

中金-AI 十年展望（五）：从 ChatGPT 到通用智能，新长征上的新变化

原创 赵丽萍 于钟海等 中金点睛 2023-02-03 07:51 发表于北京



2012 年深度学习元年以来，AI 进入学术和商业界发展新阶段，2017 年大模型路线以“通用智能”思路降低算法边际成本，逐渐成为学界与产业共识。**2022 年下半年，AIGC 及 ChatGPT 关注度大幅提升背后，我们认为其本质是弱人工智能到强人工智能的阶跃，海外及国内商业落地处在初期，但我们认为新的产业趋势值得关注。OpenAI 的技术进展及投资方向、海外 AI 独角兽业务进展、国内龙头 AI 公司的跨模态布局是重要风向标。**

本文作者：赵丽萍，于钟海，魏鹤霏

摘要

技术层面，ChatGPT 和 AIGC 的持续升温依托于大模型的技术红利。从 2012 年的深度学习元年以来，各界产生了海量的数字化需求，大模型技术路线是降低边际成本的核心，2017 年大模型（Transformer）路线逐渐成为学术界与国内外巨头的发展共识。ChatGPT 由 GPT-3.5 大模型加入基于人类反馈的强化学习训练而成，带来弱人工智能向通用智能的阶跃。

商业层面，以 OpenAI 投资方向为风向标，落地仍在探索期，跨模态打开场景空间。对于 AIGC 领域，AI 作画等跨模态应用是未来的发展趋势，全球范围内

免责声明：

1. 本附加与原报告无关；
2. 本资料来源互联网公开数据；
3. 本资料在“行业报告资源群”和“知识星球 行业与管理资源”均免费获取；
4. 本资料仅限社群内部学习，如需它用请联系版权方

合作与沟通，
请联系客服



客服微信



客服微信

行业报告资源群

1. 进群即领福利《报告与资源合编》，内有近百行业、万余份行研、管理及其他学习资源免费下载；
2. 每日分享学习最新6+份精选行研资料；
3. 群友咨询，群主免费提供相关行业报告。



微信扫码，长期有效

知识星球 行业与管理资源

知识星球 行业与管理资源 是投资、产业研究、运营管理、价值传播等专业知识库，已成为产业生态圈、企业经营者及数据研究者的智慧工具。

知识星球 行业与管理资源 每月更新5000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；

微信扫码加入后无限制搜索下载。



微信扫码，行研无忧

AIGC 独角兽已初步具备 B 端为主的变现能力，海外落地节奏显著快于国内。ChatGPT 推出后，海外已有早期合作案例，例如为 BuzzFeed 提供个性化测试、为 Amazon 解决客户和工程师技术难题等。此外，以 OpenAI 前沿投资方向为锚，我们认为文本到图片/视频的跨模态生成、垂直领域 AI 写作、智能笔记、AI 语言学习平台或为潜在落地方向。

展望未来，行业格局有望走向底层集中、垂类多点开花，MaaS 是商业模式演进的潜在方向。我们认为，ChatGPT 为代表的模式背后，成本、算力、场景、数据等多维度需求铸就高门槛，大模型路线下，未来行业格局趋向集中，泛化通用大模型能力的厂商有望呈现多强格局；关键垂类仍可能有场景、数据优势，精细调优后的垂类模型仍有差异化竞争潜力。展望未来的商业模式，我们认为以海外 Hugging Face 为代表的 Model-as-a-Service 是潜在方向，具备大模型基础的龙头有望成为 AI 开源基础设施提供商、社区生态建设者。

风险

技术进展不及预期，行业竞争加剧，商业化落地节奏不及预期。

正文

技术篇：ChatGPT 和 AIGC 持续升温背后，本质上是大模型的技术红利

ChatGPT 和 AIGC 持续升温的本质是背后的 AI 大模型进入新范式。ChatGPT 是基于 GPT-3.5 的 InstructGPT，基于人类反馈的强化学习（RLHF）训练的语言类大模型。本文梳理了 ChatGPT 背后的技术演进与应用落地节奏，并对未来 AI 行业的格局加以展望。

从 2012 年的深度学习元年开始，上一轮红利已经持续 10 年

2012 年是深度学习元年，以 Hinton 团队在 ImageNet 大赛中引入深度学习算法为重要节点。Hinton 带领的团队在 ImageNet 挑战赛中，通过构建深度神经网络 AlexNet 成功将图片识别错误率降低了 10.8pcts，以大幅超越第二名的成绩夺冠，证明了深度学习的发展潜力，开始被业界和资本关注。

图表 1：AI 从引起业界广泛关注到目前不足 10 年



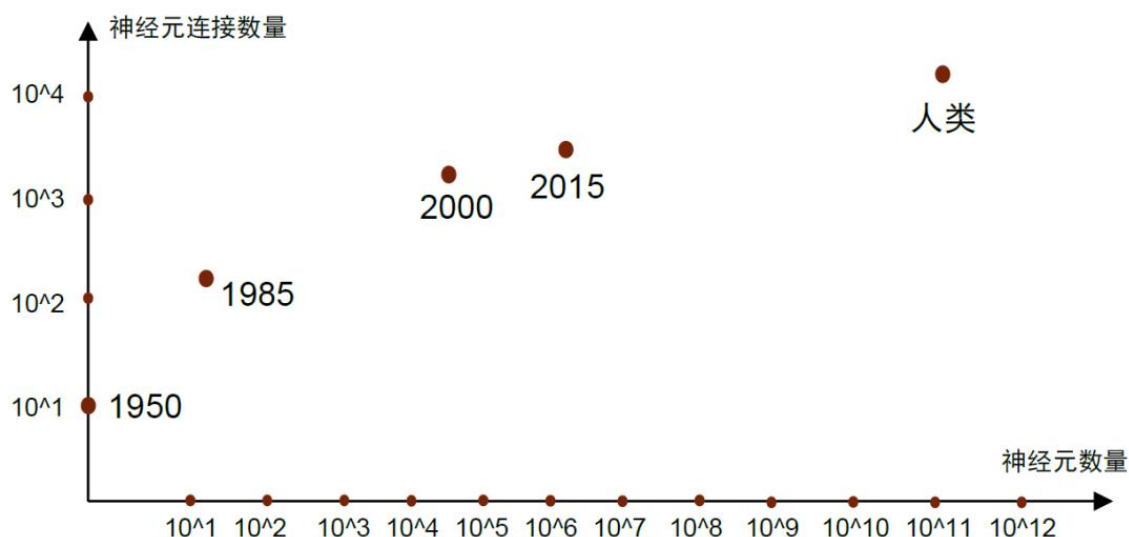
资料来源：量子位，Omdia，中金公司研究部

早年业界一度认为小模型、小算力是方向。此前的共识是通过精妙的算法和更高的模型精度，减少对算力的需求，但行业落地实践证明其短板明显：

► **精妙模型路线导致重复研发，无法让 AI 赋能千行百业。**在 AI 公司进入大量新场景时，对每个场景都重复研发大大增加了边际成本。

► **精妙模型路线无法解决长尾场景，完善 AI 商业化价值闭环。**以份额占人工智能行业近一半份额的智慧城市为例，其中的长尾场景如防火防灾、电梯事故、垃圾乱扔由于研发难度高，大部分 AI 公司难以低成本解决，严重影响到客户对整套解决方案买单的意愿。

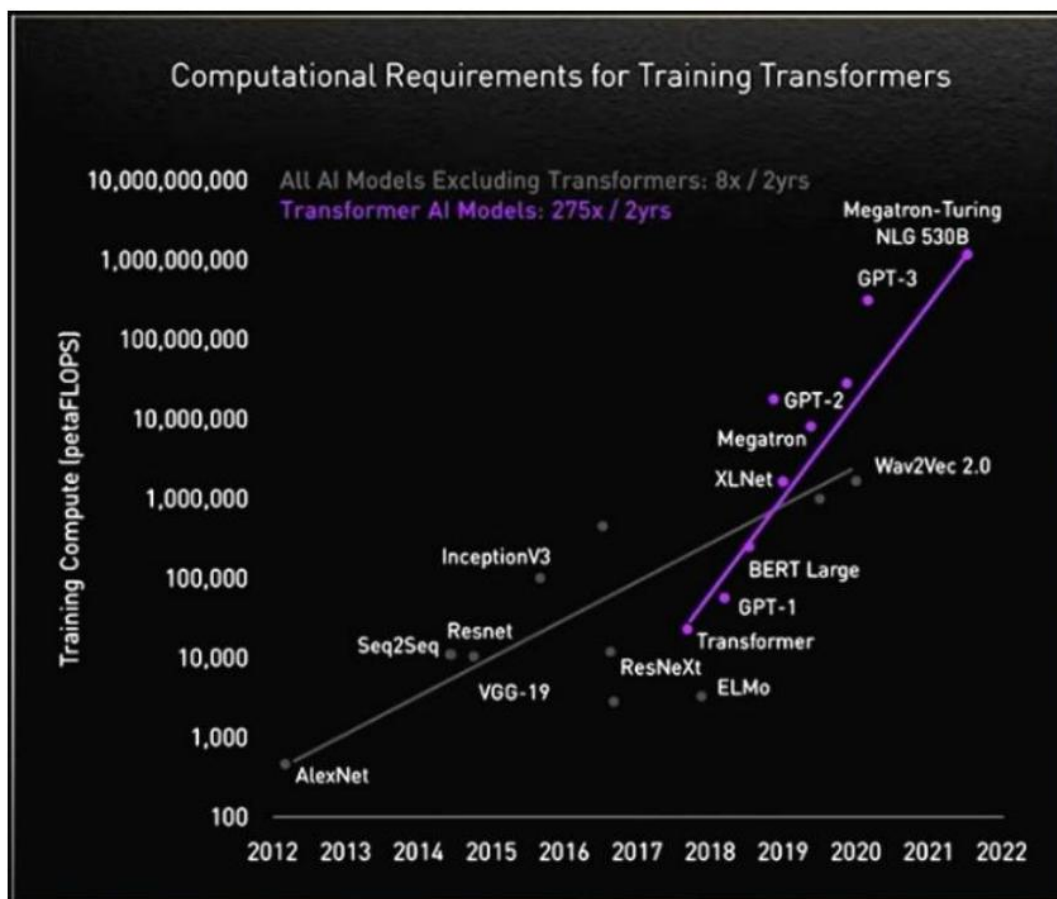
图表 2：目前 AI 模型的神经元总量与人类差距较大



资料来源：阿里云《中国企业 2020：人工智能应用实践与趋势》（2019 年），中金公司研究部

技术演进视角，大模型是实现通用 AI 的重要方向。虽然距离完全达到人类智能水平的 AI，还有很长一段路要走。但近几年在长尾场景等问题导致了对更通用的人工智能的刚需，在国内外巨头纷纷投入大量资源攻克通用人工智能难题的推动下，通用的语言模型、视觉模型甚至多模态模型也开始逐渐取得突破。2020 年 OpenAI 推出 1750 亿参数的 GPT-3 大模型是行业重要里程碑，为后续大模型迭代、加入人为干预，以及 AIGC、ChatGPT 的推出打下技术基础。

