大模型“开源、轻量、端侧”化，视频与语音加速落地

——AI 行业深度更新报告

传播文化业

评级：增持

上次评级：增持

陈筱(分析师) 021-38675863 chenxiao@gtjas.com 登记编号 S0880515040003
杨昊(分析师) 021-38032025 yanghao029514@gtjas.com S0880524020001

本报告导读：

大模型能力提升阶段性放缓之际，我们提示关注“AI 落地”进展：如大模型侧“开源”“轻量”“端侧”化趋势显著，视频、音频等领域 AI 自 6 月以来更新频出。

投资要点：

- **继续看好 AI 技术发展对内容产业的推动作用。**随着 AI 大模型开源化、轻量化，以及视频和语音等模态的快速进步，部分应用场景有望发生变化，可沿如下思路进行布局：1）游戏等应用改造，推荐吉比特、恺英网络、完美世界、美图公司，受益标的腾讯控股、网易、快手、巨人网络；2）教育赛道，受益标的南方传媒、皖新传媒、世纪天鸿；3）情感陪伴与社交，受益标的昆仑万维、盛天网络。
- **大模型侧：开源能力快速提升，轻量化趋势显著。**2024 年以来，大模型发展呈现三大趋势：1）开源模型发展，能力快速接近闭源产品水平；2）“轻量化”，模型“性价比”快速提升；3）端侧模型发展，AI 硬件已经开始布局。这些都意味着 AI 大模型的发展在向着落地可行方向进发。
- **AI 生成视频：能力兑现有望加速。**自从 2024 年 2 月 OpenAI sora 演示视频放出，AI 视频领域的行业标准被显著提高，而经历 4 个多月的积累后，6-7 月国内外多个团队交出“类 sora”产品的首份答卷：国内有多次迭代、面向全球、快速商业化的快手可灵，从文本大模型发家的独角兽企业智谱；海外则有持续保持高生成质量的 Runway Gen3 Alpha、有 3D 生成相关经验帮助的 Luma。这些产品的生成效果都比“前 sora 时代”的产品有显著提升，预示着 AI 视频领域的预期或加速兑现。
- **语音功能：或助推 AI 产品迭代。**GTP-4o 的高级语音功能已经在 7 月底开始小范围测试，这一功能使得 AI 可以从用户语音中获得情感、语调等更丰富的信息，回应时也可以体现出语调等更丰富的表达，且交互延迟小、可被打断，AI 语音交互的体验预计有显著提升。该领域 AI 技术的发展有望对教育、情感陪伴等应用场景的使用有改进效果，对人机交互体验提升将有所帮助。
- **风险提示：**AI 应用推进放缓，AI 相关商业化落地不及预期，生成式内容监管风险。

相关报告

传播文化业《多个知名 IP 获批进口版号，GPT-4o 语音功能推进》2024.08.04
传播文化业《GPT-4o 语音、视频模式测试，可提升教育、情感陪伴体验》2024.08.01
传播文化业《快手可灵推出付费会员，PixVerseV2 全面升级》2024.07.28
传播文化业《快手可灵全球上线并升级，AI 视频工具或迎加速发展》2024.07.25
传播文化业《《抓娃娃》引燃观影热情，多款头部影片待映》2024.07.20

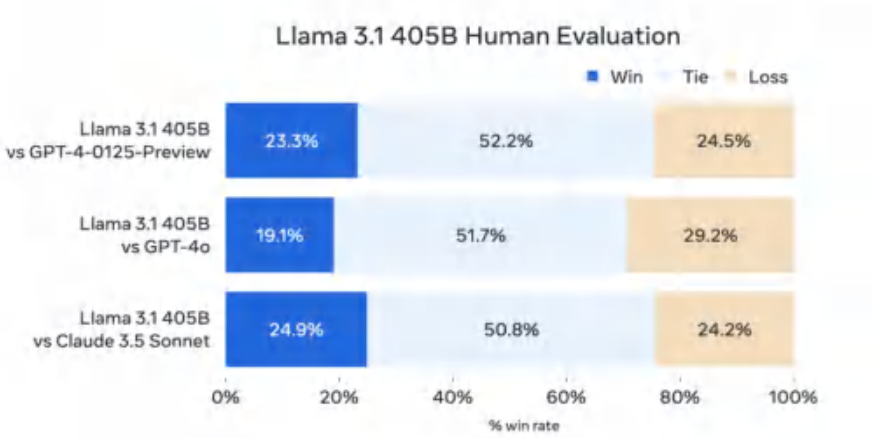
目录

1. 大模型侧：开源能力快速提升，轻量化趋势显著

1.1. 趋势一：开源模型发展，能力快速接近闭源产品水平

开源模型 Llama3.1 发布，追平 GPT-4o 和 Claude 3.5 Sonnet。2024 年 7 月 23 日，Meta 推出 Llama3.1，将上下文长度扩展到 128K，增加了对八种语言的支持，共包括 8B、70B 和 405B 三个尺寸。其 405B 的版本从性能上已经可媲美 GPT-4o 和 Claude 3.5，而其 8B 和 70B 版本都均超越同等尺寸的其他开源模型。

