

证券研究报告|行业深度

传媒

行业评级 强大于市 (维持评级)

2024年3月14日



OpenAI 的飞轮：

“AI 新品” & “巨量融资” & “算力”

证券分析师：

杨晓峰 执业证书编号：S0210524020001

研究助理：

马梓燕 执业证书编号：S0210124030014

陈熠晔 执业证书编号：S0210123080069

请务必阅读报告末页的重要声明

➤ 一、OpenAI的“融资规划和实际重置成本”

- 1、OpenAI融资的理由是支撑未来日活用户数空间，及满足未来“多模态&高算力”的AI新品上线后推理算力需求。融资额度决定了AI模型的移动终端占比和高精度模型占比的可能关系。
- 2、融资预计将用于重置OpenAI产业链，其中晶圆厂和电力为重置核心。OpenAI将采用“切入芯片生产上游”和“能源转向更低成本”的途径对成本进行重置。
- 3、目前，“高性能模型”的日活算力成本较高，只有对实际成本进行重置后，才能更好匹配用户付费能力，进而显著走通商业模式。单次推理算力成本与一次生成内容的tokens、模型参数量和迭代次数成正比。高精度模型的成本预计将是目前模型的数十倍。

➤ 二、OpenAI的“AI新品规划&算力需求”

OpenAI未来的产品线有文生3D、更好的语音模型（交互方式）、更好的网页浏览（网页Agent）、个性化（PC Agent）和GPT-5。3D模型方面，OpenAI已发布Shape E、Point E两款文生3D模型，“3D数据量”或带来涌现；Agent领域预计有两款新品+Android主屏幕小组件上线，语音向聊天式对话&视频音频生成方向改进。

➤ 三、OpenAI与竞对资源竞争：融资&算力&能源

- 1、在一级市场，OpenAI和其他AI初创公司抢夺投资。竞争的核心是主权基金对AI产业的配置。
- 2、OpenAI与Meta、谷歌争夺产业资源：在相近时间紧密发布多款“同类AI产品”，并均在能源&芯片领域积极布局。

➤ 投资建议:

➤ 建议关注: 一、短期视角: 从OpenAI新品节奏视角的映射逻辑:

1、Sora落地: 1) IP产业链: 荣信文化、掌阅科技、中文在线、中信出版、奥飞娱乐等; 2)、主旋律, 人民网; 3) AI电影产业链: 上海电影、光线传媒等; 4)、文生视频, 超讯通信和因赛集团。

2、文生3D: 3D素材, 丝路视觉、凡拓数创; 游戏公司, 紫天科技、宝通科技、恺英网络、巨人网络、三七互娱、吉比特、完美世界、盛天网络、游族网络、星辉娱乐、姚记科技、神州泰岳。

3、Agent: 昆仑万维。

4、港股内容平台: 阅文集团、哔哩哔哩、快手。

➤ 建议关注: 二、中期视角: 从OpenAI与竞对的竞争来看, 行业进入GPT4冲刺期, **AI推理将出现拐点式增长**, 同时AI新平台或诞生。

1、AI推理算力: 海外算力, 英伟达和AMD, 及与之相关的光模块如中际旭创等; 国内算力, 华为昇腾算力产业链、寒武纪等。

2、AI新平台: 字节AI工厂模式暂时领跑, 字节产业链如省广集团; 同时关注腾讯、阿里和百度的最新进展。

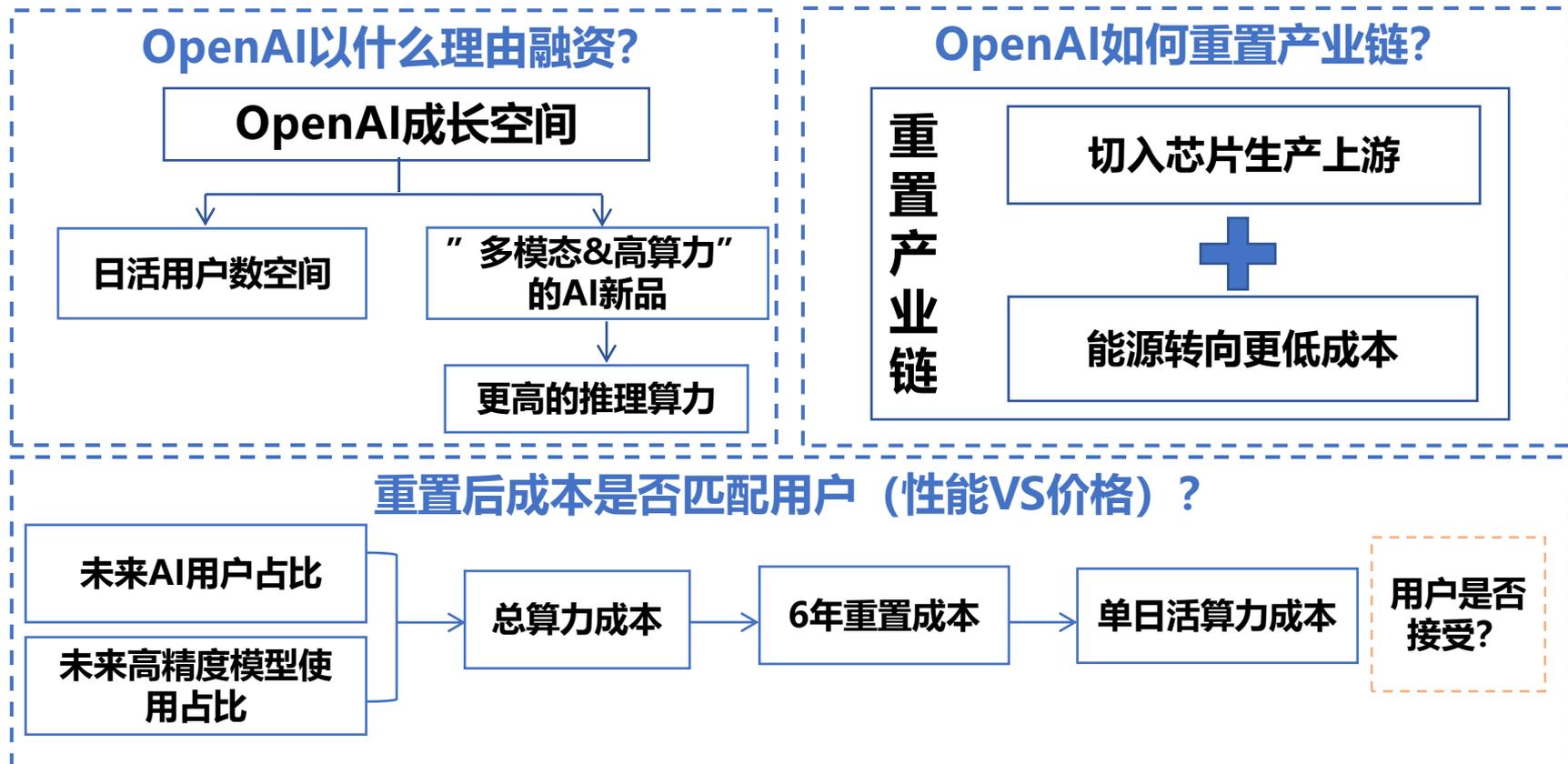
➤ 建议关注: 三、长期视角: 从OpenAI的未来发展路径来看, 核心是通过“重置产业链”降低“高性能&高算力”的AI新模型的单用户算力成本, 促使消费者能在一定付费下使用最新AI模型。因此在此背景下长期视角的产业链环节**主要是晶圆厂和电厂**, 晶圆厂包括台积电、中芯国际、英特尔。

➤ 风险提示: 海外 AI融资进度不及预期, 海外AI产业重置进度不及预期

- OpenAI的“融资规划和实际重置成本”
- OpenAI的“AI新品规划&算力需求”
- OpenAI与竞对资源竞争：融资&算力&能源
- 投资建议&风险提示