图表 3：大规模预训练模型对于算力需求呈指数增长

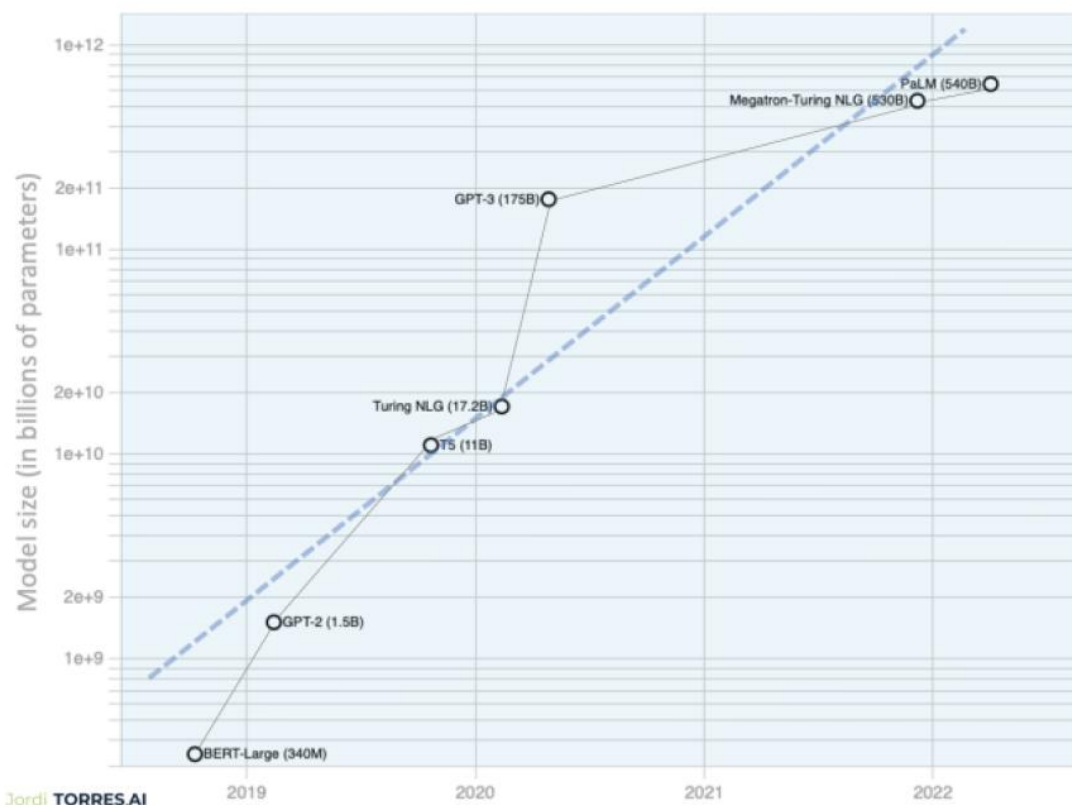


资料来源：英伟达官网，中金公司研究部

大模型将 AI 从感知提升到理解的维度

过去的深度学习在 CV（计算机视觉）领域大放光彩，但本质上还是更多的解决感知问题。CNN（卷积神经网络）已经能够较好的解决常规的人脸识别、物体识别问题，但在 NLP（自然语言处理）领域，对于物体、概念的抽象理解、语义的识别和理解，乃至推理和逻辑仍然是巨大的挑战。Transformer 算法（大模型路线）在 NLP 领域成果卓越，催化 NLP 的能力从简单的感知，向深层次的理解乃至推理不断发展。

图表 4：模型的参数大小在不断提升，性能也在与时俱进



资料来源: Jordi Torres.AI, 中金公司研究部

图表 5: AI 的能力也在从感知向理解、推理、生成攀爬



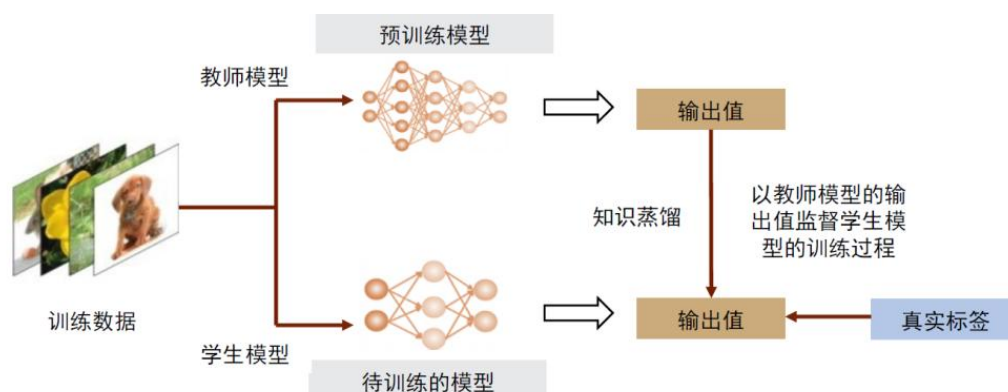
资料来源：中金公司研究部

如何简单的理解大模型（大规模预训练模型）

基于具备一定通用性的大模型，通过少量的增量训练蒸馏出小模型，是解决长尾问题的关键技术架构。从模型训练到部署，需要通过剪枝、量化、蒸馏等模型压缩技术实现更高的经济性及快速推理。以蒸馏为例，可以将结构复杂、参数规模庞大的大模型，压缩成结构简单、易于部署的小模型，相比于直接生产的小模型，大模型蒸馏出的小模型具有更强的泛化能力，边际成本大幅降低。

大模型+小模型的方式能有效降低 AI 落地边际成本。由于避免了“手工作坊”式的 AI 生产方式，不需要每做一个项目就派出大量专家花数月驻场收集数据、调试模型、训练模型，并且对长尾场景的解决为客户创造了更多价值。用足够多的数据和足够大的算法去训练一个足够大的通用模型，再通过量化、剪枝、知识蒸馏等模型压缩方法把大模型变小，高效的进行模型生产，并且由于算法足够多，能够覆盖各种长尾场景，大大降低了复制成本。因此我们认为，大模型是未来 AI 行业的必然趋势。

图表 6：蒸馏技术是类似于老师-学生传递知识的过程



资料来源：ICCV2019，华为云，中金公司研究部

2020 年超大模型 GPT-3 发布，衍生出 DALL·E 与 CLIP 模型。随着 GPT-3 在多个自然语言处理基准上展现超强性能，该模型延伸至图像领域，衍生出匹配文字和图像的 CLIP 模型，以及由文字生成图像的 DALL·E 模型。2022 年 4 月，OpenAI 推出 DALL·E 2，基于 CLIP 和 Diffusion 模型，分辨率提升了 4 倍，准确率更高，业务更广：除了生成图像，还能二次创作。

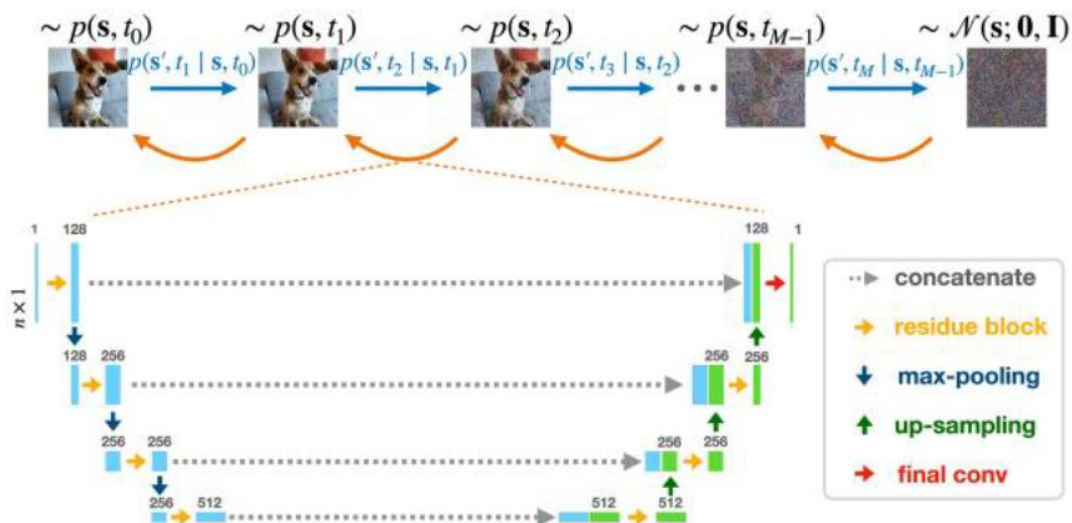
图表 7：GPT-3 衍生出 DALL·E 与 CLIP 模型



资料来源：OpenAI 官网，中金公司研究部

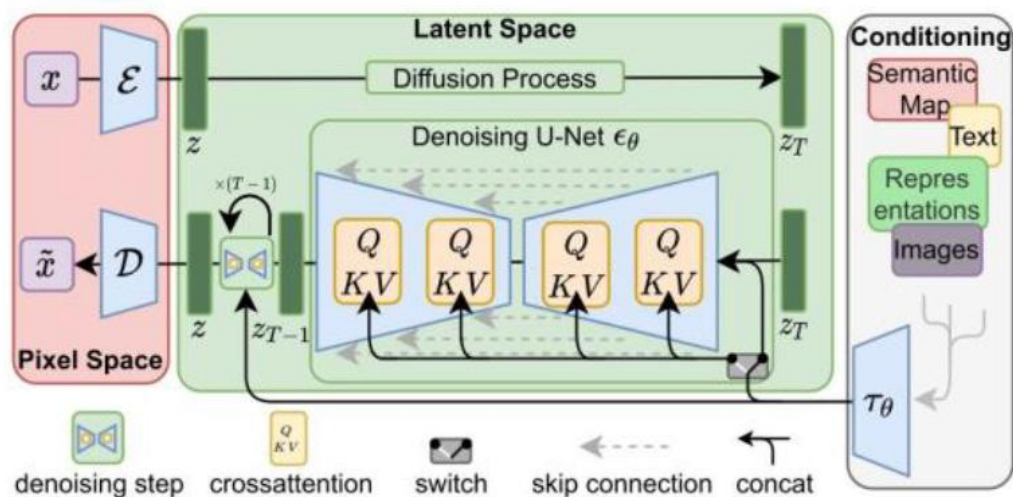
Stable Diffusion 由 Stability AI 公司提出并于 2022 年 7 月开源，是 AIGC 热度提升的重要驱动力。Stable Diffusion 的重要贡献在于使用空间降维解决内存和模型推理时长痛点，采用 Latent Diffusion Model (LDM) 方式，在模型复杂度和图像质量之间达到平衡。Stable Diffusion 不仅使用户仅在消费级显卡上就能够快速生成高分辨率、高清晰度图像，而且建立开源生态，大大降低用户的使用门槛。

图表 8：Diffusion 模型原理示意图



资料来源：Stability.AI，中金公司研究部

图表 9：Stable Diffusion 模型原理示意图



资料来源：CVPR 2022 《High-Resolution Image Synthesis with Latent Diffusion Models》，中金公司研究部

ChatGPT：基于 GPT-3.5 版本的 InstructGPT，开放接口，风靡全球

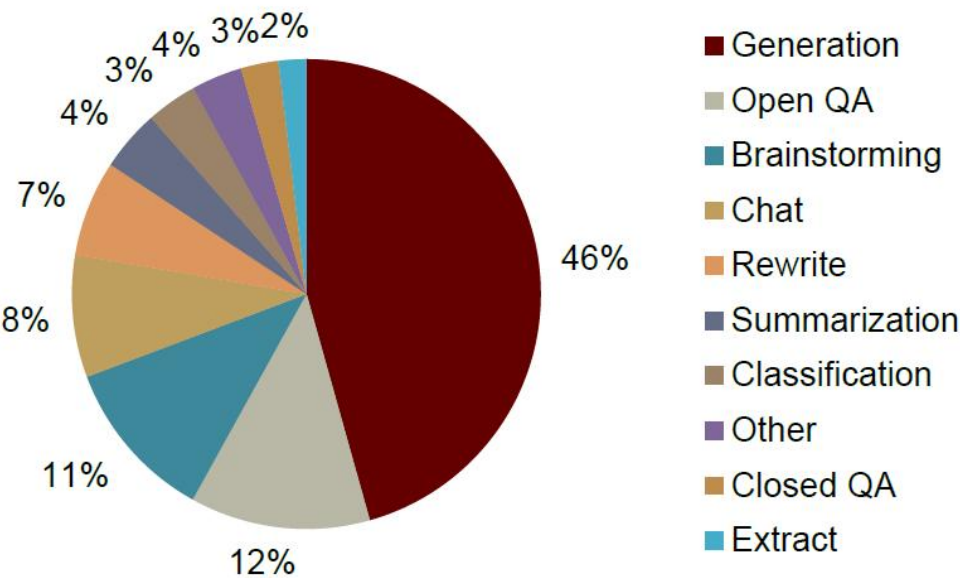
2022 年 11 月 30 日，OpenAI 发布了语言大模型 ChatGPT，通过大规模预训练(GPT-3.5)和自然语言生成技术实现多轮对话问答。该模型以对话的形式与用户进行自然交互，核心方法是基于人类反馈的强化学习(RLHF)，能够实现“回答后续问题、承认错误、质疑不正确的前提和拒绝不适当的请求”的能力。

