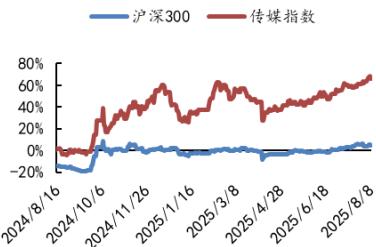


海外大模型新版本发布，关注 AI 应用

行业评级：增持

报告日期：2025-08-10

行业指数与沪深 300 走势比较



分析师：金荣

执业证书号：S0010521080002

邮箱：jinrong@hazq.com

分析师：刘京松

执业证书号：S0010524090001

邮箱：liujs@hazq.com

分析师：赵亮

执业证书号：S0010525070001

邮箱：zhaoliang@hazq.com

相关报告

- H2O 恢复在华销售，监管约谈外卖平台 2025-07-21
- AI 系列专题跟踪——视频及图像生成模型 2025-07-15
- 模型与应用共振，关注 AI 应用 2025-07-13

主要观点：

● 本周行情回顾

从指数表现来看，本周(2025-08-04 至 2025-08-08)，上证指数周涨跌幅为 2.11%，创业板指周涨跌幅为 0.49%，沪深 300 周涨跌幅为 1.23%，中证 1000 周涨跌幅为 2.51%，恒生科技周涨跌幅为 1.17%，纳斯达克指数周涨跌幅为 3.87%；从板块表现来看，传媒指数周涨跌幅为 1.95%，恒生互联网科技业周涨跌幅为 2.37%，中证海外中国互联网 50 指数周涨跌幅为 1.56%，人工智能指数周涨跌幅为 0.69%。

● OpenAI 发布 GPT-5

8月7日，OpenAI宣布推出人工智能模型GPT-5，该模型在编码、创意写作以及复杂查询推理方面具备更强的能力。GPT-5将在未来一周内分批向所有ChatGPT用户开放，包括免费用户。Plus会员可获得更多使用量，Pro会员则可以访问GPT-5 Pro版本，该版本具有扩展推理能力，能够提供更全面、更准确的答案。GPT-5是OpenAI首个“一体化”的AI系统，整合了GPT系列（大语言模型）和o系列（推理模型），具备调度子模型的能力。作为OpenAI的最强模型，GPT-5在编程能力、创意写作、健康领域等关键领域实现了显著提升。（中华网）

● 微软全线产品接入 GPT-5

OpenAI发布GPT-5模型后，微软随即宣布将其集成至全线Copilot产品体系。具体包括：Copilot对话助手启用上月曝光的智能模式，Microsoft 365 Copilot办公生产力套件全面升级，Azure AI Foundry企业级AI开发平台，GitHub Copilot编程辅助工具性能跃升。此次多平台同步部署标志着GPT-5进入规模化商用阶段，Copilot将率先启用具有突破性的智能交互模式。（腾讯网）

● 投资建议

1. 海外 AI：

8月7日，OpenAI宣布推出人工智能模型GPT-5，该模型在编码、创意写作以及复杂查询推理方面具备更强的能力。GPT-5是OpenAI首个“一体化”的AI系统，整合了GPT系列（大语言模型）和o系列（推理模型），具备调度子模型的能力。OpenAI发布GPT-5模型后，微软随即宣布将其集成至全线Copilot产品体系。8月6日，马斯克透露，xAI公司计划在下周开源Grok 2模型。8月6日，谷歌正式发布Genie 3，这是一款通用型世界模型，能够生成前所未有的多样化交互式环境。建议关注近期海外大模型及AI应用更新进展，相关公司：Meta、Adobe、Microsoft、Apple、Nvidia、AMD、Amazon等。

2. 国内 AI：

腾讯AI Lab全新推出的Cognitive Kernel-Pro，一款全开源、多模块、层次化的智能体框架，为深度研究智能体的开发与训练提供了突破性解决方案。8月4日，腾讯混元发布四款开源小尺寸模型：0.5B、1.8B、4B、7B。这四款模型消费级显卡即可运行，适用于笔记本电脑、手机、智能座舱、智能家居等低功耗场景，支持垂直领域低成本微调。阿里通义千问宣布推出Qwen3-30B-A3B模型的全新版本，命名为Qwen3-30B-A3B-Instruct-2507。相关公司：百度集团、阿里巴巴-SW、美图公司、腾讯控股、美团-W、快手等。