1 OpenAI的飞轮：成长空间→融资框架→用户付费

图表：OpenAI飞轮研究框架



1.1 OpenAI 万亿融资规模可能的组合：AI用户空间 VS 单用户算力成本

➤ 未来总算力成本(按市场价值计算)=未来日活用户数量*未来每日活用户年算力成本。

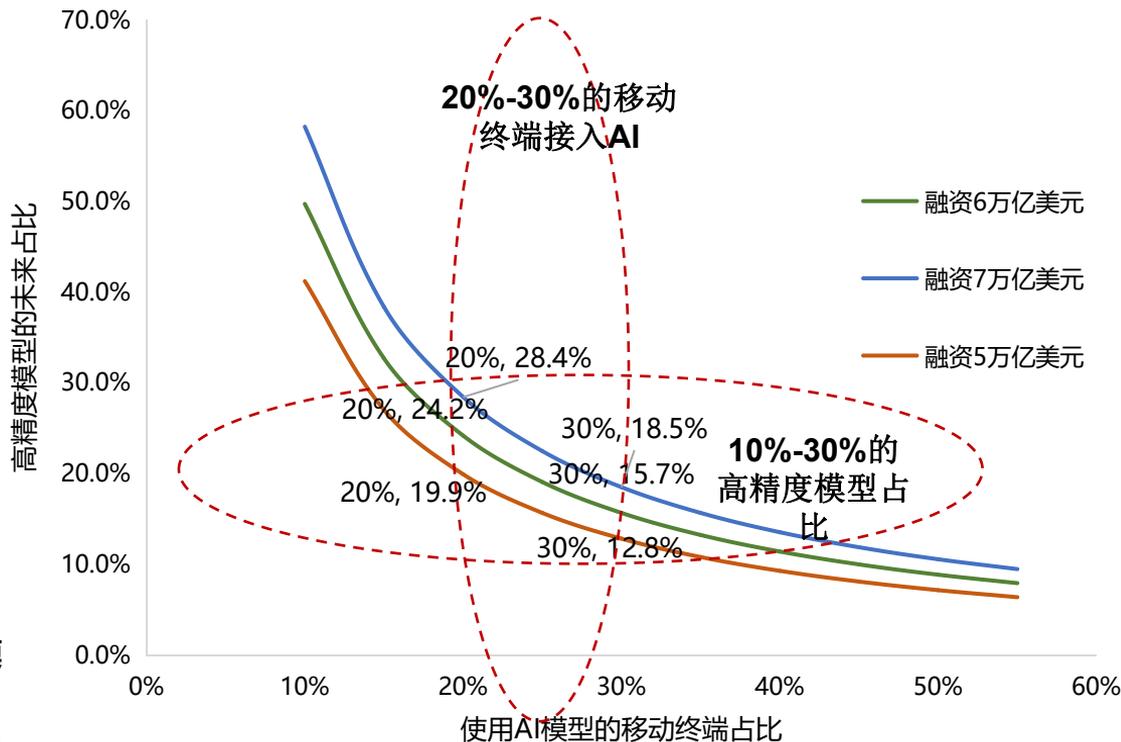
未来AI用户日活：目前AI产业可类比智能手机产业，闭源OpenAI与苹果的形势相同，开源模型则可类比安卓，预期未来AI用户渗透率对标移动互联网。假设按移动互联网用户规模估算空间，对标IOS在移动互联网中的渗透率。

未来每日活年算力成本：假设以未来高精度模型Sora为案例，并假设其占一定比例背景下的单用户算力成本。

➤ 据《华尔街见闻》，OpenAI的融资计划5-7万亿美元，而数据中心一般折旧6年，则每年折旧约0.83~1.16万亿美元。

➤ 分析框架：在不同融资额度下，可计算每年数据中心的折旧作为算力成本。由未来总算力成本公式，可以推得OpenAI在不同融资下，向潜在投资者描述的未来的可能组合，使用AI模型的移动终端占比（决定未来AI用户数量）和高精度模型占比（决定未来每日活用户算力成本）的可能组合关系。

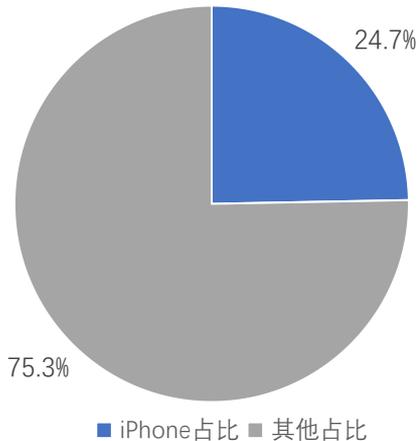
图表：不同融资下，使用AI模型的移动终端占比和高精度模型占比



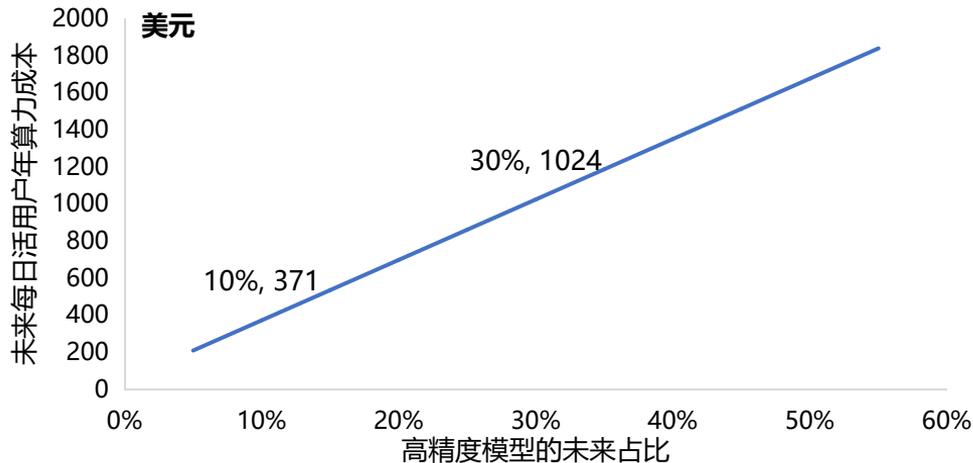
1.2 OpenAI用户市场空间：对标移动互联网中苹果市场份额

- **未来日活用户数量方面：对标移动互联网渗透率，如果OpenAI能像苹果IOS在智能手机的渗透率，在20%~30%的情况下，未来约有12-18亿的AI日活用户。**据IDC报道，最新全球苹果手机市场份额为24.7%。根据GSMA，预计2030年移动终端约60亿台。由未来日活用户数量=使用AI模型的移动终端占比*移动终端总数，对标移动互联网渗透率，若全球有30%的移动终端接入AI，则未来可达18亿的AI日活用户；若有20%接入，则有12亿AI日活用户。
- **未来每日活用户年算力成本方面：未来日活用户中，部分用户会选择更高精度模型。未来每日活用户算力成本=高精度模型使用占比*高精度模型算力成本+现有模型使用占比*现有模型算力成本，其中现有模型使用占比=1-高精度模型使用占比。**
- **OpenAI通过融资计划资金重置产业链，可以降低算力提供成本，支撑起日活用户数的增长以及高精度模型应用的空间。**

