GPT 会是 AI 算法模型的终极答案吗？

前瞻研究全球人工智能 AI 行业系列报告 5 | 2023.4.7

中信证券研究部



陈俊云
前瞻研究首席分析师
S1010517080001



许英博
科技产业首席分析师
S1010510120041



贾凯方
前瞻研究分析师
S1010522080001



刘锐
前瞻研究分析师
S1010522110001

核心观点

ChatGPT 在全球 AI 产业、技术领域带来的良好示范效应，有望推动 AI 算法模型结束当前的技术路线分叉，并不断向以 GPT 为主导的大语言模型（LLM）靠拢，加速全球 AI 产业“工业化”时代到来。中期维度，大语言模型领域“暴力美学”预计仍将是我们不断逼近通用人工智能的最可能路线，基于此基准假设，模型算法架构将遵从渐进的学术研究步伐，高质量数据集、工程实践能力、核心人才、资本将成为大模型研发领域的核心竞争壁垒。而面对当前 GPT 等大语言模型在问答可靠性、推理能力等层面缺陷，复杂问题分解、链接外部知识&工具等增强语言模型（ALM）方式料将是可行的方式，比如 ChatGPT 插件等。作为中期最具确定性的产业方向之一，我们持续看好全球 AI 领域的投资机会，并建议持续聚焦芯片、算力设施、模型架构&工程实践、应用场景等核心环节。在美股市场，我们建议持续关注：英伟达、台积电、微软、谷歌、AMD、Arista、博通、Marvell、百度、Adobe、Snowflake 等。

- **报告缘起：**ChatGPT 在全球市场的成功，以及由此引致的全球主要科技企业在 AI 领域的军备竞赛，正在加速全球 AI 产业的发展速度。AI 算法模型作为产业的核心环节，其本身的技术路线，将直接决定 AI 产业链最终形态，以及各产业环节的分工协作方式、价值分配结构等，对 AI 算法模型技术路线的系统分析、前瞻性判断，是构建 AI 产业中期投资逻辑的核心基础。在本篇专题中，我们主要回答：AI 算法模型是否会朝 GPT 方向进行收敛，算法模型核心竞争壁垒、算力是否会成为 AI 产业潜在约束因素，以及如何引入新的知识&工具，以克服当前大语言模型的潜在缺陷等。

- **技术路线：大语言模型（LLM）有望实现主导，并不断向 GPT 方案靠拢。**在过去近 10 年里，人工智能模型经历了从小模型到以 Transformer 为代表的大模型的转变，我们根据行业发展的历史总结为以下四个阶段。1）2017 年前，以深度学习为核心的小模型占据主流：这类模型以 LSTM 及 CNN 模型作为典型的特征抽取器，根据特定领域标注数据训练，在一些任务上可以达到接近人类的水准。但是硬件算力不够导致针对其他领域重新训练成本过高，让其通用型任务的完成情况很差。 2）2017 年到 2020 年，以谷歌 Bert 为代表的双向预训练+Fine Tuning（微调）的 Transformer 模型横空出世成为了人工智能的主流方向。谷歌研究团队发布的《Attention is all you need》论文开创了 Transformer 模型，重新统一了自然语言模型(NLP)的研究范式。这类模型以预训练的方式学习语言学特征，大幅简化了过去 NLP 繁琐的研究种类。Transformer 作为特征提取器效果好于 CNN、LSTM 模型等，让 AI 首次能在语言任务的部分场景中追平人类。3）2020 年到 2022 年，产业界与学术界继续沿着 Transformer 的道路前景，但开始尝试不同于 Bert 的其他 Transformer 架构，通过预训练无监督学习的方法不断增加模型大小成为了这一时段的主流。以 1750 亿参数的 GPT-3 为代表，各大互联网巨头不断尝试增加模型体积以获得更好的效果。GPT-3 为代表的自回归+Prompting 的方法开始展现产品化能力的优越性，与 Fine-tuning 方法的 Bert 模型成为两条道路。4）2022 年至今，ChatGPT 的成功证明了 GPT 模型的 Prompting 道路的正确性，同时也强调了数据质量的重要性。ChatGPT 最重要的成功是在产品化上更进一步：ChatGPT 在模型精度上并没有飞跃性的突破，但从 Few Shot prompt（需要输入范例示范）转换到 Instruct（用人类语言描述想做什么）更加贴合用户的习惯。

- **竞争壁垒：数据、工程实践、人才、资本等。**回顾过去近十年人工智能模型的发展，我们发现进入人工智能模型研发的门槛越来越高。从 Bert 开始到 GPT-3 再到谷歌的 PALM，网络中的公开语言数据源已经在被尽可能地利用（论坛、新闻、

维基百科等等），而模型的进一步优化对数据质量的要求也越来越高。我们认为优质的私有数据源将在未来 3-5 年里人工智能模型的精度优化上发挥更重要的作用。爬虫难以爬取的社交媒体数据、保存在后台的用户偏好数据、代码托管数据等仍具有开发空间，这也将成为数据获取端的优势，而结合这些私有数据生成的合成数据将进一步放大这些公司数据端的优势。此外随着模型体量仍然在成倍数增加，大量的前期资本投入是阻碍新入者的重要因素。在过去五年内，我们看到人工智能模型的独角兽都在寻找互联网科技大厂作为其背后的依靠，主要因为：1）能够得到充足而短时间内不求产出的资金支持，2）依靠互联网大厂的公有云等相关资源进一步降低成本，3)获取互联网大厂长期以来所积累的优质数据源。这也是我们看到为什么在过去五年内人工智能头部厂商逐渐从开源走向闭源，利用其资源优势来打造差异化的 AI 模型，而落后者更希望通过开源的模式以求缩小与头部厂商的距离。研发团队的工程能力是决定公司在大语言模型竞争力的另一个重要因素。随着语言模型的体积不断增加，在研究方法上现今千亿量级的模型与之前几十亿量级的小模型发生了本质变化，个体工程师没有能力通过自有资源积累对大模型的学习经验。一个合格的大模型研发团队需要依靠大公司的资源支持才能积累对模型调试、优化、实际部署等各个环节足够的经验。优秀的工程人才稀缺以及培养难度让没有经验的新公司进入更为困难，如何招募到一个大模型领域经验丰富的专家并在其影响力的基础上招募来自于大公司的有一定工程经验的团队将成为新入公司的核心问题。

- **算力成本：单位成本持续保持快速下降趋势。**根据我们之前系列报告的测算，尽管人工智能模型的总研发投入持续在提高，但主要原因来自于模型体积的倍数式增长，而人工智能模型的单位成本保持下降趋势。在推理端以相同模型为例，在 2020 年根据 Stanford 讲师 Mourri 的测算，训练 GPT-3 的成本约为 460 万美元。而至 2022 年底，根据我们的重新测算训练同样的 GPT-3 模型在两年后成本仅需要 150 万美元，成本下降了 66%以上。在模型推理端以 ChatGPT 为例，根据我们在 2022 年底的测算，ChatGPT 生成一次回答的成本约在 1-2 美分之间。而从 3 月 2 日 OpenAI 对 ChatGPT API 的定价出，0.2 美分每 1000tokens 的定价对应为百亿参数级别的模型。结合 ChatGPT 在训练过程中所使用的 Reward Model（奖励模型，记录了人类偏好）应为千亿参数级别的模型，这也意味着 OpenAI 将其在千亿参数模型中所学习到的能力通过类似于蒸馏（distillation）的方式浓缩进了百亿参数的模型中，快速降低了成本消耗。根据目前用户在社交媒体的反馈结果看，这个百亿参数模型的能力已经追平甚至超越了先前的大尺寸模型。因此无论是模型推理端或训练端，我们都观察到单位成本在持续保持下降的趋势，而总成本的持续提高来源于模型大小的持续增加以及用户数量的增加。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。
- **未来展望：通过插件等扩展大语言模型（LLM）能力边界。**目前小公司虽然没有能力研发底层的大语言模型，但其可以通过微调的形式服务垂直领域的需求。我们认为 OpenAI 三月份在 ChatGPT 中推出的插件接口是为了接入小公司的微调成果的一次尝试：第三方厂商将可以自有的数据源及部分算法作为 ChatGPT 之上的插件，ChatGPT 可以调用插件中的外部 API 接口来获得信息。插件提供者使用 OpenAPI 标准编写 API，然后这个 API 会被编译成一个 prompt 向 ChatGPT 解释如何使用 API 来增强其答案。当用户选择对应插件后，如果 ChatGPT 认为应该从 API 获取信息，它将发出请求并在尝试回应之前将信息添加到上下文中。通过插件接口这一模式简化了小公司对大语言模型微调的难度，并丰富了基于大语言模型之上的环境生态。同时通过复杂问题分解、链接外部知识&工具（比如插件）等增强语言模型方式，亦能有效解决现有大语言模型在问答可靠性、信息及时性等方面的固有缺陷。
- **风险因素：**AI 核心技术发展不及预期风险；科技领域政策监管持续收紧风险；私有数据相关的政策监管风险；全球宏观经济复苏不及预期风险；宏观经济波动导

致欧美企业 IT 支出不及预期风险；AI 潜在伦理、道德、用户隐私风险；企业数据泄露、信息安全风险；行业竞争持续加剧风险等。

投资策略：ChatGPT 在全球 AI 产业、技术领域带来的良好示范效应，有望推动 AI 算法模型结束当前的技术路线分叉，并不断向以 GPT 为主导的大语言模型（LLM）靠拢。中期维度，大语言模型+“暴力美学”预计仍将是我们不断逼近通用人工智能的最可能路线。作为中期最具确定性的产业方向之一，我们持续看好全球 AI 领域的投资机会，并建议持续聚焦芯片、算力设施、模型架构&工程实践、应用场景等核心环节。在美股市场，我们建议持续关注：英伟达、台积电、微软、谷歌、AMD、Arista、博通、Marvell、百度、Adobe、Snowflake 等。

