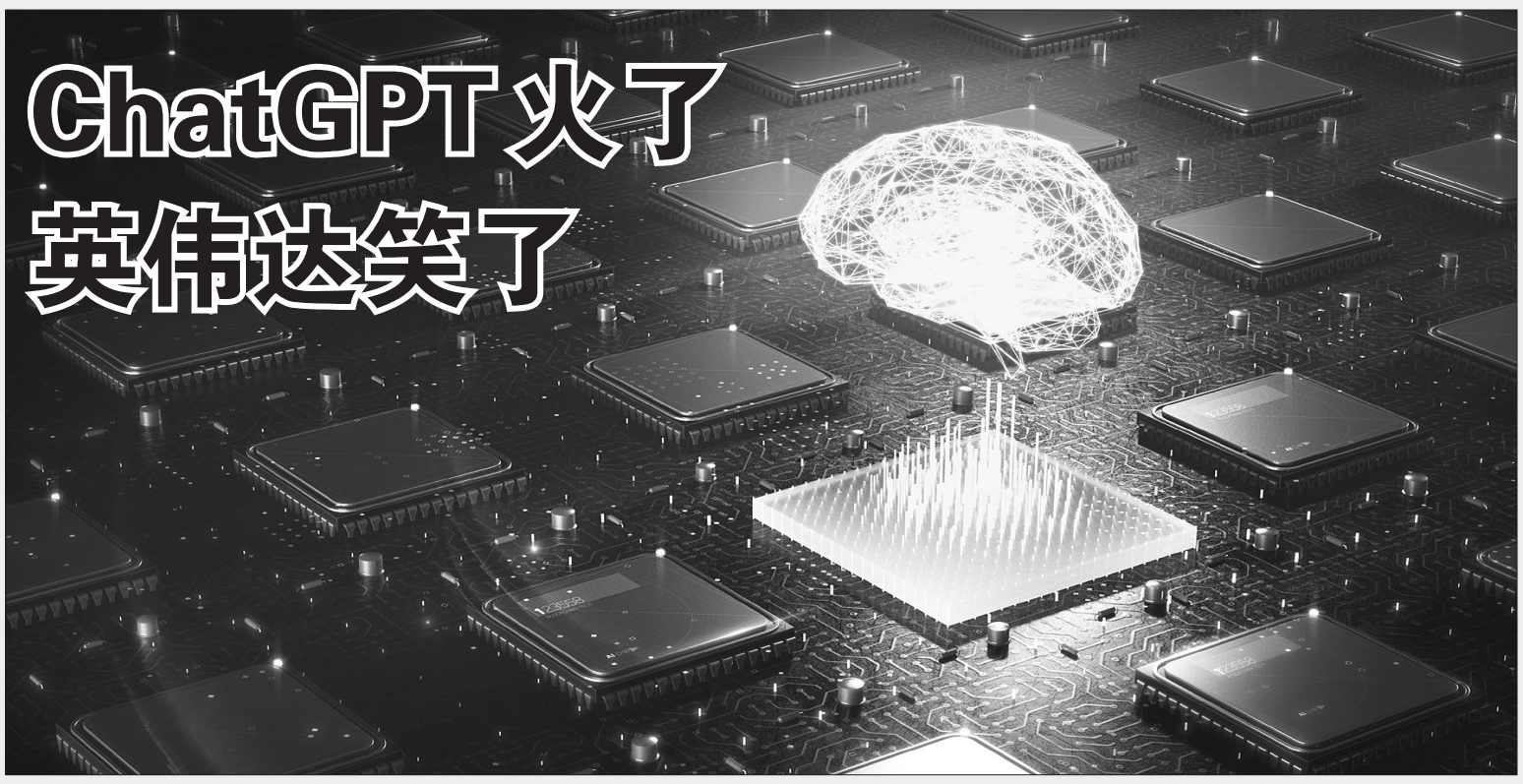

# ChatGPT火了 英伟达笑了

本报记者 张心怡

英伟达又“赚麻了”。1月3日——美股第一个交易日，英伟达的收盘价为143美元，一个月后的2月3日，英伟达股票的收盘价已经来到211美元，一个月涨了47%。华尔街分析师预计，英伟达在1月份的股价表现预计将为其创始人黄仁勋增加51亿美元的个人资产。半导体企业股价的起起伏伏本属常态，可今时不同往日，半导体市场正在经历罕见的下行周期。英伟达股价此时的增长意味着，美股市场从它身上看到了逆境中的希望。