图表：全球智能手机苹果市场份额（数据截至23Q4）

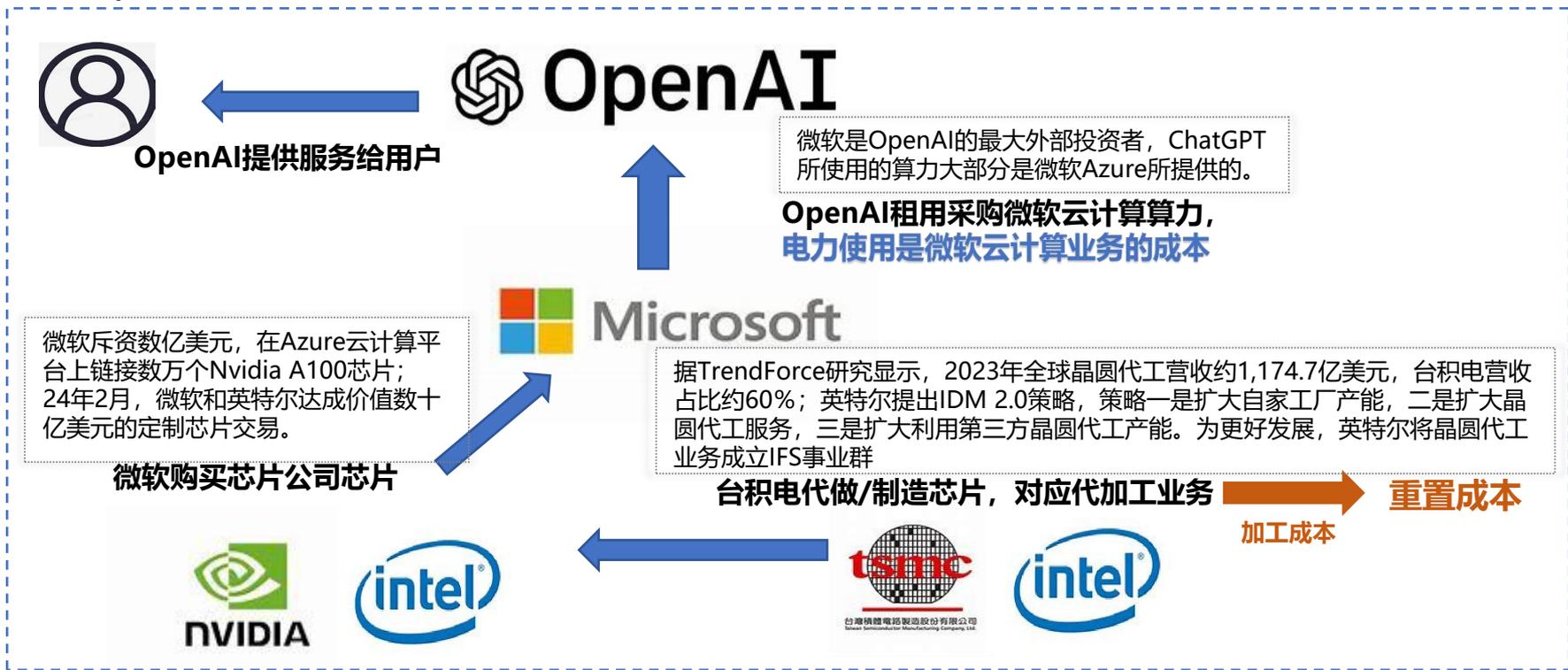


图表：不同高精度模型占比下的每日活用户每年算力成本



2 OpenAI实际重置路线图：渗透到上游

图表：OpenAI重置成本图解



2.1 单次推理成本：Sora算力是GPT4的几倍？

图表：Sora和GPT-4的单次提问算力影响因素对比

	生成内容	推理参数量
GPT-4	一次对话：0.1万tokens	3*175B
Sora	10秒视频：16.7万tokens	3B

- **从单次Query算力成本的影响因素来看，推理算力与一次生成内容的tokens、模型参数量和迭代次数成正比。**根据扩散Transformer论文一作作者杨赛宁推算，Sora参数量约30亿。视频可以看成是一帧帧图片组成，图片模型（如Stable Diffusion）的迭代次数约为25-50次，取中枢估计视频模型迭代35次。根据昆仑万维 AI Infra 负责人，在视频领域，一分钟视频约为1Mtokens。考虑到Sora的视频长度约为10秒，即对应16.7万tokens。
- **假设GPT4和Sora的推理参数量为3*1750亿和30亿，在Sora一次推理迭代35次的情况下，其推理成本是GPT-4模型的33倍。**

图表：不同迭代次数下，Sora和GPT-4的单次算力成本比

Sora迭代次数	Sora/GPT-4
20	19.05
25	23.81
30	28.57
35	33.33
40	38.10
45	42.86
50	47.62

2.2 为什么成本重置才能匹配“高性能模型”：算力成本VS用户付费

- 根据财务数据推算，预期微软AI云计算规模23年Q4为8.39亿美元，该部分规模主要是推理需求，不包括训练需求，由于目前主要是OpenAI在进行推理，因而预期其中大部分主要是OpenAI的推理算力。
- 在ChatGPT精度经过两次下降后（每次降低7%），目前GPT-4的单用户算力成本为每日活每年100.2美元。假设OpenAI日活月活比例30%背景下，由于付费是月付费，因而GPT4会员的市场定价为20美元/月，除以0.3，乘以12个月，约为800美元 > 100.2美元。
- 如果Sora的单日活算力成本为33倍的GPT-4，即为每日活年算力成本3308美元。如果按市场价格算，Sora的月付费至少大于82美元/月，才能覆盖Sora会员单日活的成本3308美元，对于用户来说费用偏高。因此只有通过成本重置降低成本进而降低付费价格才能显著走通商业模式。

图表：微软云计算业务和AI云计算规模（季度）

微软	<u>23Q4</u> 23 Jun	<u>24Q1</u> 23 Sep	<u>24Q2</u> 23 Dec
AI云计算规模 (百万美元)	129	388	839
AI云计算占比	1%	3%	6%
毛利率	72%	72%	73%

图表：不同精度模型的每日活用户年算力成本

模型	每日活年算力成本 (美元)
GPT-4	260.5
GPT-4两次精度下降后 (每次精度降低7%)	100.2
Sora模型	3308.1

3 OpenAI实际重置成本分析：晶圆厂和电力是核心

- 根据华尔街日报，OpenAI的大额债务融资主要用于两个方面的建设：芯片（减少芯片产业链环节）、电力（AI设施需要消耗大量的电力）
- 芯片方面：如果OpenAI通过与芯片厂合资建厂生产，可以节省中间产业链环节的成本。OpenAI的业务运行需要付费租用微软的云服务器，微软的成本为建设云端采购英伟达的GPU等芯片设备的采购费用，而英伟达支付芯片代做费用，将制造工作交给台积电完成。
- 电力方面：如果OpenAI通过与电厂直接合作，可以直接使用成本电价。根据美国上市能源公司的能源成本占比，计算中心的芯片若直接使用成本电价，则可以将电力花费降至原来价格的48%。

图表：OpenAI芯片产业链各环节的毛利率、管理费用率和研发费用率

产业链环节	业务毛利率	管理费用率	研发费用率
微软云计算	71.98%	7.57%	12.75%
英伟达计算中心	71.49%	4.84%	24.06%
台积电HPC业务	50.50%	2.41%	7.92%

图表：能源转向更低成本

产业链环节	业务毛利率	管理费用率
爱克斯龙(EXC)	64.68%	8.48%
新纪元能源(NEE)	75.99%	10.99%
Nextera Energy(NYSE:NEP)	51.50%	19.16%

3.1.1 OpenAI产业链拆解：从OpenAI到晶圆厂

- **假设1：** OpenAI成功实现合资建立芯片后，减少了云计算公司，芯片设计公司、芯片代做/制造公司的利润。**芯片的6年重置成本=现有算力成本（芯片部分）*（1-微软云计算盈利占比）*（1-英伟达计算中心盈利占比）*（1-台积电HPC业务盈利占比）*6年。**
- **假设2：** 算力成本中，电力和芯片的成本约各占一半。微软的云计算服务是电力消耗的环节，OpenAI重置产业链之后，**将电能转化为使用低成本能源**，则可减少服务器的运行成本。
- **假设3：** 根据前述测算结果，按每年折旧1万亿美元作为总算力成本。
- **结论：** 对于该产业链完全重置之后，6年的总重置成本为1.21万亿美元，将会带来成本的显著减少。

图表：从产业链计算OpenAI的6年重置总成本

OpenAI 算力成本 (亿美元/年)	微软云计算盈利 占比	英伟达计算中心 盈利占比	台积电HPC盈利 占比	6年重置成本-芯 片 (亿美元)	6年重置成本-电 力 (亿美元)	6年总重置成本 (亿美元)
10000.00	51.7%	42.6%	40.2%	4980.97	7079.96	12060.93

