核心观点

- **近期重点关注腾讯全新的C端AI应用腾讯元宝和6月端侧AI密集落地：**
- **1、腾讯元宝发布，覆盖AI搜索/总结/写作等多功能。**其中AI搜索独家覆盖高质量的微信公众号内容，AI总结具备类似Kimi的长文本能力。此次发布反映腾讯正式发力面向C端的AI大模型产品，有望首先在微信中形成社交裂变，快速积累大量用户群体。
- **2、6月端侧AI密集落地，关注苹果WWDC大会对AI应用的催化。**据彭博社，苹果将发布集成一系列AI功能的iOS18、具备实时语音交互能力的Siri2.0和AI应用商店，其中Siri2.0有望控制所有App，从而成为 每一个App的智能助手。
- 相关公司：
- 1）苹果iOS 18在海外市场，将率先升级AI功能，海外AI应用若接入苹果生态有望受益。在此，我们建 议关注有海外AI应用布局的国内公司，包括：汤姆猫（AI手游）、万兴科技（Filmora）、昆仑万维（ Linky、Opera）、焦点科技（AI麦可）、易点天下（KreadoAI）、美图（X-Design）。
- 2）Copilot+PC等端侧AI设备陆续发布，边缘端算力提升，端侧应用丰富度和使用体验有望提升，端 侧AI应用相关公司或将受益，包括：万兴科技、视觉中国（投资标的AiPPT.cn接入联想AI PC）。

核心观点

- 此外，AI行业应用端和模型端均出现显著变化：
- **1）B端API卷价格。**国内阿里、腾讯、百度、科大讯飞等公司大模型官宣降价，降幅最高超90%；海 外轻量级模型GPT-4o和Gemini 1.5Flash相比普通版本，价格降幅最高50%。API成本下降，推动AI应用爆发。我们拆解Perplexity AI的财务模型，假设大模型API成本降至目前的5%，全年搜索成本有望下降 33%，推动毛利润转正。
- **2）C端产品拼流量。**目前国内AI应用的用户渗透率不足5%，3日留存率在50%以下，C端AI应用通过 广告投放，寻求流量破圈。其中Kimi、豆包投放规模最大。从增速看，今年B站的AI广告是去年同期3-4倍；从投放方式看，已从普通的短视频、直播、搜索投放，拓展至达人和用户运营。
- **3）技术革新：原生多模态模型兴起，实现低延迟语音交互。**继Gemini后，GPT-4o成为又一重磅的原生多模态模型，通过统一处理多种模态数据，实现低延迟人机交互，有望提升AI应用在陪伴/社交、游戏NPC、教辅、实时翻译的体验。
- **相关公司：**
- **1）应用：昆仑万维、值得买、掌阅科技、世纪天鸿、焦点科技、易点天下、盛天网络、汤姆猫、南方传媒等。**
- **2）语料侧：值得买、视觉中国、华策影视、捷成股份、中文在线、中国科传、中国出版等。**

目录

1.1.1 腾讯元宝发布，覆盖AI搜索/总结/写作多功能

- **腾讯上线元宝AI。**腾讯基于混元大模型，推出AI助手“腾讯元宝App”，具备AI搜索、总结、写作等效率功能，以及绘画、创意头像、AI识图等休闲娱乐功能。截至5月31日上午10点，该产品排名iOS中国区效率（免费）榜第9名。
- **腾讯正式发力面向C端的AI大模型产品。**据腾讯科技，此前混元大模型主要接入现有产品。目前有超过600各腾讯的业务和场景使用大模型重构自己的产品，混元日调用量达2亿。近期已推出智能体平台【元器】和首个原生C端应用【元宝】，意味着混元大模型正式直接拓展C端场景。