重点公司盈利预测、估值及投资评级

公司	市值（亿美元）	估值方式	估值（自然年）			
			2021A	2022E	2023E	2024E
英伟达	6640	PE	79.4	59.8	45.1	35.3
AMD	1492	PE	27.1	30.3	21.6	18.0
台积电	4678	PE	14.1	16.8	13.7	12.3
亚马逊	10360	EV/EBITDA	NA	13.9	10.6	8.1
Arista	499	PE	34.5	27.3	24.5	22.6
博通	2622	PE	15.9	14.7	14.0	13.2
Marvell	344	PE	18.9	26.8	17.6	14.2
微软	21166	P/FCF	32.9	34.1	26.9	22.0
谷歌	13379	PE	22.3	20.3	17.3	14.7
百度	505	PE	16.8	14.7	12.9	12.0
Adobe	1752	P/FCF	24.3	22.0	19.3	18.1
Salesforce	1953	P/FCF	30.9	26.5	19.9	16.4
Service Now	949	P/FCF	43.9	36.9	29.2	23.4
Snowflake	476	P/S	23.0	16.5	12.1	9.0
Confluent	64	P/S	11.0	8.4	6.6	5.1
MongoDB	149	P/S	11.6	9.9	8.2	6.3
Palo Alto	588	P/FCF	34.5	30.0	24.5	20.8

资料来源：彭博，中信证券研究部 注：股价为 2023 年 4 月 5 日收盘价；预测数据来自彭博一致预期

目录

报告缘起 6
技术路线：大语言模型（LLM）有望实现主导，并不断向 GPT 方案靠拢 6
竞争壁垒：数据、工程实践、人才、资本等 11
算力成本：单位成本持续保持快速下降趋势 15
未来展望：通过插件等扩展大语言模型（LLM）能力边界 18
风险因素 20
投资策略 20

插图目录

图 1：人工智能模型发展的关键阶段 6
图 2：在这一阶段模型精度较难以超越人类水平 7
图 3：Transformer 模型的提取能力实现了质的飞跃 7
图 4：在语言任务上 Bert 模型第一次在部分场景追平人类 8
图 5：从 2018 年到 2022 年，模型参数量增加了 5000 倍 8
图 6：Prompt 简化了使用端接口，是更好的产品思维 9
图 7：产品化能力的进步让 ChatGPT 取得了史无前例的用户增长速度 9
图 8：以 Siri 为代表的传统模式 10
图 9：ChatGPT 如何回答红黑树问题 10
图 10：Real Toxicity 数据集（InstructGPT 牺牲部分精度） 11
图 11：TruthfulQA 数据集（InstructGPT 在 QA 取得更好成绩） 11
图 12：自 GPT-3 以来发布的部分大语言模型 11
图 13：目前的大模型已经在穷尽所有能找到的公开网络数据 12
图 14：奔驰通过自身的私有数据在英伟达 Omniverse 上进一步生成适用于自动驾驶的合成数据 13
图 15：世界一流大语言模型初创公司几乎都背靠强力公有云厂商的支持 13
图 16：理想情况下足够好的优化可以降低 90%以上的成本 14
图 17：Deepmind 对基础研究员的招聘要求为博士生或博士生相应的从业经验 15
图 18：对于相同的数据集的训练成本，近五年持续保持下降趋势 15
图 19：我们从 Scaling Laws 论文中给出的公式出发进行推算 16
图 20：参考 HGXA100 在 FP16 精度下的峰值算力为 5PFLOPS 16
图 21：斯坦福大学讲师 Mourri 曾在 2020 年推算使用 Tesla V100 显卡训练 GPT-3 需要 460 万美元 16
图 22：通过优化算法结构能够显著改善推理任务的内存占用 17
图 23：未来大语言模型生态格局预测 18
图 24：OpenAI 目前给出的插件 Demo 18
图 25：ChatGPT 通过插件调用外部 API 并通过 Slack、teams 等界面反馈给用户 19
图 26：Wolfram 通过插件告诉 ChatGPT 如何使用自己的 API 接口，并在用户需要时如何调用 20

表格目录

表 1：部分互联网巨头及其对应的私有数据 12
表 2：参考英伟达 Megatron-LM 175B 的实际算力为理论峰值的 51.4% 16
表 3：ChatGPT 推理任务测算表 17
表 4：重点推荐公司盈利预测 21

报告缘起

ChatGPT 在全球市场的成功，以及由此引致的全球主要科技企业在 AI 领域的军备竞赛，正在加速全球 AI 产业的发展速度。AI 算法模型作为产业的核心环节，其本身的技术路线，将直接决定 AI 产业链具体形态，以及各产业环节的分工协作方式和价值分配结构，因此对 AI 算法模型技术路线进行系统的分析、前瞻性的判断，就成为了一个难以回避的问题。在本篇报告中，我们将着重对如下问题进行展开分析，并在此基础上，就中期 AI 算法模型技术路线、产业核心竞争要素等进行分析推演：

1) 可预见时间范围内，AI 算法模型技术路线会朝什么方向发展，是继续走向分叉还是快速收敛，GPT 为代表的大语言模型在其中会扮演什么样的角色？

2) AI 算法模型的核心竞争壁垒将来自哪些领域，数据、算力设施、人才、工程实践？

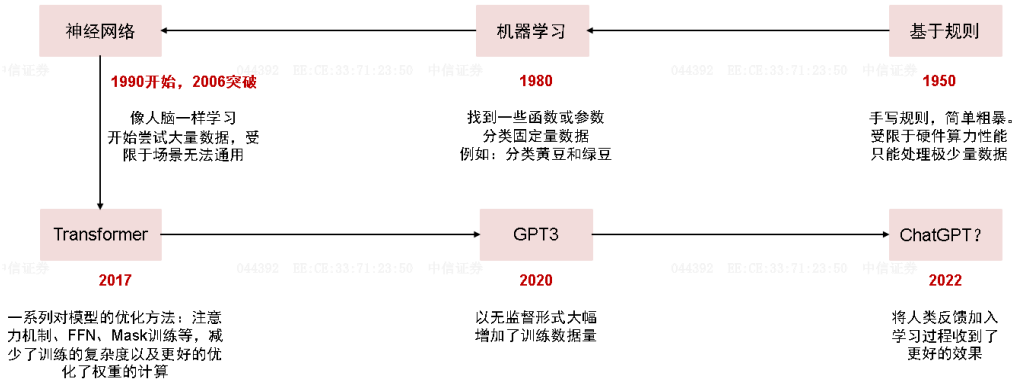
3) 算力成本是否会成为当下 AI 产业向前发展的核心瓶颈，单位算力成本下降曲线？

4) 面对当前 GPT 等大语言模型的各类突出缺陷，有何有效的解决方式，ChatGPT+插件会成为一种主流的方式吗？

技术路线：大语言模型（LLM）有望实现主导，并不断向 GPT 方案靠拢

AI 模型发展回顾：过去 10 年里，AI 模型经历了从小模型到以 Transformer 为代表的大模型的转变。从人工智能这个概念在 1950 年被提出开始，已经经历了从规则学习到机器学习再到神经网络的关键转变，本轮我们看到的人工智能技术突破都是基于在神经网络技术上的不断进步。而细分近十年来模型发展的进步，我们又观察到从以 CNN、DNN 为代表的传统神经网络小模型到以 Transformer 为代表的神经网络大模型转变的趋势。

图 1：人工智能模型发展的关键阶段

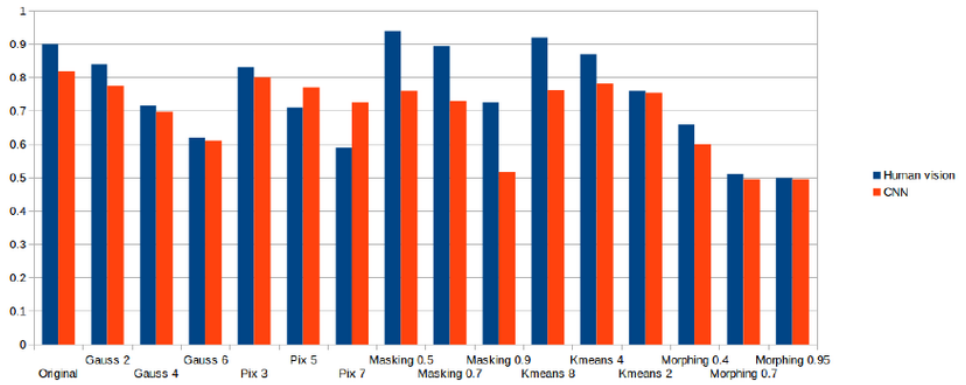


资料来源：OpenAI，中信证券研究部绘制

我们根据全球 AI 行业近 10 年的发展历史，并将其总结为以下四个阶段：

1）**2017 年前**：以深度学习为核心的小模型占据主流：这类模型以 LSTM 及 CNN 模型作为典型的特征抽取器，根据特定领域标注数据训练，在一些任务上可以达到接近人类的水准。但是硬件算力不够导致针对其他领域重新训练成本过高，让其通用型任务的完成情况较差。

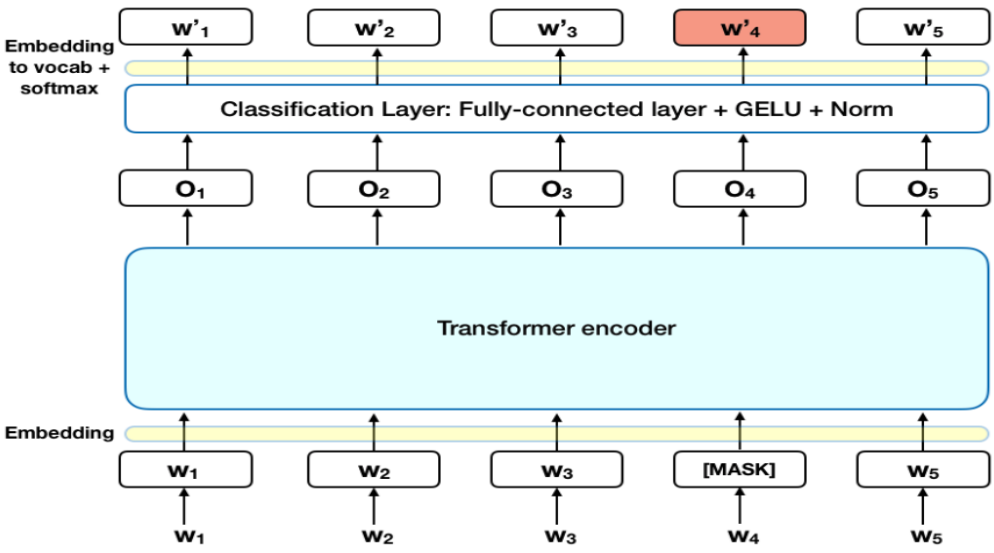
图 2：在这一阶段模型精度较难以超越人类水平



资料来源：OpenAI

2）**2017 年到 2020 年**：以谷歌 Bert 为代表的双向预训练+Fine Tuning（微调）的 Transformer 模型横空出世成为了人工智能的主流方向。谷歌研究团队发表的《Attention is all you need》论文（作者：Ashish Vaswani，Noam Shazeer，Niki Parmar 等）开创了 Transformer 模型，重新统一了自然语言模型(NLP)的研究范式。这类模型以预训练的方式学习语言学特征，大幅简化了过去 NLP 繁琐的研究种类。Transformer 作为特征提取器效果好于 CNN、LSTM 模型等，在信息容纳能力、并行计算等方面明显占优，让 AI 首次能在语言任务的部分场景中追平人类。