3.1.2 重置成本测算：与晶圆厂的不同合作方式

- 对于晶圆厂，OpenAI可以选择与其合作：1、完全重置晶圆厂环节；2、重置晶圆厂50%；3、不重置晶圆厂环节。在不同的情况下，6年的总重置成本分别是1.21万亿美元，1.37万亿美元，1.54万亿美元。
- **晶圆厂的格局：OpenAI积极与英特尔、台积电、三星等技术领导者展开合作。**据英国《金融时报》，Altman 正在寻求开发自己的 Nvidia 人工智能和 HPC GPU 替代品，并与台积电和中东投资者进行谈判，以启动一家新的芯片合资企业；而据《韩国中央日报》援引业内人士消息，Sam Altman 访韩与韩国芯片巨头三星、SK探讨合作，主要讨论了高带宽内存（HBM）芯片。另外，Intel在IFS Direct Connect活动上将IFS更名为Intel Foundry，Sam Altman出席活动，未来双方或积极合作。

图表：从与晶圆厂不同的合作方式看OpenAI的6年重置成本

晶圆厂的重置比例	微软云计算	英伟达计算中心	6年重置成本-芯片 (亿美元)	6年重置成本-电力 (亿美元)	6年总重置成本 (亿美元)
100%	52%	43%	4981.0	7080.0	12060.9
50%	52%	43%	6652.9		13732.8
0%	52%	43%	8324.8		15404.8

3.2 OpenAI的能源产业链：核能新技术是重要手段之一

- **人工智能对电力的消耗巨大，互联网巨头紧密关注未来AI所带来的能源方面问题，且正在持续推动将核能为其数据中心提供能源。**在2024年1月17日的达沃斯世界经济论坛上，OpenAI的首席执行官Sam Altman指出，未来的AI发展离不开能源方面的突破，解决这一问题的方向是发展环境友好型能源，尤其是**核聚变或更便宜的太阳能和储能**。同时，据《华尔街日报》报道，微软正在计划利用下一代核反应堆，即所谓的小型模块化反应堆（SMR），来支持其数据中心和AI项目。
- **美国核监管委员会2023年1月刚刚首次认证了SMR设计，允许公用事业公司在申请新电厂许可证时选择该设计。**

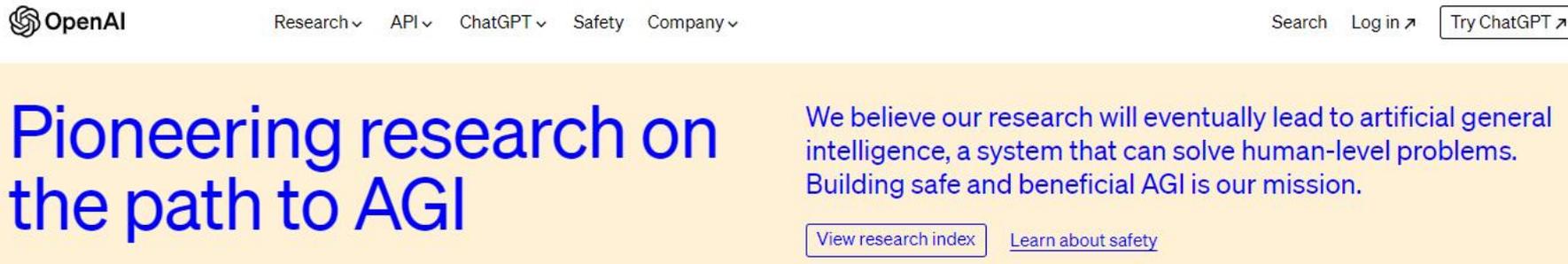
图表：OpenAI和微软的能源布局

公司	动作
OpenAI	2015年，奥特曼向核聚变初创公司Helion投资了950万美元
	2021年，OpenAI首席执行官Sam Altman宣布向Helion Energy投资3.75亿美元，成为该公司的主要投资者和董事会主席。
	Oklo 是 OpenAI 首席执行官 Sam Altman 支持的核初创公司之一，他将人工智能和廉价的绿色能源描述为实现“丰富”未来的相辅相成的要素。
微软	2023年5月，微软同意从2028年开始将在Helion购买电力。
	2023年6月，在美国拥有最多核电站的联合能源公司与微软达成协议，将在风电和光伏供应不足的情况下，向微软弗吉尼亚州数据中心提供“精确到秒”的电力供应。
	2023年9月，微软发招聘通知，招聘“核技术首席项目经理”，将负责成熟和实施全球小型模块化反应堆以及微型反应堆能源战略。
	2023年12月，微软聘任了两位核能相关经理，负责利用核能为公司的全球数据中心供电。两位人员在各自岗位上都拥有丰富的核能行业经验。
	2023年，微软内部部署团队，专门使用美国核能监管方面的文件来训练AI，目标是训练出一个能够快速减少核电站建设审批时间和成本的专用大模型
	微软在第28届联合国气候变化大会（COP28）上发布了关于先进核能和核聚变能源的政策简报。

- OpenAI的“融资规划和实际重置成本”
- OpenAI的“AI新品规划&算力需求”
- OpenAI与竞对资源竞争：融资&算力&能源
- 投资建议&风险提示

- OpenAI阐述使命是确保通用人工智能（通常比人类更聪明的人工智能系统）造福全人类。OpenAI的AGI使命主要包含两方面：1、一个可以解决人类问题的系统（A system that can solve human-level problems）；2、构建安全并且有价值的AGI（Building safe and beneficial AGI）
- 通用人工智能 AGI (artificial general intelligence)：通用人工智能AGI亦被称为“强AI”，指的是在任何可以想象的人类专业领域内，具备相当于人类智慧程度的AI。一个AGI可以执行任何人类可以完成的智力任务，甚至可以说AGI是一种在大多数具有经济价值的工作上超越人类的系统。