3. 海外半导体:

8月5日，世界半导体贸易统计组织 WSTS 宣布，今年1-6月全球半导体市场规模达3460亿美元，同比增长18.9%，其中，逻辑半导体市场规模提升37%、存储半导体增长20%、传感器增长16%、模拟和微型器件均小幅增长4%，而分立器件和光电器件分别出现4%和0.5%的环比下滑。同日，鸿海2025年7月实现6138.7亿新台币营收，同比增长7.3%，环比增长13.6%。8月7日，美国总统特朗普表示，将对半导体等进口商品征收100%的关税，但会豁免那些将生产迁回美国的公司。8月8日，台积电发布2025年7月营收报告，7月营收3231.7亿元新台币，同比增长25.8%，环比增长22.5%。相关公司：Nvidia、AMD、博通、台积电、美光、英特尔等。

4. 智能驾驶:

8月8日，特斯拉Robotaxi LLC已被德州许可与监管局登记为运输网络公司许可方，与Uber、Lyft等公司并列。美国交通部表示，美国汽车安全监管机构已批准豁免Zoox公司生产的定制无人驾驶汽车的联邦车辆安全标准。相关公司：特斯拉、小米集团-W、小鹏汽车-W、小马智行、比亚迪、Google。

5. 海外消费:

8月7日，Shopify公布2025年第二季度财务业绩。Shopify二季度营收26.8亿美元，同比增长30.7%；商品交易总额(GMV)达到878亿美元，同比增长29%；运营利润为2.91亿美元，净利润从去年同期的1.71亿美元跃升至9.06亿美元；调整后每股收益为0.35美元；自由现金流达4.22亿美元，自由现金流利润率达到16%，连续八个季度保持两位数水平。相关公司包括：Amazon、Shopify、拼多多等。

6. 本地生活:

8月5日美团宣布，为坚定落实“反内卷”，促进行业健康发展，即日起正式启动中小商户发展扶持计划，向中小商户发放助力金，预计到今年年底，美团助力金将新增覆盖超10万家餐饮小店，单店助力金最高5万元。去年12月美团推出首期10亿助力金计划，回访调研结果显示近半数商家反馈“小店订单量显著增长”，四成商家的收入明显提升。建议重点关注主流外卖平台补贴力度情况，相关公司：美团-W、阿里巴巴-W、京东集团-SW。

7. 电商:

淘宝上线全新大会员体系，打通饿了么、飞猪、盒马等阿里系资源，包含购物、外卖、旅行、出行等权益。此外，淘宝APP改版，飞猪被提升至一级入口，与今年重点发力的淘宝闪购并列。此举解决了此前各业务会员体系割裂、数据孤岛的问题，降低了跨业务获客成本，提升了生态内流量互导效率，通过高频业务（如外卖、即时零售）带动低频高价值业务（如旅游、酒店）。相关公司：阿里巴巴-W。

8. 产业互联网:

Canalys的最新数据显示，今年二季度小米手机成功超越苹果，成为欧洲市场份额排名第二的厂商。从增幅来看，小米手机是Top3厂商里，唯一实现了正增长的品牌，市场份额达23%，业务同比增长11%。建议重点关注小米手机在欧洲市场的高端化进展，相

关公司：小米集团-W。

9. 影视：

上周（2025/07/28-08/03），全国电影市场实现票房收入 17.5 亿元，较上上周 11.88 亿元增长显著，票房整体呈显著上升状态，今夏暑期档观影热潮持续升温。8 月 7 日，电影《浪浪山小妖怪》官方微博发文宣布，累计票房破 3 亿。相关公司：光线传媒，万达电影，上海电影，哔哩哔哩，快手等。

10. 游戏：

Mobilegamer.biz 发布了 2025 年上半年的全球手游收入排行榜。《王者荣耀》稳居榜首，今年上半年收入高达 10.05 亿美元。两款 4X 策略游戏《LastWar》和《Whiteout Survival》（无尽冬日）的持续崛起，《LastWar》的收入从去年上半年的 4.38 亿美元猛增至今年上半年的 8.95 亿美元，而《无尽冬日》的 IAP（应用内购买）收入同样实现了翻倍以上的增长，从 3.69 亿美元增至 8.34 亿美元。。相关公司包括：腾讯控股、网易-S、恺英网络、三七互娱、吉比特等。