图表 10: ChatGPT 拥有智能问答、多轮对话的能力



资料来源: ChatGPT, 中金公司研究部

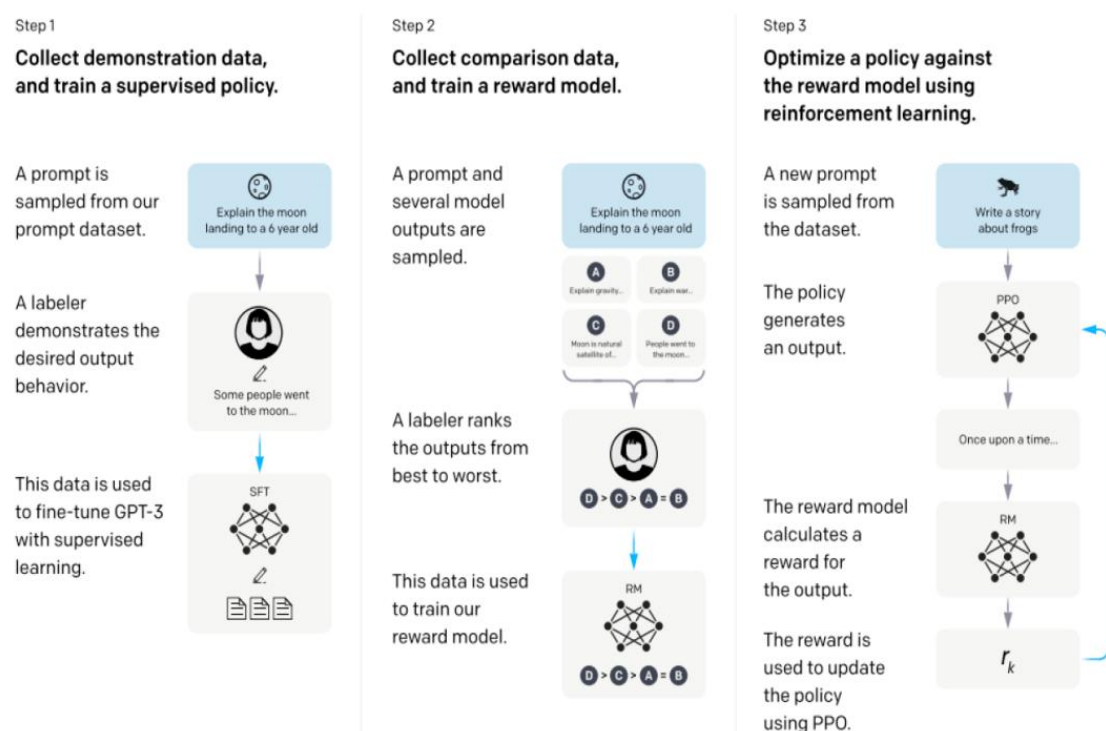
图表 11: InstructGPT 公开 API 后用户的使用功能和频率



资料来源: 《Training language models to follow instructions with human feedback》 (OpenAI,2022) , 中金公司研究部

ChatGPT 基于 InstructGPT 模型思路，在 GPT-3.5 基础上进行微调，在数据源选择与数据标注领域进行优化。ChatGPT 仅仅在训练数据和微调模型 (GPT-3.5) 方面与 InstructGPT 不同，其余的训练流程和方法二者相同。相较于 GPT-3 的训练集只有文本，ChatGPT 的训练集基于文本和代码，新增代码理解和生成的能力。目前，OpenAI 仍未公布 ChatGPT 的论文，但根据其主页可以得知它的方法与 InstructGPT 相同，本文从 InstructGPT 展开技术路径：

图表 12：InstructGPT 主要训练流程分为三步，搭建有监督 SFT 模型、强化学习模型



资料来源：《Training language models to follow instructions with human feedback》（OpenAI,2022），中金公司研究部

► 第一步：OpenAI 采集了基于 prompt（即大量的提示文本）训练方式的数据集，提示训练模型“问题在此开始”，并详细说明了下游任务。将此数据集交给约 40 人的专业团队标注人类希望得到的回答，并以该训练集进行监督学习微调

GPT-3（参数数量为 175B），得到微调后的模型 SFT。（ChatGPT 微调的模型是未公开的 GPT-3.5，比 GPT-3 参数量级更大）

- ▶ 第二步：将第一步中的数据和 GPT-3 训练集的一部分数据混合，使用微调模型 SFT 进行预测，对每个输入得到 N 个结果，此时专业人员会根据选项的好坏对其进行排序，并把排序后的数据用来训练奖励模型(RM)。
- ▶ 第三步：使用 PPO 的强化学习方法更新参数，使用 SFT 再预测一次数据集的结果通过第二步的奖励模型进行打分，计算奖励(reward)。最后将奖励分数通过 PPO 返回 SFT 进行训练。

大模型不仅对应于 ChatGPT，更是整个 AIGC 领域的技术基础

图表 13：AIGC 关键技术突破时间轴所示，GAN 带来生成模型雏形，GPT 等 NLP 大模型是通用智能的技术基础



资料来源：OpenAI 官网，《Denoising Diffusion Probabilistic Models》(Ho et al. 2020)，中金公司研究部

生成对抗网络 GAN 不断演进，助力 AI 生成图像逐步完善，为 AIGC 兴起的原始思路。 GAN (Generative Adversarial Networks，生成对抗网络) 于 2014 年提出，主要原理为，将两个神经网络（生成网络和判别网络）相互对立，以提高模型输出结果的质量。通过 GAN，计算机可以根据输入的文字描述、图像分割、草图、风格等生成实际不存在的图像，也可以将已有图片根据输入目标转化，

肉眼几乎无法分辨真假。AI 能够将输入的简易指令转化为图像等复杂生成结果，具备支撑 AIGC 的技术条件。

GAN 是 AIGC 发展的基础框架，但技术方面仍有多处不足。在 AIGC 由学术界实验室向应用导向的过程中，生产者对于图像生成的质量、分辨率提出更高的要求，GAN 的不足之处体现在：一方面容易生成随机图像，对结果的掌控力较差，另一方面是图像分辨率较低，能够支撑宏伟图景、难以满足细节要求，此外，生成过程中依赖对原有图像数据的模仿，缺乏创新性。基于以上不足，AIGC 发展在学术界、应用界均遇发展瓶颈。

Transformer 大模型的计算并行性奠定了网络规模数据集训练的模型中的最高地位。2017 年谷歌提出 Transformer 模型，其并行化的语言处理方式使得并行计算的效率得到大幅的提升。基于 Transformer 模型，OpenAI 于 2018 年首次提出了 GPT 模型，能够在无监督的情况下进行训练，在大语料场景下能够显著的改善模型效果，正式将 NLP（自然语言处理）带入预训练时代。

CLIP 模型（Contrastive Language-image Pre-training）由 OpenAI 提出并在 2021 年开源，图像-文本对应能力为 AIGC 提供落地基础。CLIP 是以文本为监督信号来训练可迁移的视觉模型，是多模态领域的重要推动力，主要价值在于具备强大的特征提取能力，基于互联网搜集的数十亿张图片及文本信息进行训练，能够实现将任意图像和文本信息配对，为 AIGC 的主流应用由文本生成图片和视频奠定多模态应用基础。

Diffusion 扩散模型在 2022 年以多维度技术优势进一步推动 AIGC 应用。2015 年，Diffusion 模型最早在 ICML 的论文《Deep Unsupervised Learning using Nonequilibrium Thermodynamics》被提出，2020 年由 Jonathan Ho 等人

在《Denoising Diffusion Probabilistic Models》中提出的 DDPM 模型引起了学术界更多关注。模型本质分为前向扩散、反向生成两阶段，分别完成对图像逐步添加高斯噪声-随机噪声、去噪声的过程，相较 GAN 模型在数据量需求上更小，在生成任务效果上大幅超越 GAN 等传统模型。

DALL·E 及升级版 DALL·E 2 项目基于 CLIP 和 Diffusion 大模型开发，AI 具备依据文字进行创作的能力，AIGC 作画领域落地进入快车道。 DALL·E 系统由微软注资的 OpenAI 于 2021 年 1 月推出，并于 2022 年 4 月更新至 DALL·E 2，该系统基于 CLIP 和 Diffusion 关键模型建立，具备三种功能：1) 根据文本提示生成图像，2) 以给定图像生成新图像，3) 以文本编辑图像元素。2022 年 7 月，DALL·E 2 开启邀请制公测，生成图像在关注宏伟场景的同时关注人物关系细节，是 AIGC 早期落地的重要标杆事件。

Stable Diffusion 于 2022 年 7 月建立完全开源生态，是 2022 年 AIGC 落地门槛降低、应用热度进一步提升的又一重要驱动力。 AIGC 在 2022 年已经具备了 CLIP 开源文本-图片多模态模型基础、LAION 开源数据库、Diffusion 大模型算法框架创新，Stable Diffusion 的重要贡献在于使用空间降维解决内存和模型推理时长痛点，以及 2022 年 7 月构建的完全开源生态。部署在国内的二次元创作 Novel AI 模型即是基于 Stable Diffusion 模型发展而来，作画方式更为多元，包括文本生成图像、原画改写、简笔画生成等模式，出图质量较高，深受二次元爱好者的认可。至此，开源生态推动 AIGC 的数据、模型与算力问题初步解决，直接降低了使用者的门槛，渗透进多个垂直领域。

展望未来，AI 技术逐步进入无监督学习时代，支撑 AIGC 的产业化发展。 2012 年以前，AI 模型大多为针对特定场景进行训练的小模型，无法拓展至通用场景，

且仍需要人工调参和大量数据来提升模型精度。随着技术发展，AI 可以在图像、文本等多维度上实现融合互补，在无监督情况下自动学习不同任务、并快速迁移到不同领域。例如，AI 驱动虚拟人可以利用现实人脸及声音等多维度数据生成形象，GAN、Diffusion 模型可以通过文字、图像等数据进行多模态创作。我们认为，AI 技术已呈现出能稳定支持内容生产的发展趋势，未来有望突破“小作坊”式生产，助力 AI 内容生产进入工业化时代。

商业篇：以 OpenAI 投资方向为风向标，落地仍在探索期，跨模态打开场景空间

AIGC：跨模态应用是未来发展趋势，落地尚在早期

在深度学习模型支撑下，早期 AIGC 在文本生成领域开启内容创作落地，逐渐向音频生成、图像生成等领域推广。深度学习带来 AI 在学术和应用落地领域的分水岭，大模型进一步将应用接近认知智能。2014 年起，AIGC 在文本理解、结构化协作领域小范围应用，按照特定模式提取情感语义，或按照框定模板生成结构化内容。在 2018 年 NLP 领域 BERT、GPT 系列大模型出现后，非结构化协作等高自由度创作具备落地空间。此外，AIGC 在音频-音频生成、图像-图像生成和视频-视频生成创作等领域跨越落地门槛，在单模态发展中呈现多点开花局面。