而这一希望之火的来源，就是当前科技圈的“顶流”——ChatGPT。这款由OpenAI推出的聊天机器人在推出仅两个月后，月活跃用户就达到了1亿人，成为历史上用户增长最快的消费应用。由于ChatGPT属于生成式AI，被誉为“AI芯片”第一股的英伟达应声而涨，在ChatGPT商业化模式尚未明确的初期阶段就斩获了一波红利。而美股市场如此看好英伟达，既有英伟达从显卡厂商成长为AI芯片霸主的历史原因，也在于ChatGPT当前阶段与英伟达生态的契合性。

## 缘何成为“AI芯片第一股”？

20世纪90年代，3D游戏的快速发展和个人电脑的逐步普及，彻底改变了游戏的操作逻辑和创作方式。1993年，黄仁勋等三位电气工程师看到了游戏市场对于3D图形处理能力的需求，成立了英伟达，面向游戏市场供应图形处理器。1999年，英伟达推出显卡GeForce 256，并第一次将图形处理器定义为“GPU”，自此，“GPU”一词与英伟达赋予它的定义和标准在游戏界流行起来。

那时，三位创始人可能没有想到：让英伟达股价飞升的不是游戏显卡，而是AI芯片；让英伟达市值超越英特尔的不是游戏显卡，而是AI芯片。

在30年后，半导体遭遇罕见逆风的2022年，撑起英伟达财报表现的不是游戏显卡，而是包含AI加速、高性能计算、超算等业务的数据中心业务。至2023财年第三季度（截至2022年10月30日），数据中心业务的营收已经是游戏业务的两倍有余，对英伟达的营收贡献高达64.6%。

有意思的是，让英伟达踏上AI这条路的，不是GPU硬件，而是软件编程平台：CUDA（统一计算架构）。在21世纪初，CPU难以继续维持每年50%的性能提升，而内部包含数千个核心的GPU能够利用内在的并行性继续提升性能，且GPU的众核结构更加适合高并发的深度学习任务。这一特性逐渐被深度学习领域的开发者注意。但是，作为一种图形处理芯片，GPU难以像CPU一样用C语言、Java等高级程序语言，极大地限制了GPU向通用计算领域发展。英伟达很快注意到了这种需求。为了让开发者能够用英伟达GPU执行图形处理以外的计算任务，英伟达在2006年推出了CUDA平台，支持开发者用熟悉的高级程序语言开发深度学习模型，灵活调用英伟达GPU算力，并提供数据库、排错程序、API接口等一系列工具。虽然当时方兴未艾的深度学习并没有给英伟达带来显著的收益，但英伟达一直坚持投资CUDA产品线，推动GPU在AI等通用计算领域前行。

6年后，英伟达终于等到了向AI计算证明GPU的机会。在21世纪10年代，由大型视觉数据库ImageNet项目举办的“大规模视觉识别挑战赛”是深度学习的标志性赛事之一，被誉为计算机视觉领域的“奥赛”。在2010年和2011年，ImageNet挑战赛的最低差错率分别是29.2%和25.2%，有的团队差错率高达99%，深度学习的前景因不确定性蒙上了一层阴影。

2012年，来自多伦多大学的博士生Alex Krizhevsky，用120万张图片训练神经网络模型，和前人不同的是，他选择用英伟达GeForce GPU为训练提供算力。在当年的ImageNet，Krizhevsky的模型以约15%的差错率夺冠，震惊了神经网络学术圈。

这一标志性事件，证明了GPU对于深度学习的价值，也打破了深度学习的算力枷锁。自此，GPU被广泛应用于AI训练等大规模并发计算场景。数据显示，在2010—2011年，ImageNet挑战赛中没有任何团队使用GPU，2012年Krizhevsky首开先河后，2013年参赛团队使用的GPU达到了60块，2014年进一步提升至110块。

除了学术圈，科技企业也纷纷向英伟达伸出橄榄枝。2012年，英伟达与谷歌人工智能团队打造了当时最大的人工神经网络。到2016年，Facebook、谷歌、IBM、微软的深度学习架构都运行在英伟达的GPU平台上。2017年，英伟达GPU被惠普、戴尔等厂商引入服务器，被亚马逊、微软、谷歌等厂商用于云服务。2018年，英伟达为AI和高性能计算打造的Tesla GPU被用于加速美国、欧洲和日本最快的超级计算机。

与英伟达AI版图一起成长的，是股价和市值。2016年初，英伟达的股价在30美元左右。而2018年10月，英伟达股价来到292美元的高位，一度被资本市场誉为“AI芯片第一股”。2020年7月，英伟达市值首次超越英特尔，成为当时美国最大的芯片公司。