图1：Llama3.1 性能上追平 GPT-4o 和 Claude 3.5 Sonnet



数据来源：Meta

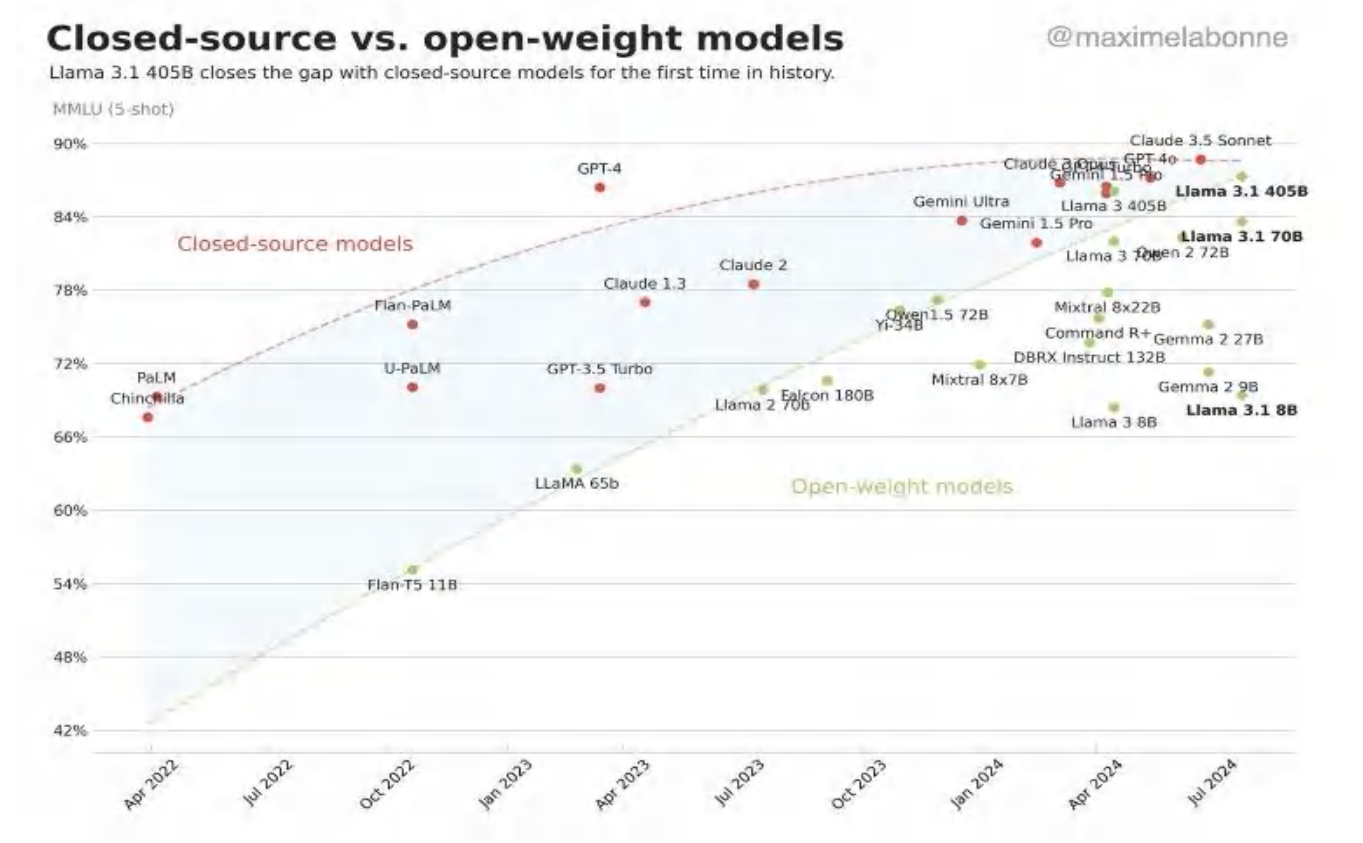
图2：Llama 8B 和 70B 能力超越同尺寸其他开源模型

Benchmark	Llama 3.1 8B	Gemma 2 9B IT	Mistral 7B Instruct	Llama 3.1 70B	Mixtral 8x22B Instruct	GPT 3.5 Turbo
MMLU	73.0	72.3	60.5	86.0	79.9	69.8
MMLU PRO	48.3	-	36.9	66.4	56.3	49.2
IFEval	80.4	73.6	57.6	87.5	72.7	69.9
HumanEval	72.6	54.3	40.2	80.5	75.6	68.0
MBPP EvalPlus	72.8	71.7	49.5	86.0	78.6	82.0
GSM8K	84.5	76.7	53.2	95.1	88.2	81.6
MATH	51.9	44.3	13.0	68.0	54.1	43.1
ARC Challenge	83.4	87.6	74.2	94.8	88.7	83.7
GPQA	32.8	-	28.8	46.7	33.3	30.8
BFCL	76.1	-	60.4	84.8	-	85.9
Nexus	38.5	30.0	24.7	56.7	48.5	37.2
ZeroSCROLLS/QuALITY	81.0	-	-	90.5	-	-
InfiniteBench/En.MC	65.1	-	-	78.2	-	-
NIH/Multi-needle	98.8	-	-	97.5	-	-
Multilingual MGSM	68.9	53.2	29.9	86.9	71.1	51.4

数据来源：Meta

顶尖的开源模型趋近顶尖的闭源模型，Llama3.1 或标志行业转折点。整体来看，开源大型语言模型在功能和性能方面仍大多落后于闭源模型，但开源模型的成长性有更高的斜率，随着时间推进，开源模型的能力在快速赶上，如在 MMLU 的测试维度中，最新的 Llama3.5 405B 就已经非常接近 Cluade 3.5 Sonnet。开源模型更为开放，在学习和成长上来源丰富，其与闭源模型的差距有望持续缩小，甚至超越。

图3：开源模型能力快速接近闭源产品



数据来源：maximelabonne，36Kr

2024 年以来开源模型频现，能力不断刷新。7 月，Mistral AI 发布最新模型 Mistral Large 2，参数 123B，用不到三分之一的参数量性能比肩 Llama 3.1 405B，也不逊于 GPT-4o、Claude 3 Opus 等闭源模型。2024 年以来推出的开源模型不在少数，性能上足以媲美当前领先的闭源模型。

表 1：推荐公司盈利预测与估值情况表

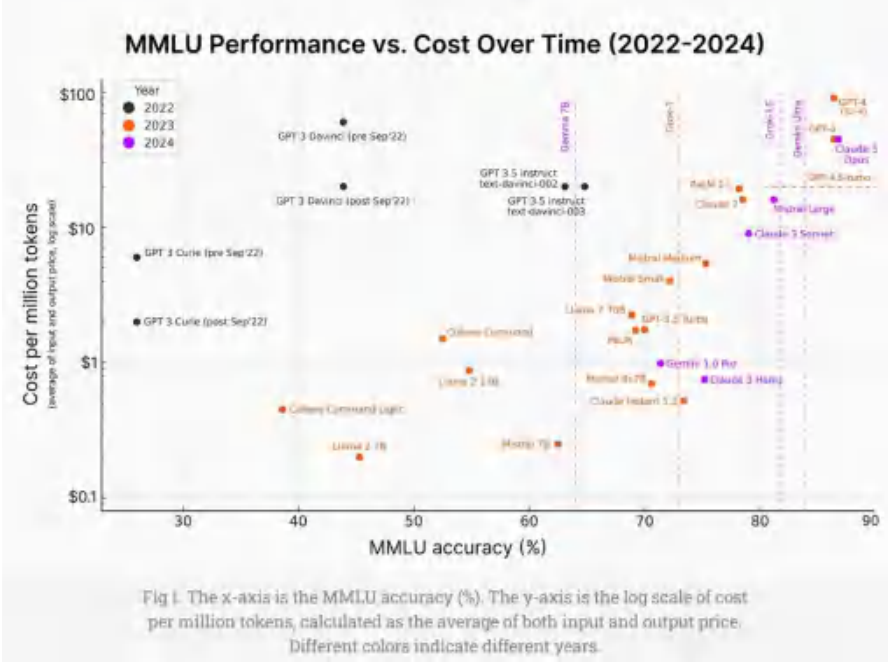
产品名	厂商	发布时间	参数量	模型能力水平
Gemma	谷歌	2 月	2B、7B	大幅超过 Llama2
Phi-3-mini	微软	4 月	3.8B	优于 Llama 8B
Llama 3	Meta	5 月	8B、70B	接近 GPT-4
DeepSeek v2	DeepSeek	5 月	236B	GPT-4 Turbo
Llama 3.1	Meta	7 月	8B、70B、405B	GPT-4o
Qwen2-72B	阿里巴巴	6 月	72B	超过 Llama3-70B
Mistral Large 2	Mistral AI	7 月	123B	Llama 3.1 405B、GPT-4o、Claude 3 Opus

数据来源：智东西，机器之心，量子位，国泰君安证券研究

1.2. 趋势二：“轻量化”，模型“性价比”快速提升

大模型性价比逐年提升，优秀轻量模型层出不穷。成本更低的模型往往表现也更弱，但是随着相关研究推进，2022-2024 年在同等成本下的大模型表现逐年提升，2024 年轻量模型赛道也吸引了各家机构的关注，各类轻量模型层出不穷。

图4：轻量级模型更具性价比



数据来源：semaphore

表 2: 2024 年以来领先轻量级通用语言模型不断出现

机构	模型	参数规模	上下文长度
面壁智能	MiniCPM	1.2B,2.4B	4096,128K
	MiniCPM-S	1.2B	未披露
阿里巴巴	Qwen1.5	0.5B,1,8B,4B,7B	32K
	Qwen2	0.5B,1.5B,7B	32K,128K
Google	Gemma 1	2B,7B	8192
	Gemini 1.5 Flash	未披露	1M
	Gemma 2	2.6B	8192
Anthropic	Claude 3 Haiku	未披露	200K
商汤	InternLM2	1.8B,7B	200K
	InternLM2.5	7B	200K，1M
Meta	Llama3	8B	8K
苹果	OpenELM	270M,450M,1.1B,3B	2048
	DCLM	7B	2048，8192
微软	Phi-3	3.8B,7B	4K、8K、128K
HuggingFace	SmolLM	135M,360M,1.7B	2048
OpenAI	GPT-4o mini	未披露	128K