图表 14：AIGC 应用全景图一览

分类	具体分支	应用领域	代表公司案例
文本生成	文本理解	话题解析、文本情感分析	科大讯飞、阿里巴巴和微软亚洲研究院在文本理解挑战赛的完全匹配得分均超过人类得分
	结构化写作	新闻撰写	Automated Insights开发的Wordsmith可以生成评论文章
	非结构化写作	营销文案、剧情续写	Jasper平台为社交媒体、广告营销、博客等产出标题、文案、脚本、文章等
	交互性文本	客服、游戏	OpenAI与Latitude推出的游戏AI Dungeon，可根据输入的动作或对话生成个性化内容
音频生成	语音克隆	地图导航	百度地图可根据输入音频，生成专属导航语音
	语音机器人	客服、销售、培训	思必驰拥有外呼机器人、呼入机器人、陪练机器人等产品
	音乐生成	播客、电影、游戏	OpenAI的MuseNet可利用10种乐器共同生成4分钟音乐作品
图像生成	图像编辑与融合	设计、电影	谷歌的Deep Dream Generator可上传图像并选择风格，生成新图像
	2D图像生成3D模型	游戏、教育、产品测试	英伟达的GANverse3D可利用汽车照片生成3D模型，并在NVIDIA Omniverse中行驶
视频生成	画质增强修复	视频插帧、视频细节增强、老旧影像的修复与上色	当虹科技的画质增强修复技术帮助视频画质提升
	切换视频风格	电影风格转换、医学影像成像效果增强	腾讯天衍工作室在结直肠内镜项目中切换视频风格，优化医学影像视觉效果
	动态面部编辑	AI换脸	Akool的faceswap平台拍摄样本视频便可编辑、替换模特面部
	视频内容创作	制作电影预告片、赛事精彩回顾	IBM的Watson制作了20世纪福克斯的科幻电影《Morgan》的预告片
跨模态生成	文本生成图像	传媒、娱乐	OpenAI的DALL·E 2可通过输入文字生成高仿真图像
	文本生成视频	电影、短视频创作	Meta的Make-A-Video输入文本可生成数秒的视频
	图像/视频生成文本	搜索引擎、问答系统	谷歌的MUM模型支持多模态复杂信息搜索
	文本生成代码	Copilot	OpenAI的Codex模型可将自然语言翻译成代码

资料来源：量子位智库，中国信通院，Tom Mason，OpenAI 官网，中金公司研究部

DALL·E 2 是一个由 OpenAI 开发的人工智能模型，它具备三种功能：1) 以文本提示生成新图像；2) 以给定图像生成新图像；3) 以文本编辑图像元素。DALL·E 2 的使用方法很简单，用户只需在软件提示框中发送文字提示或图像修改文字，或选择给定垫图，AI 便会根据输入生成结果。用户可以随意输入想要表达的内容和所希望的艺术风格，DALL·E 2 可以**模仿许多知名艺术家的风格，并准确地描绘动物和人物，并表达他们之间的关系**（例如打架或聊天）。

DALL·E 2 目前采取付费购买次数的商业模式：加入 Open Beta 项目后，首月可以获得 50 个免费点数，每一个点数对应一次绘图，之后每个月可以免费补充

15 个，测试第一阶段的价格是 15 美元 115 个点数。花费点数生成图像后，用户将获得使用 DALL·E 创建的图像商业化的完全使用权，包括重印、销售和商品化的权利。

图表 15: DALL·E 2 通过文字添加图片元素

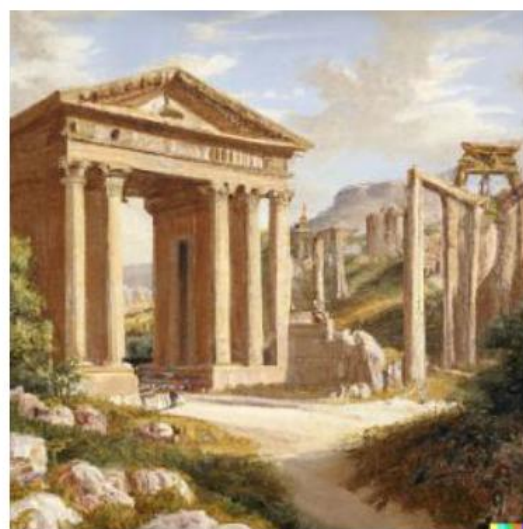


资料来源：DALL·E 2 官网，中金公司研究部

图表 16: DALL·E 2 通过文字输入生成图片结果示例



Prompt: A close-up view of a man meeting the devil, by Caravaggio



Prompt: A view of the gateway to the Great Temple at Baalbec, by Paolo Veronese

资料来源：DALL·E 2 官网，中金公司研究部

全球范围内 AIGC 独角兽已初步具备 B+C 端的变现能力，但海外公司明显领先于国内。AIGC 应用层可分为图像、文本、视频、音频，海外 AIGC 企业如 Jasper、Grammarly 等已经积累大批用户群体，年收入超千万美元；大部分企业都至少启动了 A 轮融资，融资金额相对较大。

图表 17：全球 AIGC 应用层企业梳理

应用层	公司名称	公司简介	融资阶段	金额
图像	 Midjourney	美国一款AI绘画工具，专注于设计与人工智能方向，目前拥有近600万用户和相当可观的收入。	未融资	-
	 stability.ai	英国的一家AI图像生成器研发商，主要推出一款免费且开源的文本到图像生成器。Stable Diffusion全渠道的用户数量每天超过1,000万	A轮	估值达10亿美元
文本	 copy.ai	一家美国自动营销副本解决方案提供商，目前已经有超过200 万的个人和团队在使用，包括微软、eBay、雀巢。	A轮	ARR超1,000万美元
	 Jasper	美国一家AI内容提供商，帮助撰写营销推广文案。2021年Jasper付费用户数累计突破七万，收入4,000万美元	A轮	估值达 15 亿美元
	 grammarly	Grammarly是一家语法检查器服务提供商，普通用户超过3,000万人、团队用户超3万，2021年收入7,200万美元。	C轮	估值达130亿美元
视频	 Hour One	Hour One是以色列一家AI合成视频提供商，为40多位客户提供服务，生成了超过 100,000 个视频。	A轮	融资2,000万美元
	 synthesia	一家英国基于人工智能的视频合成工具提供商，面向包括企业和与个人创作者在内的两个客户群体。	B轮	B轮融资5,000万美元
音频	 WELLSAID	WellSaid Labs 美国一家人工智能语音技术服务商。	A轮	融资1,000万美元
	 DEEPMUSIC.AI	灵动音科技 一家基于人工智能技术的音乐创作服务商。	战略投资	2018年完成数千万元A轮融资

资料来源：鲸准，中金公司研究部

注：统计时间截至 2023 年 1 月 29 日

而国内相关技术企业进展较慢，参与方以应用层为主。要提高国内 AIGC 独角兽发展速度，必须加快基础层建设，提高核心竞争力，同时找到更好的商业模式并具备变现能力。

图表 18：国内 AIGC 应用层企业梳理

应用层	公司名称	公司简介	融资阶段	金额
虚拟人	 小冰	小冰公司 一家跨平台人工智能机器人研发商，小冰框架目前已覆盖中国、日本、印度尼西亚等国6.6亿在线用户、10亿台第三方智能设备和9亿内容观众。	B轮	B轮融资10亿元
信息处理	 出门问问	出门问问 主要在信息处理和存储支持服务领域进行经营。	D轮	D轮融资数亿美元，前三轮融资金额达到7,160万美元
内容	 TIAMAT	TIAMAT 是一个生成式AI平台，为个人用户和企业客户提供创意生产、广告、数字藏品等场景形式下的服务，内测高活跃用户数千人、月度新增数万人。	天使轮	累计融资3,250万元
文本+图像	 TURING AI	图灵瞬息 一家AI文字赋能平台，基于人工智能技术，给文字赋能，让其变成漫画、视频	未融资	-
视频	 GliCloud	GliaCloud 一个AI视频制作平台，受到有影响力的出版商信赖，如抖音、优酷。	天使轮	融资50万美元
游戏	 独咕	Storyha AI 一家沉浸式互动AIGC游戏服务商。	未融资	-

资料来源：鲸准，中金公司研究部

注：统计时间截至 2023 年 1 月 29 日

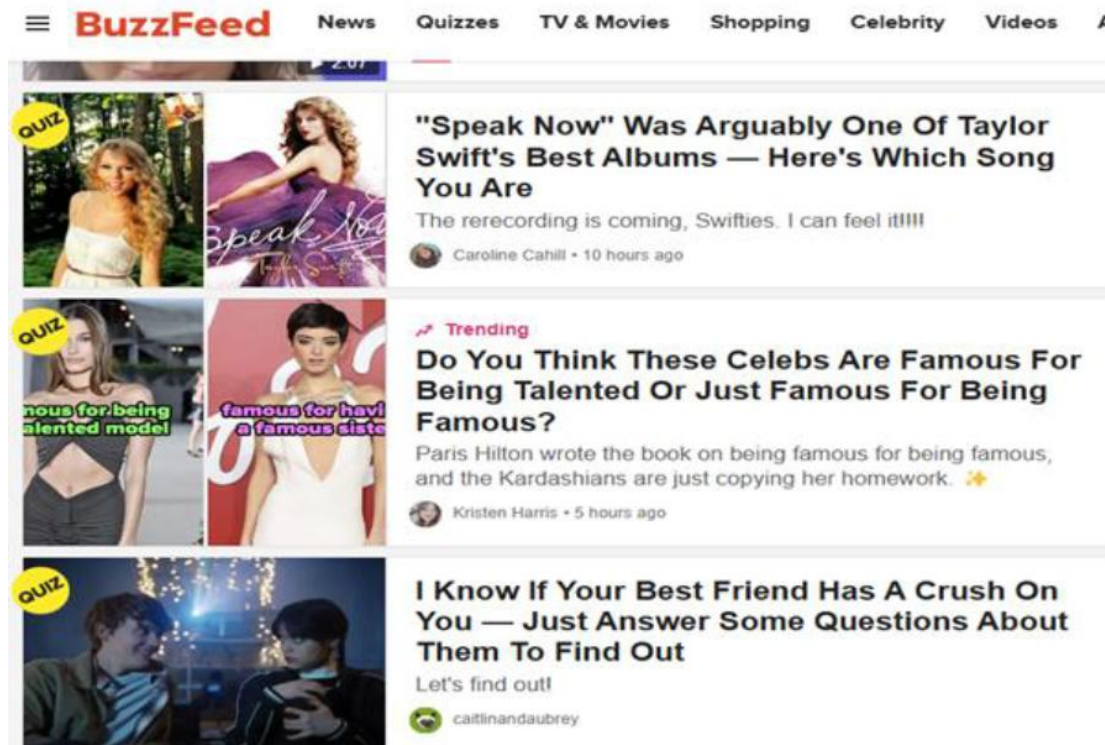
ChatGPT 应用场景广泛，海外已有龙头落地成功案例

ChatGPT 具备强大自然语言处理、多轮问答能力，可落地于广泛的应用场景。

ChatGPT 将在未来投入商业化领域，通过强化的上下文理解能力，优化银行、电商等客服机器人，生成商品描述和营销文案，以及协助翻译工作，大大提高服务质量。目前，ChatGPT 已经成功在海外市场得到了应用，为 BuzzFeed 提供了个性化测试，并帮助 Amazon 解决了客户和工程师面临的技术难题。

► **个性化测试:**2023 年 1 月,美国新闻聚合网站 BuzzFeed 官宣将依托 OpenAI 为其用户互动栏目 Quizzes 制作和个性化各种小测试。ChatGPT 会生成一系列提问，再根据个人的回答产生用户独有的、可分享的文章。不久前，Meta 向 BuzzFeed 支付了数百万美元，以让 BuzzFeed 为 Meta 的平台生成内容，并培训平台上的创作者。

图表 19：BuzzFeed 利用 ChatGPT 个性化 Quizzes 推送内容



资料来源：BuzzFeed 官网，中金公司研究部

► **嵌入微软全系列产品：**2023 年世界经济论坛期间，微软首席执行官萨蒂亚纳德拉（Satya Nadella）在接受《华尔街日报》采访上时表示：微软将扩大对 OpenAI 技术的访问并把 ChatGPT 嵌入到微软各大系列产品，例如 Bing、Office、Outlook 等。届时，ChatGPT 将在搜索引擎、日常办公、收发邮件中发挥其智能的语言理解和问题解答能力，使生产效率更上一层楼。