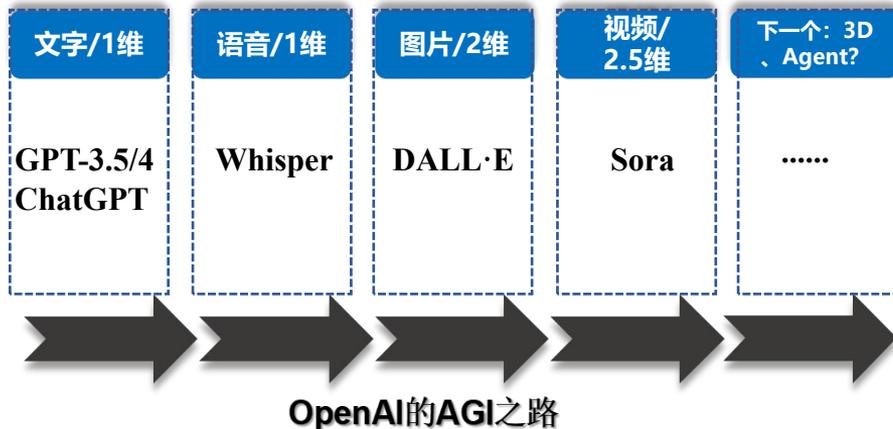
图表：OpenAI将AGI目标写在官网



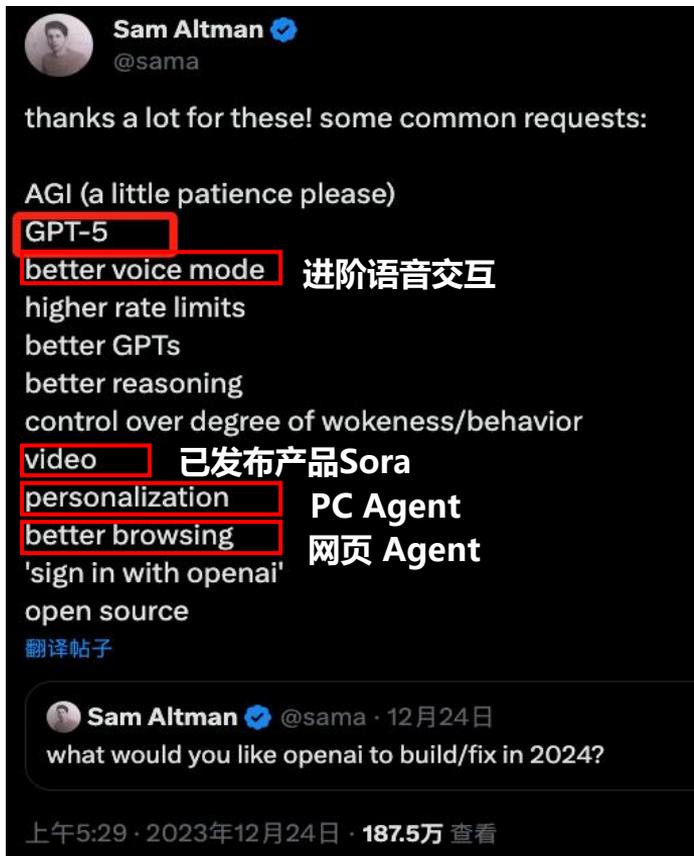
2 AI预测产品线：文生3D、Agent、语音和GPT-5

- 在奥特曼23年12月底的发言中，可以看到未来的产品线写有GPT-5、更好的语音模型（交互方式）、更好的网页浏览（网页Agent）、个性化（PC Agent）。
- 目前Video模型（即Sora）已发布。
- 同时OpenAI连续两次发表文生3D论文，包括Shape-E、Potin-E。
- 目前OpenAI的产品线沿AGI之路发展，已有包含文字、语音、图片、视频的模式发布。未来预计在3D模型、Agent，语音交互方面有更多进展。

图表：OpenAI向AGI的发展现状



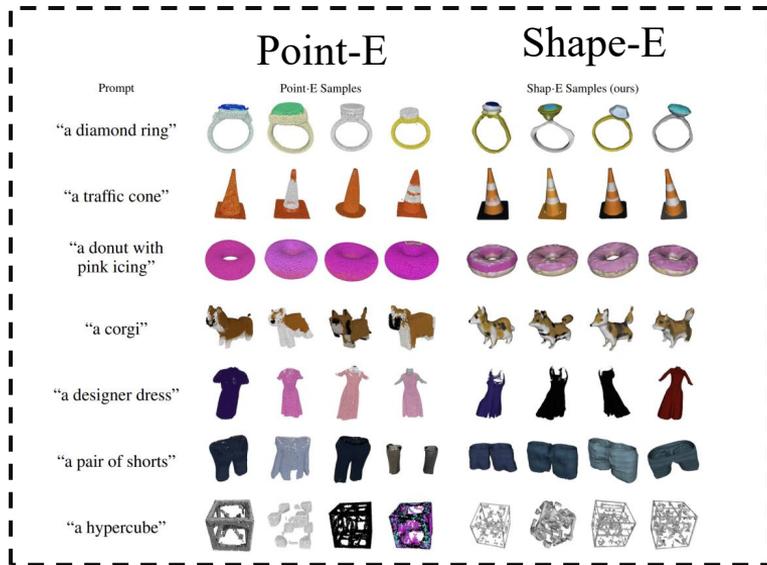
图表：奥特曼爆料OpenAI产品线



2.1.1、文生3D：已发布Point-E和Shape-E，“3D数据量”或带来涌现

- **OpenAI在文生3D领域已发布Point-E、Shape-E模型，预计未来将有更高质量模型出现。** 2022年12月OpenAI发布Point-E，可以产生3D点云，或代表三维形状的离散数据点集合。**2023年5月，OpenAI发布Shape-E模型**，相比基于点云的显式生成模型Point-E，收敛速度更快，在更高维的表示输出空间中实现了更好的样本质量。
- **预计千万级3D数据集逐渐变多，亿级数据集也将使3D生成领域迎来爆发。** 类比于2D生成发展路径，3D生成出现了具有统治地位的NeRF模型，2023年7月发布的Objaverse-XL数据集使得3D资产数据集破千万。

图表：OpenAI的Shape-E表现比Point-E更好



图表：Objaverse-XL破千万，比上一版本大一个量级

Animation Count



Armature Count



2.1.2、文生3D算力需求：约50-60倍GPT-4的算力需求

- ▶ 采用3D模型Zero123，单场景3D资产生成需要迭代30000次左右，在RTX 3090显卡（24GB）上运行约3.3-4.2小时，模型一次生成的算力约为GPT-4一次推理的49.5-63倍。总优化时间（小时）=迭代次数/每秒迭代次数/60seconds/60minutes。根据我们实机测试的结果，在Zero123算法下，设定迭代30000次时生成的3D成品相对可行。利用3090显卡（24GB）且显存使用率约50%（即12GB左右）时运行文生3D模型，对处理过的单场景图片，根据测算，每秒迭代次数约2.35次。对于需要迭代30000次的单场景来说，生成一个3D模型需要3.3-4.2小时。
- ▶ GPT-4运行在A100（24个80GB的GPU显存），一次推理约需3秒完成，则文生3D的算力为GPT-4的49.5-63倍。
 （ $(3.3h * 60min * 60sec * 24GB) / (24个 * 80GB * 3sec) = 49.5$ ）。
- ▶ 以2D为例，迭代次数越多，生成质量越高，对象越复杂。一般情况下，2D场景迭代20-50次左右即可实现要求。

图表:不同迭代数和迭代速度下算力时间花费

单场景不同迭代次数下所需GPU (3090) 小时数				
迭代次数 \ 每秒迭代次数 (次)	25000	30000	35000	40000
1	6.9	8.3	9.7	11.1
1.5	4.6	5.6	6.5	7.4
2	3.5	4.2	4.9	5.6
2.5	2.8	3.3	3.9	4.4
3	2.3	2.8	3.2	3.7
3.5	2	2.4	2.8	3.2
4	1.7	2.1	2.4	2.8

2.2、Agent：预计两款新品+Android主屏幕小组件

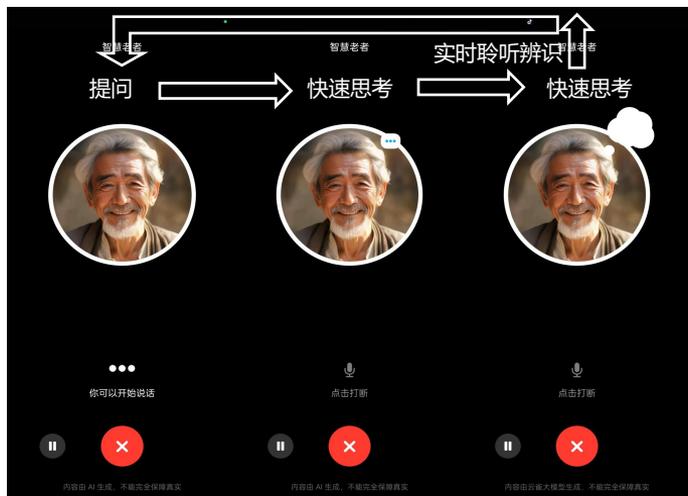
- **AI Agent指能够感知、理解环境，做出决策，采取行动，并不断学习和进化的AI系统。**决策基于它对环境的理解，比如它的目的地在哪，它的位置在哪，周围有哪些障碍等。另外，AI Agent可以“学习”：通过与环境交互，不断改进自己的决策过程。
- **“更好地网页浏览能力”对应网页端Agent：据The Information今日报道，OpenAI正在开发两款Agent软件。**
 - 1) **一款Agent通过有效接管客户的设备来自动执行复杂任务，**客户可要求Agent将数据从文档传输到电子表格进行分析，或者自动填写费用报告并输入到会计软件；
 - 2) **另一款Agent将处理基于网络的任务，**例如收集有关公司的一组公共数据、在一定预算下创建行程或预订机票。目前还不清楚OpenAI计划何时发布其Agent产品。
- **“个性化”规划对应个人 Agent：据The Verge援引知名Android专家迈克尔·拉赫曼 (Mishaal Rahman) 报道，ChatGPT为安卓平台推出了全新主屏幕小组件功能，**用户点击并按住安卓手机主屏幕上的空白位置就可以调出小组件菜单。目前这一功能仍处于测试阶段，尚未在Google Play Store的最新版本中提供，具体的推出时间并不确定。