腾讯元宝App

腾讯内部已有600+业务接入混元

资料来源：七麦数据，腾讯科技，中信建投

1.1.1 腾讯元宝发布，覆盖AI搜索/总结/写作多功能

- **腾讯元宝的AI搜索独家覆盖微信公众号内容，AI总结提供类似Kimi的长文本能力：**
- **1）AI搜索：**元宝的AI搜索接入了微信搜一搜和搜狗搜索，搜索资料来源包括微信公众号、腾讯新闻、搜狗等。微信公众号具有大量高质量内容，据新榜，23年微信公众号产生超4.48亿篇文章，阅读数10万+文章超25万篇，已成为众多用户获取信息的首选渠道。
- AI搜索功能为微信公众号创作者带来更多的曝光和流量；开启付费功能的内容也会作为搜索的资料来源，腾讯元宝不会展示完整内容，但用户可以通过资料来源链接打开相应内容，从而提高优质内容的可见度，为创作者带来潜在的收益。

腾讯元宝App的AI搜索功能

询问“如何看待AI搜索的市场空间”，AI对搜索结果进行结构化输出

资料来源主要是公众号内容，包括我们团队的AI搜索深度报告

结果末尾提供内容推荐和相关搜索关键词，推荐了我们团队的AI搜索深度报告

资料来源：腾讯元宝，中信建投

1.1.1 腾讯元宝发布，覆盖AI搜索/总结/写作多功能

- **腾讯元宝的AI搜索独家覆盖微信公众号内容，AI总结提供类似Kimi的长文本能力：**
- **2）AI总结：**具备超长的上下文窗口。支持图片、文件、语音等多种输入格式，能一次性解析最多10个PDF/word/txt文件，或总结多个网页内容。用户看到长篇公众号内容后，可以将链接发给元宝，让元宝进行总结。

腾讯元宝App的AI总结功能

我们让AI基于【上海电影】深度报告，回答问题：请总结上海电影IP授权业务的竞争优势和市场空间

答案逻辑清晰、完整度和准确度较高，性能与Kimi相近

资料来源：腾讯元宝，中信建投

1.1.2 怎么看腾讯元宝对其他AI助手产品的影响？

- **1、国内AI应用渗透率低，各个产品均处于破圈期。**据腾讯科技、Quest Mobile，目前中国移动互联网用户规模12.32亿，其中大学生、职场用户等核心用户规模数千万；而国内AI头部产品日活跃用户规模仅数百万，在移动互联网全体用户、核心用户中的渗透率仅1%、10%。

- **2、不同产品进行差异化竞争，并非彼此替代关系：**头部产品中，秘塔AI搜索和Kimi的访问量主要来自网页端，凭借优质的AI性能、更结构化的输出结果（例如秘塔可以自动生成思维导图、表格等），已在办公、研究场景中建立了高粘性的用户群体和口碑；天工AI搜索具备搜索、音乐、写作、智能体等丰富的产品功能，应用场景广阔，近期日活突破百万。而腾讯元宝是基于腾讯生态的全新AI助手，独家覆盖微信公众号等优质数据源，有望成为微信用户的Copilot。

- **3、对标豆包，看好腾讯元宝依托微信生态快速破圈。**豆包是字节体系内孵化出的AI助手，通过抖音、今日头条等产品引流，已稳定成为国内用户规模最大的AI助手类App，并拓展了大量下沉用户。元宝与微信生态紧密结合，有望首先在微信中形成社交裂变，快速积累大量用户群体。

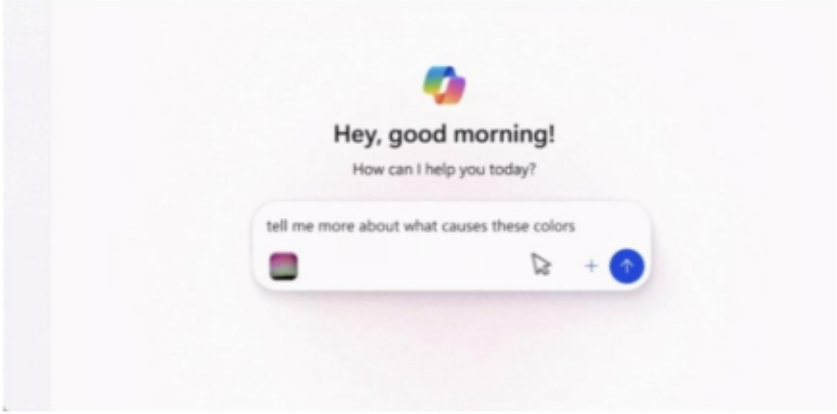
- **相关公司：**
- **值得买：**据腾讯科技，6月元器平台的智能体将允许分发至微信公众号和小程序。值得买具有稀缺的消费数据储备，已向Kimi开放数据API接口，联合推出AI导购智能体，并与百川、智谱等其他大模型公司接触。微信小程序具有庞大的用户基础，有望为值得买的AI导购智能体提供新的渠道。