● 风险提示

消费恢复不及预期风险、用户增长不及预期风险、行业竞争格局剧变风险、宏观经济波动风险、政策及监管趋严风险、科技创新落地进展不及预期风险。

正文目录

1 本周行情回顾	6
2 行业一周要闻	9
2.1 AI+领域	9
2.2 互联网	25
2.3 传媒	25
3 公司公告	28
风险提示	28

图表目录

图表 1 纳斯达克指数走势	6
图表 2 恒生互联网科技指数及恒生科技指数走势	6
图表 3 传媒指数走势	6
图表 4 人工智能指数走势 (931071.CSI)	6
图表 5 A股传媒个股涨跌幅前十 (2025-08-04 至 2025-08-08)	7
图表 6 港股个股本周行情表现 (2025-08-04 至 2025-08-08)	8
图表 7 美股个股本周行情表现 (2025-08-04 至 2025-08-08)	8
图表 8 A股上市公司官宣的AI应用进展 (2024-01-01 至 2025-08-08)	11
图表 9 海外主要大模型迭代情况 (截至 2025-08-09)	16
图表 10 国内主要大模型迭代情况 (截至 2025-08-09)	19
图表 11 国内游戏畅销榜 TOP 10 (2025-08-08)	27
图表 12 国内游戏畅销榜 TOP 10 (2025-08-02)	27
图表 13 公司本周公告 (2025-08-04 至 2025-08-08)	28

1 本周行情回顾

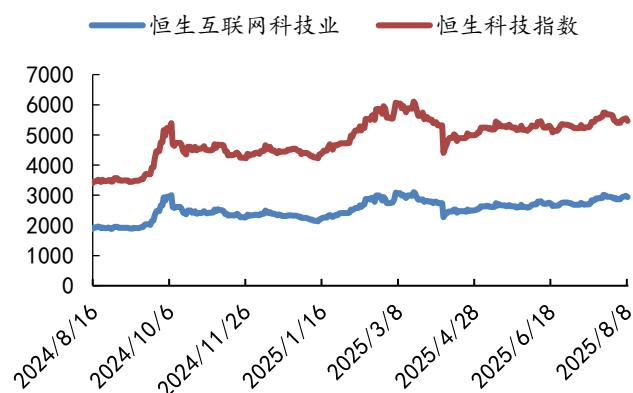
从指数表现来看，本周(2025-08-04 至 2025-08-08)，上证指数周涨跌幅为 2.11%，创业板指周涨跌幅为 0.49%，沪深 300 周涨跌幅为 1.23%，中证 1000 周涨跌幅为 2.51%，恒生科技周涨跌幅为 1.17%，纳斯达克指数周涨跌幅为 3.87%；从板块表现来看，传媒指数周涨跌幅为 1.95%，恒生互联网科技业周涨跌幅为 2.37%，中证海外中国互联网 50 指数周涨跌幅为 1.56%，人工智能指数周涨跌幅为 0.69%。

图表 1 纳斯达克指数走势



资料来源：iFinD，华安证券研究所

图表 2 恒生互联网科技指数及恒生科技指数走势



资料来源：iFinD，华安证券研究所

图表 3 传媒指数走势



资料来源：iFinD，华安证券研究所

图表 4 人工智能指数走势 (931071.CSI)



资料来源：iFinD，华安证券研究所

从 2025-08-04 至 2025-08-08，本周港股涨幅前三分别为：心动公司 (+36.17%)，IGG (+33.5%)，东方甄选 (+29.28%)；本周港股跌幅前三分别为：东软教育 (-11.14%)，创梦天地 (-7.56%)，中国儒意 (-5.75%)；本周美股涨幅前三分别为：富途控股 (+13.59%)，苹果 (+13.33%)，BOSS 直聘 (+12.97%)；本周美股跌幅前三分别为：虎牙 (-6.4%)，冬海集团 (-4.28%)，赛富时 (-4.08%)；本周 A 股传媒涨幅前三分别为：巨人网络 (+23.59%)，吉比特 (+11.59%)，安克创新 (+9.65%)；本周 A 股传媒跌幅前三分别为：昆仑万维 (-7.32%)，值得买 (-6.62%)，中文在线 (-5.82%)。

图表 5 A股传媒个股涨跌幅前十 (2025-08-04 至 2025-08-08)