图 3：Transformer 模型的提取能力实现了质的飞跃



资料来源：OpenAI

图 4：在语言任务上 Bert 模型第一次在部分场景追平人类

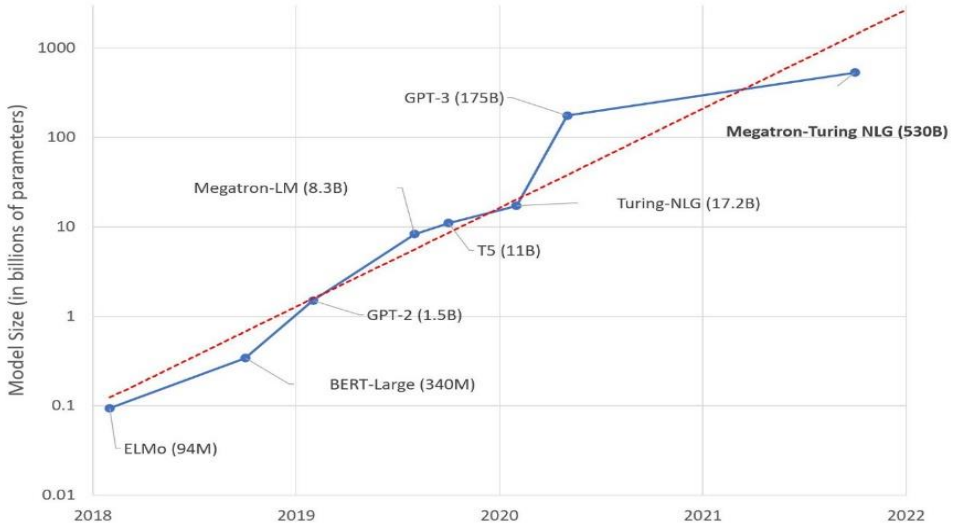
Model	Accuracy
BERT	**0.9387**
Voting Classifier	0.9007
Logistic Regression	0.8949
Linear SVC	0.8989
Multinomial NB	0.8771
Ridge Classifier	0.8990
Passive Aggresive Classifier	0.8931

curacy retrieved by the different methodologies in the IMDB experiment
dation set

资料来源：《Attention is All you need》Ashish Vaswani，Noam Shazeer，Niki Parmar 等

3）**2020 年到 2022 年**：产业界与学术界继续沿着 Transformer 的道路前景，但开始尝试不同于 Bert 的其他 Transformer 架构，通过预训练无监督学习的方法不断增加模型体积成为了这一时段的主流。以 1750 亿参数的 GPT-3 为代表，各大互联网巨头不断尝试增加模型体积以获得更好的效果。

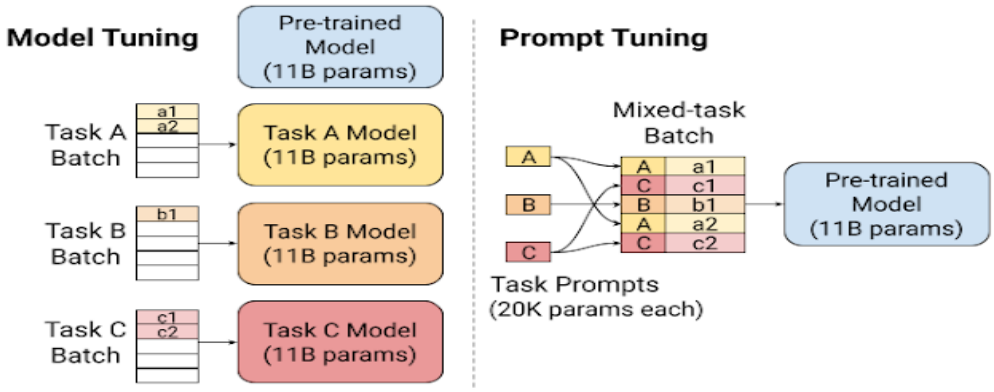
图 5：从 2018 年到 2022 年，模型参数量增加了 5000 倍



资料来源：FAIR，中信证券研究部

GPT-3 为代表的自回归+Prompting 的方法开始展现产品化能力的优越性，与 Fine-tuning 方法的 Bert 模型成为两条道路。相比于 Bert，GPT-3 的自回归+Prompting 在产品化能力展现出了两个优点：1）Fine-tuning 对于小公司更难部署，从产品化的角度上更加困难；2）Prompting 的方法更符合我们对以人类的方式使用 AI 的愿望，并实现了人、模型之间的自然交互。谷歌等巨头在这一阶段也意识到了 Prompting 方法的重要性，逐渐开始转向。国内 AI 研究在这一段时间内出现了明显落后，仍然沿着 Bert 模型的方向继续前进，对 GPT-3 的研究很少；同时受制于中文优质语料的缺乏，模型大小也较难提升。

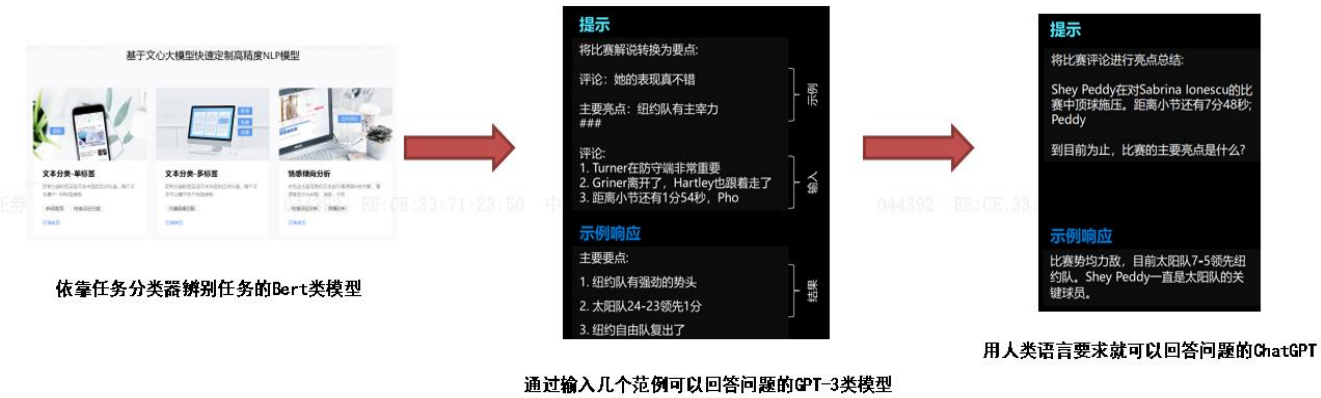
图 6：Prompt 简化了使用端接口，是更好的产品思维



资料来源：Google 官网，中信证券研究部

4）**2022 年至今**：ChatGPT 的成功证明了 GPT 模型的 Prompting 道路的正确性，同时也强调了数据质量的重要性。ChatGPT 最重要的成功是在产品化上更进一步：ChatGPT 在模型精度上并没有飞跃性的突破，但从 Few Shot prompt（需要输入少量范例示范）转换到 Instruct（用人类语言描述想做什么）更加贴合用户的习惯。

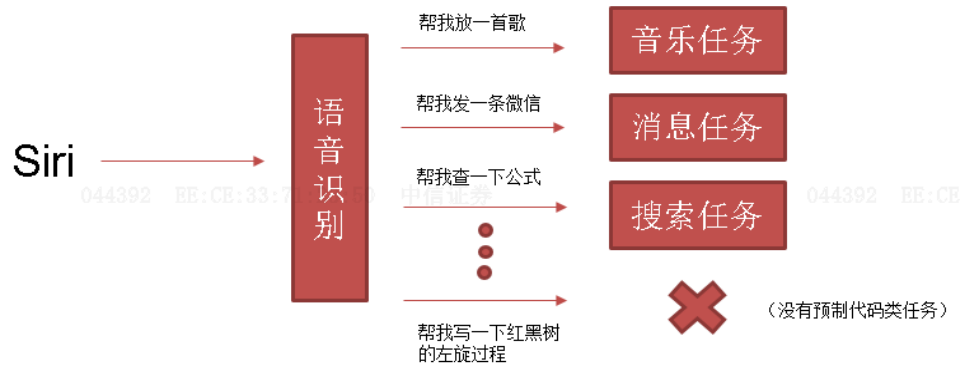
图 7：产品化能力的进步让 ChatGPT 取得了史无前例的用户增长速度



资料来源：百度官网，微软官网，中信证券研究部

总结：产品逻辑上，Prompting 虽然牺牲了部分精度，但无需用任务区分器区别不同的任务，更接近于大众所理解的“通用人工智能”。此前大众此前接触的人工智能如·Siri、小爱、小度音箱等，通过以任务分类的形式运行，准备不同任务的标注数据分别进行训练，是传统的 Bert 类模型。简单来说，将预先设置好的任务类型放于模型背后，使用者通过描述任务类型系统来匹配对应的模块，缺点是使用者的指令需要清晰且无法执行没有预先设置的任务类型。而 GPT 类模型面对用户不同的输入，模型可以自行判断给了用户更好的体验，这也更接近于大众理解的“通用人工智能”。除了应用于文字领域，目前 LLM（大语言模型）已经被应用于图像、音视频等场景的问题求解中，并不断取得理想的效果，而在 LLM 模型的技术路线上，GPT 在内容生成、理解上的良好表现，以及更为可行的人、模型交互方式（prompt 提示词），亦使得 LLM 技术路线不断向 GPT 方向进行收敛。

图 8：以 Siri 为代表的传统模式



资料来源：中信证券研究部绘制

图 9：ChatGPT 如何回答红黑树问题

红黑树是一种自平衡二叉搜索树，它的平衡性是通过节点的颜色属性和旋转操作来保证的。其中左旋操作是一种重要的操作，可以使红黑树保持平衡。

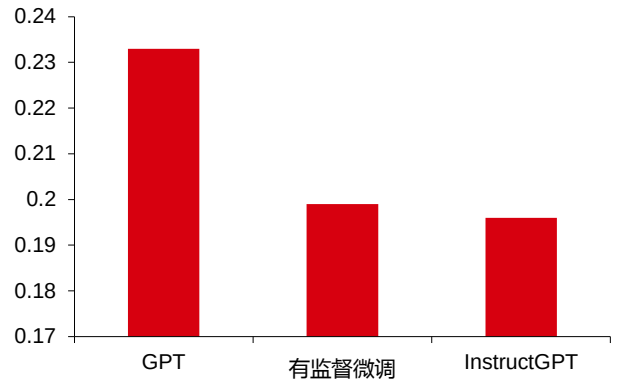
假设要对一棵红黑树进行左旋，旋转的节点为节点x，x的右子节点为节点y。左旋过程如下：