目录

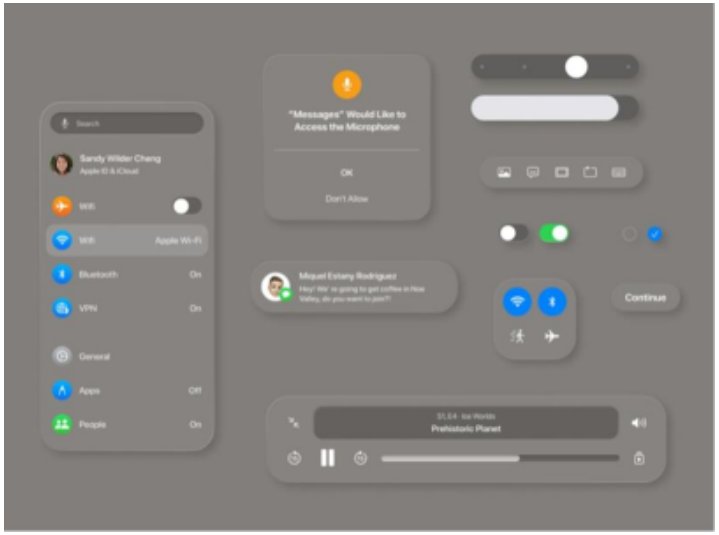
1.2 WWDC大会举办在即，6月端侧AI密集发布

- **微软发布Copilot+ PC，将于6月18日上市 。**5月21日微软发布AI电脑Copilot+PC，具备超40TOPS的AI算力，无缝接入GPT-4o等大模型。电脑配备AI Agent，只需在键盘轻点Copilot按键即可完成交互，包括查找使用记录、生成并优化AI图 像、实时字幕等。
- **苹果WWDC 大会将于6月10-14日举行，旨在展示 iOS、iPadOS、macOS等前沿创新。**据IT之家、CNET、彭博，苹果或将发布iOS18、Siri2.0和全新的AI应用商店。