## ChatGPT的最大受益者？

在AI计算领域的长期储备，让英伟达在ChatGPT尚处于商业化探索的早期，就率先受益，在股市斩获颇丰。接下来，ChatGPT的火热有望进一步体现在英伟达的销售额上。IDC亚太区研究总监郭俊丽向《中国电子报》表示，从算力来看，ChatGPT至少导入了1万颗英伟达高端GPU，总算力消耗达到了3640PF-days。

“我们预计，ChatGPT很可能推动英伟达相关产品在12个月内销售额达到35亿美元至100亿美元。”郭俊丽说。ChatGPT引起了全球用户的极大兴趣，在于它能够满足各种各样的需求。解释名词概念、写作文、作诗、填写表单、编写SQL查询并且执行，甚至可以编写代码。而支撑这种多元化功能的，是AI大模型技术。一位AI从业者向《中国电子报》记者表示，大模型技术涉及AI开发、推理、训练的方方面面。所谓模型的“大”，主要是参数量大、计算量大，需要更大体量的数据和更高的算力支撑。对于GPU厂商来说，大模型是值得期待的算力红利。

可英伟达真的能够在这一波算力红利中独占鳌头吗？如今，在通用GPU领域，AMD一直是独立GPU的第二大供应商，且一直保持着高速成长步伐。2022年，AMD数据中心事业部的营业额实现了高达64%的同比增长。英特尔一直是全球最大的集成显卡供应商，在宣布重返独立GPU市场后，推出了面向数据中心和AI的Xe HP架构以及面向高性能计算的Xe HPC架构。与此同时，乘着AI东风崛起的一批中小GPU企业，也对新的市场机会虎视眈眈。显然，英伟达并不是AI开发者们的唯一选择。

那么，美股市场对英伟达的信心来源于什么？首先是GPU平台的通用性。一位互联网从业者向记者表示，小模型是做一个任务就训练一个模型，而大模型要具备一定的通用性。如果说对小模型的训练是一堂课，那么对大模型的培训就相当于九年义务教育。

而CUDA平台加持的英伟达GPU，就是以通用性见长。“英伟达通用性高，支持AI的能力强。当一个新的AI热点出现，其成长过程中会出现哪些新型应用是难以在初期预测的，通用性强的芯片平台是更加安全的选择。因此，AI开发者往往会优先选择英伟达的GPU。等这个AI热点成熟了，方向相对明确了，再去研发自己的芯片。”Gartner研发副总裁盛陵海向记者表示。

英伟达的另一道护城河，是其AI生态的黏性。CUDA几乎只支持英伟达的Tesla架构GPU，不容易迁移，有利于AI开发者与英伟达软硬件长期绑定。“在AI领域，英伟达的GPU占据绝对的领导地位；在训练领域，英伟达的GPU产品的市场份额超过80%。再配合CUDA软件工具，实现对GPU等硬件芯片的捆绑，构筑了行业壁垒。”赛迪顾问集成电路产业研究中心总经理滕冉向记者表示。

## 这波红利能吃多久？

在英伟达的发展史中，令其股价飙升的热点有很多。有些是技术畅想，比如AI和元宇宙，在为全社会带来想象空间的同时，也倒逼英伟达推出了新的产品和平台。但也有一些是单纯的GPU走量，比如“挖矿”，虽然短期内急速拉升了英伟达和AMD的显卡销量，但也对芯片供应秩序和GPU厂商财务表现的稳定性造成了伤害，成了昙花一现的短期红利。

ChatGPU带来的算力红利，又能持续多久？这个问题，可以分两个层次来看。第一个层次，在于ChatGPU是不是一项能够长期发展的颠覆性技术。在黄仁勋本人看来，答案是肯定的。在近期参加的Berkeley Haas商学院Dean's Speaker系列谈话中，黄仁勋表示人工智能领域出现了ChatGPT，相当于手机领域出现了iPhone。“ChatGPT的出现对人工智能领域的意义，类似手机领域‘iPhone’的出现。这一刻在科技领域具有里程碑的意义，因为现在大家可以将所有关于移动计算的想法，汇集到了一个产品中。例如，通过API接口，可以将ChatGPT连接到数据表、Powerpoint、绘图程序、照片编辑程序等，一切都可以变得更完善。”黄仁勋说。

但也有观点认为，ChatGPT并没有为底层技术带来变革。Meta首席AI科学家、图灵奖得主杨立昆表示，ChatGPT并没有为底层技术带来创新，更多的是一个组合得很好的产品。盛陵海向记者表示，目前来看，ChatGPT还是基于现有的数据进行组合式回答，而不是去创造新的内容。ChatGPT要长远发展，需要持续向生产工具演变，比如短期内可以用来提升搜索引擎的正确率，而不是仅仅停留在与用户对话。