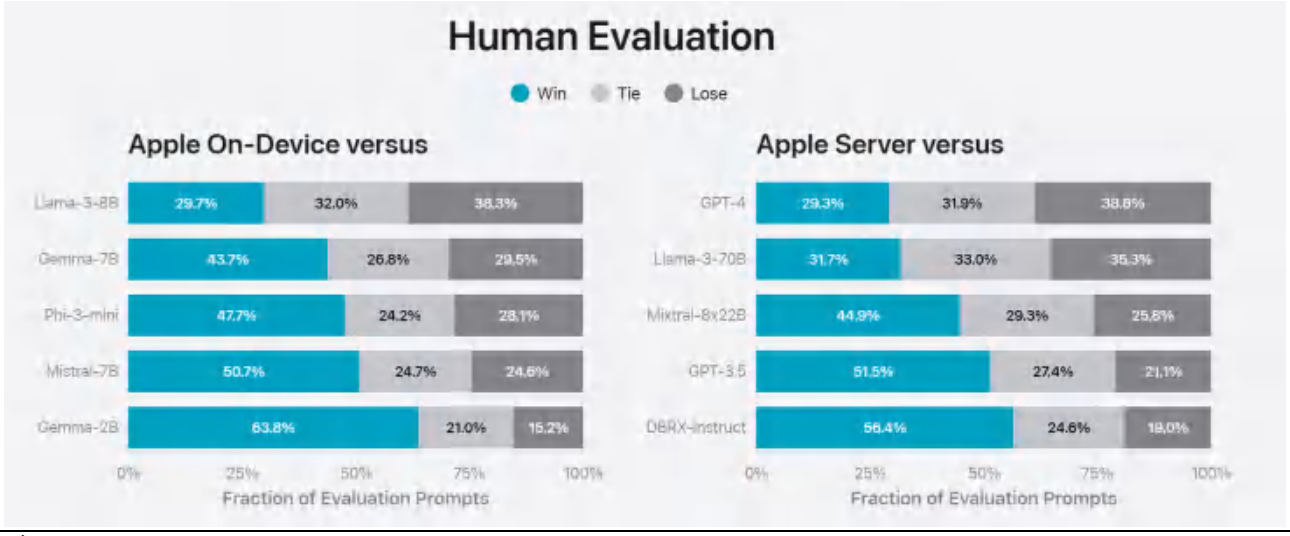
数据来源：智东西，国泰君安证券研究

1.3. 趋势三：端测模型发展，AI 硬件已经开始布局

人工评估显示 Apple Intelligence 优于同等大小模型及大模型。苹果 AI 包括两大基础语言模型：约 30 亿参数的端侧模型 AFM-on-device，和在云服务器中运行的更大参数模型。其中，端侧模型具有 30 亿参数，能够完成文本撰写和润色、优先处理和总结通知、创建图像，以及执行应用内操作以简化跨应用的交互。从苹果在 iPhone 15 Pro 上的测试结果来看，如果用户向模型发送 1000 个 token 的 prompt，模型将需要 0.6 秒开始响应，之后它将以每秒 30 个 token 的速度生成结果。相比于 Gemma-2B、Mistral-7B、Phi-

3B-Mini 和 Gemma-7B 等大多数竞争对手模型，苹果 AI 的模型更受人类评分者的青睐。

图5：人工评估显示 Apple Intelligence 模型比其他竞争对手模型更受人类评分者青睐

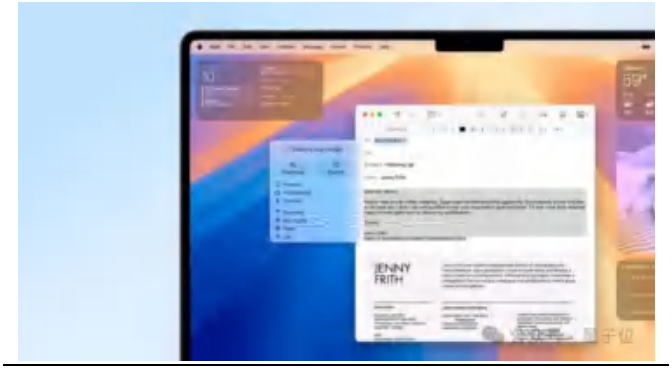


数据来源：Apple

Apple Intelligence 支持摘要、写作帮助、工具使用和代码等广泛功能。2024 年 7 月 30 日，iOS 18.1 Beta 版上线，同时发布了苹果 AI 的部分功能：

1）文本生成，只要使用标准输入文本系统，在第三方应用程序当中也能够实现文本总结、校对和重写，另外结合 iOS 18 Beta 的语音备忘录中已经上线的音频转录功能，文本生成系统还可以为录音生成摘要；
2）Siri：新的 Siri 可以理解两个查询之间的上下文，而无需重述正在谈论的内容；
3）**相册：**相册更新后，用户可以用自然语言搜索特定照片，甚至是视频当中的具体时刻。

图6：Apple Intelligence 可文本总结、校对及重写等



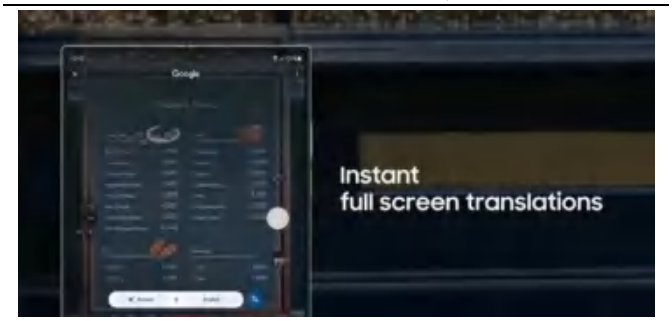
数据来源：量子位

图7：用户可用自然语言搜索特定照片

数据来源：量子位

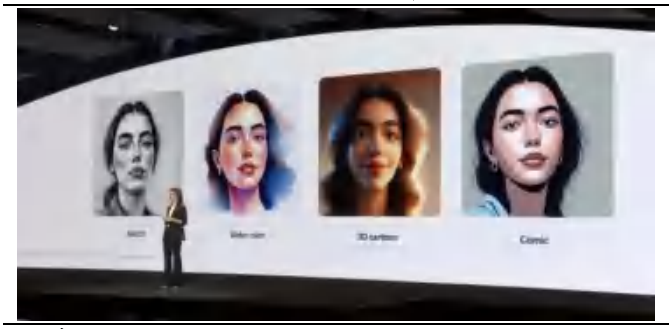
三星公布系列 AI 功能，包括图像、翻译等。2024 年 7 月 10 日，三星新品发布会上重点展示 AI 系列功能。在具体 AI 功能上，在三星折叠屏手机 Galaxy Z Fold6 和 Galaxy Z Flip6 上搭载了多项 AI 功能，如 AI 画圈即搜、AI 翻译、AI 图像生成等。其中，AI 画圈即搜可以让用户直接在相机取景框中圈出感兴趣的物体，AI 就会自动识别并提供相关信息；AI 翻译则可以实时翻译多种语言，并支持面对面交流时使用折叠屏的外屏显示翻译结果；AI 图像生成功能允许根据用户的几笔笔迹， 生成精美的图片。此外，三星 AI 也具备 AI 改写等功能。

图8：三星 AI 翻译能够实现实时全屏翻译



数据来源：三星

图9：三星 AI 图像生成可将相册中照片转成动漫风格

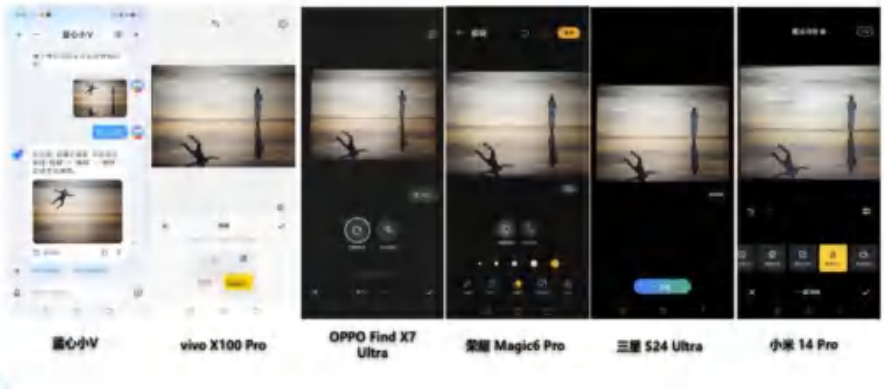


数据来源：三星

Vivo X100 Pro 在主流厂商旗舰手机中综合得分最高。2024 年 4 月，移动发布《主流旗舰手机 AI 功能评测报告》，其中出厂内置蓝心小 V 的 vivo X100 Pro 在五款主流厂商旗舰手机（OPPO Find X7 Ultra、vivo X100 Pro、荣耀 Magic6 Pro、三星 S24 Ultra、小米 14 Pro）中整体表现最佳，提供了文生图、图片作诗等图像 AI 功能，给用户 AI 体验最佳。

1）图片功能上，vivo X100 Pro 文生图的表现相对优秀；
2）文字功能上，vivoX100Pro 的蓝心小 V 在文字创作和总结摘要功能方面都明显领先其他手机；
3）识屏功能上，vivo X100 Pro 在屏幕朗读和识别人物、物品内容的 AI 功能上表现优秀；
4）语音功能上，vivo X100 Pro 整体表现较好，AI 字幕识别语言种类较多，面对面翻译和语音助手方面的 AI 功能表现较为优秀。

