

电子

报告日期：2023 年 02 月 08 日

ChatGPT 的背后：指数级增长的芯片与半导体需求

——科技中观策略研究系列

投资要点

□ **ChatGPT 热潮席卷全球。** ChatGPT (Chat Generative Pre-trained Transformer) 是由 OpenAI 于 2022 年 12 月推出的对话 AI 模型，一经面世便受到广泛关注，其 2023 年 1 月月活跃用户达到 1 亿，是史上月活用户增长最快的消费者应用。主要以问答类为主，但与其他问答类 AI 产品不同，ChatGPT 具备训练集中的所有知识，具有语言生成能力，可以实现拟人化交流，而不只是像天猫精灵、小爱同学等其他 AI 产品的一问一答模式。在问答模式的基础上 ChatGPT 可以进行推理、编写代码、文本创作等等，这样的特殊优势和用户体验使得应用场景流量大幅增加。

□ **▲ 芯片需求=量↑x 价↑**

1) 量：AIGC 带来的全新场景+原场景流量大幅提高

① **技术原理角度：** ChatGPT 是基于 GPT3.5 架构开发的对话 AI 模型，历经 GPT-1/2/3 迭代，在 GPT3.5 模型后，开始引入代码训练和指令微调，加入 RLHF 技术（人类反馈强化学习），从而实现能力进化。GPT 作为知名的 NLP 模型，基于 Transformer 技术，随着模型不断迭代，层数也越来越多，对算力的需求也就越来越大。

② **运行条件角度：** ChatGPT 完美运行的三个条件：训练数据+模型算法+算力。其中训练数据市场广阔，技术壁垒较低，投入足够的人力物力及财力后即可获得；基础模型及模型调优对算力的需求较低，但获得 ChatGPT 功能需要在基础模型上进行大规模预训练，存储知识的能力来源于 1750 亿参数，需要大量算力。所以，算力是 ChatGPT 运行的关键所在。

2) 价：对高端芯片的需求将拉动芯片均价

采购一片英伟达顶级 GPU 成本为 8 万元，GPU 服务器成本通常超过 40 万元。对于 ChatGPT 而言，支撑其算力基础设施至少需要上万颗英伟达 GPU A100，一次模型训练成本超过 1200 万美元。从芯片市场角度出发，芯片需求的快速增加会进一步拉高芯片均价。目前 OpenAI 已推出 20 美元订阅模式，初步构建了优质的订阅商业模式，未来继续扩容的能力将会大幅提升。

□ **“背后英雄”系 GPU 或 CPU+FPGA 等算力支撑**

1) GPU 可支撑强大算力需求。 具体来说，从 AI 模型构建的角度：第一阶段是用超大算力和数据构建预训练模型；第二阶段是在预训练模型上进行针对性训练。而 GPU 由于具备并行计算能力，可兼容训练和推理，所以目前被广泛应用，ChatGPT 训练模型中就已导入至少 1 万颗英伟达 GPU（曾经风靡一时的 AlphaGO 也只需要 8 块 GPU），推理部分使用微软的 Azure 云服务，也需要 GPU 进行运作。所以，ChatGPT 的火热兴起对 GPU 的需求可见一斑。

2) CPU+FPGA 拭目以待。 从深度学习的角度来看，虽然 GPU 是最适合深度学习应用的芯片，但 CPU 和 FPGA 也不可忽略。FPGA 芯片作为可编程芯片，可以针对特定功能进行扩展，在 AI 模型构建第二阶段具有一定的发挥空间。而 FPGA 想要实现深度学习功能，需要与 CPU 结合，共同应用于深度学习模型，同样可以实现庞大的算力需求。

3) 云计算依靠光模块实现设备互联。 AI 模型向以 ChatGPT 为首的大规模语言模型发展，驱动数据传输量和算力提升。伴随数据传输量的增长，光模块作为数

行业评级：看好(维持)

分析师：陈杭

执业证书号：S1230522110004
chenhang@stocke.com.cn

研究助理：安子超

anzichao@stocke.com.cn

相关报告

- 1 《2023 半导体未来十大产业趋势预测》2023.01.21
- 2 《如何判断本轮消费类芯片的库存周期拐点？》2023.01.17
- 3 《近期全球半导体产业政策梳理》2023.01.10
- 4 《从全球成熟工艺占比，看本土半导体设备发展》2023.01.08
- 5 《北方华创在半导体设备行业的产业格局》2022.12.31
- 6 《华大九天（301269）：国产 EDA 变局中开新局》2022.12.27
- 7 《华为产业链深度系列研究：鲲鹏生态助力华为涅槃》2022.12.26
- 8 《华为产业链深度系列研究：华为造车的三条路》2022.12.11
- 9 《华为产业链深度系列研究：华为战略及破局关键》2022.12.10
- 10 《站在新科技牛市起点 —— 复盘篇：2019-2021 年十倍股成因盘点》2022.12.06

据中心内设备互联的载体，需求量随之增长。此外，伴随算力提升能耗增长，厂商寻求降低能耗方案，推动低能耗的光模块发展。

- **结论：** ChatGPT 作为一种新兴的超智能对话 AI 产品，无论是从技术原理角度还是运行条件角度，ChatGPT 都需要强大算力作为支撑，从而带动场景流量大幅增加，此外，ChatGPT 对于高端芯片的需求增加也会拉动芯片均价，量价齐升导致芯片需求暴涨；面对指数级增长的算力和数据传输需求，可以提供 GPU 或 CPU+FPGA 芯片厂商、光模块厂商即将迎来蓝海市场。
- **看好：** 看好国内 GPU、CPU、FPGA、AI 芯片及光模块产业链玩家。
 - ① GPU：海光信息、景嘉微
 - ② CPU：龙芯中科、中国长城
 - ③ FPGA：安陆科技、复旦微电、紫光股份
 - ④ AI 芯片：寒武纪、澜起科技
 - ⑤ 光模块：德科立、天孚通信、中际旭创
- **风险提示：** 1) 半导体下游需求不及预期；2) 研发进度不及预期。

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 买入：相对于沪深 300 指数表现 + 20% 以上；
2. 增持：相对于沪深 300 指数表现 + 10% ~ + 20%；
3. 中性：相对于沪深 300 指数表现 - 10% ~ + 10% 之间波动；
4. 减持：相对于沪深 300 指数表现 - 10% 以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 看好：行业指数相对于沪深 300 指数表现 + 10% 以上；
2. 中性：行业指数相对于沪深 300 指数表现 - 10% ~ + 10% 以上；
3. 看淡：行业指数相对于沪深 300 指数表现 - 10% 以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层

北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层

深圳地址：广东省深圳市福田区广电金融中心 33 层

上海总部邮政编码：200127

上海总部电话：(8621) 80108518

上海总部传真：(8621) 80106010

浙商证券研究所：<https://www.stocke.com.cn>