1. 将节点y的左子节点赋值给节点x的右子节点（即将节点y的左子节点变为节点x的右子节点）。
2. 如果节点y的左子节点非空，将节点x设置为节点y的左子节点的父节点（即将节点y的左子节点的父节点设置为节点x）。
3. 将节点x的父节点设置为节点y的父节点。
4. 如果节点x的父节点为NULL，则将节点y设置为根节点。
5. 如果节点x是其父节点的左子节点，则将节点y设置为节点x父节点的左子节点。
6. 如果节点x是其父节点的右子节点，则将节点y设置为节点x父节点的右子节点。
7. 将节点y的左子节点设置为节点x。
8. 将节点x的父节点设置为节点y。

经过上述步骤，节点x和节点y的位置发生了变化，但是红黑树的性质仍然得到了保持。在左旋过程中，节点x的右子节点y变为了节点x的父节点，同时节点y的左子节点变为了节点x，从而保持了红黑树的左子树比右子树高度相等的性质。同时，节点x和节点y的颜色和节点y的右子节点颜色不会发生变化，因此也保持了红黑树的平衡性质。

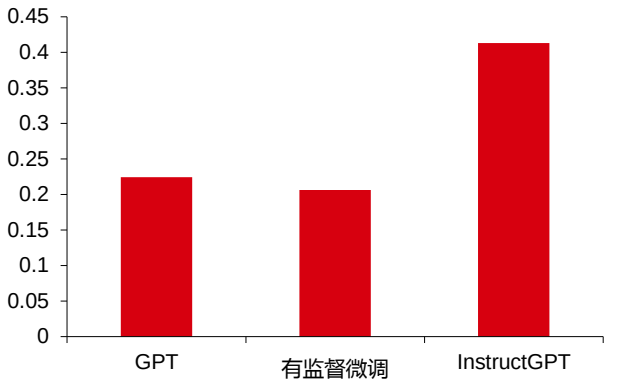
资料来源：openAI，中信证券研究部

图 10：Real Toxicity 数据集（InstructGPT 牺牲部分精度）



资料来源：OpenAI，中信证券研究部

图 11：TruthfulQA 数据集（InstructGPT 在 QA 取得更好成绩）

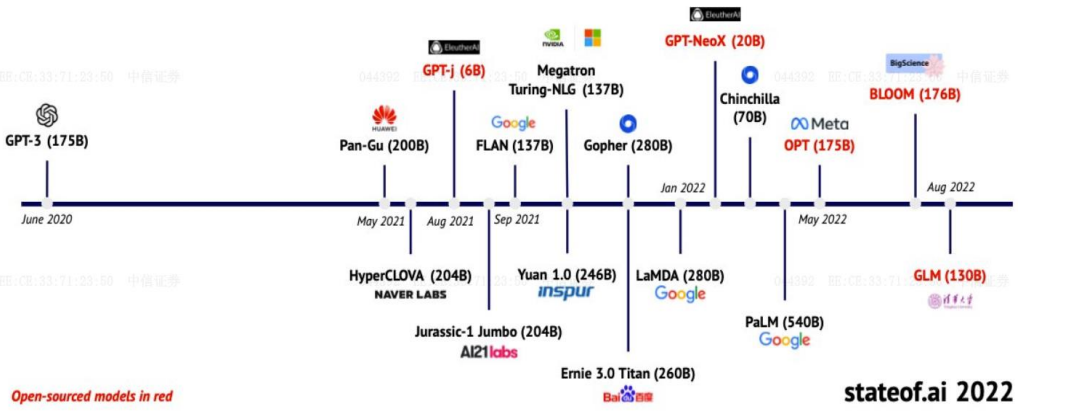


资料来源：OpenAI，中信证券研究部

竞争壁垒：数据、工程实践、人才、资本等

回顾过去近十年人工智能模型的发展，我们总结发现，新厂商进入大语言模型研发的门槛越来越高。从 2017 年谷歌提出 Bert 模型至今，我们可以看到大语言模型的迭代主要都是由老牌的互联网大厂或是巨头全力支持的初创公司（OpenAI、Deepmind）推动。这个赛道中的参与者数量十分稀少，而原因就来自于极高的准入门槛。基于上文的分析，中期维度，伴随 AI 模型技术路线朝着以 GPT 为主的 LLM（大语言模型）方向收敛，我们总结认为，AI 算法模型的核心技术壁垒将主要集中到：数据、工程实践、资本、核心人才等层面。

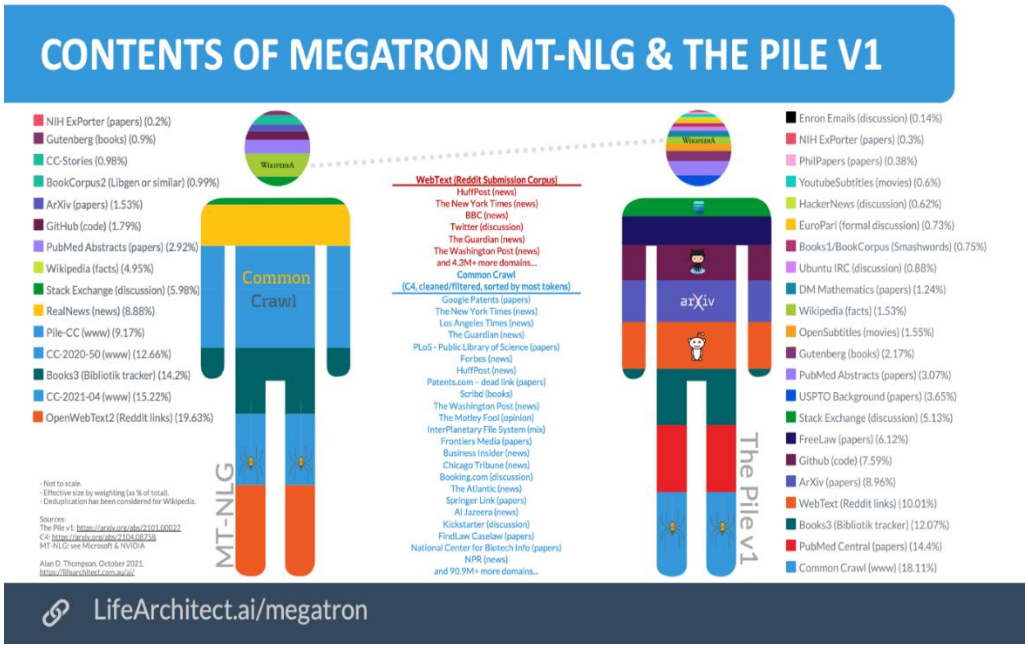
图 12：自 GPT-3 以来发布的部分大语言模型



资料来源：《stateof.ai 2022》Nathan Benaich，Ian Hogarth

数据：目前公开互联网数据已经被很大程度利用。从 Bert 开始到 GPT-3 再到谷歌的 PALM，网络中的公开语言数据源已经在被尽可能地利用（论坛、新闻、维基百科等等），利用爬虫等手段可以大量爬取公开的数据源。但模型的进一步优化仍需要更多数据，这要求模型开发商有能力接触到优质的私有数据来源，从而才能在模型的数据底层取得差异性的优势。

图 13：目前的大模型已经在穷尽所有能找到的公开网络数据



资料来源：LifeArchitect，中信证券研究部

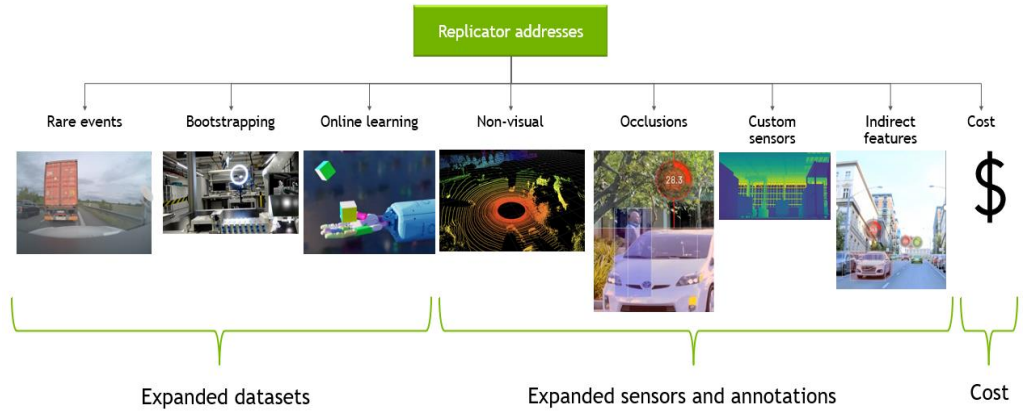
我们判断，优质私有数据源，以及通过私有数据进一步生成的合成数据，将在未来 3-5 年里人工智能模型的精度优化上发挥更为重要的作用。爬虫难以爬取的社交媒体数据、保存在后台的用户偏好数据、代码托管数据等仍具有开发空间，这也将成为数据获取端的优势。如 Google 对用户搜索以及 Youtube 视频的数据源、微软在 Github 中收集到的私有数据都将成为决定模型能力的差异化因素。此外，结合这些私有数据生成的合成数据将进一步放大这些公司数据端的优势。例如英伟达推出的 Omniverse 平台就与奔驰展开合作：奔驰可以通过将自有行车记录数据输入 Omniverse 生成用于自动驾驶的合成数据，这部分数据可以模拟在现实世界中较为少见的特殊情景并对自动驾驶模型做进一步的补充。

表 1：部分互联网巨头及其对应的私有数据

公司	产品	对应的私有数据
微软	Github、Office、Teams 等	代码托管、办公数据及用户偏好
谷歌	搜索引擎、Youtube 等	视频数据、用户搜索偏好等
Meta	Facebook、Instgram 等	爬虫难以爬取的社交媒体数据
亚马逊	电商	电商中的用户偏好数据
百度	搜索引擎、贴吧等	用户偏好以及部分无法被爬虫的论坛数据
腾讯	社交媒体、办公软件等	私有的社交媒体数据
阿里	电商	电商中的用户偏好数据

