参与了名为Pinch-Bench的新型基准测试,该测试专门评估模型对OpenClaw的控制能力,Nemotron 3 Super在该测试中位列第一。

在技术层面,英伟达公开了训练该模型所采用的多项创新方法,涵盖提升模型推理能力、长上下文处理能力及强化学习响应能力的架构与训练技巧。

Catanzaro表示:“英伟达正在对开源模型开发给予远比以往更重视,我们正在取得大量进展。”

在生态层面,英伟达已与谷歌云Vertex AI、甲骨文云基础设施、戴尔技术、HPE等主流云服务商及硬件厂商达成合作,亚马逊AWS Bedrock及微软Azure的接入亦在筹备中。CodeRabbit、Factory、Greptile等软件开发智能体公司,以及生命科学机构Edison Scientific和Lila Sciences,也已宣布将该模型整合至其智能体工作流。

## 重新定义路线

长期以来,英伟达的核心优势集中在芯片硬件领域,全球AI芯片市占率超过80%,但在AI模型层的话语权相对较弱,此前大模型的技术标准、训练范式大多由OpenAI、Meta等厂商定义。

此次英伟达下场自研顶级开源模型,核心目的就是要从底层定义AI模型的技术路线,让自家的硬件架构、软件栈成为整个AI行业的事实标准,通过开源模型拉动算力需求。若Nemotron成为企业智能体AI的主流基础

模型,大规模运行该模型所需的GPU基础设施仍将倚重英伟达——在模型层推进开放的同时,巩固硬件层的需求锁定。

金融分析师预测,如果英伟达在巩固其硬件霸主地位的同时,能在基础模型市场成功攫取10%的份额,此举有望在三年内为公司每年额外贡献高达500亿美元的营收。Bryan Catanzaro表示,推动开源生态发展完全符合英伟达的核心利益,此次巨额投资并非盲目跟风,而是经过长期行业研判后的战略抉择。

当地时间周二,英伟达CEO黄仁勋还发表了一篇罕见的关于人工智能的长篇博客文章,这是他自2016年以来发表的第七篇公开长文,文章系统阐释了AI产业的底层逻辑,黄仁勋在文中定义了AI的“五层架构”。他表示,当前AI产业仍处于极早期发展阶段,尽管行业已投入数千亿美元,但AI的真正潜力尚未被完全发掘,未来仍需数千万美元的持续投资来完善底层基础设施。

黄仁勋指出,AI已成为当今塑造世界的最大力量之一,它并非单一的聪明应用程序或模型,而是如同电力和互联网一样至关重要的基础设施,运行在真实的硬件、能源和经济基础之上,能够吸收原材料并转化为规模化的智能,未来每家公司都将使用AI,每个国家都将建设AI基础设施。

针对AI发展带来的就业担忧,黄仁勋认为,AI非但不会削减岗位,反而会创造大量新的就业机会,尤其是在基础设施和熟练技术工种领域,支持AI基础设施建设所需的劳动力极其庞大,AI工厂需要电工、水管工、钢铁工人、网络技术人员、安装工和操作员等,这些都是高技能、高薪的岗位,且目前供不应求。AI正在填补全球范围内卡车司机、护士、会计等岗位的巨大劳动力缺口,而非制造失业。

北京商报记者 赵天舒

## Focus

# 本田上市以来首现年度亏损

受全球电动化转型受阻及市场激烈竞争的双重夹击,本田汽车遭遇上市以来首次年度亏损。3月12日,本田汽车宣布,预计2025财年(2025年4月至2026年3月)将录得上市以来的首次净亏损,亏损额最高达6900亿日元(约合人民币247亿元)。财报预告显示,本田2025财年预计经营亏损2700亿至5700亿日元,而此前预期为盈利5500亿日元;预计净亏损4200亿至6900亿日元,与此前预计的3000亿日元盈利形成巨大反差。

回顾本田近三年的财务表现,其业绩曲线呈现“过山车”式波动。2023财年(2023年4月至2024年3月)是本田的高光时刻,实现营收20.43万亿日元,营业利润达1.38万亿日元,创下历史新高,净利润达1.11万亿日元。2024财年(2024年4月至2025年3月)业绩急转直下。

财报显示,该财年营收虽微增至21.69万亿日元,但净利润骤降至835.8亿日元,同比暴跌24.5%。其中中国市场销量下滑影响明显:2024年本田在华销量仅85.37万辆,同比下滑31%。

从销量数据看,本田在中国市场已连续四年下滑。2021年至2024年,其销量分别为156.15万辆、137.31万辆、123.42万辆和85.37万辆,跌幅逐年扩大。