周涨幅前十				周跌幅前十			
股票代码	股票名称	市值 (亿元)	周涨跌幅	股票代码	股票名称	市值 (亿元)	周涨跌幅
002558.SZ	巨人网络	576.75	23.6%	300418.SZ	昆仑万维	419.79	-7.3%
603444.SH	吉比特	279.75	11.6%	300785.SZ	值得买	67.59	-6.6%
300866.SZ	安克创新	726.33	9.6%	300364.SZ	中文在线	184.02	-5.8%
002602.SZ	世纪华通	1,052.30	9.5%	000681.SZ	视觉中国	143.90	-5.4%
300315.SZ	掌趣科技	163.00	9.3%	002400.SZ	省广集团	134.93	-3.9%
002517.SZ	恺英网络	417.67	7.5%	002291.SZ	遥望科技	61.09	-3.4%
000665.SZ	湖北广电	68.57	5.1%	300133.SZ	华策影视	143.61	-3.1%
300113.SZ	顺网科技	140.34	4.6%	600556.SH	天下秀	92.20	-2.7%
002027.SZ	分众传媒	1,132.27	4.4%	300058.SZ	蓝色光标	247.27	-2.6%
002624.SZ	完美世界	296.43	4.3%	300027.SZ	华谊兄弟	75.74	-2.2%

资料来源：iFinD，华安证券研究所

图表 6 港股个股本周行情表现 (2025-08-04 至 2025-08-08)

周涨幅前十				周跌幅前十			
股票代码	股票名称	市值(亿港元)	周涨跌幅	股票代码	股票名称	市值(亿港元)	周涨跌幅
2400.HK	心动公司	354.67	36.2%	9616.HK	东软教育	22.17	-11.1%
0799.HK	IGG	63.72	33.5%	1119.HK	创梦天地	19.69	-7.6%
1797.HK	东方甄选	309.11	29.3%	0136.HK	中国儒意	472.31	-5.8%
9992.HK	泡泡玛特	3733.38	14.3%	6068.HK	光正教育	2.72	-4.6%
8083.HK	中国有赞	50.13	13.4%	1810.HK	小米集团-W	13337.57	-4.0%
0302.HK	中手游	16.47	11.1%	1357.HK	美图公司	510.96	-3.5%
1060.HK	阿里影业	352.54	10.3%	1896.HK	猫眼娱乐	86.01	-3.0%
9899.HK	云音乐	581.08	9.0%	1890.HK	中国科培	29.21	-2.7%
2013.HK	微盟集团	79.91	8.9%	2423.HK	贝壳-W	1662.50	-2.1%
1024.HK	快手-W	3416.14	6.1%	3690.HK	美团-W	7382.14	-1.1%

资料来源：iFinD，华安证券研究所

图表 7 美股个股本周行情表现 (2025-08-04 至 2025-08-08)

周涨幅前十				周跌幅前十			
股票代码	股票名称	市值(亿美元)	周涨跌幅	股票代码	股票名称	市值(亿美元)	周涨跌幅
FUTU.O	富途控股	224.04	13.6%	HUYA.N	虎牙	7.20	-6.4%
AAPL.O	苹果	34,036.43	13.3%	SE.N	冬海集团	873.88	-4.3%
BZ.O	BOSS直聘	100.25	13.0%	CRM.N	赛富时	2,299.28	-4.1%
WMG.O	华纳音乐	165.29	10.8%	DIS.N	迪士尼	2,021.42	-3.6%
MNSO.N	名创优品	62.32	8.5%	CMCSA.O	康卡斯特	1,169.32	-2.6%
TME.N	腾讯音乐	342.77	7.7%	ADSK.O	欧特克	622.06	-2.1%
EA.O	艺电	419.45	6.7%	MSFT.O	微软	38,804.10	-0.4%
GDS.O	万国数据	75.06	6.6%	ATVI.O	动视暴雪	742.89	0.0%
GOOGL.O	谷歌A	24,359.73	6.5%	YY.O	欢聚	0.00	0.0%
ATHM.N	汽车之家	33.92	5.6%	SHOP.N	Shopify	0.00	0.0%