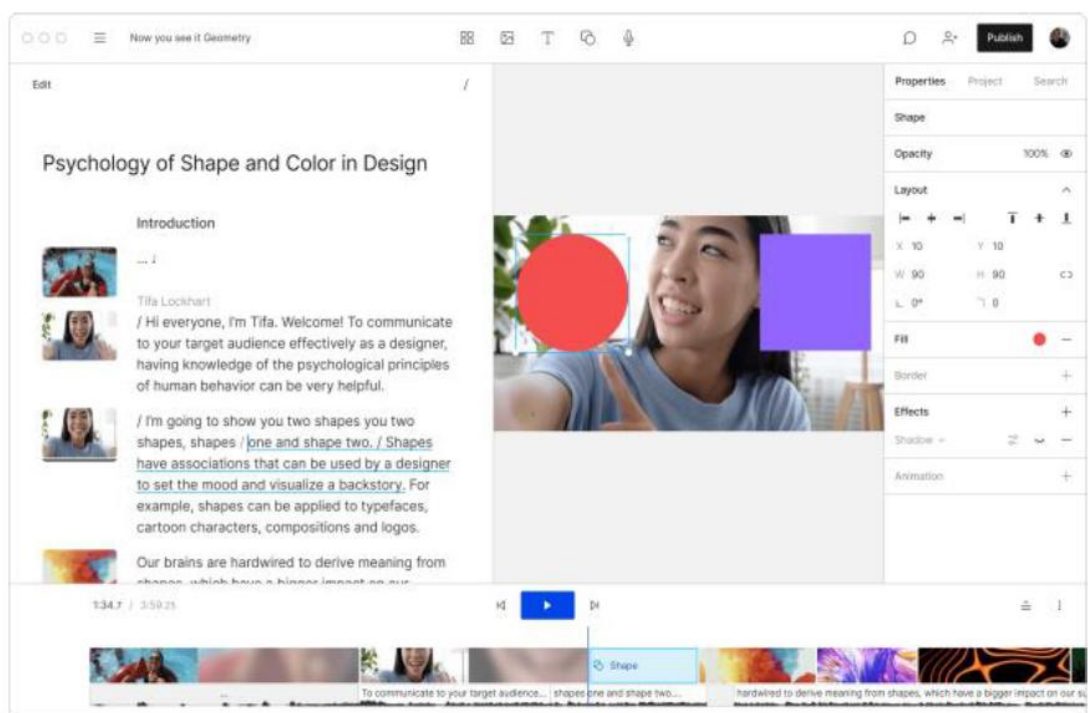
OpenAI 通过战略投资，产业赋能，加速 ChatGPT 在实际应用中落地。作为 AI 行业龙头 OpenAI，公司不止自身持续开发出惊动业界的 AI 模型如 GPT 系列模型，而且创立基金投资于 AI 创业公司，不断拓宽 AI 应用边界，重塑人类的创造力、法律、生产力、教育等方面能力，这四大领域也构成了 OpenAI 在 2022 年的四大投资风向标。

Descript：AI 驱动多媒体创新，重新定义音视频剪辑软件

音视频转文字，把繁琐的剪辑工作变得像编辑文档。 Descript 是一个像文档一样工作的编辑音视频的软件，包括转录文字、播客、屏幕录制等功能。其最大的特色功能是可以将音视频内容转录成文本，将文本中的每个词和媒体文件的时间戳对应，允许用户通过编辑这个转录的文本信息来编辑原始音视频。Descript 主要功能如下：

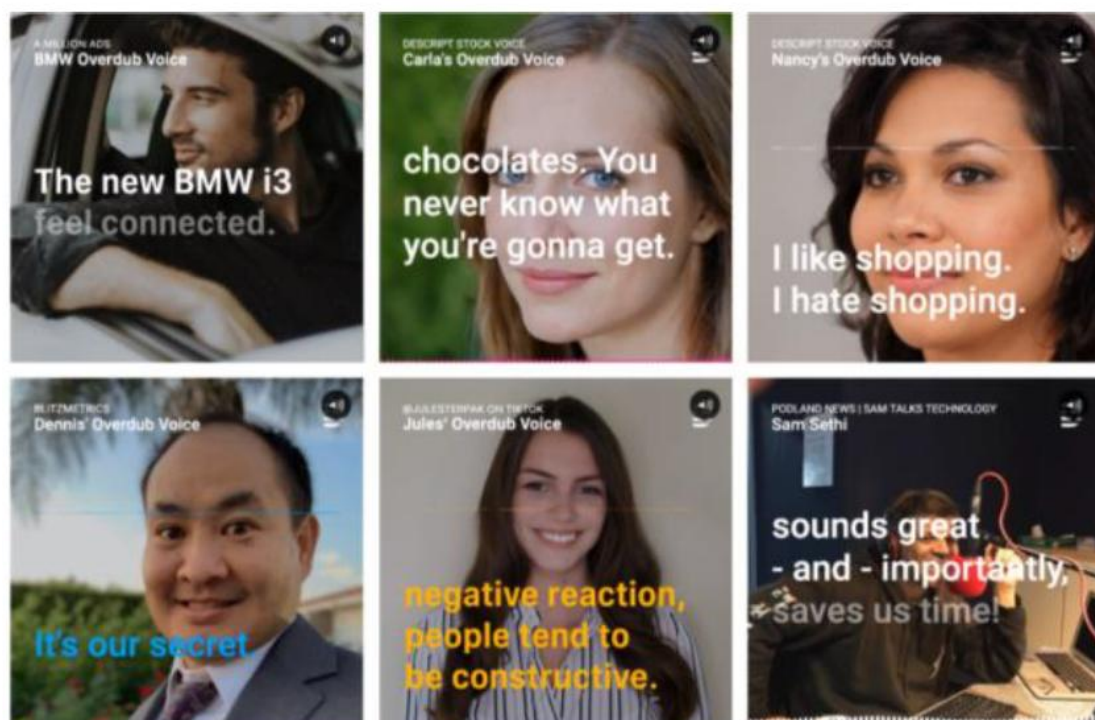
- ▶ **转录：**将音视频中的内容转录成文本，自动识别不同的讲话。按照分钟数计费，目前支持 22 种语言；也可使用人工转录。
- ▶ **屏幕录制：**支持快速录制、编辑和分享屏幕录制的视频，远程录制允许用户在使用 Zoom 等会议软件时调用 Descript。
- ▶ **音视频编辑：**1) 删除：当用户删除词句，音频中对应的词也会被删除；2) 生成(收购加拿大初创 Lyrebird 后推出 Overdub 功能)：当用户通过打字添加文本，AI 可以生成该用户的声音；3) 一键检测和删除语气词：例如 “um”、“uh” 等大量重复使用的单词。

图表 20：Descript 操作界面：使用脚本编辑录音，添加场景来排列视觉效果



资料来源：Descript 官网，中金公司研究部

图表 21：Overdub：创建文本到语音的转换模型，或选择超逼真的库存语音



资料来源：Descript 官网，中金公司研究部

Descript 定位于协同编辑工具，根据团队用户数和每月转录时长进行收费。免费版每月有 1 小时的转录时长；Creator 版本每月 10 小时转录时长，定价 12 美元；Pro 版本每月 30 小时转录时长，定价 24 美元。**Descript 目前已完成 4 轮融资，总融资额达 1 亿美元。**2022 年 11 月，OpenAI 领投 C 轮 5,000 万美元融资，估值达 5.5 亿美元。

Harvey: AI 协助律师撰写文件，提升法律工作流程效率

Harvey 有望作为法律的自然语言接口，大幅减少简单重复的法律流程工作。

Harvey 是面向律师群体的 AI 助手工具，为法律工作者提供统一的、直观的操作界面。同时借助强大的生成式语言模型（主要为 GPT-3）的能力，可以在律师的工作流程中帮助处理一系列繁杂的工作，如研究、起草、分析法律相关条文，使律师能够将更多时间花在工作中高价值部分。

Harvey 目前仍处于内部测试阶段，但已受到 OpenAI 的关注和青睐。2022 年 11 月，Harvey AI 获得 OpenAI 的 500 万美元投资，Google AI 负责人 Jeff Dean、Mixer Labs 联合创始人 Elad Gil 以及一批天使投资人跟投。**本轮融资后，Harvey 将正式向律师客户群体推出服务。**

Harvey 背靠 OpenAI 和微软，享受更多资源领先竞争对手。市场格局中，Casetext 同样利用人工智能技术进行案例搜索，进行法律研究、摘要起草等工作。据 OpenAI 表示，Harvey 将受益于与 OpenAI 的关系，能够提前访问微软的新 OpenAI 系统和 Azure 资源。

Mem: AI 学习用户工作、习惯，提供智能化笔记

Mem 作为一款轻量笔记工具，致力于打造“世界上第一款自动组织的笔记产品”，提高用户的工作效率与生产潜力。Mem 主打快速记录与内容搜索，利用人工智

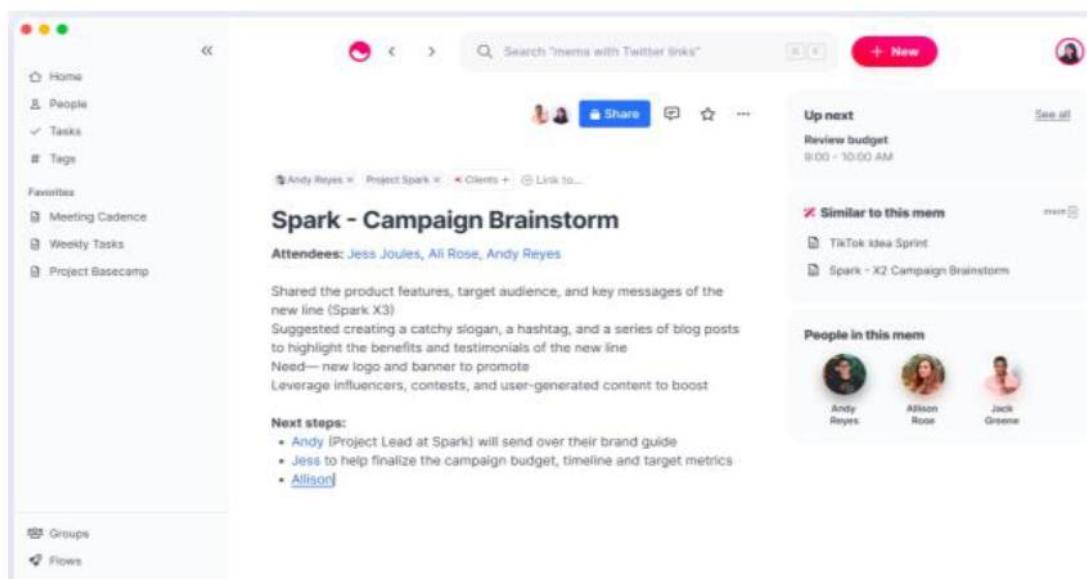
能技术提供预测、自动整理笔记内容，并且允许用户添加话题标签、标记其他用户、添加定期提醒等。

目前，Mem 为 Twitter 推出了 Mem It 应用，允许用户保存推文串，获得 AI 生成的内容摘要，并看到类似推文的建议。Mem 的内置工作助手 Mem X 也得到进一步完善，增添了智能写作和智能编辑等新功能，利用人工智能根据提示生成文本、总结文件、为文件生成标题，并让用户使用自然语言命令来编辑或格式化文本。

Mem 采用 Freemium（免费增值）的定价模式，从个人用户向团队和企业版进行拓展。付费版 Mem X 套餐价格是 10 美元/月，在免费版基础上增加了 AI 的能力，能够自动整理和归类信息；付费团队和企业版则增加更多存储空间、团队管理能力以及安全性等内容，团队版的定价是 15 美元/月，企业版需根据具体情况定价。在 OpenAI 的支持下，Mem 计划进一步开发人工智能驱动的功能和体验。2022 年 11 月，OpenAI 领投 A 轮 2,350 万美元融资，本轮融资中估值 1.1 亿美元，累计融资总额达 2,900 万美元。