资料来源：各公司官网，中信证券研究部

图 14：奔驰通过自身的私有数据在英伟达 Omniverse 上进一步生成适用于自动驾驶的合成数据



资料来源：英伟达官网、中信证券研究部

资本投入：随着模型体积仍在指数级增长，大量的前期资本投入是阻碍新入者的重要因素。在过去五年内，我们看到人工智能模型的独角兽都在寻找互联网科技大厂作为其背后的依靠，极少有仅靠 PE、VC 融资的大语言模型初创公司出现。**原因我们可以总结为以下几点：**

图 15：世界一流大语言模型初创公司几乎都背靠强力公有云厂商的支持



资料来源：Reuters，中信证券研究部

1）能够得到充足而短时间内不求回报的资金支持；

2）依靠互联网大厂的公有云等相关资源进一步降低成本；

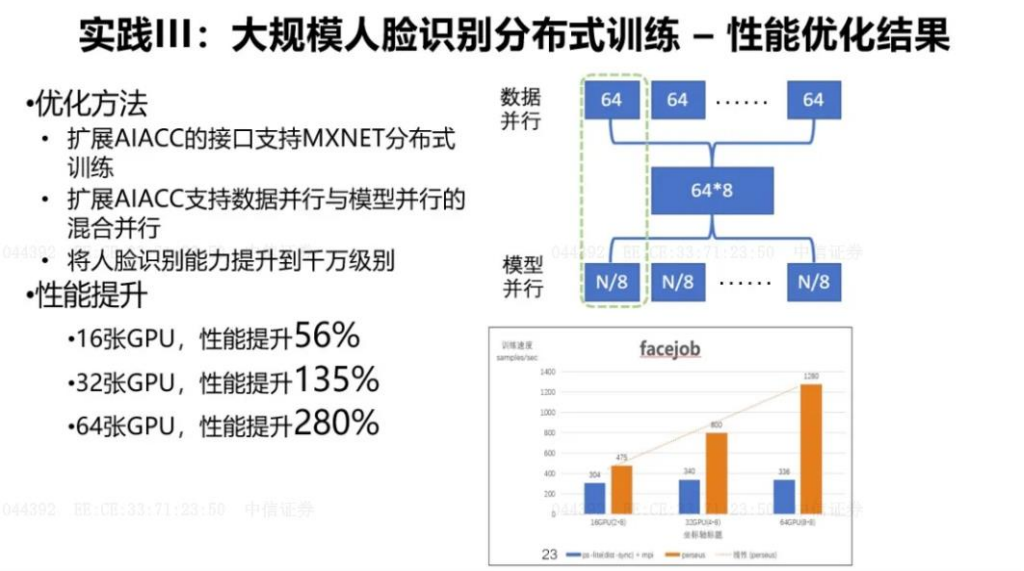
3）获取互联网大厂长期以来所积累的优质数据源。

这也是我们为什么看到在过去五年内人工智能头部厂商逐渐从开源走向闭源，利用其资源优势来打造差异化的 AI 模型，而落后者更希望通过开源的模式以求缩小与头部厂商的距离的原因。

工程实践：研发团队的工程能力是决定公司在大语言模型竞争力的另一个重要因素。工程能力决定了 AI 研发团队对模型的部署、迭代以及维护的速度，这也变相增加了模型的成本。当前的大语言模型参数量在 2000 亿以上，整个预训练过程长达最少一个月，这其

中包含了大量的工程实施问题。如何优化工程细节，以及缺乏实践工程经验导致的试错成本是目前新进入者的一个重要困难。

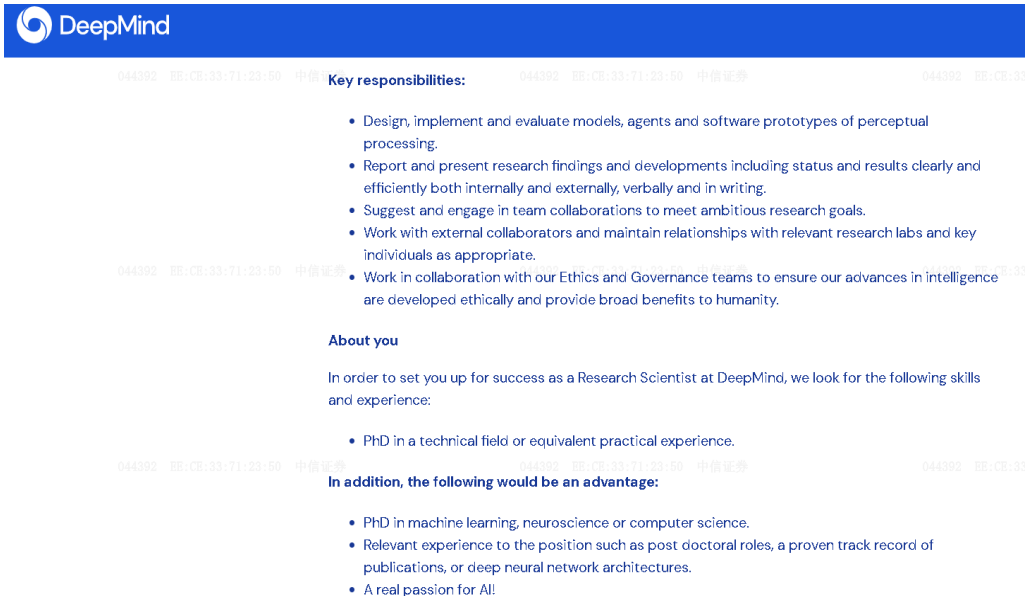
图 16：理想情况下足够好的优化可以降低 90%以上的成本



资料来源：阿里云官网

核心人才：大语言模型与以往小模型研究范式的区别决定了，目前具备足够工程能力的人才十分稀缺，大语言模型研发拥有经验丰富的领头人极为关键。以往的人工智能模型开发主要围绕 LSTM、CNN 等小模型，这些小模型的调试、使用方法等往往都包含在了学校课程中，个体工程师也可以通过自有资源进行摸索并丰富自己的经验。但随着语言模型的体积不断增加，在研究方法上现今千亿量级的模型与之前几十亿量级的小模型发生了本质变化，个体工程师没有能力通过自有资源积累对大模型的学习经验。一个合格的大模型研发团队需要依靠大公司的资源支持才能积累对模型调试、优化、实际部署等各个环节足够的经验。根据 OpenAI 在其招聘官网中所公布的细节，其基本要求博士生毕业获有 3 年以上在大语言模型经验的研究生。优秀的工程人才稀缺以及培养难度让没有经验的新公司进入更为困难，如何招募到一个大模型领域经验丰富的专家并在其影响力的基础上招募来自于大公司的有一定工程经验的团队将成为新入公司的核心问题。

图 17：Deepmind 对基础研究员的招聘要求为博士生或博士生相应的从业经验


DeepMind

Key responsibilities:

- Design, implement and evaluate models, agents and software prototypes of perceptual processing.
- Report and present research findings and developments including status and results clearly and efficiently both internally and externally, verbally and in writing.
- Suggest and engage in team collaborations to meet ambitious research goals.
- Work with external collaborators and maintain relationships with relevant research labs and key individuals as appropriate.
- Work in collaboration with our Ethics and Governance teams to ensure our advances in intelligence are developed ethically and provide broad benefits to humanity.

About you

In order to set you up for success as a Research Scientist at DeepMind, we look for the following skills and experience:

- PhD in a technical field or equivalent practical experience.

In addition, the following would be an advantage:

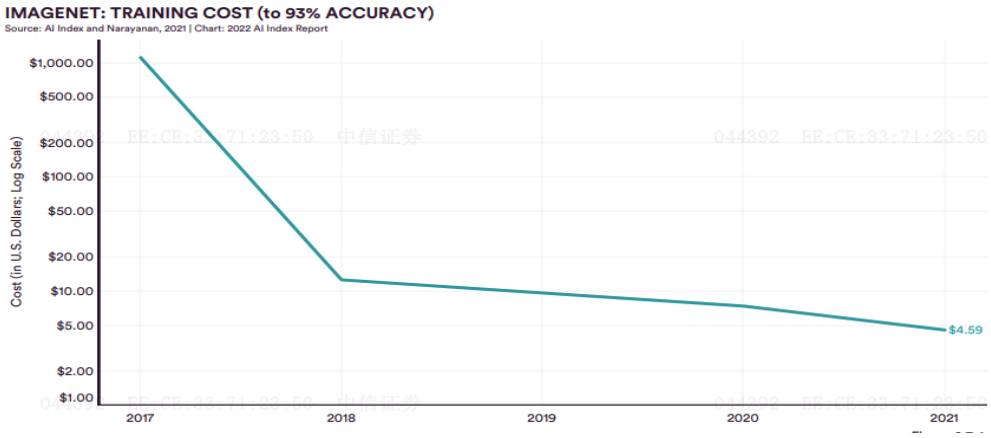
- PhD in machine learning, neuroscience or computer science.
- Relevant experience to the position such as post doctoral roles, a proven track record of publications, or deep neural network architectures.
- A real passion for AI!