未来OpenAI的语音技术方式有待改进：语音延迟、复杂环境主体辨认等。预计语音生成的技术进步也将助力Video视频技术。24年3月10日，Pika 最新推出的功能Sound Effects，可以给视频无缝生成音效，可支持提示词生成音效或自动配音。字节推出的AI语音聊天软件可以使用户和AI达到实时的聊天。

端到端Transformer在模拟声波时需要解决的问题：

- 识别物体：每个物体具体信息、物体间的高阶互动
- 识别声音环境：识别环境，特别是复杂环境下的不同声音
- 检索物体和环境：检索典型声音模式
- 创造声音：通过学习得到的物理规则来组合和调整声音模式的参数

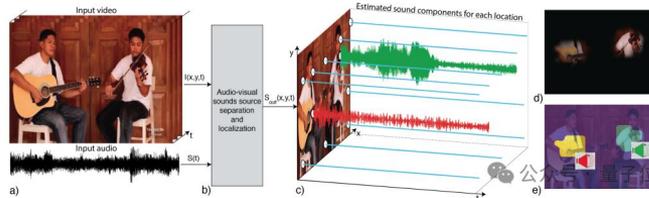
图表：豆包--实时语音对话交互模型



图表：端到端Transformer模拟声波



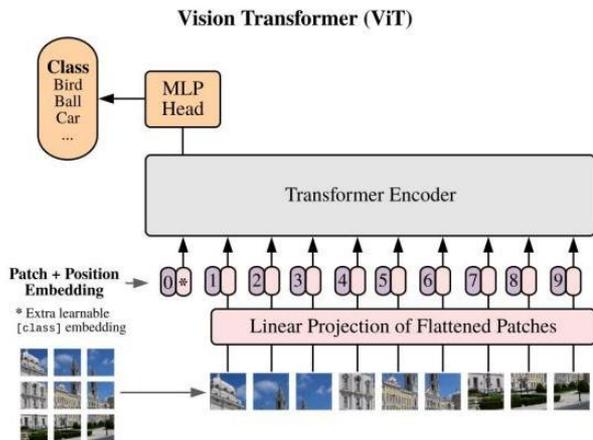
We introduce PixelPlayer, a system that, by watching large amounts of unlabeled videos, learns to locate image regions which produce sounds and separate the input sounds into a set of components that represents the sound from each pixel. Our approach capitalizes on the natural synchronization of the visual and audio modalities to learn models that jointly parse sounds and images, without requiring additional manual supervision.



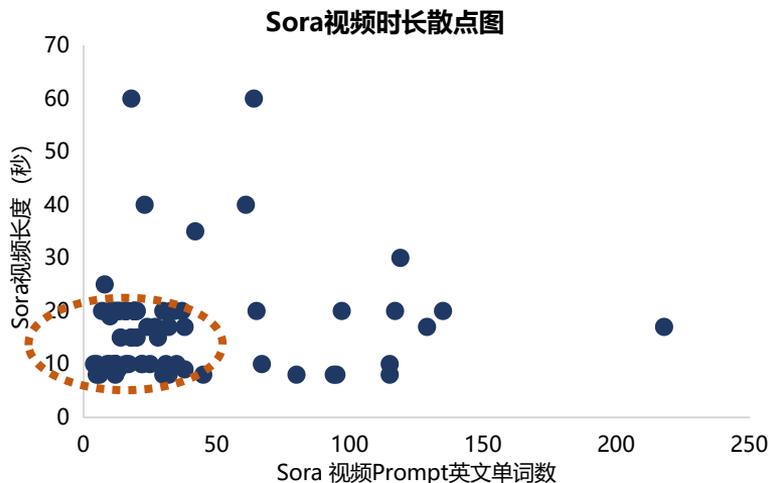
2.4、文生视频Sora：推理算力和使用效果

- **Sora生成视频的长度大多在10-20秒区间。** Sora模型继承了DALL·E3的画质和遵循指令能力，能够从文本说明中生成长达60秒的视频，10s、20s占比较多。
- **Sora是扩散模型，更是扩散模型+Transformer的视觉大模型，推理Sora参数量约30亿。** 目前，Sora能够提供具有多个角色、特定类型的动作和详细背景细节的场景。Meta首席AI科学家杨立昆称，纽约大学助理教授谢赛宁作为一作的扩散Transformer论文是Sora的基础。谢赛宁也积极发长文分析Sora基于DiT架构、可能用到谷歌NaViT技术，推算Sora参数量约30亿。

图表：Sora的原理：U-net→Transformer



图表：Sora 生成视频的特点

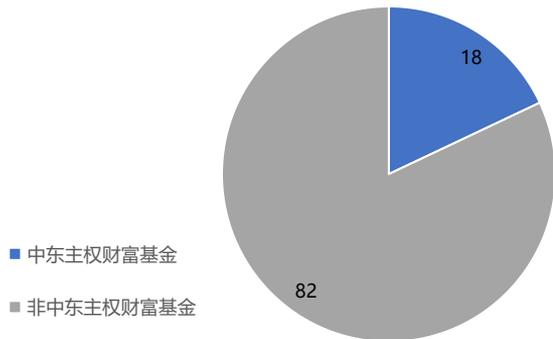


- OpenAI的“融资规划和实际重置成本”
- OpenAI的“AI新品规划&算力需求”
- OpenAI与竞对资源竞争：融资&算力&能源
- 投资建议&风险提示

1、OpenAI的一级市场融资：核心是主权基金对AI的配置

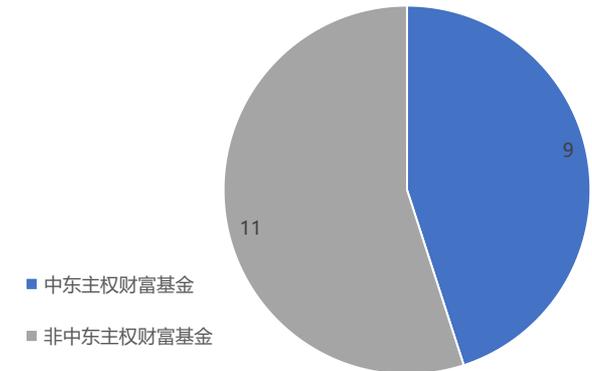
图表：主权财富基金全球前100名中东投资者占比

主权财富基金全球前100名占比



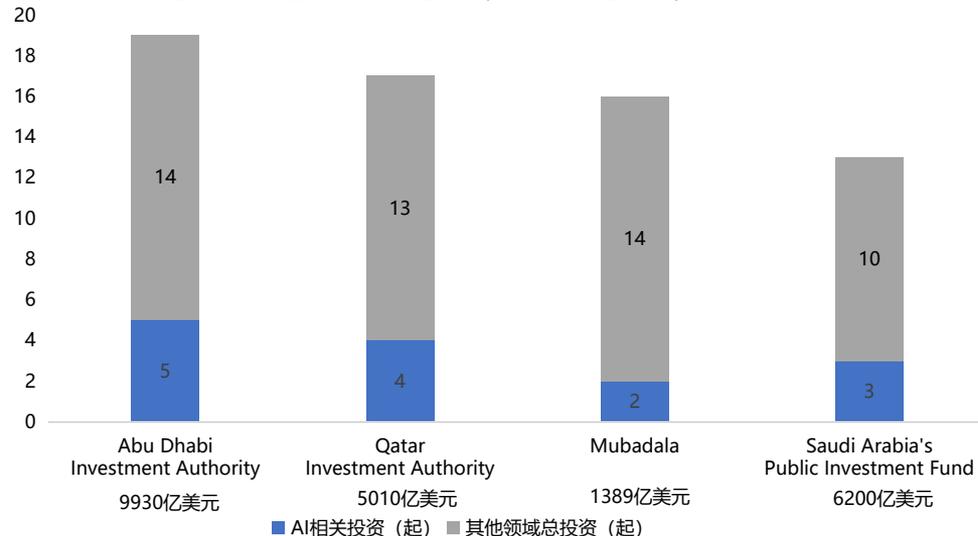
图表：主权财富基金全球前20名中东投资者占比

主权财富基金全球前20名占比



- 根据SWFI统计，在全球资产前二十名的主权财富基金中，来自中东的主权财富基金有9个。全球排名前100的主权财富基金，中东有18家，总计资产额达到4.43万亿美元。
- AI领域初创公司获中东投资者青睐。总资产最多的中东主权财富基金Abu Dhabi Investment Authority（总资产9930亿美元），自23年2月份进行了19起投资，对AI领域的投资就有5起。其他中东大型资产基金也多次投资AI初创公司。