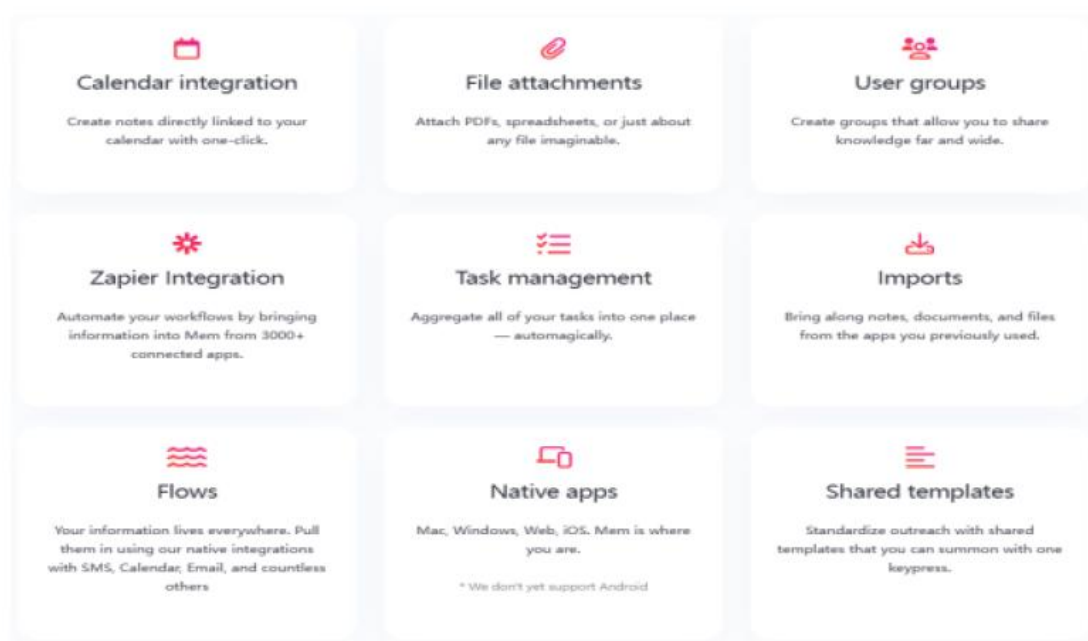
从 AI 个性化角度出发，Mem 与市场中常见的工作管理软件形成差异化竞争格局。工作协助平台 Glean 于 2022 年 5 月完成 1 亿美元的 C 轮融资，为 SaaS 企业提供一致的搜索体验，市值估值达 10 亿美元；Atlassian 类似维基的协作工作平台 Confluence 适合于远程操作的团队工作，能够有效的融合知识与协作。**Mem 致力于生成式知识管理方法，其个性化机器学习模型超越简单的搜索与记录。**

图表 22：Mem 操作界面：利用人工智能技术实时整理笔记



资料来源：Mem 官网，中金公司研究部

图表 23：Mem 主要功能：九大核心功能打造自动组织的笔记



资料来源：Mem 官网，中金公司研究部

Speak: AI 驱动语言学习，向学习者提供实时反馈

Speak 是具有人工智能功能的英语学习平台。 Speak 利用 AI 能力，开发了一系列对话内容，模拟了日常的各种场景和互动，通过和 AI 导师的互动来提高用户

口语。其最大的特点是 AI 导师可以基于对话内容，提供实时的反馈，包括发音、语法、词汇等。

Speak 主要通过提供自动续订的月度和年度订阅的模式盈利。Speak 在韩国有近 10 万付费用户，2022 年开设约 1,500 万节课，ARR 超千万美元。定价方面，Speak 按年度订阅约 100 美元/年，按月付 22 美元/月。

Speak 将训练新的语种，进一步拓展至韩国以外的市场。2022 年 11 月，OpenAI 领投 B 轮 2,700 万美元融资。下一步，公司计划拓展新的语种与市场（例如日本）并投资于利用文本生成模型的功能（例如 GPT-3），加速开发新的对话语言体验。

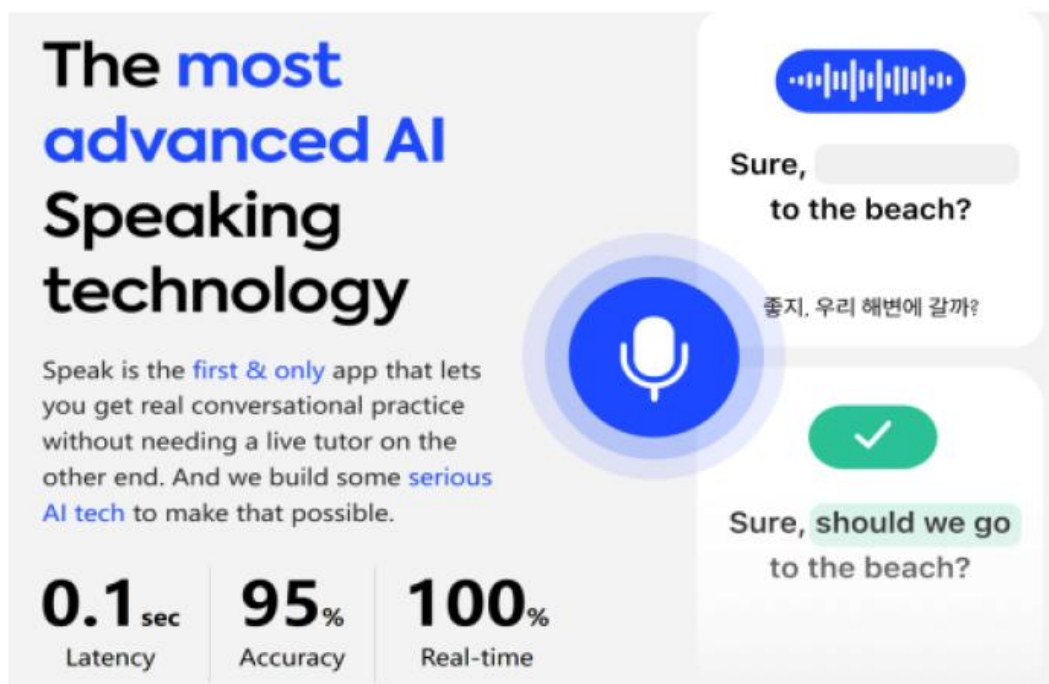
市场格局方面，Duolingo 是成熟的语言学习软件巨头。作为全球用户量最大的语言学习平台，Duolingo 提供超过 40 门语言课程，目前月度活跃用户 4,000 万，付费用户 190 万，ARR 达 3.6 亿美元。

图表 24: Speak: 提供真正的交互式口语学习体验产品



资料来源：Speak 官网，中金公司研究部

图表 25: 最先进的 AI 语音技术：0.1s 延迟，95%准确率，100%实时



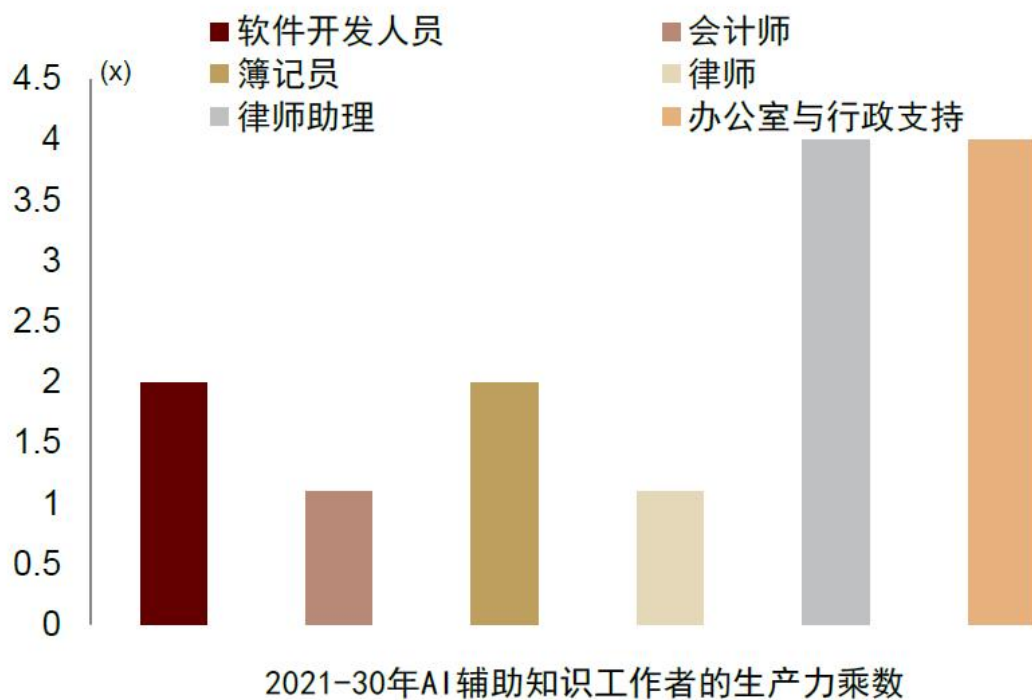
资料来源：Speak 官网，中金公司研究部

AI 还将不断增强人类，在更多低脑力领域实现效率倍增

AI 将人类从简单重复的任务中解放出来，专注于高价值、创新型的工作。通过解决密集型劳动、重复性劳动的痛点，AI 正在逐步取代部分人工工作，给社会带来降本增效的价值。在未来，随着 AI 技术的不断提高，它将在更多的低脑力领域提供支持，从而帮助人们实现更高效的工作。

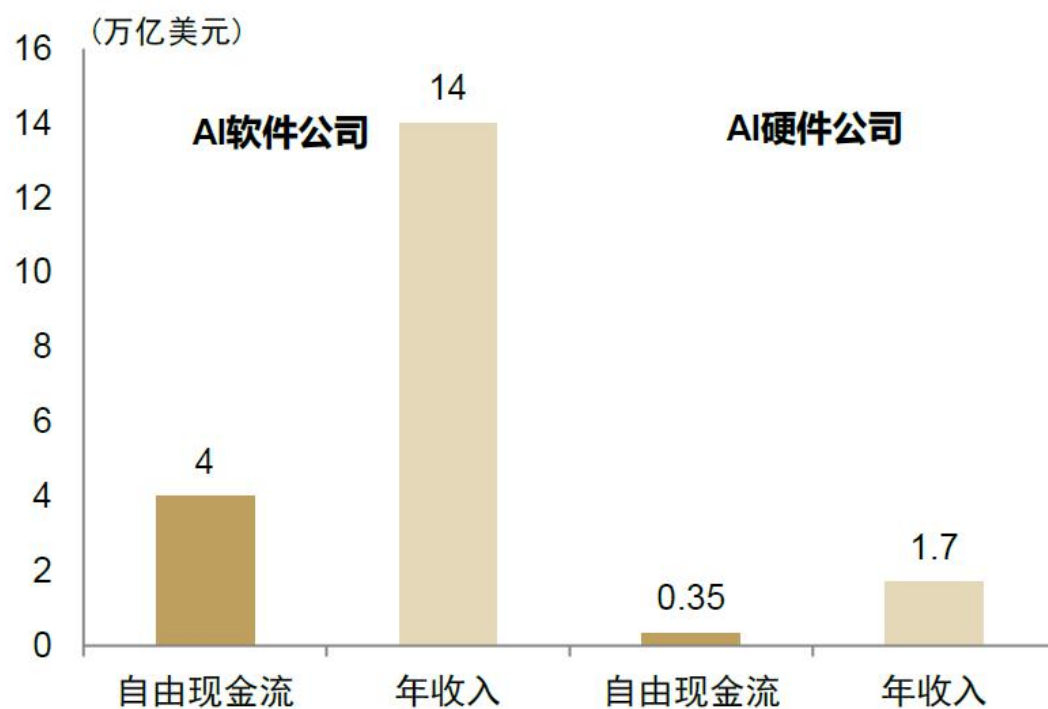
除了消除重复性任务，AI 同样起到简化决策与提供新洞察等作用，提升决策的速度和效率。据 Gartner 预测，截至 2023 年，超过 1/3 的大型企业将使用决策智能实现结构化决策，随着决策智能成为业务流程的核心部分，决策制定得比以前更快、更容易，而且成本更低。例如，在医疗领域，AI 技术已经成为支持临床诊断、疾病预测、个体化治疗等的重要工具；在金融领域，AI 技术支持了金融机构快速、高效的风险评估、客户服务等。

