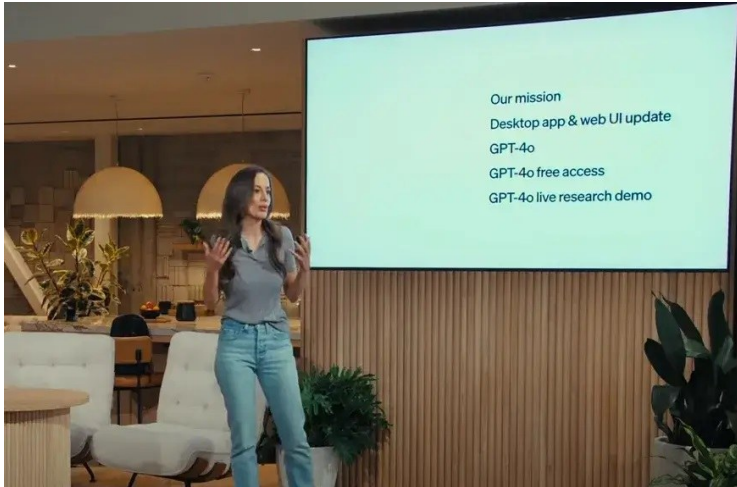


人工智能 2024：模型端 OpenAI 走下神坛，应用端商业化渐清晰



在经历过2023年的狂飙后，人工智能相关产业仍然是2024年最火热的行业之一。

将时间拨回到两年之前，彼时，ChatGPT刚刚问世就引起了广泛的关注。不过，虽然当时大家惊艳于LLM带来的震撼，但也没有想到在短短两年之后，人工智能就已经以包括Chat在内的各种不同形式融入了我们的日常，并正以前所未有的速度、广度和深度重塑生产生活方式。

而这些变化，则绝大多数发生在过去的一年当中。

那么，站在2025年的起点，就让我们一起回顾一下过去一年中各大人工智能企业，在模型端、应用端取得了哪些进展？大厂之间的竞争格局发生了怎样的变化，商业化又以怎样的形式铺开呢？

第一部分：模型端 OpenAI 走下神坛，各家大模型“百花齐放”

据多家媒体表示，目前几乎所有的人工智能龙头企业，都在新模型开发上的付出与回报的差额正在逐渐扩大。原因则是由于算力与可使用数据的不足，被业界认为是大模型预训练第一性原理的 Scaling Law (认为模型的性能同计算量、模型参数量、数据大小成幂律关系)，已经事实性受阻。

而在 Scaling Law 失效被大规模讨论之前，其实 OpenAI 给大家的惊喜程度，已经放缓。

2024年中，OpenAI 保持了一贯的领先者姿态，GPT 快速迭代。

5月13日，OpenAI 的 CTO Mira Murati，在三十分钟的时间里发布了4项功能，而其中最重要的就是 GPT-4o。相比于 GPT-4 需要将音频与文本进行相互转换后才能处理，GPT-4o 作为一个原生的多模态模型，所有输入和输出都由同一神经网络处理，在发布的时候可以认为是进行了一波“屠榜”。

9月13日，OpenAI 官宣发布 o1，包括两种版本 o1-preview 和 o1-mini。o1 在响应用户的请求之前，会像人类“深思熟虑”一样，进行很长时间的内部思维链思考，并可以在这个过程中尝试不同解决问题的方法甚至自我纠正错误。这也让其在数学、物理、化学、编程等领域，拥有更强的能力。

12月20日，OpenAI 在其连续12天的直播中，发布了最新的推理模型 o3 及其轻量版 o3-mini。在编码、数学和科学问答上相比于 o1 有了很大的提升，并成功完成了对 ARC-AGI 基准的突破，完成了人工智能在适应新任务的能力上重大飞跃。

不过，虽然从 4o 到 o1 再到 o3 都有着长足的进步，但或许是因已有基准较高，所以在实际的使用体感上，再也没有 ChatGPT 问世时的惊喜。

此外，每一代最新模型在成本方面也实在算不上经济。据 ARC-AGI 发起者在测试报告中写道，以 o3 为例，用户能够以每项任务大约 5 美元（折合人民币约 36 元）的价格来支付人工解决 ARC-AGI 任务，而在低推理模式下，o3 完成每个任务需要花费 17-20 美元（折合人民币约 124~145 元）。

这样的成本，无论对于个人用户还是企业用户来说，都实在是“不可承受之重”。

也就在 OpenAI 的新玩意没有给出大家预料中的惊喜的时候，其他 AI 企业的大模型进度却在飞速赶上。

3月，Anthropic 推出了 Claude 3 系列，其中的不同版本根据场景不同进行了特定的优化；6月，Claude 3.5 问世，各项能力均有不俗提升；10月，升级版的 Claude 3.5 Sonnet 和全新的 Claude 3.5 Haiku 同大家见面，在价格与生成速度不变的情况下，实现了全面性能提升。

Meta 则于 9 月推出了 Llama 3.2 AI 模型，并于次月发布了这两个模型的量化版本，使其真正能用得到的场景快速增加。

谷歌的 Gemini 也在母公司的支持下快速迭代，并应用于生产场景中——据谷歌首席执行官桑达尔·皮查伊在最新季度财报电话会议上发言，Gemini API 的使用量在过去六个月中飙升了 14 倍，展现出了强势的发展势头。

在国内，以阿里、百度、字节、腾讯为首的大厂和新兴的“六小虎”们，也同样进行着快速迭代，譬如 Kimi 的 o1，就很好地对标了 GPT-o1，在深度思考方面取得了长