图10：vivoX100 Pro 的蓝心小 V 可同步消除路人水面的倒影



数据来源：中国移动，国泰君安证券研究

2. AI 生成视频：能力兑现有望加速

2.1. sora 打破以往时长限制，树立行业标准

2.1.1. sora 的“高度一致性”、“60s 时长”为行业树立全新的标准

sora 推出前，大多数 AI 视频产品采用“逐帧预测”思路，视频动态相对保守，时长也多在 3 内，通过延伸功能难以保持一致性；而自从 2024 年 2 月，OpenAI 放出 sora 的技术报告与演示视频，其展现出高度的“一致性”，时长方面超过 5 秒甚至长达 60 秒，画面动态有明显提升、穿梭镜头、光影表现也非常出众，这使得“AI 生成视频”赛道的行业标准被迅速拔高。

表 3: OpenAI sora 的出现突破了以往的行业标准

产品/模型	发布方	产品形态	发布时间	可生成时长	特色
Runway Gen2	Runway	网页产品（闭源）	2023.3	3 秒为主，可延伸至 16 秒	多动态笔刷功能：可以在图像中绘制最多 5 个独特区域，以独立于其他区用控制其运动
Pika 1.0	Pika Labs	网页产品（闭源）	2023.11	3 秒为主，可延伸至 15 秒	可进行视频画幅拓展；可以圈定换物；动画风格效果突出
Stable Video Diffusion (SVD)	Stability AI	开源	2023.11	<4 秒	开源模型，基于 stable diffusion
Emu Video	meta	发布产品官网	2023.11	4 秒	使用了分解式生成方法，先生成一张图像，再以该图像和文本作为条件生成视频
Sora	**OpenAI**	**发布技术报告（闭源）**	**2024.2**	**最长可达 1 分钟**	**可变的持续时间、分辨率、宽高比，在生成时长、连贯性等方面都有显著的优势。**

数据来源：各公司产品官网，Github，国泰君安证券研究

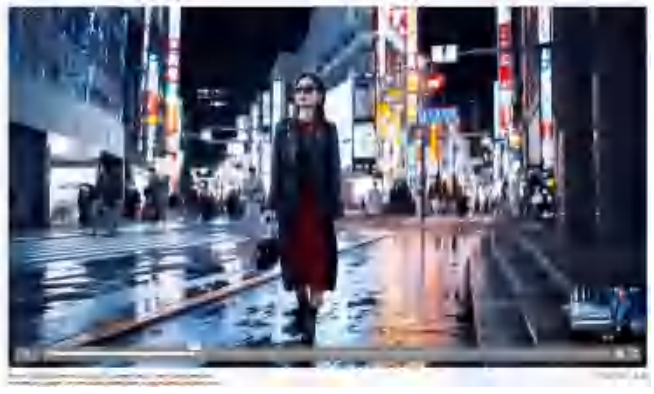
2.1.2. 采用 DiT 思路，大规模训练下体现出“涌现”能力

采用多视频编码形式训练，效果突出。Sora 最早于 2024 年 2 月 16 日公开，OpenAI 一次给出多达 48 个由 sora 直接生成、未经修改的视频，最长的时长可达 59 秒，远高于主流模型的 3-4 秒的时长和 15-16 秒的极限时长，风格上涵盖写实、动画、剪纸、3D、风景、微观、细节特写等多种。不同于传统视频类 AI 采用单帧画面逐帧预测的方式，OpenAI 在 sora 的训练中采用了 diffusion Transformer 的思路，参考 LLM 的 token 思路，将多种视频编码成统一格式，形成 visual patches，然后借助 DALLE3 中采用重新标注技术，对视频形成标注，基于大规模训练数据达成了较为稳定的一致性。**根据 OpenAI CTO 穆拉蒂 4 月接受采访的消息，sora 预计在 24 年末上线。**

图 11： OpenAI 采用将视频 token 化的思路进行训练

数据来源：OpenAI

图 12： OpenAI sora 展示视频画面，时长达到 59 秒



数据来源：OpenAI

图 13：sora 体现出了一些模拟能力上的“涌现”

数据来源：OpenAI，国泰君安证券研究

2.2. 6 月以来多家“AI 视频”产品推出，产业呈现加速发展

国内外多个知名团队在 6-7 月推出“AI 视频”产品。包括国内大厂如快手（可灵）、AI 大模型公司如智谱（清影）、创业团队如爱是科技（PixVerse）、生数科技（Vidu），海外团队如 Luma（Dream Machine）、Runway（Gen-3 Alpha），整体呈现加速迭代趋势。

表 4：6-7 月有多个团队推出“AI 视频”产品

产品/模型	发布方	产品形态	发布时间	可生成时长	特色
Luma Dream Machine	Luma	网页产品（闭源）	2024.6	5 秒	速度快，120 秒即可生成 120 帧；动作逼真，流畅，融入电影级别的摄影技巧和戏剧张力；角色一致性极强，能够模拟物理世界；运镜自然，可匹配场景情感
可灵	快手	网页、App 端功能	2024.7	单次文生视频时长已增至 10 秒；最长 3 分钟、30fps 的 1080P 视频	大幅度的合理运动；长达 2 分钟的视频生成；模拟物理世界的特性；强大的概念组合能力；电影级的画面生成；支持自由的输出视频宽高比
清影	智谱 AI	网页、App 端、小程序	2024.7	生成时长 6 秒、1440x960 清晰度的高精视频	快速生成，30 秒内即可完成 6 秒视频的生成；能够准确理解和执行复杂的 prompt；生成的视频能够较好地还原物理世界中的运动过程；画面调度

					具有灵活性
PixVerse V2	爱诗科技	网页产品	2024.7	单片段可以实现 8 秒、多片段可以实现 40 秒视频生成	在时空建模上，采用自研注意力机制，超越传统架构，增强时空感知。文本理解方面，利用多模态模型提升信息对齐和表达能力。优化 flow 模型，通过加权损失提高训练效率
Vidu	生数科技与清华大学联合发布	网页产品	2024.7	提供 4s 和 8s 两种时长选择，分辨率最高 1080P	开放了文生视频、图生视频两大核心功能。效果方面，Vidu 在延续高动态性、高逼真度、高一致性等基础上，新增了角色一致性、动漫风格、文字与特效画面生成等能力
Gen-3Alpha	Runway	网页产品	2024.7	5 秒或 10 秒视频	支持图像启动视频生成并增加精细控制，是一个集视频图像训练于一体的系统，提供多种控制模式以调整视频结构、风格和动态

数据来源：各公司产品官网，Github

2.3. 快手可灵：已有多次升级，面向全球并尝试商业化

自 6 月初发布以来，已多次迭代。快手可灵 AI 自 2024 年 6 月 6 日发布后不断升级，6 月 21 日新增图生视频和视频续写功能，7 月 24 日，国际版 1.0 上线，全球用户可用邮箱注册，同日国内方面则是再次升级，升级后画面构图、色调、以及美观程度显著提升；运动表现也明显提升。国内方面还开始尝试商业化，用户每日登录免费得 66 点“灵感值”，可用于兑换平台内指定的功能使用权或增值服务，单个视频生成需要耗费 10 点“灵感值”。

图 14：6 月初至今快手可灵动作频频

数据来源：快手可灵，国泰君安证券研究

快手可灵具备多项出色功能：

1）最长 10 秒的文生视频；

2）5 秒时长的图生视频：新增运镜控制、自定义首尾帧等功能；

3）最长 3 分钟的视频续写：单次让视频运动延续 4.5 秒，支持连续多次的续写，最长可生成 3 分钟的视频。

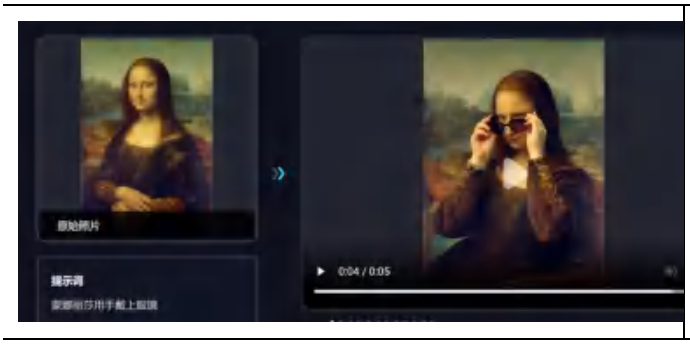
表 5：快手可灵的六大优势能力

优势	具体内容
大幅度的合理运动	可灵采用 3D 时空联合注意力机制，能够更好地建模复杂时空运动，生成较大幅度运动的视频内容，同时能够符合运动规律。
长达 2 分钟的视频生成	得益于高效的训练基础设施、极致的推理优化和可扩展的基础架构，可灵大模型能够生成长达 2 分钟的视频，且帧率达到 30fps。
模拟物理世界特性	基于自研模型架构及 Scaling Law 激发出的强大建模能力，可灵能够模拟真实世界的物理特性，