资料来源：iFinD，华安证券研究所

2 行业一周要闻

2.1 AI+领域

1. 海外 AI

● OpenAI 发布 GPT-5

8月7日，OpenAI宣布推出人工智能模型GPT-5，该模型在编码、创意写作以及复杂查询推理方面具备更强的能力。GPT-5将在未来一周内分批向所有ChatGPT用户开放，包括免费用户。Plus会员可获得更多使用量，Pro会员则可以访问GPT-5 Pro版本，该版本具有扩展推理能力，能够提供更全面、更准确的答案。GPT-5是OpenAI首个“一体化”的AI系统，整合了GPT系列（大语言模型）和o系列（推理模型），具备调度子模型的能力。作为OpenAI的最强模型，GPT-5在编程能力、创意写作、健康领域等关键领域实现了显著提升。（中华网）

● 微软全线产品接入 GPT-5

OpenAI发布GPT-5模型后，微软随即宣布将其集成至全线Copilot产品体系。具体包括：Copilot对话助手启用上月曝光的智能模式，Microsoft 365 Copilot办公生产力套件全面升级，Azure AI Foundry企业级AI开发平台，GitHub Copilot编程辅助工具性能跃升。此次多平台同步部署标志着GPT-5进入规模化商用阶段，Copilot将率先启用具有突破性的智能交互模式。（腾讯网）

● 马斯克宣布 xAI 下周开源 Grok 2 模型

8月6日，马斯克透露，xAI公司计划在下周开源Grok 2模型。马斯克旗下的xAI公司于2024年8月13日发布了Grok-2和Grok-2 mini两款AI模型的Beta版本。这些模型在推理能力方面进行了显著优化。与此同时，X平台还向Premium和Premium+订阅用户开放了新功能，允许用户直接在平台上生成图片。（新浪科技）

● 谷歌正式发布 Genie 3

8月6日，谷歌正式发布Genie 3，这是一款通用型世界模型，能够生成前所未有的多样化交互式环境。Genie 3是首个支持实时交互的世界模型，用户只需输入文本提示，Genie 3就能以每秒24帧的速度实时生成可供自由探索的动态世界，并在720p分辨率下保持数分钟的画面一致性。

相较于前两代产品，它不仅能在720p、24 fps的规格下即时生成可交互的三维场景，还把单次交互时长从几十秒延长到数分钟，并首次加入“可提示的世界事件”和“视觉记忆”等关键特性，使得同一房间里的涂鸦、家具位置等细节在玩家多次往返时依然保持一致，真实感大幅提升。（腾讯网）

● 谷歌推出 Gemini 2.5 Deep Think 模型

8月1日，谷歌宣布，Gemini 2.5 Deep Think模型现已正式上线，现在在Gemini应用中向Google AI Ultra订阅用户开放。这款新模型在多个关键基准测试中超越了OpenAI的o3和xAI的Grok 4。（IT之家）

2. 国内 AI

● 腾讯 AI Lab 开源可复现的深度研究智能体

腾讯AI Lab全新推出的Cognitive Kernel-Pro，一款全开源、多模块、层次化的智能体框架，为深度研究智能体的开发与训练提供了突破性解决方案。在GAIA基准全集上，Cognitive Kernel-Pro超越开源免费框架SmolAgents，性能逼近依赖付费工具的智能体，展现出卓越的综合能力。在GAIA-text上，训练的8B模型超越WebDancer和WebSailor-7B。（腾讯网）

● 腾讯混元 0.5B、1.8B、4B、7B 模型发布并开源

8月4日，腾讯混元发布四款开源小尺寸模型：0.5B、1.8B、4B、7B。这四款模型消费级显卡即可运行，适用于笔记本电脑、手机、智能座舱、智能家居等低功耗场景，支持垂直领域低成本微调。目前已在 Github 和 Huggingface 等开源社区上线，Arm、高通、Intel、联发科技等多个消费级终端芯片平台宣布支持部署。（快科技）

● 阿里通义千问发布 Qwen3 新模型

阿里通义千问宣布推出 Qwen3-30B-A3B 模型的全新版本，命名为 Qwen3-30B-A3B-Instruct-2507。这一非思考模式的最新模型在技术架构上实现重要突破，长文本理解能力显著提升至 256K 规模。该模型采用创新的参数激活策略，仅需激活 3B 参数即可达到与 Gemini 2.5-Flash 和 GPT-4o 等主流开源模型相当的性能水平。（腾讯网）

● 华为开源 7180 亿参数大模型

8月5日，华为轮值董事长徐直军正式宣布 CANN 全面开源开放，还开源了三款盘古模型，参数规模为 1B、7B 和 718B。其中参数规模 7180 亿的盘古 Ultra MoE 模型最早在今年 5 月 30 日首次官宣。（腾讯网）