在芯片层面，尚未看到针对ChatGPT推出的新产品。但ChatGPT作为明星产品，引发的是全社会对于生成式AI和大模型技术的关注，而后者对于芯片用量的更大需求、芯片规格的更高要求，已经是较为明朗的趋势。“未来大模型将成为AI技术领域重要的生产工具，需要更强的训练、推理能力，支撑海量数据模型且高效地完成计算，这些要求会对芯片的算力、存储容量、软件栈、带宽等技术有更高的要求。”郭俊丽表示。

第二个层次是，在ChatGPT等生成式AI发展的不同时期，英伟达的蛋糕份额是否会有所变化。

对于中小企业来说，一旦想明白了要用ChatGPT做什么，按照业务特点定制AI芯片，是更经济的选择。郭俊丽表示，随着ChatGPT技术不断成熟推进、算法不断优化普及，ASIC将更具备竞争优势。

而头部企业普遍想从芯片层、框架层、模型层，一直做到应用层。因而，无论国际的谷歌、微软、亚马逊，还是国内的百度、阿里，都推出了自己的算力芯片。为了让芯片层更加贴合自己的框架模型，科技企业会不断提升软硬件的契合度，进一步提升自研芯片的比例。

因此，当ChatGPT发展到成熟期，其算力底座有可能从英伟达独占鳌头的局面逐渐向“百家争鸣”的割据战倾斜，从而压缩英伟达在该领域的盈利空间。但那个时期，可能又会有下一个AI热点出现，开启下一轮从通用GPU平台进行早期探索的循环。毕竟芯片企业的起起落落，概念股票的跌跌涨涨，都源于人们对于技术进步和美好生活的期待。只要想象力不会终止，就永远有新的发现令市场瞩目，新的热点供企业追逐。

# CINNO Research：2022年中国半导体投资1.4万亿元

**本报讯** 2月8日，市场研究机构CINNO Research举办“新兴科技产业新春策略研讨会”。CINNO Research表示，根据统计，2022年中国的半导体、光电显示、线路板、消费电子与新能源等5大新兴科技板块共实现投资11.7万亿元。其中，新能源行业整体投资金额达9.2万亿元，占比最大，达到79%；半导体投资1.4万亿元，占比13%；线路板与光电显示占比各约3%，消费电子约2%。

“双碳”战略带来新能源的市场发展机遇，储能、锂电池、光伏成为三大主要引擎。2022年涉足锂电池、光伏产业投资的企业分别超400家和300家，市场竞争态势日趋激烈。新能源汽车带动了动力电池的投资。新能源锂电池行业整体投资金额近2.2万亿元，锂电池模组与正极材料项目投资金额各超8000亿元，占锂电行业投资规模的3/4；负极材料、电解铜箔分别占比8%和6%。锂电隔膜投资规模也达到近800亿元，成为2022年中国膜材料最大投资赛道之一。

半导体投资中，芯片设计产业成为主力，投资规模超5600亿元。半导体材料的资金投入规模超2800亿元，封测投资超1300亿元，半导体设备投资超400亿元。

战略性新兴产业是以重大技术突破和重大发展需求为基础，对经济社会全局和长远发展具有重大引领带动作用，知识技术密集、物质资源消耗少、成长潜力大、综合效益好的产业。投资是拉动新兴产业发展的重要手段。但是，CINNO Research也指出，提升国内企业创新能力，提升基础研究与供应链协同创新，建设国内统一大市场是新兴产业未来的重要发展方向。

（陈炳欣）

# SK海力士10年来首次出现季度亏损

**本报讯** 2月1日，韩国存储芯片巨头SK海力士在官网公布了其截至2022年12月31日的2022财年及第四季度财务报告。SK海力士表示，从2022年下半年开始，存储芯片的市场需求不断减少，产品价格大幅下跌，第四季度经营业绩由盈转亏。2022财年第四季度合并收入为7.6986万亿韩元(约合人民币423.4亿元)，营业亏损为1.7012万亿韩元(约合人民币93.6亿元)，净亏损为3.5235万亿韩元(约合人民币193.8亿元)。这是SK海力士自2012年第三季度以来首次出现季度营业亏损。