	生成符合物理规律的视频。
强大的概念组合能力	基于对文本-视频语义的深刻理解和 Diffusion Transformer 架构的强大能力，可灵能够将用户丰富的想象力转化为具体的画面，虚构真实世界中不会出现的的场景。
电影级的画面生成	基于自研 3D VAE，可灵能够生成 1080p 分辨率的电影级视频，无论是浩瀚壮阔的宏大场景，还是细腻入微的特写镜头，都能够生动呈现。
支持自由的输出视频宽高比	可灵采用了可变分辨率的训练策略，在推理过程中可以做到同样的内容输出多种多样的视频宽高比，满足更丰富场景中的视频素材使用需求。

数据来源：快手可灵，国泰君安证券研究

图 15：快手可灵图生视频功能



数据来源：快手可灵

图 16：快手可灵视频续写功能

数据来源：快手可灵

快手可灵推出会员体系，国内 AI 进入“付费时代”。2024 年 7 月 24 日，Kling AI 宣布上线会员体系，目前分为黄金、铂金、钻石三个类别，区别主要在获得灵感值的数量，按照享受时长可选择月卡、季卡、半年卡、年卡等多种套餐。以月卡为例，三档会员价格分别为 66 元、266 元和 666 元，分别可生成约 66 个、300 个或 800 个标准视频。

表 6：快手可灵的付费方案差异主要体现在灵感值数量上

	非会员	黄金会员	铂金会员	钻石会员
定价（元/月）	0	66	266	666
灵感值每月	-	660	3000	8000
每 100 灵感值定价（元）	-	10	8.867	8.325
可生成内容	-	约生成 3300 张图片或 66 个高性能视频	约生成 15000 张图片或 300 个高性能视频	约生成 40000 张图片或 800 个高性能视频
功能	登录每日送灵感值	登录每日送灵感值；**图片、视频去水印；高表现视频生成；视频延长功能；视频专家大师运镜；**	登录每日送灵感值；图片、视频去水印；高表现视频生成；视频延长功能；视频专家大师运镜；**新功能优先体验**	登录每日送灵感值；图片、视频去水印；高表现视频生成；视频延长功能；视频专家大师运镜；**新功能优先体验**

数据来源：快手可灵，国泰君安证券研究

2.4. 智谱“清影”：AI 大模型团队的“视频”领域尝试

国产大模型独角兽公司智谱发布“清影”，为首家入局“AI 视频领域”的 AI 大模型公司。2024 年 7 月 26 日，作为首家入场文生视频的国产大模型独角兽的智谱 AI 对视频生成模型进行全新升级，正式上线了 AI 视频生成功能“清影”，支持文生视频、图生视频。清影此次面向所有用户全量上线，无需

预约，人人可用。智谱大模型开放平台 bigmodel.cn 也部署了“清影”。企业和开发者可通过 API 调用式，体验并使用“清影”的文本生成视频和图像生成视频功能。

“清影”的生成速度和指令遵循程度较为亮眼。

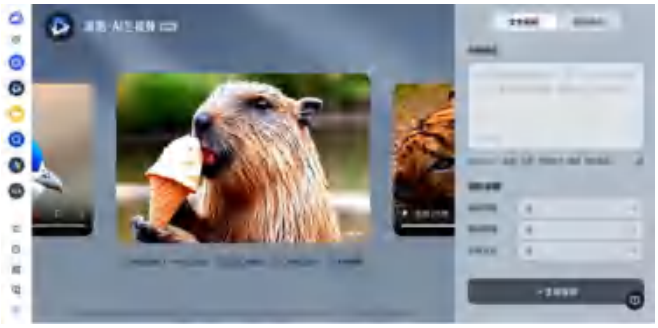
1）快速生成：仅需 30 秒即可完成 6 秒视频的生成。

2)高效的指令遵循能力:即使是复杂的 prompt,清影也能准确理解并执行。

3）内容连贯性：生成的视频能够较好地还原物理世界中的运动过程。

4）画面调度灵活性：镜头能够流畅地跟随画面中的主要物体或人移动。

图 17：“清影”使用界面

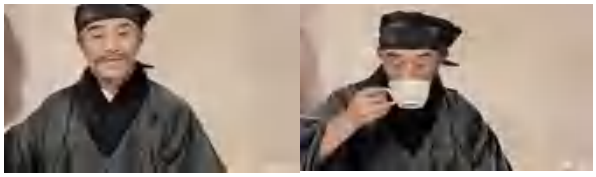
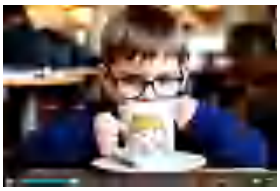


数据来源：智谱 AI

图 18：土豆逐渐变成薯条，贴合 Prompt 中“广告感”要求

数据来源：智谱 AI

表 7 :清影基于智谱自研的视频生成大模型 CogVideoX 的核心技术特点

核心技术特点	具体内容	成果图
内容连贯性提升	智谱 AI 自主研发了一套高效的三维变分自编码器结构（3D VAE），将原始视频数据压缩至原始大小的 2%，显著降低了训练成本和难度。结合 3D RoPE 位置编码模块，该技术强化了时间维度上帧间关系的捕捉，建立了视频中的长期依赖关系。	
可控性增强	智谱 AI 打造了一款端到端的视频理解模型，能够为大量视频数据生成描述，从而增强了模型对文本的理解和对指令的遵循能力，确保生成的视频更加符合用户需求，并能处理超长复杂的 prompt 指令。	
多维度融合的 Transformer 架构	采纳了一种将文本、时间、空间三维一体融合的 transformer 架构，通过 Expert Block 实现文本与视频模态的对齐，并通过 Full Attention 机制优化模态间的交互效果。	 无结构版本  有结构版本

数据来源：清影，国泰君安证券研究

作为国内 AI 大模型独角兽公司，智谱的产品之前已涵盖文本、写作、图片、代码等多形态。智谱已经构建了健全的大模型产品矩阵，包括 GLM-4 系列模型、CodeGeeX、CogView、CogVLM 等，旗下 AIGC 产品包括智谱清言、写作蛙、智谱手语等，不仅服务于普通用户，也适用于大模型技术极客、专业工程师以及寻求专业大模型技术服务的企业。

图 19：智谱 AI 产品矩阵

数据来源：智谱 AI、国泰君安证券研究

2.5. Runway Gen-3 Alpha：视频领域“老将”，继续画质领跑

Runway Gen-3 Alpha 保持一贯的“高质量”水准。7 月 2 日，Runway Gen-3 Alpha 正式向用户开放使用，生成速度快、质量高，截止 7 月底版本，Gen-3 Alpha 的“文生视频”可以提供如 Motion Brush、Advanced Camera Controls、Director Mode 等功能，生成的视频画幅比例支持 16:9。

2024 年 7 月 31 日，Gen-3 Alpha 正式推出图生视频功能，支持生成的视频最长为 11 秒。这一功能让用户能够轻松将任意图像转化为视频的首帧，无论是单独使用图像还是结合文本提示，都能迅速生成视频。

图 20：Gen-3 Alpha 可在 60-90 秒内生成 5 或 10 秒 720p 视频

数据来源：Runway

图 21：Gen-3Alpha 生成视频保持了质量高、细节精细的特点

数据来源：Runway

表 8: Gen-3Alpha 提供的功能及效果

功能	具体内容
Motion Brush	用户可以涂选特定区域或主题，选择运动方向和调节运动强度，从而对生成的内容拥有更多的控制。
Advanced Camera Controls	用户可以通过小数调整摄像机移动，以获得更精准和有意图的专业摄影效果。
Director Mode	允许用户调整镜头角度、运动轨迹、光照效果等，实现专业级的视频制作效果。

数据来源：Runway，国泰君安证券研究

2.6. Luma Dream Machine：3D 资产经验助力“AI 视频”拓展

Dream Machine 视频质量及生成速度俱佳。2024 年 6 月 13 日，Luma AI 首发其视频生成模型 Dream Machine，可以通过文字或图片生成高质量的逼真视频。此外，API 对全球免费开放，每个用户每月有 30 次免费生成的额度，每条视频时长为 5 秒。