图表：大型中东投资者的投资类型分布（自23年2月至今）



1.1 一级市场的主要竞争对手：Anthropic

- **OpenAI竞对Anthropic获得来自中东投资共13.5亿美元的投资。**目前Anthropic的Claude3是OpenAI的重要竞争对手，其中Anthropic和Inflection AI分别已累计获得76亿美元和15亿美元的资金投入。其中，Anthropic更是得到了中东投资者ICS FO、Seven Rivers Capital LTD、XTX Ventures的投资，三者于2023年5月，均跟投Anthropic 4.5亿美元。

图表：OpenAI的竞对公司获融资情况，及投资Anthropic的中东投资者

初创公司	总融资金额 (美元)	主要投资者 (按投资金额排序)
Anthropic	7.6B	Index Ventures 、James McClave、Jaana Tallinn、Center for Emerging Risk Research等， 中东投资者 ICS FO、Seven Rivers Capital LTD、XTX Ventures
Inflection AI	1.5B	NVIDIA 、Bill Gates、Reid Hoffman、Microsoft等
Cohere	434.9M	Founders Circle Capital 、Raquel Urtasun、Radical Ventures、Geoffrey Hinton、Fei-Fei Li等
Adept AI	415M	Addition 、Greylock、Joe Morrissey、Chris Re、Spark Capital等
Character AI	150M	SVA 、A.Capital Ventures、Nat Friedman、Elad Gil等
Jasper	131M	NVIDIA 、Bill Gates、HubSpot Ventures、Bossanova Investimentos、Founders Circle Capital等
Pika	55M	Homebrew 、Alex Chung、Clément Delangue、Nat Friedman等

- **XTX Ventures**是全球前列的外汇/贵金属外汇交易所 XTX Markets的投资分部，专注于早期创业公司（种子到A轮）应用机器学习，还投资了智能教育公司Graide、健康科技公司BlueSkeye AI等等，在2023年5月23日投资4.5亿美元给Anthropic。
- **ICS FO**专注于成长型投资，关注信息技术、互联网电子商务、软件等，在2023年5月23日投资4.5亿美元给Anthropic。
- **Seven Rivers Capital LTD**的总部在迪拜，在2023年5月23日投资4.5亿美元给Anthropic。

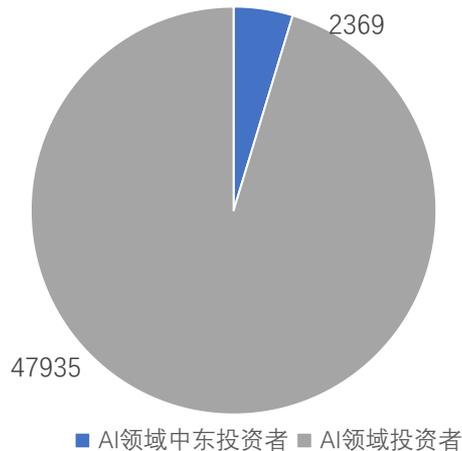
1.2、一级市场的主权投资基金

图表：全球排名前100主权财富基金：中东投资者

排名	全球排名前100主权财富基金 (中东)	资产额 (亿美元)
1	Abu Dhabi Investment Authority	9930.0
2	Kuwait Investment Authority	8030.0
3	Public Investment Fund	7766.6
4	Qatar Investment Authority	5010.0
5	Investment Corporation of Dubai	3410.7
6	Turkey Wealth Fund	2792.6
7	Abu Dhabi Developmental Holding Company	1900.0
8	National Development Fund of Iran	1565.0
9	Mubadala Investment Company	1389.8
10	Emirates Investment Authority	870.0
11	Dubai Investment Fund	800.0
12	Oman Investment Authority	463.4
13	Mumtalakat Holding	176.4
14	Sovereign Fund of Egypt	120.0
15	Sharjah Asset Management	19.0
16	Israeli Citizens Fund	10.1
17	Palestine Investment Fund	9.6
18	Bahrain Future Generations Reserve Fund	6.8
总计	44270亿美元	

- 在全球排名前100的主权财富基金中，18家主权财富基金的总资产4.43万亿美元，占比较高，且持有资金较大。而在全球，关注AI领域的投资者/机构有大约50304个，其中有2369个来自中东，约占4.7%。

图表：AI领域的中东投资者占比



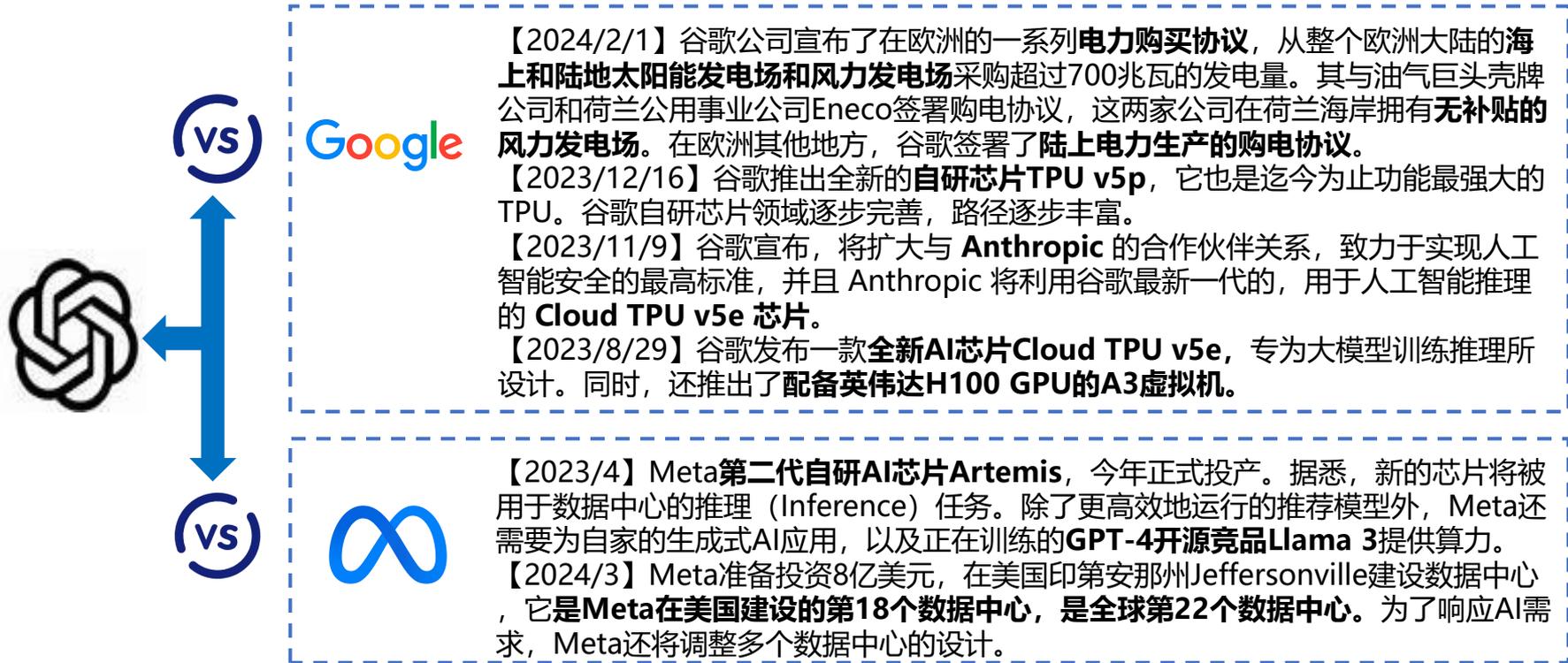
2、OpenAI的产业资源竞争：相近时间发布“同类AI产品”