图表 26：AI 在执行环节、低级别决策环节具备替代人工的潜力



资料来源：ARK，中金公司研究部

图表 27：ARK 测算：2030 年，AI 软硬件公司年收入总和有望达 15.7 万亿美元



资料来源：ARK，中金公司研究部

格局和展望篇：行业格局有望走向底层集中、垂类多点开花，MaaS 是商业模式演进的潜在方向

算法边际成本是 AI 竞争核心焦点，碎片化需求使得成本居高不下

低边际成本是大模型的优势，碎片化的需求成为降低成本难的关键因素。目前国内大部分 AI 工作是以项目制的形式进行的，落地还停留在“手工作坊”阶段，存在重复造轮子情况，边际成本高。而且目前大部分场景数据规范性差、长尾，且采集方法落后，导致工作量较大，严重影响 AI 大模型进展。

碎片化的本质原因在于现阶段 AI 模型的通用性低，单个模型只适用特定任务。

例如，在工厂场景下检测零部件、在医疗图像中检测病理特征，虽然本质上都是检测，项目过程中都需要重新收集、标注数据、训练模型。由于客户需求多样，以至于几乎每个项目都要重复进行这一流程，研发流程难以复用，重度依赖人力，边际成本很高。

► **需求端：**全社会的数字化是人工智能的重要目标，同时也意味着大量的建模需求。随着数字信息世界、物理世界融合，产生的数据量是以前的成千上万倍，监控和管理这些信息牵制了大量人力，未来需要大量的 AI 模型来处理这些信息。而大量的模型需求需要较高的 AI 模型生产效率、较低的算法边际成本。

► **供给端：**AI 技术相对较新，引起业界关注仅 10 年，人才培养不足，导致人才稀缺。从事 AI 技术与研究开发的人员有限，难以满足市场大量模型需求。此外，技术快速发展，技术人员和训练数据集也需要不断学习更新。

图表 28：弱人工智能阶段长尾场景种类繁多导致项目碎片化



资料来源：量子位，中金公司研究部

AI 行业不会向碎片化的方向发展，而是强者愈强，用大模型结合底层全栈自研来解决边际成本问题。在大模型路线下，AI 模型的边际成本还会受益于三个因素的影响大幅降低：

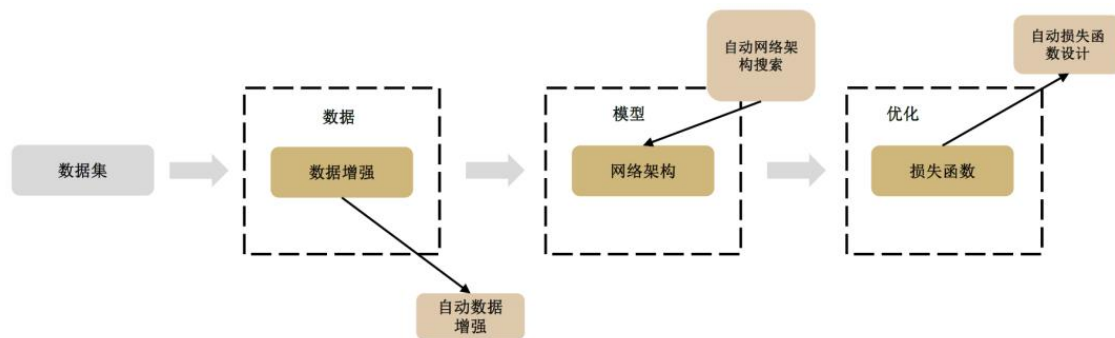
► **底层基础设施可复用，降低边际成本。**以商汤科技的 AI 大装置为例，公司多年累积的硬件、框架和 AI 算法和落地经验结合起来，一起融合到 AI 大装置，能尽可能的减少重复研发。

► **模型研发流程可复用，大模型提升 AI 通用性。**在大模型压缩制造小模型的工业化生产方式下，AI 公司可以生产大量的、覆盖不同场景的模型。这样，在遇到新场景的情形下，可以通过将原有的模型模块化组装，快速制造新模型，无需针对新场景再次定制化生产。

► **研发流程自动化（AutoML），开发门槛降低，人员成本降低。**AutoML 能自动执行 AI 流程中的大部分工程性任务，减少了 AI 模型生产过程中对人工的需

求量，且由于机器学习门槛的降低，不再需要招聘深度理解 AI 工程的专家，使得开发人员的成本降低。

图表 29: AutoML 在数据、模型和优化三个环节减少对深度学习专家的依赖



资料来源：CSDN，中金公司研究部

复刻 ChatGPT 为代表的模式较难，多维度需求铸就高门槛

ChatGPT 基于 GPT-3.5 微调而来，该超大泛化模型对成本、组织力、工程壁垒、大算力和底层框架都提出极高要求。随着模型的不断扩大，模型训练和维护成本也在不断增加；在技术层面上，需要先进的底层技术和高效的框架来保证模型训练的稳定性和效率，而且构建的模型必须要有很高的泛化能力，这需要研发团队有相当高的技术水平和工程能力。

► **训练成本：**对于**训练基础模型 GPT-3**为例，训练一次所要消耗的成本高达近**8400 万元人民币**。这不仅需要资本，也需要公司的组织力和战略支持。依赖于**小规模团队的试验探索**来创新不再可能。

► **标注数据：**由于 ChatGPT/InstructGPT 模型训练使用的是 prompt 方法，在训练的第一步 OpenAI 雇佣标注师对 **1.3 万条数据**进行人工回答，同时在第二部分需要对模型给出的 **3.3 万条答案进行排序**。在论文附录中可以看到 OpenAI 对于人工标注数据的生成有诸多细节和心得。

► **强大算力**：ChatGPT 训练所耗费的算力大约为 **3640 PF-days**，即假设**每秒运算一千万亿次**，需要连续**运行 3640 天**。训练大模型需要强大的算力，而优化大规模算力背后的**技术底层框架需要深度定制和自主研发能力**，这些因素都构成了诸多**工程壁垒**。

除了显性的成本外，**场景、流量、数据壁垒和社会包容度**所形成的**隐性成本**同样**不可小觑**。不同的场景需要专门的数据集，数据的获取和标记需要额外的成本；数据的质量、隐私和安全构建起数据壁垒；此外，公众对不同规模公司发布的新模型的包容度是不同的。

► **用户真实体验**：除了标注人员需要标注，也需要**用户通过测试 API** 提出实际使用中的问题获取 prompt，得到更多 User-Based 数据，缩小和用户需求的差距，不断迭代优化，该过程需要更**高昂的数据清洗成本和时间成本**。

► **基于场景反馈**：ChatGPT/InstructGPT 的核心创新点在于 RLHF，这不仅局限于标注师的反馈，也需要实际用户使用中的反馈。例如 **Midjourney** 在用户使用中生成 4 张图像，用户的点选就会作为 **reward 反馈**传回公司进行迭代**强化学习**。找到场景、拥有流量，使得广大爱好者能够众包式的迭代模型，也是商业模式上的挑战。

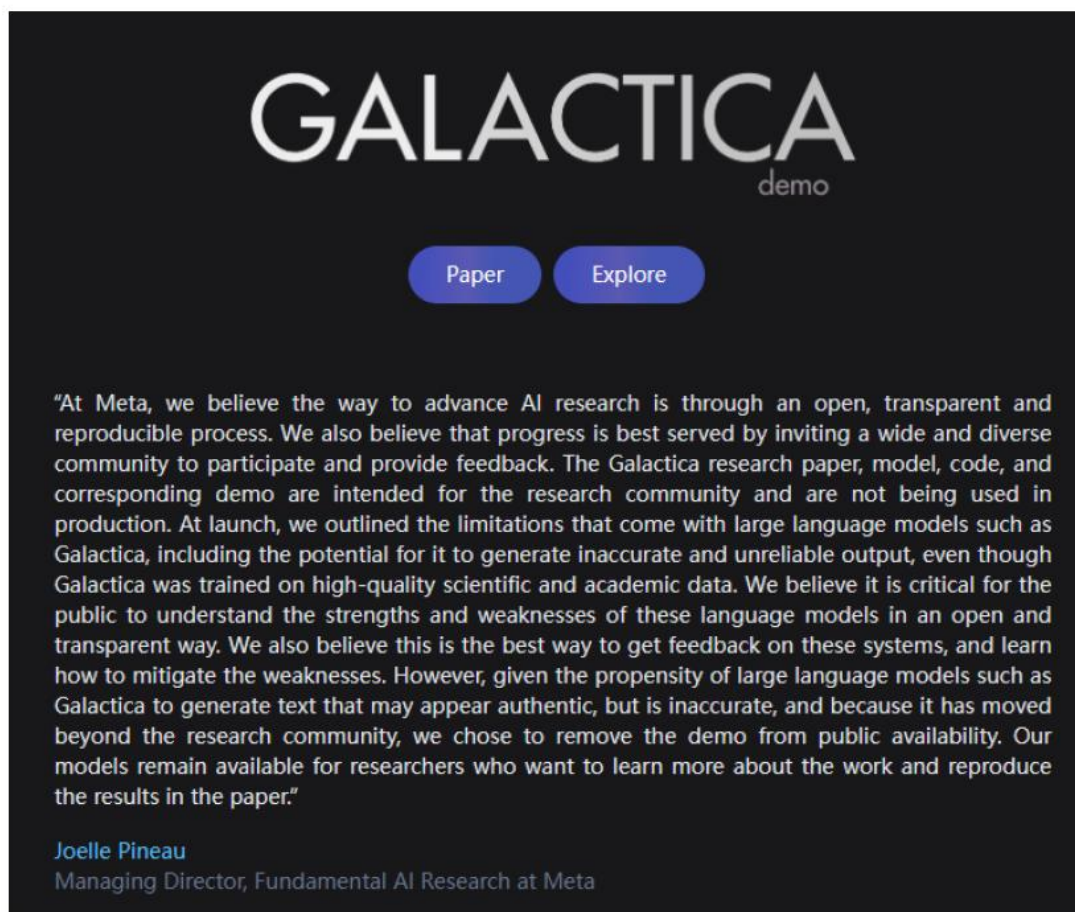
► **社会包容程度**：对于初创公司，这种内测模型更容易发布，用户包容度更高，但随着模型表现出色，**用户期待值更高，未来再通过免费内测形式会遇到更多问题**。例如谷歌、Meta 都曾发布语言模型测试，但因数据质量层次不齐，输出结果有明显偏误，被迅速下架。