1）从视频质量上，Dream Machine 可生成堪比电影级别的视频效果，在画

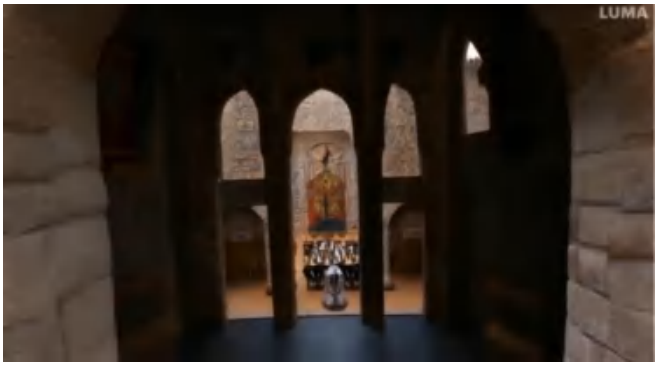
面的清晰度、色彩的饱和度和及光影的变化俱佳。与此同时，Dream Machine 还支持自由变换摄像机视角，实现追踪、环绕和俯视等效果。

2）从生成速度上，Dream Machine 能够在 120 秒内生成一个包含 120 帧的高质量视频，目前单个视频最长为 5 秒，用户等待时间大大缩短，创作效率提升。

图 19：Dream Machine 可生成堪比电影级别的视频效果

数据来源：Dream Machine

图 20：Dream Machine 可生成优质的穿梭镜头视频



数据来源：Dream Machine

Dream Machine 新增“关键帧”和“Loops”功能。6 月 29 日，Dream Machine 新增“关键帧”功能，并向所有用户免费开放使用。关键帧相当是一个视频生成可视化控制功能，可以帮助大模型框定生成的内容，同时在文本的引导下生成内容的效果、场景切换、运镜也更精准，极大地满足了媒体制作人员的需求。7 月 22 日，Luma AI 正式推出了其 Dream Machine 平台的新功能“Loops”。此创新举措显著提升了内容创作与数字营销的便捷性，用户得以轻松构建无限循环的视频内容，无需顾虑画面中的剪辑痕迹或过渡不连贯问题。

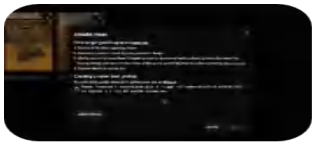
表 9: Dream Machine 的主要功能和特点

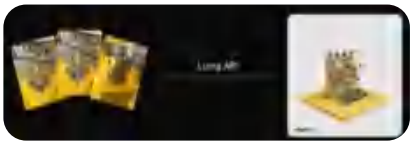
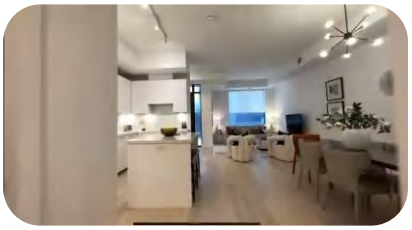
功能	具体内容
关键帧	允许用户仅需上传初始和结束的两张图片，并简要描述所需生成的特效效果，即可便捷地创作出迷你影片。
Loops	使得用户能够利用文本描述、图片或关键帧来生成无缝且连续的视频循环。
物理模拟	支持物理模拟，确保生成的视频在时间和空间上保持连贯性、一致性，如重力下落、碰撞、光影变化等。
角色一致性	视频中的人物、动物或物体在不同时间、不同场景下都能保持高度一致性，解决了 AI 生成视频中常见的“跳跃感”问题。
运镜流畅自然	能够根据场景和情绪需要，自动生成各种电影级别的运镜手法，大大降低了视频创作的门槛

数据来源：Dream Machine，国泰君安证券研究

Luma 有长期的“AI 生成 3D”相关积累。Luma 团队自 2022 年以来就长期从事各类 3D 视频和 3D 资产的生成，在 3D 重建方面积累丰富，这一经验对于团队开发视频生成类 AI 工具有所帮助，特别是在保持一致性方面。

表 10: Luma 的过往产品大多与 3D 资产有关

产品名称	时间	具体功能	效果图
网页版 Luma	2022.10.2	用户可以根据网页中的拍摄指导上传视频素材，生成 3D 视频	

Imagine 3D	2022.12.14	允许用户通过文本描述生成 3D 模型	
NeRF Reshoot	2023.1.7	成为市场上首个提供 NeRF + App 解决方案的公司	
Luma AR	2023.3.21	允许用户在实景照片中标记 AR 视频路线，并自动生成视频	
视频转 3D API	2023.3.27	使开发人员能够将 Luma 的 3D NeRF 模型集成到其他应用程序中	
Unreal Engine plug-in v0.3	2023.7	引入质量控制，提取 NeRF 的特定区域，微调渲染质量以匹配特定用例	
视频生成 APP Flythroughs	2023.8	可以模拟生成无人机一镜视频	

数据来源：Luma，国泰君安证券研究

3. 语音功能：或助推 AI 产品迭代

3.1. 以 GPT-4o 语音为代表，相比传统 TTS 信息更多

3.1.1. GPT4-o：无延迟对话、理解和表达情感

根据官方演示效果，GPT-4o 能实现：

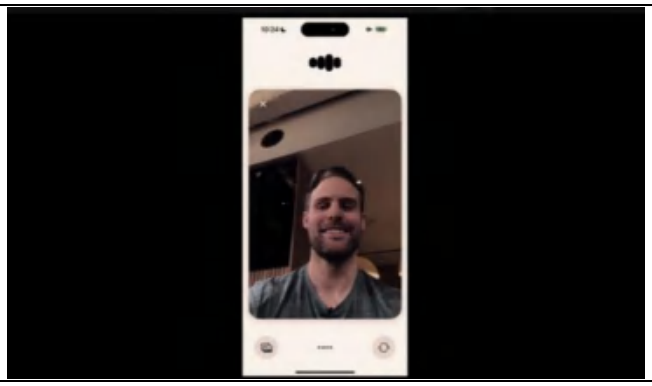
1）实时响应：GPT-4o 能够在最短 232 毫秒、平均 320 毫秒的时间内响应音频输入，反应速度几乎与人类对话时相同，实现用户的自然交流。与演示者即时交流；

2）真实情感模拟：GPT-4o 具备识别用户情绪变化的能力，甚至能够察觉人的喘息声和呼吸频率。此外，它还能表现出类似人类的情感反应，对于愉悦、悲伤和愤怒情绪采用合适的语气回应；

3）通过摄像头实时捕捉和分析环境中的视觉信息：GPT-4o 根据摄像头输入实时互动解答问题，无需传统的上传图片步骤，实现直接通过打开摄像头来实时观察周围发生的事情。

相比下，之前的 GPT 只能进行单轮次的语言对话、单张照片输入，也无法理解和表达语言情绪，语音沟通的实现是通过“语音转文字”、“文字理解（GPT4）”、“文字转语音”的方式进行文本信息的处理。GPT-4o 所有输入和输出由同一个神经网络处理，能够同时理解文本、图像、音频等，并能将其任何组合作为输入或输出。