资料来源：Deepmind 官网

算力成本：单位成本持续保持快速下降趋势

根据我们之前系列报告的测算，尽管人工智能模型的总研发投入持续在提高，但主要原因来自于模型体积的倍数式增长，而人工智能模型的单位成本保持下降趋势。根据斯坦福大学 AI index 2022 的报告，2017 年训练 93%精确度的 Imagenet 需要花费 1112.6 美元，而在 2021 年底仅需 4.6 美元，单位成本仅为四年前的 223 分之 1。

图 18：对于相同的数据集的训练成本，近五年持续保持下降趋势

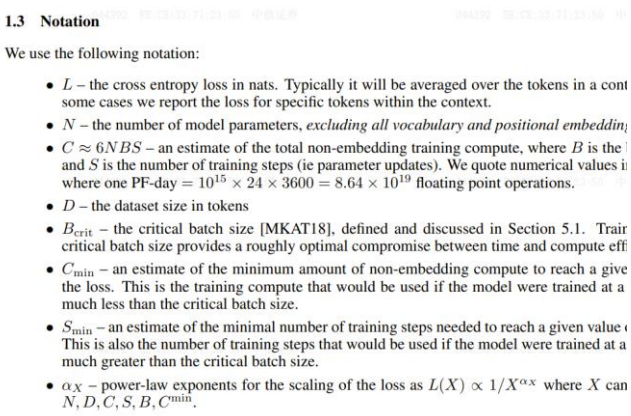


资料来源：斯坦福 AI Index 2022

大语言模型训练端成本：两年内下降约 66%。因为单次模型训练耗时较长，一般情况下不太会反复预训练大语言模型，因此我们在测算训练环节成本时，假定预训练单次，且训练过程中没有出现错误时的成本。实际情形中，考虑到训练过程中出现工程类错误的可能性，实际成本会高于我们计算的理想情况成本。假设参数量为 175B、500B Tokens 的情况下，根据 OpenAI 的“Scaling Laws for Neural Language Models”论文（作者：Jared

Kaplan,Sam McCandlish, Tom Henighan 等），在使用 256 个 HGX A100 服务器(2048 个 A100 GPU 卡）的情况下，Model FLOPs Utilization（MFU）假设为 Megatron-LM 的 51.04%，我们推测单次训练时长约为 30.71 天，换算为 1509702.7 GPU 小时。假设训练使用成本价约为 1 美元/GPU 小时的情况下，耗费服务器端成本约为 151 万美元。而在 2020 年根据 Stanford 讲师 Mourri 的测算，当时训练 GPT-3 的成本约为 460 万美元，我们可以看到训练相同模型的成本在两内下降了约 66%。

图 19：我们从 Scaling Laws 论文中给出的公式出发进行推算



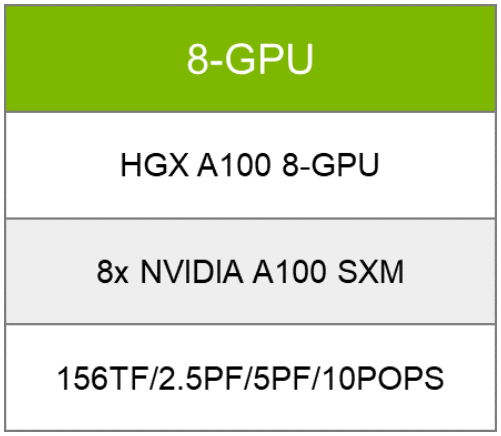
1.3 Notation

We use the following notation:

- L – the cross entropy loss in nats. Typically it will be averaged over the tokens in a cont some cases we report the loss for specific tokens within the context.
- N – the number of model parameters, *excluding all vocabulary and positional embeddin*
- $C \approx 6NBS$ – an estimate of the total non-embedding training compute, where B is the and S is the number of training steps (ie parameter updates). We quote numerical values i where one PF-day $= 10^{15} \times 24 \times 3600 = 8.64 \times 10^{19}$ floating point operations.
- D – the dataset size in tokens
- B_{crit} – the critical batch size [MKAT18], defined and discussed in Section 5.1. Trair critical batch size provides a roughly optimal compromise between time and compute eff
- C_{min} – an estimate of the minimum amount of non-embedding compute to reach a give the loss. This is the training compute that would be used if the model were trained at a much less than the critical batch size.
- S_{min} – an estimate of the minimal number of training steps needed to reach a given value This is also the number of training steps that would be used if the model were trained at a much greater than the critical batch size.
- α_X – power-law exponents for the scaling of the loss as $L(X) \propto 1/X^{\alpha_X}$ where X can $N, D, C, S, B, C^{\text{min}}$.

资料来源：Scaling Laws for Neural Language Models（Jared Kaplan,Sam McCandlish, Tom Henighan 等）

图 20：参考 HGXA100 在 FP16 精度下的峰值算力为 5PFLOPS



资料来源：英伟达官网

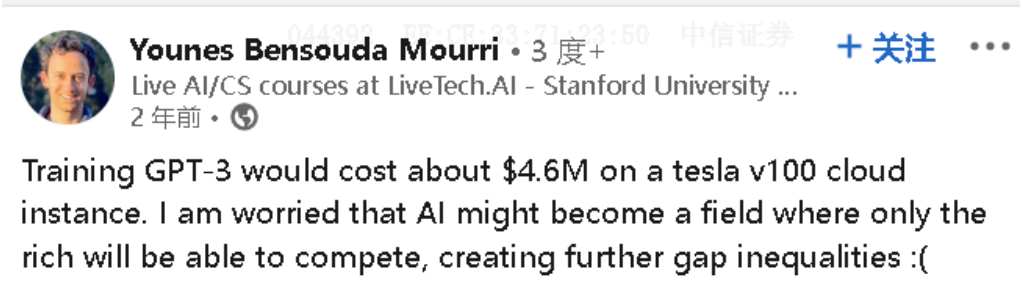
表 2：参考英伟达 Megatron-LM 175B 的实际算力为理论峰值的 51.4%

模型大小	MFU	HFU
22B	41.50%	43.70%
175B	51.40%	52.80%
530B	56.00%	57.00%
1T	56.30%	57.00%

资料来源：英伟达 Megatron 模型 Readme，中信证券研究部

成本下降速度：根据斯坦福大学讲师 Mourri 在 2020 年使用 Tesla V100 显卡作为基准的估算，当时训练 GPT-3 的成本约为 460 万美元。由此可见成本在两年内下降了约 66%，不过考虑到训练时间成本仍然高昂，反复预训练大语言模型仍不会成为首选。

图 21：斯坦福大学讲师 Mourri 曾在 2020 年推算使用 Tesla V100 显卡训练 GPT-3 需要 460 万美元



资料来源：Younes Bensouda Mourri@LinkedIn

大语言模型推理端成本：通过蒸馏、稀疏化处理等方式持续优化，中短期内仍存在数倍的优化空间。以 ChatGPT 为例，根据我们在 2022 年底的测算，ChatGPT 生成一次回答的成本约在 1-2 美分之间。而从 3 月 2 日 OpenAI 对 ChatGPT API 的定价出，0.2 美分

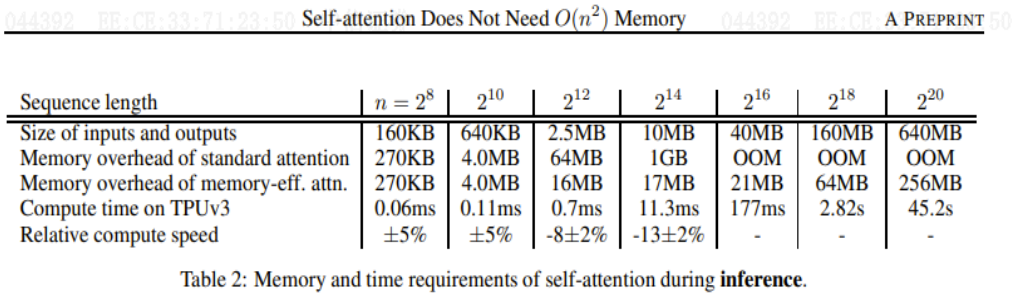
每 1000tokens 的定价对应为百亿参数级别的模型，结合 ChatGPT 在训练过程中所使用得 Reward Model（奖励模型，记录了人类偏好）应为千亿参数级别的模型，这意味着 OpenAI 将其在千亿参数模型中所学习到的能力通过类似于蒸馏（distillation）的方式浓缩进了百亿参数的模型中，通过缩减参数量的形式快速降低了成本消耗。根据目前用户在社交媒体的反馈结果看，这个百亿参数模型的能力以及追平甚至超越了先前的大尺寸模型。

表 3：ChatGPT 推理任务测算表

假设：模型 Layers	96	假设：模型 Hidden Size	12288	假设：Batch Size	1
8XA100 80GB 算力 PetaFLOPS	5	FLOPs Utilization	30%	生成 1 个 token 需要内存（Bytes）	2.61E+11
8XA100 80GB 带宽 TB/s	4.8	带宽损耗	15%	带宽限制生成一个 Token 的时间（s）	0.064
加载参数权重需要的时间（s）	0.015	加载权重时 Input 长度	367	算力计算一个 Token 的时间（s）	1.7*10^-4
算力密度临界值	320	每次回答需要生成 Token	200	单次请求生成时间（s）	12.22
8xA100 服务器成本（美元/小时）	8	单次生成成本（美分）	2.7		

资料来源：中信证券研究部测算

图 22：通过优化算法结构能够显著改善推理任务的内存占用



Self-attention Does Not Need $O(n^2)$ Memory A PREPRINT

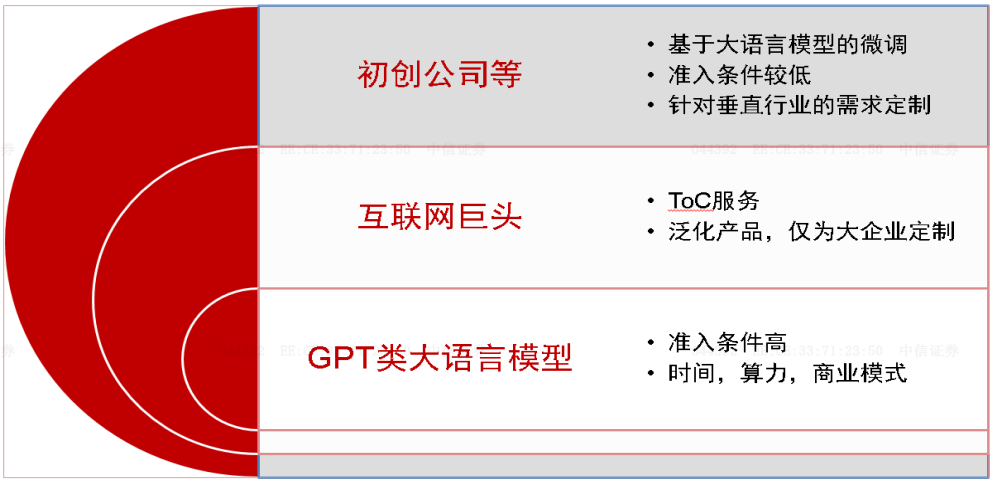
Sequence length	$n = 2^8$	2^{10}	2^{12}	2^{14}	2^{16}	2^{18}	2^{20}
Size of inputs and outputs	160KB	640KB	2.5MB	10MB	40MB	160MB	640MB
Memory overhead of standard attention	270KB	4.0MB	64MB	1GB	OOM	OOM	OOM
Memory overhead of memory-eff. attn.	270KB	4.0MB	16MB	17MB	21MB	64MB	256MB
Compute time on TPUv3	0.06ms	0.11ms	0.7ms	11.3ms	177ms	2.82s	45.2s
Relative compute speed	±5%	±5%	-8±2%	-13±2%	-	-	-

Table 2: Memory and time requirements of self-attention during **inference**.