图表：谷歌、Meta紧密发布产品，与OpenAI竞争产业资源



3、OpenAI&谷歌&Meta的产业资源布局

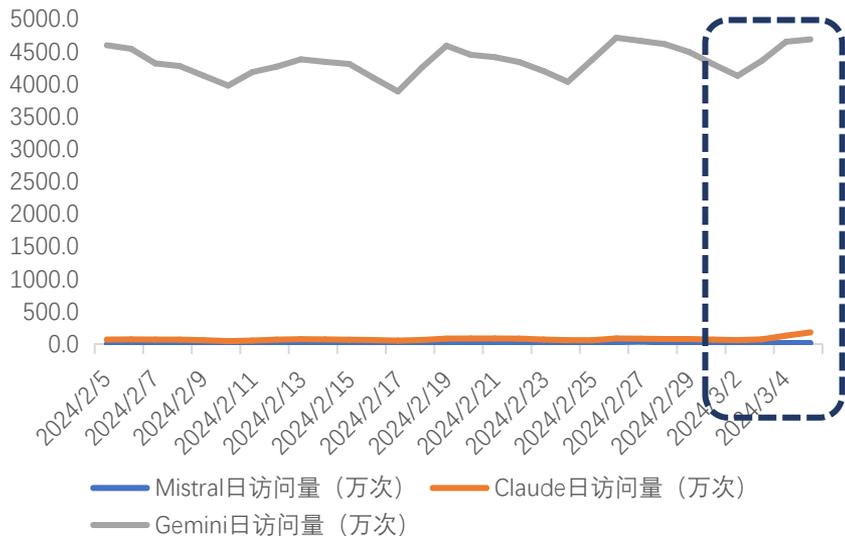
图表：谷歌、Meta在能源、芯片领域积极布局



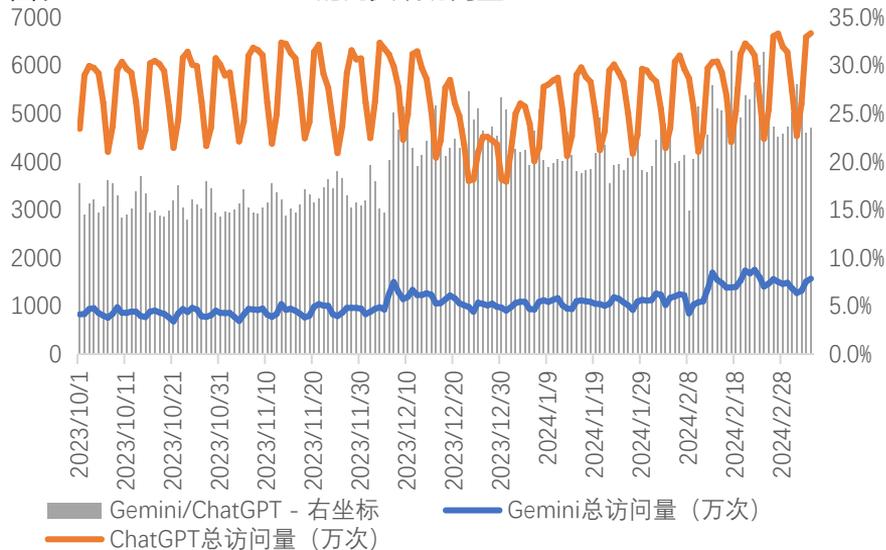
4、AI推理竞争：OpenAI 新品窗口期或为7月，GPT4冲刺潮已来

- **达GPT-4 水平的大模型从0突破到5，OpenAI的AI代差优势在7月LLama3开源后将减弱，因而新产品在7月前发布有助于融资。**2月27日，法国Mistral AI（曾开源8×7B MOE）发布对标GPT-4的Mistral Large；为了巩固开源优势，Meta宣布于7月初开源LLaMA3（对标GPT-4），同时Claude3发布，并且性能接近GPT4。
- **谷歌类GPT-4模型Gemini Ultra推动用户超OpenAI 20%：**根据Similarweb，在网页端，Gemini日活用户达1508万，接近OpenAI网页端的20%。其中一部分原因是，目前绑银行卡成为Gemini Ultra会员后，前2个月免费，第3个月开始收费，价格与GPT-4会员接近。

图表：Mistral&Claude&Gemini的网页端访问量



图表：ChatGPT&Gemini的网页端访问量



- OpenAI的“融资规划和实际重置成本”
- OpenAI的“AI新品规划&算力需求”
- OpenAI与竞对资源竞争：融资&算力&能源
- 投资建议&风险提示

➤ **建议关注：一、短期视角：从OpenAI新品节奏视角的映射逻辑：**

1、Sora落地：1) IP产业链：荣信文化、掌阅科技、中文在线、中信出版、奥飞娱乐等；2)、主旋律，人民网；3) AI电影产业链：上海电影、光线传媒等；4)、文生视频，超讯通信和因赛集团。

2、文生3D:3D素材，丝路视觉、凡拓数创；游戏公司，紫天科技、宝通科技、恺英网络、巨人网络、三七互娱、吉比特、完美世界、盛天网络、游族网络、星辉娱乐、姚记科技、神州泰岳。

3、Agent：昆仑万维。

4、港股内容平台：阅文集团、哔哩哔哩、快手。

➤ **建议关注：二、中期视角：**从OpenAI与竞对的竞争来看，行业进入GPT4冲刺期，**AI推理将出现拐点式增长**，同时AI新平台或诞生。

1、AI推理算力：海外算力，英伟达和AMD，及与之相关的光模块如中际旭创等；国内算力，华为昇腾算力产业链、寒武纪等。

2、AI新平台：字节AI工厂模式暂时领跑，字节产业链如省广集团；同时关注腾讯、阿里和百度的最新进展。

➤ **建议关注：三、长期视角：**从OpenAI的未来发展路径来看，核心是通过“重置产业链”降低“高性能&高算力”的AI新模型的单用户算力成本，促使消费者能在一定付费下使用最新AI模型。**因此在此背景下长期视角的产业链环节主要是晶圆厂和电厂**，晶圆厂包括台积电、中芯国际、英特尔。

➤ **风险提示1：** 海外AI融资进度不及预期

- 融资环境变化
- 一级市场融资竞争激烈
- 融资谈判结果不及预期

➤ **风险提示2：** 海外AI产业重置进度不及预期

- 海外AI产业重置时间不及预期
- AI产业重置落地效果不及预期

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

一般声明

华福证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，该等公开资料的准确性及完整性由其发布者负责，本公司及其研究人员对该等信息不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，之后可能会随情况的变化而调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。投资者应自行决策，自担投资风险。

本报告版权归“华福证券有限责任公司”所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

特别声明

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。投资者请勿将本报告视为投资或其他决定的唯一参考依据。

投资评级声明

类别	评级	评级说明
公司评级	买入	未来6个月内，个股相对市场基准指数涨幅在20%以上
	持有	未来6个月内，个股相对市场基准指数涨幅介于10%与20%之间
	中性	未来6个月内，个股相对市场基准指数涨幅介于-10%与10%之间
	回避	未来6个月内，个股相对市场基准指数涨幅介于-20%与-10%之间
	卖出	未来6个月内，个股相对市场基准指数涨幅在-20%以下
行业评级	强于大市	未来6个月内，行业整体回报高于市场基准指数5%以上
	跟随大市	未来6个月内，行业整体回报介于市场基准指数-5%与5%之间
	弱于大市	未来6个月内，行业整体回报低于市场基准指数-5%以下

备注：评级标准为报告发布日后的6~12个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中，A股市场以沪深300指数为基准；香港市场以恒生指数为基准；美股市场以标普500指数或纳斯达克综合指数为基准（另有说明的除外）。

诚信专业 发现价值

联系方式

华福证券研究所 上海

公司地址：上海市浦东新区浦明路1436号陆家嘴滨江中心MT座20楼

邮编：200120

邮箱：hfyjs@hfzq.com.cn