图 21： GPT-4o 可理解用户表情及情绪变化



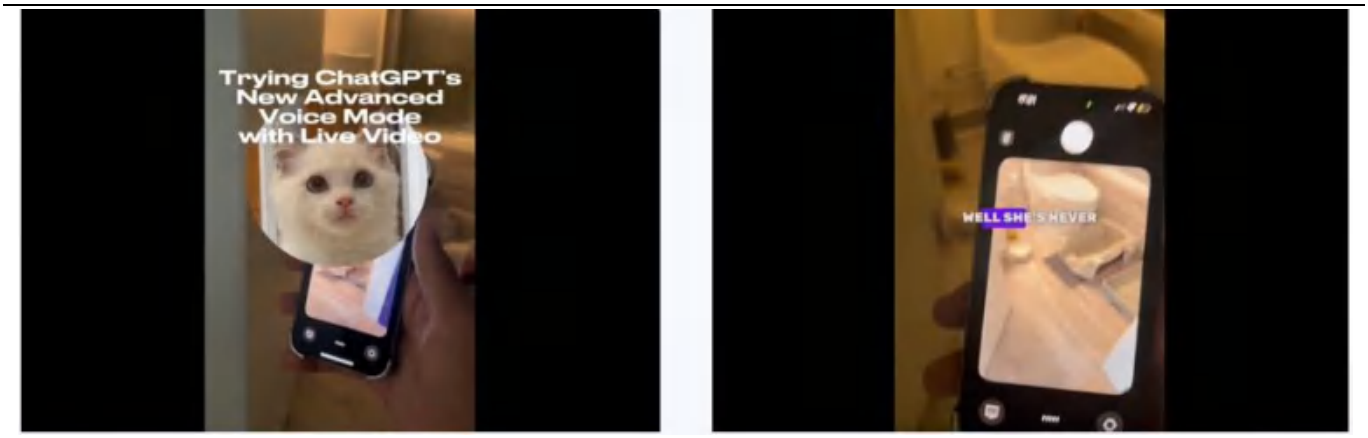
数据来源：OpenAI

图 22： GPT-4o 实时观看并解答数学问题

数据来源：OpenAI

GPT-4o 高级语音、视频等功能开始测试，语音功能后续有望向所有付费用户开放。北京时间 2024 年 7 月 31 日凌晨， OpenAI 宣布开始向一小部分 ChatGPT Plus 用户推出高级语音模式，根据测试用户反馈来，部分用户利用 GPT-4o 进行口语练习，GPT-4o 将针对用户发音进行实时评分，多种语言测试下都有稳定表现；情感方面，在对 GPT-4o 讲笑话时，它将提供笑声陪伴，及时给予情绪 反馈；GPT-4o 能实现在讲故事的同时创建背景声，增加沉浸感； 有用户结合视频功能向 GPT-4o 展示了宠物猫的情况，GPT-4o 也能够积极回应。本次测试将主要搜集安全、功能方面的反馈，随后，OpenAI 还会发布视频和屏幕共享新功能。语音功能预计今年秋天会向所有 ChatGPT Plus 用户开放。

图 24：GPT-4o 对用户展示的宠物猫给予积极回应

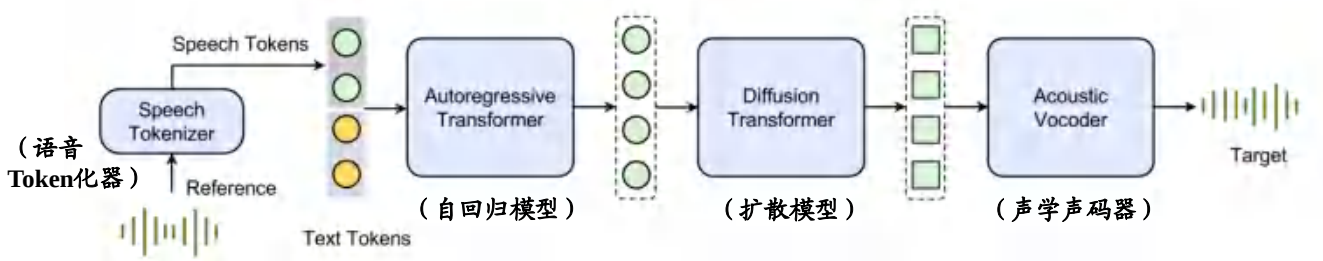


数据来源：APPSO 公众号

3.1.2. 字节跳动 Seed-TTS：可在表现力上接近人类水平

Seed-TTS 是一种基于自回归 Transformer 的模型。2024 年 6 月，字节跳动推出 AI 模型 Seed-TTS，其能够合成自然度和表现力达到人类水平的语音。Seed-TTS 模型的工作包含四个模块：1）语音 token 化器将语音信号转换成语音 token 序列，然后基于其训练一个 token 语言模型。2）在推理期间，该模型则以自回归方式生成语音 token。3）生成的 token 交由扩散模型处理，以增强其声学细节。4）其输出后，再经过声学声码器处理，得到最终声波波形。

图 25： Seed-TTS 包含四个主要模块



数据来源：字节跳动、机器之心

Seed-TTS 自然度、表现力、稳定性更佳。相比之前的模型，Seed-TTS 有两大优势：

1)Seed-TTS 合成的语音都有更好的自然度和表现力，考虑多种不同场景(包括怒吼、哭喊、声情并茂演讲等高难度场景)。使用 t-SNE 绘制 25 个说话人的真人语音和合成语音的说话人嵌入，结果显示来自同一说话人的真人语音与合成语音紧密地聚类在一起，证明 Seed-TTS 的语音生成质量很好，并且与真人语音很相似。

2)Seed-TTS 解决了基于语言模型的 TTS 系统普遍存在的不稳定问题。Seed-TTS 在稳定性上的卓越表现得益于 token 和模型设计的提升、改进过的训练和推理策略、数据增强和强化学习后训练。因此，Seed-TTS 在测试集上的表现出了显著更优的稳健性。

图 26：来自同一说话人的真人语音与合成语音紧密聚类

数据来源：字节跳动

3.1.3. ChatTTS：流畅语音合成，可预测和控制细粒度的韵律特征

2024 年 5 月，ChatTTS 文本转语音项目在 GitHub 上引起极大关注，最大模型使用了超过 10 万小时的中英文数据进行训练。在 HuggingFace 中开源的版本为 4 万小时训练且未经特定领域微调(SFT)的版本。其能够产生中英文语音、复刻逝去的人的绝版声音、支持多说话人。此外，其在韵律方面超过大部分开源模型支持细粒度控制，并允许加入笑声、说话间停顿及语气词。

图 27： ChatTTS 生成语音支持中英文混说

数据来源：哔哩哔哩

ChatTTS 可自动为文本生成韵律及停顿，但处理长文本能力略有欠缺。实际使用时，在文本框内输入文本后，ChatTTS 会自动生成韵律和停顿，还会加入一些“然后”之类的语气词。如果在输入时在文本中加入 [laugh] 和 [uv_break]，就能手动控制 ChatTTS 在说话间产生笑声。但是，当前 ChatTTS 处理长文本能力略有欠缺，生成超 30 秒音频时需要手动修复，此外，在处理长文本时分词也会出现问题。

图 28：输入[laugh]和 [uv_break]可手动产生控制生成的语音

数据来源：机器之心

3.2. 应用端：可显著提升教育和情感陪伴应用体验

3.2.1. 口语等教学场景质量有望提升

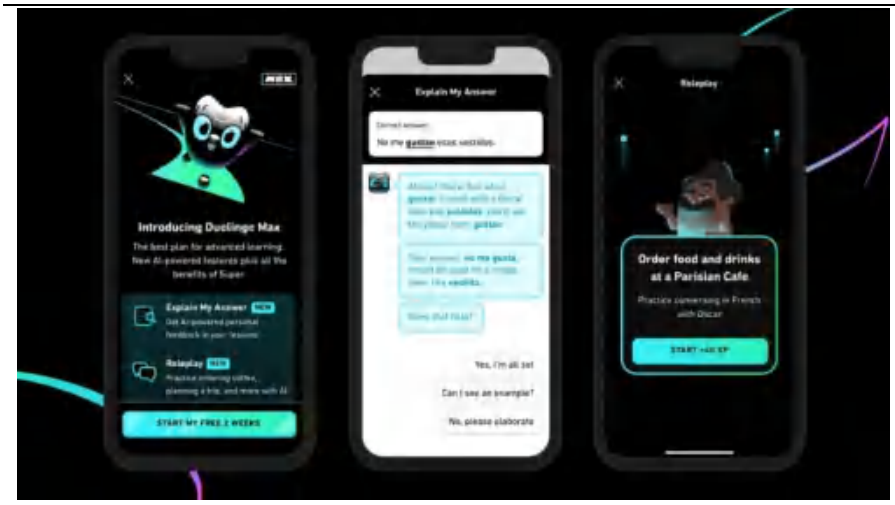
多邻国：双智能体协同作用，提供沉浸式学习体验。2023 年 3 月，多邻国嵌入 GPT-4 后新增两项功能：

1）Role Play(角色扮演)：学习者可以在与虚拟角色对话中练习口语及听力

技能，互动后，学习者将从虚拟角色处得知他们回答的准确性和复杂性、未来对话的建议等反馈。这种按需支持的对话模式能给学生们提供即时、准确的帮助。

2） Explain my Answer（解释我的答案）: 在某些练习类型后，学习者可以通过点击钮，了解为什么他们的答案正确或错误的简单解释，并可以请求示例或进一步的澄清。