3. 海外半导体

● 特朗普将对芯片进口征收 100% 关税，苹果被豁免

特朗普周三表示，将对半导体等进口商品征收 100% 的关税，但会豁免那些将生产迁回美国的公司。当天，苹果公司 CEO 蒂姆库克（Tim Cook）与特朗普在白宫椭圆形办公室宣布了一项新的 1000 亿美元投资计划。（IT之家）

● WSTS：2025 年上半年全球半导体市场规模同比增长 18.9%，达 3460 亿美元

8月4日，世界半导体贸易统计组织 WSTS 宣布，今年 1-6 月全球半导体市场规模达 3460 亿美元，同比增长 18.9%。按类别划分，上半年逻辑半导体市场规模提升 37%、存储半导体则增长 20%、传感器增长 16%、模拟和微型器件均小幅增长 4%，而分立器件和光电器件分别出现 4% 和 0.5% 的环比下滑。（IT之家）

● AMD 第二季度调整后净利润 7.81 亿美元

8月5日，AMD 公布第二季度财报显示，第二季度营收 76.9 亿美元，同比增长 32%，预估 74.3 亿美元，经调整毛利率 43%，上年同期 53%，预估 43.1%。第二季度调整后净利润 7.81 亿美元，同比下降 31%，调整后每股收益 0.48 美元，上年同期 0.69 美元，预估 0.49 美元。（IT之家）

● 台积电 7 月营收 3231.7 亿元新台币，同比增长 25.8%

8月8日，台积电发布了 2025 年 7 月营收报告，7 月营收 3231.7 亿元新台币，同比增长 25.8%，环比增长 22.5%。（IT之家）

● 联发科 2025 年 7 月营收 432.2 亿新台币

8月8日，联发科公布 2025 年 7 月份营业额数据，实现 432.2 亿新台币营收，相较去年同期下滑 5.24%，环比降幅则达到 23.41%。（IT之家）

4. 智能驾驶

● 特斯拉获批美国得州网约车牌照

8月8日，特斯拉 Robotaxi LLC 已被德州许可与监管局登记为运输网络公司许可方，与 Uber、Lyft 等公司并列。根据 9 月 1 日起生效的新州法律，所有使用自动驾驶汽车的网约车服务必须与人工驾驶的网约车一样接受监管，因此这项许可现在变得必要。（腾讯网）