导致本财年巨额亏损的直接原因,是本田对电动化战略的重新评估。公司同日宣布,取消部分美国制造电动汽车的研发和上市计划,预计因此产生的总费用和损失最高达2.5

万亿日元(约合人民币1082亿元)。

华南永昌证券援引Factset数据指出,分析师已纷纷下调对本田的盈利预期,18位分析师对本田2026年EPS预估中位数由2.21美元下修至2.18美元。市场对本田股票的评级分布为7个买入、14个持有和1个卖出,显示出市场对其前景存在分歧。

本田首席执行官三部敏宏此前坦言,各国关税政策频繁修订对业务影响巨大,严重干扰前景规划。作为本田最大市场,美国2024年贡献了140万辆销量,其中近40%来自日本出口,关税政策对其影响深远。面对困局,本田计划在2030财年前投入约10万亿日元推进电动化转型,但近期已决定将部分投资削减至7万亿日元,战略重心向混合动力车型倾斜。

面对困境,本田宣布对电动化战略进行根本性重估,将2030年全球纯电销量目标从200万辆大幅下调至70万至75万辆,并决定结束与通用

汽车的电动车合作。同时,本田宣布自2026年4月起,将汽车研发职能重新回归旗下研发子公司,试图通过强化自研来扭转颓势。

处境更为艰难的当属日产汽车。根据其发布的2025财年业绩预期,日产汽车预计全财年将出现6500亿日元的净亏损。财报显示,2025财年前9个月(2025年4月至12月),日产全球销量同比下降5.8%至226万辆,累计净亏损已达2502亿日元(约合人民币119.93亿元)。即便在第三财季实现了单季经营盈利,但盈利规模大幅收缩。

为改善经营状况,日产正推进“Re:Nissan”重组计划,包括全球关闭7家工厂、裁员2万人等举措。透过日产和本田的遭遇可以看出,在行业转型的十字路口,未能准确把握市场节奏的代价,正在以最直观的财务数字呈现出来。

不过,同为日系车企的丰田汽车在2025年延续了其全球销冠的强势地位,全年销量达966.2万辆,超过大众集团蝉联第一。财务数据显示,2025自然年(日本的财政年度为每年4月1日至次年3月31日)丰田汽车实现营收约44.7万亿日元(约合人民币2万亿元),同比增长约6.8%;实现净利润约为3.7万亿日元(约合1649亿元人民币),同比下降约26%,以此计算其单车利润为1.7万元人民币。

但丰田并未能规避外部冲击:美国关税带来了1.45万亿日元的负面影响,致使其北美市场出现营业亏损。此外,亚洲地区销量出现下滑。

北京商报综合报道

## · 图片新闻 ·

# 美国通胀降温但预期升温



11日,美国洛杉矶,一名顾客在Arco加油站给汽车加油,该站油价超过每加仑5美元。视觉中国/图

随着油价、关税等因素对物价影响的逐步显现,美联储年内首次降息节点或进一步推迟至9月。当地时间周三(11日),美国劳工统计局报告显示,2月CPI环比上涨0.3%,1月涨幅为0.2%。同比方面,CPI上涨2.4%,与预期及1月持平,反映去年同期高基数效应消退。能源价格环比上涨0.6%,同比上涨0.5%。食品价格环比上涨0.4%,同比上涨3.1%。鸡蛋价格环比下降3.8%,同比大幅下跌42.1%。

最新物价数据公布后,美联储降息预期有所降温。据芝加哥商品交易所(CME)美联储观察工具显示,数据公布前,市场预期美联储在下周3月会议上维持利率不变的概率为99%。4月会议的降息概率仅为11%,而一个月前为21%。交易员预计美联储下次降息将在9月,年底前实施第二次降息的概率约为43%。

机构普遍认为,这份数据发布时间早于近期

中东局势引发的油价暴涨,意味着能源成本上涨带来的影响很可能在未来几个月逐步显现。未来数月,高油价可能令通胀前景复杂化,因为汽油及其他能源产品涨价通常会传导至运输、航运和各类消费品。即便核心物价压力保持稳定,原油价格持续上涨也会迅速体现在整体通胀数据中。这些因素会在多大程度上影响美联储的利率决策,将在下周政策会议后变得更加清晰。

摩根士丹利表示,美联储最早可能在6月份恢复降息,但中东局势引发的油价冲击可能会推迟美联储的行动。尽管能源价格上涨可能加剧通胀,但该行经济学家仍坚持此前的预测,即美联储今年将两次降息,分别在6月和9月,每次降息25个基点。不过,他们认为美联储存在将首次降息推迟到9月甚至12月的可能性,这样一来,下一次降息可能会推迟到2027年。

北京商报综合报道