微软发布Copilot+ PC



iOS 18预计采用类visionOS的设计元素



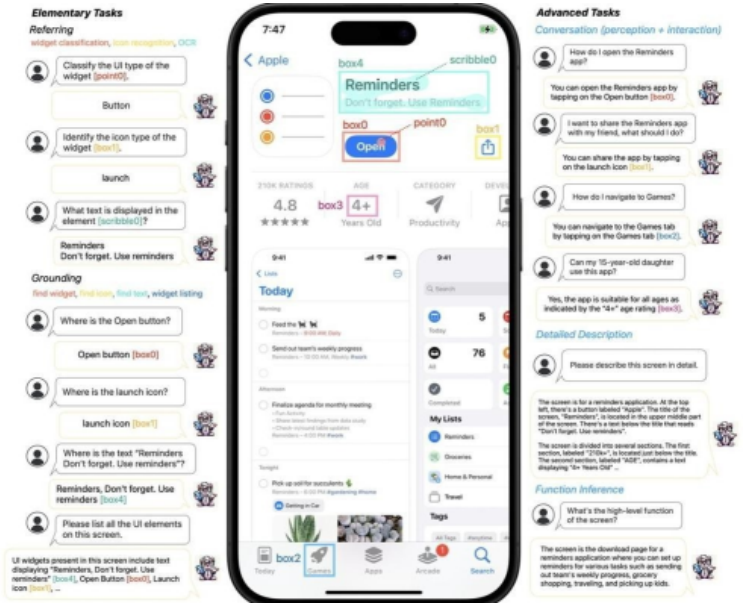
资料来源：微软官网，IT之家，中信建投

1.2 苹果或将发布iOS18和Siri 2.0

- **苹果此前已在AI方面进行大量布局。**今年3月苹果通过论文发布300亿参数的多模态大模型MM1，综合使用图像字幕、交错图像文本和纯文本数据三种不同类型的训练数据，提高其图像识别和推理能力。4月开源小模型Phi-3-mini，性能 与GPT-3.5相近，适合端侧部署。同时发布多篇端侧AI相关论文，包括如何将大模型内嵌入iPhome中。我们预计苹果 AI手机有望同时接入端侧小模型和云端大模型：
- 1）端侧小模型：预计采用自研方式，实现大模型与内存、功耗等硬件技术的高度调优；
- 2）云端大模型：据彭博，已与OpenAI达成合作，正在与谷歌Gemimi谈判，预计GPT和Gemini系列模型都有望接入。

苹果发布多篇端侧AI相关论文

AI放大图像至任何分辨率，便于识别



提取屏幕中的文字信息，便于AI响应用户需求

资料来源：36氪，中信建投

1.2 苹果或将发布iOS18和Siri 2.0

- **苹果或将在WWDC大会上发布iOS18、Siri2.0和AI应用商店。据彭博社：**
- **1）iOS18：**将集成一系列AI功能，包括语音备忘录支持语音转文字、AI修照片、Spotlight搜索更快速准确、改进Safari
  网页搜索、根据文本内容生产全新emoji、自动生成邮件和短信的回复内容、智能回顾等。
- **2）Siri2.0：**集成AI能力，支持实时语音交互，或将能控制手机中所有App。
- **3）AI应用商店：**AI应用的集中获取渠道。

- **相关公司：**
- **1）苹果iOS 18在海外市场，将率先升级AI功能，海外AI应用若接入苹果生态有望受益。在此，我们建议关注有海外AI应用布局的国内公司，包括：汤姆猫（AI手游）、万兴科技（Filmora）、昆仑万维（Linky、Opera）、焦点科技（AI麦可）、易点天下（KreadoAI）、美图（X-Design）。**
- **2）Copilot+PC等端侧AI设备陆续发布，边缘端算力提升，端侧应用丰富度和使用体验有望提升，端侧AI应用相关公司或将受益，包括：万兴科技、视觉中国。**
- 其中视觉中国21年已战略投资爱设计&AiPPT.cn，近期作为领投方，再次参与其数千万元的B1轮融资。AiPPT.cn作为联想AI智能体小天的唯一官方AIPPT应用，直接触达联想AI PC用户。

资料来源：36氪，中信建投

目录

1.3.1 B端卷成本：大模型API降价潮开启，降幅最高超99%

- **近日国内四款大模型发布API降价公告。**阿里通义千问、百度文心、讯飞星火、腾讯混元大模型均发布API降价或免费措施，具有两大共同点（**四大模型的完整API价目表详见附录**）：
- **1）部分入门级API免费，高性能级API降价幅度较低：**百度文心、讯飞星火、腾讯混元的入门级API价格均降至0，千亿参数通义2.5、百度文心3.5和4.0系列、讯飞星火Spark3.5 Max等性能较高的API降价幅度较小；
- **2）输入token的降价幅度大于输出token：**主要是用户对输入token的调用量更大。据财经网统计，输入调用量一般是输出的8倍，因为很多用户会结合长文本对模型提问。

国内大模型API降价公告

资料来源：阿里云公众号，百度公众号，科大讯飞公众号，腾讯云公众号，中信建投

1.3.1 B端卷成本：大模型API降价潮开启，降幅最高超99%

- **海外厂商也推出了API价格更低的轻量级模型。**OpenAI的GPT-4o和谷歌Gemini 1.5 Flash，分别作为GPT-4 Turbo和 1.5Pro的轻量级版本，价格变为50%和10%。
- 梳理海内外主流模型的API价格，目前每百万token的输入价格已从100元附近下探至10月以内，降幅近90%。