● **亚马逊旗下 Zoox 无人驾驶汽车获美国监管机构批准**

美国交通部周三在一份声明中表示，美国汽车安全监管机构已批准豁免 Zoox 公司生产的定制无人驾驶汽车的联邦车辆安全标准。（新浪财经）

4. A股上市公司AI应用进展（24年1月初至今）
图表 8 A股上市公司官宣的AI应用进展（2024-01-01至2025-08-08）

类型	公司名称	股票代码	日期	公司业务对AIGC技术的应用	接入的模型
1) AI+内 容制 作 b	昆仑万维	300418.SZ	2024年4月26日	董之磊明确了中文在线未来将持续深耕的三个方向：AI垂直大模型、AI内容的市场应用、AI内容柔性供应链。	ChatGPT、文 心一言、澜舟 科技自研的类 ChatGPT大语 言模型
			2024年7月2日	公司在AI大模型、AI多模态方面已积极开展技术建设并进行了商业化落地	
			2025年4月22日	公司在AI生成漫画以及动态漫已制作完成近百部作品，实现了商业化连载和付费经营。	
			2024年1月6日	AI大模型首次牵手国民级综艺 昆仑万维天工AI联合《最强大脑》加速大模型落地	
			2024年1月26日	昆仑万维大模型在优酷优图实验室联合厦门大学此前开展的多模态大语言模型测评中，综合得分排名第一	
			2024年2月6日	“天工2.0”MoE大模型：支持超15万个汉字上下文窗口 昆仑万维发布“天工2.0”涵盖AI算命等新功能 昆仑万维发布“天工2.0”MoE大模型 系国内首个免费开放的千亿级参数大语言模型AI应用 天工2.0 MoE大模型发布，新版天工AI智能助手APP向C端开放	
			2024年2月7日	昆仑万维发布天工2.0：千亿参数大语言模型来了	
			2024年2月28日	目前公司已形成AI大模型、AI搜索、AI游戏、AI动漫、AI社交六大AI业务矩阵，同时布局国内外与海外市场	
			2024年3月21日	航行318——全球首个公路旅行AI机器重磅发布	
			2024年3月25日	昆仑万维新增“AI预料”概念	
1) AI+内 容制 作 b	昆仑 万维	300418.SZ	2024年4月2日	昆仑万维宣布4月17日发布并开源“天工大模型3.0”：4000亿参数，号称性能超Grok 1.0	ChatGPT、与 奇点智源联合 开发天工3.5
			2024年4月11日	K昆仑万维升级天工3.0大模型，AI音乐迭代	
			2024/4/17	昆仑万维“天工3.0”基础大模型与“天工SkyMusic”音乐大模型正式开启公测。	
			2024/4/18	公司自主研发AI游戏《Club Koala》正式开启Alpha版本测试	
			2024/4/23	公司宣布将坚定践行“AI in AGI与AIGC”发展战略，加大研发投入。	
			2024/4/29	昆仑万维2050全球研究院发布Vitron，奠定通用视觉多模态大模型终极形态	
			2024/5/7	目前公司已有算力预计能够满足未来1-2年除视频AIGC之外的大模型算力需求。 公司记者会回答天工“3.0”采用4,000亿参数MoE架构，是目前全球模型参数最大、性能最强的开源MoE之一 公司旗下Opera通过与OpenAI合作、推出原生AI助手、集成多种AI工具等多措并举，全面拥抱人工智能，成为首个全面接入AI的独立浏览器品牌 昆仑万维荣登2024中国生成式AI企业TOP50榜单	
			2024/5/28	昆仑万维还与中国移动咪咕音乐联合发布“AI一语成歌”AI音乐视频彩铃创作平台 Opera与谷歌云重磅合作Alra接入Gemini大模型	
			2024/6/3	“全球首创”单台RTX 4090服务器推理，昆仑万维开源2千亿稀疏大模型天工MoE	
			2024/6/7	昆仑万维助力建设北京首个AIGC视听产业创新中心	
			2024/6/21	昆仑万维：2024年4月17日，公司正式将天工大模型迭代至3.0版	
			2024/6/25	昆仑万维·天工SkyMusic是中国首个音乐SOTA模型 昆仑万维携手南洋理工大学枪发Q*算法：百倍提升7B模型推理能力	
			2024/7/4	昆仑万维将与蓝色光标合作，在合作中开放大模型底层能力与APP平台分发资源，基于天工AI大模型、AI搜索、AI音乐、AI视频、AI社交等领域的AI技术能力为生成式广告发展提供技术基础。	
			2024/8/5	昆仑万维：公司携手新加坡南洋理工大学成功开发Q*算法，显著提升现有大模型推理能力 昆仑万维天工开放平台第一时间面向OpenAI API用户推出“一键搬家计划”，并为用户提供1.5亿Tokens的API调用资源与专属迁移服务技术支持。	
			2024/8/15	昆仑万维发布AI流媒体音乐平台Melodio	
			2024/8/20	昆仑万维发布全球首款AI短视频平台SkyReels 推动短剧创作进入新时代	
			2024/9/12	昆仑万维发布全新开源大模型奖励模型Skywork-Reward，登顶RewardBench	
			2024/11/18	天工大模型4.0.01版（英文名：Skywork 01）将于11月27日启动邀测	
			2024/11/20	昆仑万维推出「天工大模型4.0.140版（Skywork 40）」，实时语音对话助手Skyo上线在即	
			2024/11/21	昆仑万维：公司发布国内首个AI搜索引擎——天工AI搜索	
			2025/1/6	天工大模型4.0.01版和40版同步上线	
			2025/3/26	昆仑万维发布全球首款音乐推理大模型Mureka 01	
			2025/5/22	“AI版Office”天工超级智能体发布，加速推动内容生产与办公协作新模式 新经济观察	ChatGPT、文 心一言
			2025/5/26	全球首款Office智能体：昆仑万维天工超级智能体App上线	
			2025/5/28	Skywork 超级智能体正式上线	
			2025/6/20	昆仑万维正式发布并开源Skywork-SWE-32B	
			2025/7/9	昆仑万维发布并开源Skywork-RIV 3.0版本	
			2025/8/2	昆仑万维推出并开源全新推理大模型Skywork MindLink	
			2024/1/2	AI创作“名画秘境”万兴科技“天幕”多媒体大模型支持AI扩图	
1) AI+内 容制 作 b	万兴 科技	300624.SZ	2024/1/8	万兴科技wondershare Filmora上线文生声音功能 有“天幕”大模型技术赋能	ChatGPT、文 心一言
			2024/1/11	万兴“天幕”多媒体大模型算法通过备案 将在AIGC产品中逐步应用 “四算一体”产业联盟成立 浙江实验室助理万兴天幕多媒体大模型研发	
			2024/1/22	AI创作腊八粥里的江湖 万兴科技“天幕”大模型视频生成效果持续提升	
			2024/1/29	万兴天幕音视频多媒体大模型即将发布	
			2024/1/30	公司“天幕”大模型是以视频创意类AI技术为核心的多媒体大模型，涵盖音频、图像、视频等多模态能力，将支撑公司在影视、广告、动画等内容创作领域的应用升级与拓展 国内首个音视频多媒体大模型万兴“天幕”正式发布	
			2024/2/2	公司聚焦数字创意软件产品的销售与服务，业务涵盖视频创意、绘图创意、文档创意、实用工具等领域	