资料来源：Self-attention Does Not Need O(n2) Memory（Markus N. Rabe and Charles Staats）

总结：因此无论是模型推理端或训练端，我们都观察到单位成本在持续保持快速下降趋势，而总成本的提高，则主要来源于模型体积的增长、训练数据集增加、推理端用户调用次数增加等。当未来模型大小与日活用户数接近阈值后，我们认为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。

图 23：未来大语言模型生态格局预测

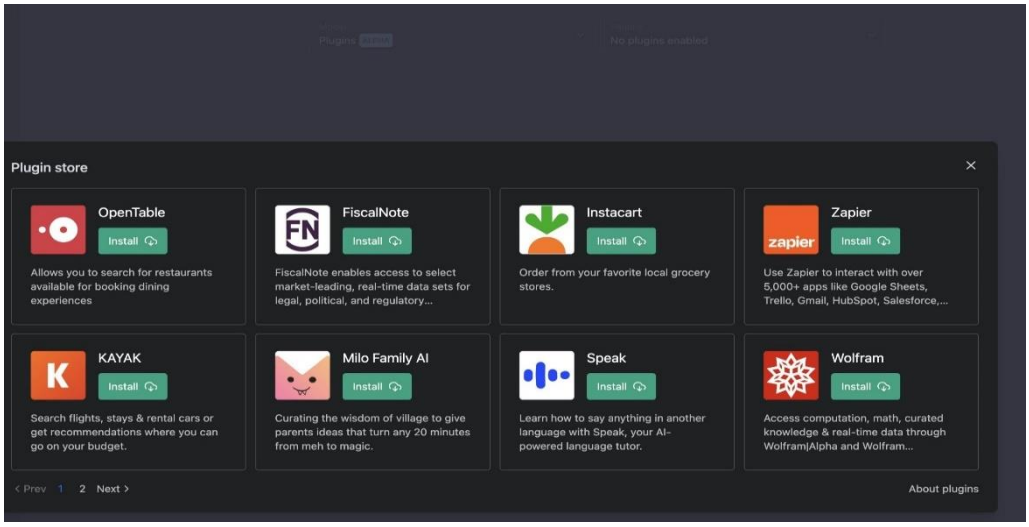


资料来源：中信证券研究部绘制

未来展望：通过插件等扩展大语言模型（LLM）能力边界

OpenAI 在 ChatGPT 中推出插件接口，进一步扩展 ChatGPT 能力。OpenAI 团队于 2023 年 3 月宣布其为插件（Plugins）接口展开了 Limited Alpha 测试，插件将 ChatGPT 连接到第三方应用程序。这些插件使 ChatGPT 能够与外部定义的 API 进行交互，增强 ChatGPT 的功能并使其能够执行各种操作。例如：获取实时信息：体育比分、股票价格、最新新闻等；获取知识库信息：公司文档、个人笔记等；代表用户执行操作：预订航班、订购食物等。插件的推出，能够有效解决 ChatGPT 在获取联网信息、回答复杂逻辑推理问题等层面的潜在缺陷，并为持续扩展 ChatGPT 的能力边界，提供了一种极为理想的实现方式，这种思路和图灵奖得主 Yann LeCun 在最新的论文《Augmented Language Models: a Survey》提出的增强语言模型（ALM）不谋而合，Yann LeCun 认为，可以通过复杂问题分解、链接外部知识&工具等所谓的增强语言模型方式，来解决现有大语言模型在问答可靠性、信息及时性等方面的固有缺陷。

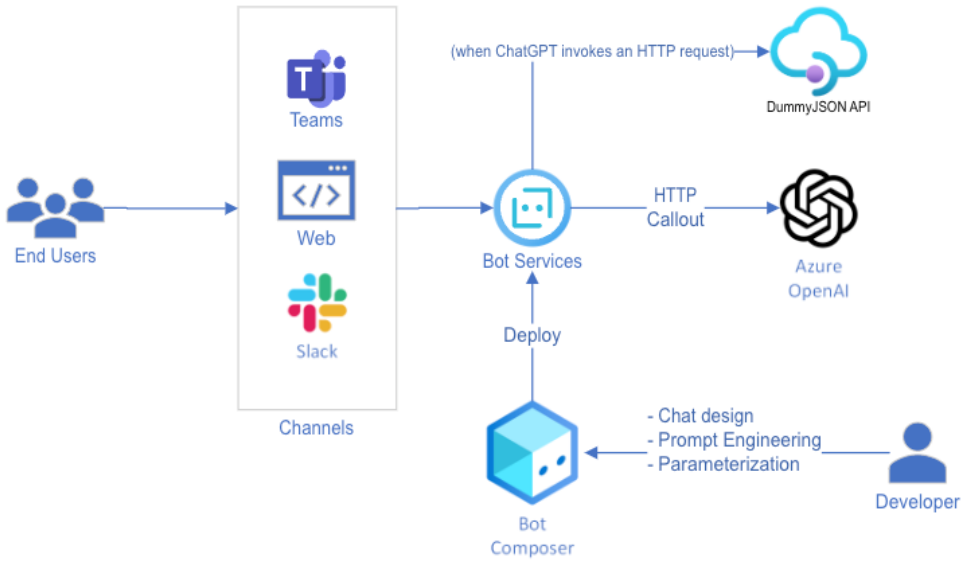
图 24：OpenAI 目前给出的插件 Demo



资料来源：OpenAI 官网

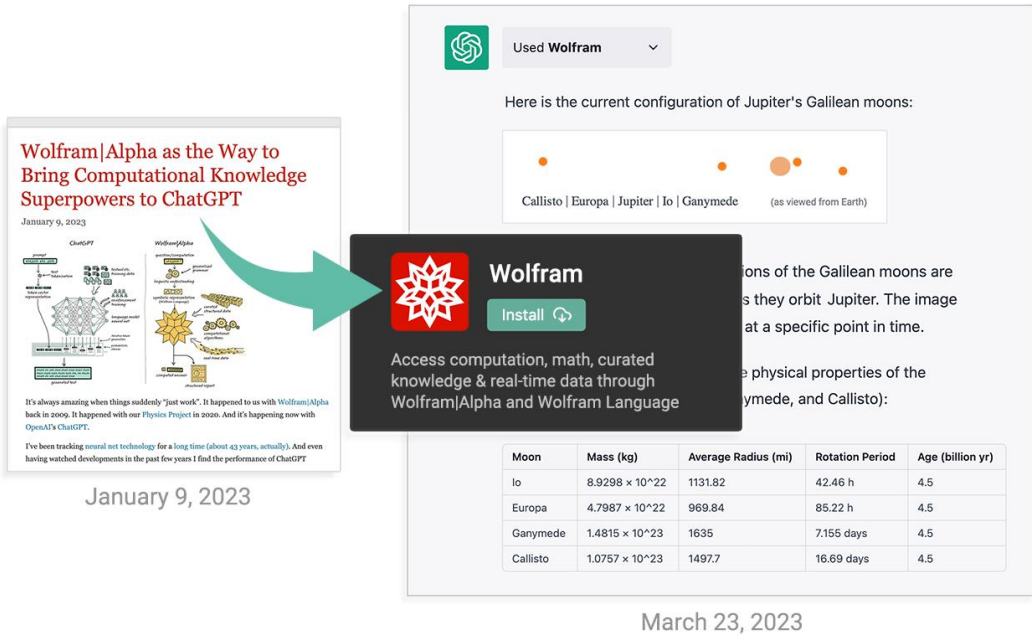
插件模式下，第三方厂商无需接触部署大语言模型，简化了微调环节的实施难度。过去第三方厂商对模型的微调需要首先将模型本地部署再根据自有数据进行调整，但随着模型体积的膨胀，本地部署对一些小公司非常困难。现在第三方厂商将可以自有的数据源及部分算法作为 ChatGPT 之上的插件，ChatGPT 可以调用插件中的外部 API 接口来获得信息。插件提供者使用 OpenAPI 标准编写 API，然后这个 API 会被编译成一个 prompt 向 ChatGPT 解释如何使用 API 来增强其答案。当用户选择对应插件后，如果 ChatGPT 认为应该从 API 获取信息，它将发出请求并在尝试回应之前将信息添加到上下文中。例如，如果用户询问："我在巴黎住几晚，应该住在哪里？"大语言模型可能会选择调用酒店预订插件 API，接收 API 响应，并生成结合 API 数据和自然语言能力的面向用户的答案。通过插件接口这一模式简化了小公司对大语言模型微调的难度，并丰富了基于大语言模型之上的环境生态。

图 25：ChatGPT 通过插件调用外部 API 并通过 Slack、teams 等界面反馈给用户



资料来源：微软官方网站

图 26：Wolfram 通过插件告诉 ChatGPT 如何使用自己的 API 接口，并在用户需要时如何调用



资料来源：Wolfram 官网

风险因素

AI 核心技术发展不及预期风险；科技领域政策监管持续收紧风险；私有数据相关的政策监管风险；全球宏观经济复苏不及预期风险；宏观经济波动导致欧美企业 IT 支出不及预期风险；AI 伦理、道德风险；企业数据泄露、信息安全风险；行业竞争持续加剧风险等。

投资策略

ChatGPT 在全球 AI 产业、技术领域带来的良好示范效应，有望推动 AI 算法模型结束当前的技术路线分叉，并不断向以 GPT 为主导的大语言模型（LLM）靠拢。中期维度，大语言模型+“暴力美学”预计仍将是我们不断逼近通用人工智能的最可能路线，在此假设下，模型算法架构将遵从渐进的学术研究步伐，高质量数据集、工程实践能力、核心人才、资本将成为大模型研发领域的核心竞争壁垒。同时我们判断，模型单位算力成本（训练、推理）亦有望持续快速下降，从而推动模型体积、应用场景等产业核心要素相互促进，不断加快 AI 产业的发展速度。面对当前 GPT 等大语言模型在问答可靠性、推理能力等层面缺陷，复杂问题分解、链接外部知识&工具等增强语言模型（ALM）方式料将是可行的方式，比如 ChatGPT 插件等。

作为中期最具确定性的产业方向之一，我们持续看好全球 AI 领域的投资机会，并建议持续聚焦芯片、算力设施、模型架构&工程实践、应用场景等核心环节。在美股市场，我们建议持续关注：英伟达、台积电、微软、AMD、Arista、博通、Marvell、百度、Adobe、Snowflake 等。