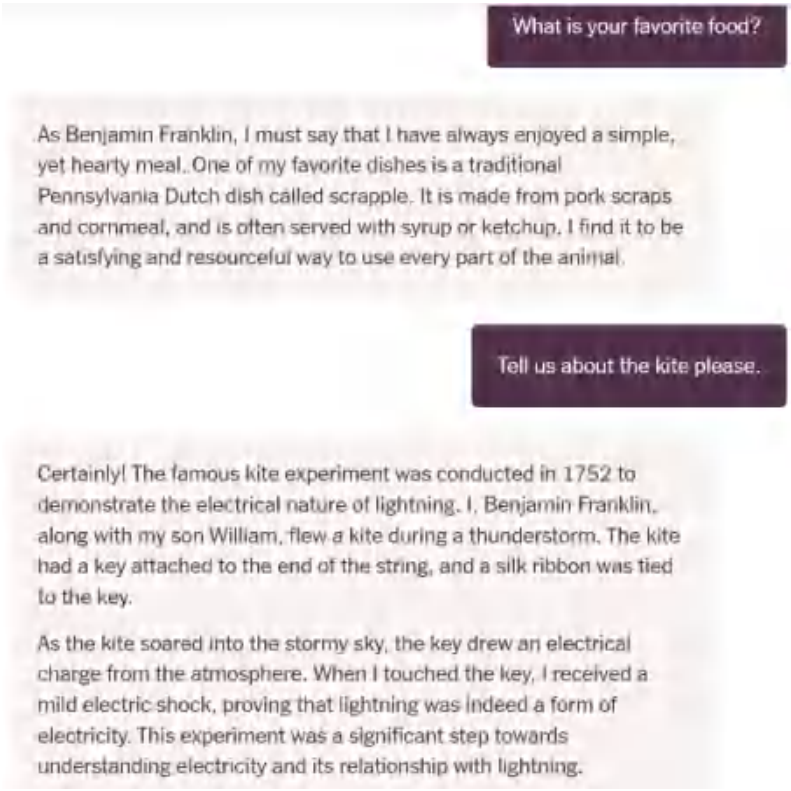
图 27: 多邻国双智能体提供沉浸式学习体验



数据来源：多邻国

可汗学院 Khanmigo: 实现学生与虚拟人物对话。可汗学院(Khan Academy) AI 在线教育工具 Khanmigo 可以作为学生的虚拟导师，也可以作为教师的课堂助手，帮助学生学习数学、科学和人文等课程。此外，Khanmigo 还可以让学生与历史人物或虚构角色进行虚拟对话，增强他们的兴趣和理解。

图 30: 可汗学院 Khanmigo 实现与虚拟人物对话



数据来源：Khanmigo

网易有道 Hi Echo ：对用户口语练习进行实时反馈与点评。Hi Echo 是由网易有道推出的 AI 英语学习应用，利用国内首个教育大模型“子曰”技术，提供 24 小时在线的虚拟人口语教练服务。2024 年 5 月 29 日，网易有道推出“英语口语练习神器”Hi Echo3.0，此次更新后，用户练习口语时将展现实时评分结果，以便及时了解自己的发音及语法问题，同时还能获得该句详细点评及地道润色。

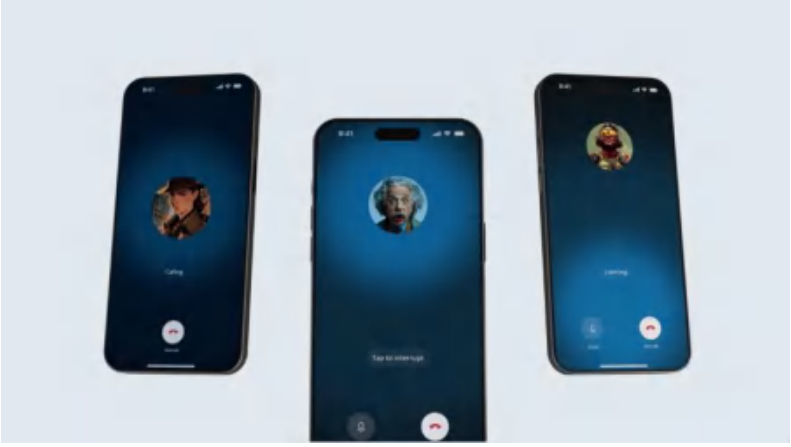
图 31： Hi Echo3.0 展现实时用户口语评分结果

数据来源：Hi Echo

3.2.2. 情感陪伴：有望增加情感认同及潜在付费点

Character AI 的“智能 NPC 语音通话”功能广受用户好评。2024 年 3 月，情感陪伴应用 Character.AI 推出了“角色声音”功能，用户可以在一对一聊天时听到角色说话。2024 年 6 月 28 日，Character.AI 宣布允许用户与人工智能角色通话。该功能目前支持多种语言，包括英语、西班牙语、葡萄牙语、俄语、韩语、日语和中文。在该功能测试期间，就有 300 多万用户拨打了 2000 多万个电话。用户只需轻点按钮，就能直接与用户生成的人工智能角色发起通话，未来，该功能可用于练习语言技能、模拟面试，或将其加入角色扮演游戏的玩法中。随着 GPT4-o 语音功能的实现，有望增加情感陪伴应用的情感认同及付费点。

图 32： Character AI 推出用户与人工智能角色通话的功能



数据来源：Character AI

4. 投资建议

继续看好 AI 技术发展对内容产业的推动作用，随着 AI 大模型开源化、轻量化，以及视频和语音等模态的快速进步，部分应用场景有望发生变化，可沿如下思路进行布局：

1）游戏等应用改造，推荐吉比特、恺英网络、完美世界、美图公司，受益标的腾讯控股、网易、快手、巨人网络；
2）教育赛道，受益标的南方传媒、皖新传媒、世纪天鸿；
3）情感陪伴与社交，受益标的昆仑万维、盛天网络。

表 11：推荐公司盈利预测与估值情况

代码	简称	股价（元）	市值（亿元）	EPS（元/股）			PE			评级
				2023A	2024E	2025E	2023A	2024E	2025E	
002624.SZ	完美世界	7.95	154.23	0.26	0.56	0.66	30.58	14.20	12.05	增持
002517.SZ	恺英网络	9.66	207.93	0.70	0.97	1.13	13.80	9.96	8.55	增持
603444.SH	吉比特	184.70	133.06	15.63	16.14	18.98	11.82	11.44	9.73	增持
1357.HK	美图公司	2.11	95.91	0.09	0.11	0.17	23.50	19.18	12.35	增持

注：数据截止 2024 年 8 月 6 日收盘
资料来源：Wind，国泰君安证券研究预测

5. 风险提示

AI 应用推进放缓，AI 相关商业化落地不及预期，生成式内容监管风险。

本公司具有中国证监会核准的证券投资咨询业务资格

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

免责声明

本报告仅供国泰君安证券股份有限公司（以下简称“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本公司的当然客户。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌。过往表现不应作为日后的表现依据。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保持在最新状态。同时，本公司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。

本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。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。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也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。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

本公司利用信息隔离墙控制内部一个或多个领域、部门或关联机构之间的信息流动。因此，投资者应注意，在法律许可的情况下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的情况下，本公司的员工可能担任本报告所提到的公司的董事。

市场有风险，投资需谨慎。投资者不应将本报告作为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向专业人士咨询并谨慎决策。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息或进而交易本报告中提及的证券。本报告不构成本公司向该机构之客户提供的投资建议，本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

评级说明

投资建议的比较标准

投资评级分为股票评级和行业评级。

以报告发布后的 12 个月内的市场表现为比较标准，报告发布日后的 12 个月内的公司股价（或行业指数）的涨跌幅相对同期的沪深 300 指数涨跌幅为基准。

	评级	说明
股票投资评级	增持	相对沪深 300 指数涨幅 15%以上
	谨慎增持	相对沪深 300 指数涨幅介于 5%～15%之间
	中性	相对沪深 300 指数涨幅介于-5%～5%
	减持	相对沪深 300 指数下跌 5%以上
行业投资评级	增持	明显强于沪深 300 指数
	中性	基本与沪深 300 指数持平
	减持	明显弱于沪深 300 指数

国泰君安证券研究所

	上海	深圳	北京
地址	上海市静安区新闸路 669 号博华广场 20 层	深圳市福田区益田路 6003 号荣超商务中心 B 栋 27 层	北京市西城区金融大街甲 9 号 金融街中心南楼 18 层
邮编	200041	518026	100032
电话	（021）38676666	（0755）23976888	（010）83939888
E-mail:	gtjaresearch@gtjas.com		