足的进步。

随之而来的，就是 OpenAI 市占率的降低。

据 Menlo Ventures 的调研数据，相比于 2023 年，在闭源模型中，OpenAI 的先发优势有所削弱，企业市场份额从 50% 降至 34%。而其丢失掉的市场份额，则主要被 Anthropic 和谷歌拿下。

要具体到案例上，最具有标志性的事件，一定是 Sora 画出的大饼。

2月，OpenAI 发布了首个文生视频模型 Sora。据介绍，Sora 可以直接输出长达 60 秒的视频，并且包含高度细致的背景、复杂的多角度镜头，以及富有情感的多个角色。从当时放出的 demo 来看，无论是精细的画质，还是长达一分钟的输出时间，都以一种降维打击的形态，锤爆了当时甚至截至目前来说所有的文生视频大模型产品。

也就在 Sora 发布 demo 到正式面世的大半年中，众多对手奋起直追，只不过到了发布前的最后一刻还是没有哪家的文图生视频能做到 Sora 放出的 demo 那般丝滑。

只不过，在 Sora 正式开放之后，大家却发现 Sora 的 demo 真的只是 demo，具体生成的内容同已有的文图生视频应用相差不大。一名 OpenAI 员工在直播中建议道，“如果你期望点击一个按钮就让 Sora 生成一部故事片，我认为这种期望是错误的。”

价格方面，想要订阅可以无限生成 1080p 视频的 pro 模式居然高达 200 刀，这样的定价对于那些只需要 Sora 的用户来说，真的会让人眼前一黑。

之后，谷歌迭代推出的 Veo2，也被人认为是超越 Sora 的文图生

视频产品；而国内的 Vidu、可灵、即梦、海螺、清影视频，也有着不输 Sora 太多的表现。毕竟，既然都是抽卡，谁不愿意抽个价格更低效果还差不多的卡呢？

而这，也是大模型竞争中的冰山一角，或许在不久后的将来，OpenAI 都只会以相对领先的“领军人”角色存在，更多的 AI 大厂会在更多的 AI 赛道中，你追我赶。

第二部分：应用端 Chatbot、AI 陪伴、图片编辑霸榜，BC 两端商业化渐清晰

随着各大人工智能企业在模型端的持续发力与成熟，作为大模型能力想要落地用户的媒介，应用端也在 2024 年迎来了最好的发展时期。

据 Sensor Tower 发布的《2024 年 AI 应用市场洞察》，1-8 月全球 AI 应用下载量同比增长 26% 至 22 亿次，内购收入激增 51% 至 20 亿美元，预计全年下载量将突破 33 亿次，收入达到 33 亿美元。

具体应用方面，ChatGPT 在收入榜、下载榜与活跃用户榜均居首位，而在下载榜 Top30 中，几乎都是 Chatbot 应用、图片编辑应用与 AI 陪伴应用。

其中，Chatbot 应用与图片编辑应用主要是作为工具存在，而以 Character.AI 与 Talkie AI 为代表的陪伴式 AI 产品的用户黏性更高，2024 年 Character.AI 平均每日会话次数高达 25 次，平均每次会话接近 4 分钟，用户每日使用时长超过 1.5 小时，远远高于 ChatGPT。

此外，在该榜单里，国内 AI 产品豆包位列下载榜第五，但据 AI 产品榜的最新数据，豆包 11 月的

MAU 已经来到了 59.98M，仅次于 ChatGPT。

除了这些面向大众的、成熟的应用之外，医疗机器人、生物医学影像、数字疗法等全新技术已经应用于尖端治疗中；教育领域中个性化学习、智能教学助手、沉浸式学习体验等技术也已经并不罕见；能源领域里，具体负责调配管理与优化、故障预测与维护的系统也已经投入日常使用……

可以说，2024 年的人工智能，正在将触角伸向几乎所有的传统或非传统行业中，并且已经有越来越多细分领域中的专业应用从中涌现。

而随着 AI 应用渗透率的提升，相关商业化也随之起步。

在 C 端市场里，AI 应用大多都采用了订阅制的变现方式。其中，除了 OpenAI 推出的 Plus 会员收费高达 200 美元之外，Pro 会员的收费价格同 Anthropic 的 Claude Pro、Google 的 Gemini Advanced 几乎完全一致，分别为 20、20、19.99 美元。

而国内大模型产品的收费方式也几乎一致。

譬如，目前国内 Chatbot 主流产品中唯一推出付费会员模式的文心一言，就推出了 49.9 元的连续订阅包月价格，开通会员后可使用最新的文心大模型 4.0 与 4.0 Turbo；而在文图生视频领域中，包括即梦、可灵在内的大多数 AI 创作应用，也同样采用了月度、季度会员兑换积分，积分使用服务的链路。

B 端市场中，由于有了云服务的“前车之鉴”，按 Tokens/API 接口调用次数收费是相对已经成熟、也是大多数大模型厂商都会选择的收费方式。而在这些 AI 大厂之间，大模型的价格战早已打响。

从今年 5 月开始，中国云厂商就开始了大模型推理算力价格战。包括字节、阿里、百度在内的老牌云厂商，已经把大模型推理算力价格打了骨折，每 token 的价格几乎每隔一段时间都要再创一次新低。

而在未来，B 端的 AI 商业化，也大概会像云时代的竞争一般，随着基础模型的优化与算力芯片的迭代，逐渐变为 AI 时代的又一场“卷王”竞争，但如何在这场不得不打的价格战中，通过提升企业客户最关心的模型能力、效果杀出重围，才是 AI 大厂们最需要关心的核心命题。

Market Share of LLMs: 2023 vs. 2024