图表 30：Midjourney 生成图像用户反馈界面



资料来源：Midjourney 官网，中金公司研究部

图表 31：Meta 发布的语言模型 GALACTICA 上线仅两天就因输出结果有偏误而光速下架



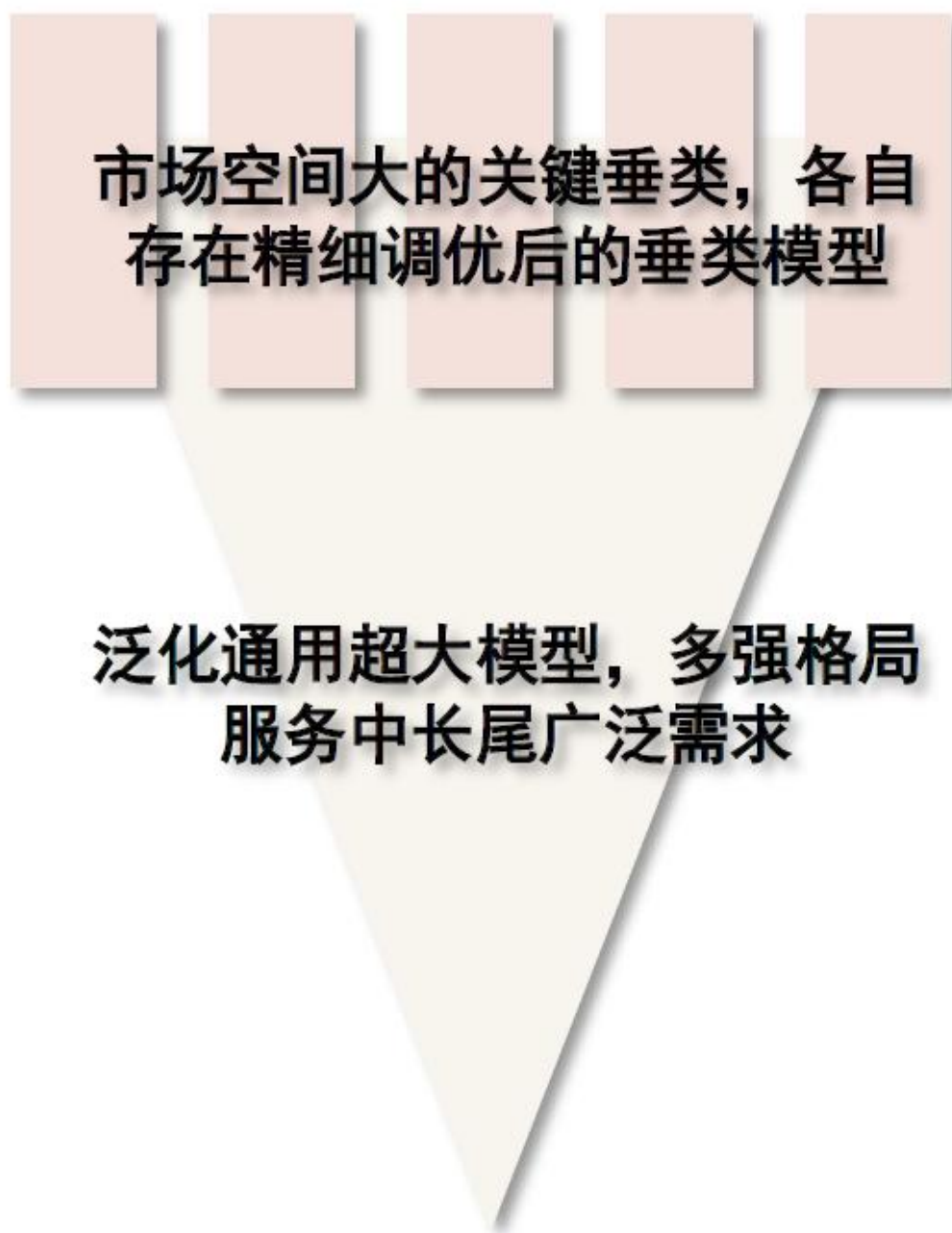
资料来源：GALACTICA 官网，中金公司研究部

产业链百花齐放，创业公司垂直领域也具备机遇

大模型路线下，行业格局必将走向集中。不同于深度学习的技术路线，大模型路线的高壁垒和“前期投入大，边际成本低”的模式决定了行业必将走向集中格局，缺少资源的中小参与者难以跟随技术发展的步伐。

在中长尾走向寡头市场，在关键垂类仍有 Know-How 的魅力。尽管行业走向集中，我们却并不认为行业会走向影视作品中一个超级人工智能统领全球的景象。在市场空间足够大的关键垂类中，Know-How 和数据壁垒仍旧能够显著增加模型的表现，同时获得足够的经济回报。而在更中长尾的场景中，有泛化多模能力的超大模型将呈现多强格局，滋养诸多的 ISV 在其基础上搭建应用，或者进一步调优。

图表 32：我们预测的未来格局



资料来源：中金公司研究部

图表 33：InstructGPT 论文显示：经过知识调优后的小参数模型效果远超未调优的大模型

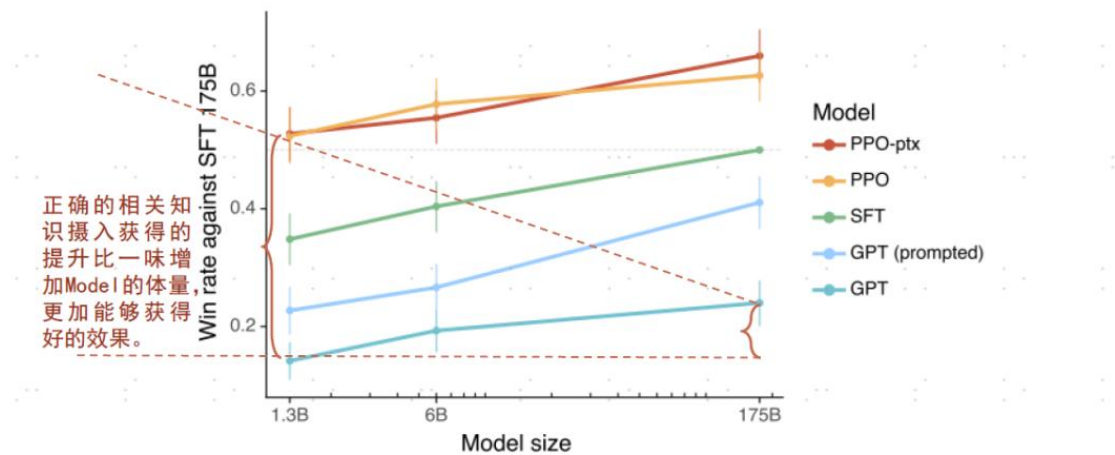


Figure 1: Human evaluations of various models on our API prompt distribution, evaluated by how often outputs from each model were preferred to those from the 175B SFT model. Our InstructGPT models (PPO-ptx) as well as its variant trained without pretraining mix (PPO) significantly outperform the GPT-3 baselines (GPT, GPT prompted); outputs from our 1.3B PPO-ptx model are preferred to those from the 175B GPT-3. Error bars throughout the paper are 95% confidence intervals.

资料来源：《Training language models to follow instructions with human feedback》（OpenAI,2022），中金公司研究部

未来商业模式探讨：Model-as-a-Service 是潜在方向

目前 ChatGPT 还没有明确的盈利商业模式，但我们认为 MaaS 是一个潜在方向。未来的商业模式可以参照 Hugging Face 践行的 **AI 开源社区模式**以及潜在方向：**模型即服务 Model-as-a-Service**，即 Hugging Face 提供的 AutoTrain 和 Inference API & Infinity 解决方案。MaaS 将模型训练、维护和部署的过程转移到云服务上。这样的方式不仅提高了效率，还降低了客户对模型开发和维护的依赖，使其能够更加专注于业务上的投入。

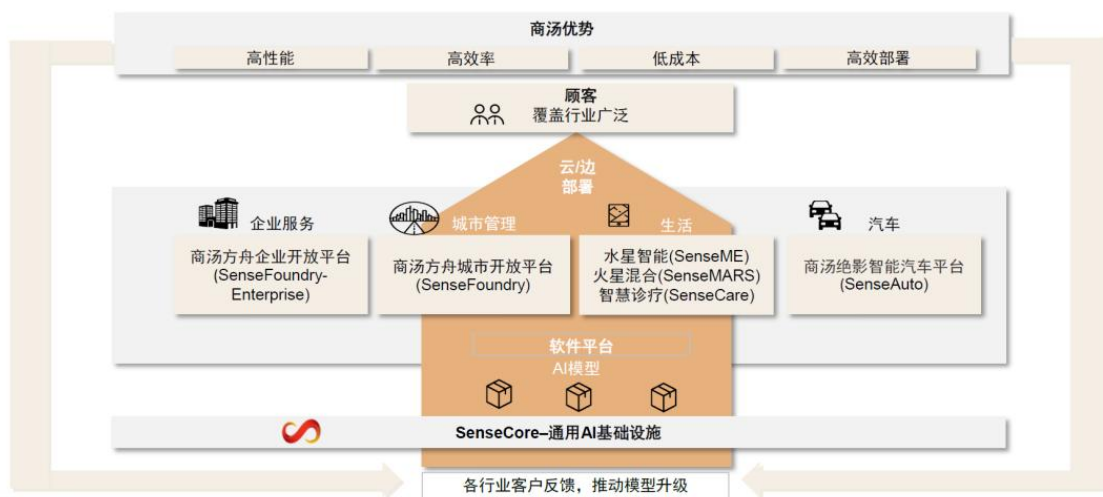
图表 34：Hugging Face 商业模式

MaaS	商业模式	具体形式	客户类型
	AutoTrain	提供端到端的自动化NLP模型解决方案，该方案门槛较低，用户只需创建任务，并上传数据，AutoTrain就能自动创建、调整、评估模型，找到最佳模型，并将模型部署到实际应用中。	没有足够多的数据科学家人才的SMB客户或传统行业客户
	Inference API & Infinity	部署客户的数据和模型在公司或客户自己的服务器上，并提供ML解决方案，采用传统的租用CPU/GPU的pay-as-you-go的收费模式。	Inference API: 对数据安全性要求低的企业 Infinity: 对数据安全性要求高的大型企业
	Private Hub	类似GitHub靠私人代码托管实现商业化，Hugging Face的Private Hub通过替用户托管模型、数据集、Pipeline收取一定的费用。	无特定客户类型，Hugging Face社区用户均有可能使用
	Expert Support	内部组建了一支世界级的机器学习专家团队，对外为客户解决ML部署和实施问题，提供相应的解决方案。	无特定客户类型，Hugging Face社区用户均有可能使用

资料来源：元宇宙官网，中金公司研究部

商汤科技是国内领先的人工智能公司，其自主研发的 AI 基础设施 SenseCore，实现了通过 AI 平台化赋能百业，商业模式类似 MaaS。与传统的 AI 技术不同，商汤把人工智能的核心底层基础设施、软件平台和应用层打造成了一个通用的生态系统，可以被各种应用程序和服务轻松使用，而不需要对模型进行专门的开发。这种模型即服务（Model-as-a-Service）的架构，可以帮助企业更快速地实现人工智能的价值。

图表 35：商汤建立了包括底层基础设施、AI 软件平台及应用层平台的 AI 生态系统



资料来源：弗若斯特沙利文，商汤科技，中金公司研究部

风险

技术进展不及预期：人工智能作为前沿新兴技术，仍处于技术的快速发展期，其进展有一定的不确定性，若技术进展不及预期，可能导致产业化进展缓慢。

行业竞争加剧：人工智能是产业的热点，未来商业价值显著，科技巨头、初创公司均在此领域布局，未来垂类及应用层的行业竞争可能会进一步加剧。

商业化落地节奏不及预期：商业化落地是人工智能能否顺利走向下一阶段的关键点，若商业化落地节奏不及预期，对人工智能的进展将带来负面影响。

文章来源

本文摘自：2023 年 2 月 2 日已经发布的《人工智能十年展望 (五)：从 ChatGPT 到通用智能，新长征上的新变化》

赵丽萍 分析员 SAC 执证编号：S0080516060004 SFC CE Ref: BEH709

于钟海 分析员 SAC 执证编号：S0080518070011 SFC CE Ref: BOP246

魏鹤霏 联系人 SAC 执证编号：S0080121070252 SFC CE Ref: BSX734

免责声明：

1. 本附加与原报告无关；
2. 本资料来源互联网公开数据；
3. 本资料在“行业报告资源群”和“知识星球 行业与管理资源”均免费获取；
4. 本资料仅限社群内部学习，如需它用请联系版权方

合作与沟通，
请联系客服



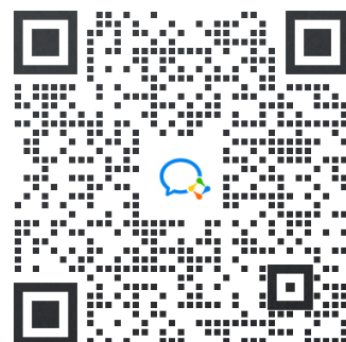
客服微信



客服微信

行业报告资源群

1. 进群即领福利《报告与资源合编》，内有近百行业、万余份行研、管理及其他学习资源免费下载；
2. 每日分享学习最新6+份精选行研资料；
3. 群友咨询，群主免费提供相关行业报告。



微信扫码，长期有效

知识星球 行业与管理资源

知识星球 行业与管理资源 是投资、产业研究、运营管理、价值传播等专业知识库，已成为产业生态圈、企业经营者及数据研究者的智慧工具。

知识星球 行业与管理资源 每月更新5000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；

微信扫码加入后无限制搜索下载。



微信扫码，行研无忧