表 4：重点推荐公司盈利预测

类别	公司	代码	市值（亿美元）	估值方法	估值（自然年）			
					2022A	2023E	2024E	2025E
算力	英伟达	NVDA	6640	PE	79.4	59.8	45.1	35.3
	AMD	AMD	1492	PE	27.1	30.3	21.6	18.0
	台积电	TSMC	4678	PE	14.1	16.8	13.7	12.3
	亚马逊	AMZN	10360	EV/EBITDA	NA	13.9	10.6	8.1
	Arista	ANET	499	PE	34.5	27.3	24.5	22.6
	博通	AVGO	2622	PE	15.9	14.7	14.0	13.2
	Marvell	MRVL	344	PE	18.9	26.8	17.6	14.2
算法模型	微软	MSFT	21166	P/FCF	32.9	34.1	26.9	22.0
	谷歌	GOOG	13379	PE	22.3	20.3	17.3	14.7
	百度	BIDU	505	PE	16.8	14.7	12.9	12.0
应用场景	Adobe	ADBE	1752	P/FCF	24.3	22.0	19.3	18.1
	Salesforce	CRM	1953.1	P/FCF	30.9	26.5	19.9	16.4
	Service Now	NOW	948.6799	P/FCF	43.9	36.9	29.2	23.4
产业配套	Snowflake	SNOW	476	P/S	23.0	16.5	12.1	9.0
	Confluent	CFLT	64	P/S	11.0	8.4	6.6	5.1
	MongoDB	MDB	149	P/S	11.6	9.9	8.2	6.3
	Palo Alto	PANW	588	P/FCF	34.5	30.0	24.5	20.8

资料来源：彭博，中信证券研究部　　注：股价为 2023 年 4 月 5 日收盘价；预测数据来自彭博一致预期

分析师声明

主要负责撰写本研究报告全部或部分内容的分析师在此声明：（i）本研究报告所表述的任何观点均精准地反映了上述每位分析师个人对标的证券和发行人的看法；（ii）该分析师所得报酬的任何组成部分无论是在过去、现在及将来均不会直接或间接地与研究报告所表述的具体建议或观点相联系。

一般性声明

本研究报告由中信证券股份有限公司或其附属机构制作。中信证券股份有限公司及其全球的附属机构、分支机构及联营机构（仅就本研究报告免责条款而言，不含 CLSA group of companies），统称为“中信证券”。

本研究报告对于收件人而言属高度机密，只有收件人才能使用。本研究报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。本研究报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。中信证券并不因收件人收到本报告而视其为中信证券的客户。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断并自行承担投资风险。

本报告所载资料的来源被认为是可靠的，但中信证券不保证其准确性或完整性。中信证券并不对使用本报告或其所包含的内容产生的任何直接或间接损失或与此有关的其他损失承担任何责任。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可跌可升。过往的业绩并不能代表未来的表现。

本报告所载的资料、观点及预测均反映了中信证券在最初发布该报告日期当日分析师的判断，可以在不发出通知的情况下做出更改，亦可因使用不同假设和标准、采用不同观点和分析方法而与中信证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。中信证券并不承担提示本报告的收件人注意该等材料的责任。中信证券通过信息隔离墙控制中信证券内部一个或多个领域的信息向中信证券其他领域、单位、集团及其他附属机构的流动。负责撰写本报告的分析师的薪酬由研究部门管理层和中信证券高级管理层全权决定。分析师的薪酬不是基于中信证券投资银行收入而定，但是，分析师的薪酬可能与投行整体收入有关，其中包括投资银行、销售与交易业务。

若中信证券以外的金融机构发送本报告，则由该金融机构为此发送行为承担全部责任。该机构的客户应联系该机构以交易本报告中提及的证券或要求获悉更详细信息。本报告不构成中信证券向发送本报告金融机构之客户提供的投资建议，中信证券以及中信证券的各个高级职员、董事和员工亦不为（前述金融机构之客户）因使用本报告或报告载明的内容产生的直接或间接损失承担任何责任。

评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即：以报告发布日后的 6 到 12 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准；韩国市场以科斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅 20%以上
		增持	相对同期相关证券市场代表性指数涨幅介于 5%～20%之间
		持有	相对同期相关证券市场代表性指数涨幅介于-10%～5%之间
		卖出	相对同期相关证券市场代表性指数跌幅 10%以上
	行业评级	强于大市	相对同期相关证券市场代表性指数涨幅 10%以上
		中性	相对同期相关证券市场代表性指数涨幅介于-10%～10%之间
		弱于大市	相对同期相关证券市场代表性指数跌幅 10%以上

特别声明

在法律许可的情况下，中信证券可能（1）与本研究报告所提到的公司建立或保持顾问、投资银行或证券服务关系，（2）参与或投资本报告所提到的公司的金融交易，及/或持有其证券或其衍生品或进行证券或其衍生品交易，因此，投资者应考虑到中信证券可能存在与本研究报告有潜在利益冲突的风险。本研究报告涉及具体公司的披露信息，请访问 https://research.citicsinfo.com/disclosure。

法律主体声明

本研究报告在中华人民共和国（香港、澳门、台湾除外）由中信证券股份有限公司（受中国证券监督管理委员会监管，经营证券业务许可证编号：Z20374000）分发。本研究报告由下列机构代表中信证券在相应地区分发：在中国香港由 CLSA Limited（于中国香港注册成立的有限公司）分发；在中国台湾由 CL Securities Taiwan Co., Ltd.分发；在澳大利亚由 CLSA Australia Pty Ltd.（商业编号：53 139 992 331/金融服务牌照编号：350159）分发；在美国由 CLSA（CLSA Americas, LLC 除外）分发；在新加坡由 CLSA Singapore Pte Ltd.（公司注册编号：198703750W）分发；在欧洲经济区由 CLSA Europe BV 分发；在英国由 CLSA （UK）分发；在印度由 CLSA India Private Limited 分发（地址：8/F, Dalamal House, Nariman Point, Mumbai 400021；电话：+91-22-66505050；传真：+91-22-22840271；公司识别号：U67120MH1994PLC083118）；在印度尼西亚由 PT CLSA Sekuritas Indonesia 分发；在日本由 CLSA Securities Japan Co., Ltd.分发；在韩国由 CLSA Securities Korea Ltd.分发；在马来西亚由 CLSA Securities Malaysia Sdn Bhd 分发；在菲律宾由 CLSA Philippines Inc.（菲律宾证券交易所及证券投资者保护基金会员）分发；在泰国由 CLSA Securities (Thailand) Limited 分发。

针对不同司法管辖区的声明

中国大陆：根据中国证券监督管理委员会核发的经营证券业务许可，中信证券股份有限公司的经营范围包括证券投资咨询业务。

中国香港：本研究报告由 CLSA Limited 分发。 本研究报告在香港仅分发给专业投资者（《证券及期货条例》（香港法例第 571 章）及其下颁布的任何规则界定的），不得分发给零售投资者。就分析或报告引起的或与分析或报告有关的任何事宜，CLSA 客户应联系 CLSA Limited 的罗鼎，电话：+852 2600 7233。

美国：本研究报告由中信证券制作。本研究报告在美国由 CLSA（CLSA Americas, LLC 除外）仅向符合美国《1934 年证券交易法》下 15a-6 规则界定且 CLSA Americas, LLC 提供服务的“主要美国机构投资者”分发。对身在美国的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所述任何观点的背书。任何从中信证券与 CLSA 获得本研究报告的接收者如果希望在美国交易本报告中提及的任何证券应当联系 CLSA Americas, LLC（在美国证券交易委员会注册的经纪交易商），以及 CLSA 的附属公司。

新加坡：本研究报告在新加坡由 CLSA Singapore Pte Ltd.，仅向（新加坡《财务顾问规例》界定的）“机构投资者、认可投资者及专业投资者”分发。就分析或报告引起的或与分析或报告有关的任何事宜，新加坡的报告收件人应联系 CLSA Singapore Pte Ltd，地址：80 Raffles Place, #18-01, UOB Plaza 1, Singapore 048624，电话：+65 6416 7888。因您作为机构投资者、认可投资者或专业投资者的身份，就 CLSA Singapore Pte Ltd.可能向您提供的任何财务顾问服务，CLSA Singapore Pte Ltd 豁免遵守《财务顾问法》（第 110 章）、《财务顾问规例》以及其下的相关通知和指引（CLSA 业务条款的新加坡附件中证券交易服务 C 部分所披露）的某些要求。MCI（P）085/11/2021。

加拿大：本研究报告由中信证券制作。对身在加拿大的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所载任何观点的背书。

英国：本研究报告归属于营销文件，其不是按照旨在提升研究报告独立性的法律要件而撰写，亦不受任何禁止在投资研究报告发布前进行交易的限制。本研究报告在英国由 CLSA （UK）分发，且针对由相应本地监管规定所界定的在投资方面具有专业经验的人士。涉及到的任何投资活动仅针对此类人士。若您不具备投资的专业经验，请勿依赖本研究报告。

欧洲经济区：本研究报告由荷兰金融市场管理局授权并管理的 CLSA Europe BV 分发。

澳大利亚：CLSA Australia Pty Ltd (“CAPL”) (商业编号：53 139 992 331/金融服务牌照编号：350159) 受澳大利亚证券与投资委员会监管，且为澳大利亚证券交易所及 CHI-X 的市场参与主体。本研究报告在澳大利亚由 CAPL 仅向“批发客户”发布及分发。本研究报告未考虑收件人的具体投资目标、财务状况或特定需求。未经 CAPL 事先书面同意，本研究报告的收件人不得将其分发给任何第三方。本段所称的“批发客户”适用于《公司法（2001）》第 761G 条的规定。CAPL 研究覆盖范围包括研究部门管理层不时认为与投资者相关的 ASX All Ordinaries 指数成分股、离岸市场上市证券、未上市发行人及投资产品。CAPL 寻求覆盖各个行业中与其国内及国际投资者相关的公司。

印度：CLSA India Private Limited，成立于 1994 年 11 月，为全球机构投资者、养老基金和企业提供股票经纪服务（印度证券交易委员会注册编号：INZ000001735）、研究服务（印度证券交易委员会注册编号：INH000001113）和商人银行服务（印度证券交易委员会注册编号：INM000010619）。CLSA 及其关联方可能持有标的公司的债务。此外，CLSA 及其关联方在过去 12 个月内可能已从标的公司收取了非投资银行服务和/或非证券相关服务的报酬。如需了解 CLSA India“关联方”的更多详情，请联系 Compliance-India@clsa.com。

未经中信证券事先书面授权，任何人不得以任何目的复制、发送或销售本报告。

中信证券 2023 版权所有。保留一切权利。