SK海力士官网显示，其2022财年合并收入为44.6481万亿韩元(约合人民币2455.6亿元)，营业利润为7.0066万亿韩元(约合人民币385.4亿元)，净利润为2.4389万亿韩元(约合人民币134亿元)。2022财年营业利润率为16%，净利润率为5%。2022年，SK海力士在服务器和PC市场的高容量DRAM产品供应量有所提升。另外，在发展势头良好的AI、大数据、云端等领域所需的DDR5和HBM等产品的销售额也在逐步提升。特别是数据中心用的SSD（固态硬盘）的销售收入与去年相比，增加了4倍。

SK海力士表示，虽然2022年销售额保持增长，但由于半导体市场情况持续低迷，2022年的营业利润较2021年相比有所下滑。为了减少损失，SK海力士2023年的投资规模相较于2022年的19万亿韩元(约合人民币1045亿元)将减少50%。但仍计划继续投资DDR5、LPDDR5、HBM3等主力产品的量产和未来高成长领域。

SK海力士一直在加速研发新产品，以刺激客户淘汰旧产品，扩大市场需求。1月12日，SK海力士宣布，其研发的第四代10纳米级DDR5服务器DRAM获得了英特尔近期上市的全新第四代Xeon服务器处理器兼容认证。SK海力士表示，英特尔的第四代Xeon服务器处理器是存储器半导体行业反弹的关键，SK海力士将把握此次机会，尽早克服存储器半导体的低迷市况。1月25日，SK海力士宣布，开发出当前速度最快的移动DRAM（内存）“LPDDR5T”，并已向客户提供了样品。据介绍，本次产品的速度比现有产品快13%，运行速度高达9.6Gbps。

SK海力士预测，虽然今年上半年的半导体市场情况可能持续低迷，但下半年会出现好转，IT企业将增加价格大幅下降的存储芯片的使用量，市场需求将逐渐回升。

（许子皓）

# AMD 2022年营收同比增长44% 净收入大幅下降

**本报讯** 当地时间1月31日，AMD发布2022年第四季度及全年财报。信息显示，AMD 2022年第四季度营业额为56亿美元，同比增长16%。按照GAAP（美国通用会计准则）标准，AMD第四季度净收入2100万美元，同比下降98%，每股收益同比下降99%。按照非通用会计准则，AMD第四季度净收入与去年同期持平，每股收益同比下降25%。

财报显示，AMD在2022年第四季度的营收增长，主要得益于嵌入式和数据中心业务的增长，部分被较低的客户端和游戏业务营业额所抵消。从各部门营收来看，数据中心事业部营业额为16.55亿美元，同比增长42%，主要得益于霄龙服务器处理器的销售增长。客户端事业部营业额为9.03亿美元，同比下降51%，主要原因是PC市场疲软，以及PC供应链的库存调整，导致处理器出货量减少。游戏事业部营业额为16.44亿美元，同比下降7%，主要原因是游戏图形业务销售下降，部分被半定制产品营业额的增长所抵消。嵌入式事业部营业额为14亿美元，同比增长1868%，主要由于并入了赛灵思嵌入式业务的营业额。

对于2022年第四季度净收入和每股收益的大幅下降（按照GAAP标准），AMD在财报中表示，主要原因是与收购赛灵思相关的无形资产摊销。基于非GAAP标准，本季度摊薄后，每股收益为0.69美元，同比下降25%，主要由于客户端业务营业额较低。

2022年全年，AMD总营业额达236亿美元，同比增长44%，部分被较低的客户端业务营业额所抵消。在AMD和赛灵思合并的基础上，2022年并表营业额为241亿美元，同比增长20%。

分部门来看，2022年AMD数据中心事业部营业额为60.43亿美元，同比增长64%。客户端事业部营业额为62.01亿美元，同比下降10%，是唯一同比负增长的业务部门。游戏事业部营业额为68.05亿美元，是营收最高的业务部门之一。嵌入式事业部营业额为45.52亿美元，同比增长1750%，是增幅最大的业务部门之一。

AMD预计2023年第一季度营业额约为53亿美元，上下浮动3亿美元，同比下降约10%。与去年同期相比，客户端和游戏业务预计下降，部分被嵌入式和数据中心事业部的增长所抵消。AMD预计2023年第一季度非GAAP标准的毛利率约为50%。

（张心怡）

社长：张立 社址：北京市海淀区紫竹院路66号赛迪大厦18层 邮编：100048 每周二、五出版 周二8版 周五8版 零售4.50元 全年定价420元 广告部：010-88558848/8808 发行部：010-88558777 广告许可证：京石工商广登字20170003号 发行单位：中国电子报社 印刷单位及地址：经济日报印刷厂 北京市西城区白纸坊东街2号