主流大模型API价格与性能

模型	输入	输出	中文综合能力 AlignBench	英文综合能力（MT-Bench）
	元/百万tokens			
DeepSeek-V2	1	2	7.91	8.97
Claude 3 Haiku	2	9	6.42	8.39
Gemini 1.5 Flash（Prompts<=12.8万个）	3	4	-	-
豆包通用模型pro-128k	5	9	-	-
Qwen1.5 72B（通义千问）	5	5	7.19	8.61
Gemini 1.5 Flash（Prompts>12.8万个）	5	8	-	-
abab-6.5s（MiniMax）	10	10	7.34	8.69
GPT-3.5	11	14	6.08	8.21
Mixtral 8x22B	14	43	6.49	8.66
Claude 3 Sonnet	22	109	6.70	8.47
Moonshot-v1（月之暗面）	24	24	7.22	8.59
Gemini 1.5 Pro（Prompts<=12.8万个）	25	76	7.33	8.93
LLaMa 3 70B	27	82	7.42	8.95
abab-6.5（MiniMax）	30	30	7.97	8.82
GPT-4o	36	108	-	-
Gemini 1.5 Pro（Prompts>12.8万个）	51	152	7.33	8.93
GPT-4 Turbo-1106	72	217	8.01	9.32
GLM-4（智谱清言）	100	100	7.88	8.60
Claude 3 Opus	109	543	7.62	9.00
ERNIE-4.0（文心一言）	120	120	7.89	7.69
GPT-4-0613	217	434	7.53	8.96

API输入价格增加

资料来源：各产品官网，中信建投

1.3.1 B端卷成本：以Perplexity为例，API降价贡献利润弹性

- **以Perplexity为例，假设大模型API成本降至目前的5%，合计搜索成本有望下降33%，推动毛利润转正。**据新浪财经，目前Perplexity年收入约2000万美元，以200美元/年的订阅价格测算，付费用户数约10万人。Perplexity的成本分为两部分： 1）搜索引擎API：主要调用Bing的API，官网价格为每1000次10-20美元；2）大模型API成本：默认模型预计系基于 GPT-3.5 Turbo的微调模型，订阅会员可选GPT-4 Turbo等高端模型，由此计算出单次搜索成本约0.02美元。
- 假设去年搜索次数由23年的5亿次增加至10亿次，对应搜索成本2300万美元，毛利润仅-314万美元。若大模型API成本降 至目前的5%，预计全年搜索成本下降33%，推动毛利润转正至459万美元。

API价格下降对Perplexity毛利润的影响测算

	API降价前	API降价后	备注
收入			
付费用户数（万）	10	10	目前年度经常性收入为2000万美元，考虑B端业务
ARPPU（美元/年）	200	200	近期才推出，收入主要依靠C端业务。按订阅价格
年收入（万美元）	2000	2000	200美元/年（20美元/月），付费用户数约10万人。
成本			
搜索引擎API（美元/次）	0.0150	0.0150	Bing Search API的价格为10-20美元/1千次
输入token	50	50	1）默认模型预计系基于GPT-3.5 Turbo的微调模
输入成本（美元/1k token）	0.0053	0.0003	型，订阅会员可以选择GPT-4 Turbo的高端模型。
输出token	500	500	假设综合API成本为GPT-3.5 Turbo和GPT-4 Turbo
输出成本（美元/1k token）	0.0158	0.0008	的平均值
大模型API成本（美元/次）	0.0081	0.0004	2）假设大模型API成本降至目前的5%
单次搜索成本（美元/次）	0.0231	0.0154	
全年搜索次数（亿次）	10	10	23年搜索次数5亿次，预计今年翻倍
合计搜索成本（万美元）	2314	1541	合计搜索成本下降33%
毛利润（万美元）	-314	459	毛利润转正

资料来源：新浪财经，微软官网，OpenAI官网，Multiplatform.AI，Exploding Topics，中信建投