			公司打造的万兴“天幕”大模型算法已通过《互联网信息服务深度合成管理规定》备案 天幕大模型主要基于公司在数字创意领域二十多年的产品开发与运营经验而打造的音视频多模态大模型，具备多模态、垂直解决方案、算力数据及应用本土化等特色	
		2024/2/20	公司视频创意产品 Filimora V13 版本已有文本成片功能，万兴播客/Virbo 已有文生音视频的能力，且已落地服务全球多个国家和地区用户	
		2024/3/7	公司音视频多模态大模型“天幕”，聚焦数字创意垂类创作场景，以音视频生成式 AI 技术为基础，将视频模型、音频模型、图像模型和语言模型能力广泛赋能给公司各类创意产品	
		2024/3/24	万兴科技旗下万兴喵影率先接入 Kimi 其他产品对接测试	
		2024/3/26	Kimi 打开大规模量产空间 万兴“天幕”聚焦音视频应用落地	
		2024/3/29	Kimi 爆红 万兴科技美图公司等攻坚商业化落地	
		2024/4/22	万兴科技旗下墨刀入选 3 月 AI 工具速增榜 万兴科技旗下万兴播客获评 2024“最值得期待 AIGC 产品”	
		2024/4/23	万兴科技副总裁宋伟宣布万兴天幕音视频多模态大模型将于 4 月 28 日正式公测	
		2024/4/29	万兴“天幕”正式公测 将逐步开放 API 接口，AI 视频可长达 60 秒	
		2024/5/10	万兴科技旗下墨刀入选“3 月访问量环比增速最快的 AI 工具”排行榜	
		2024/5/13	万兴科技：公司天幕多模态音视频大模型已赋能支持《歌手 2024》片头部分视频画面生成 万兴科技：目前公司天幕多模态大模型正优先面向专业创作者和 B 端企业进行测试，以真实用户反馈赋能技术持续迭代	
		2024/5/22	AI 创新力突出 万兴科技上榜 2024AIGC 赋能行业创新企业 TOP100	
		2024/5/23	万兴科技携万兴“天幕”大模型等 AI 创新成果亮相文博会	
		2024/5/28	万兴科技亮相欧洲最大科技创新盛会 VivaTech 开启 AI 创作提效新模式	
		2024/6/7	公司旗下 Filimora 即将发布已支持 NVIDIA RTX Video 的新版本，可帮助开发者原生集成 AI，进行采样、锐化、降低压缩失真和高动态范围 (HDR) 转换	
		2024/6/17	万兴科技推出 AI 家装设计平台 Wondershare Decoritt	
		2024/6/20	华为云联合万兴科技等共同发布“凌云出海联盟 3.0”	
		2024/6/24	万兴科技旗下 HiPDF 海外版焕新 全新上线 AI 文档翻译/校对/格式转换功能	
		2024/6/25	万兴科技：公司使用的微软 Azure OpenAI 企业服务 OpenAI 终止对中国提供服务影响	
		2024/6/26	万兴科技与深圳卫视达成战略合作 万兴“天幕”将赋能节目生产	
		2024/6/28	万兴喵影新增对 RTX Video HDR 的支持及 RTX 加速的 AI 功能	
		2024/7/19	万兴喵影受邀亮相 Chinajoy AI 赋能游戏玩家视频创作	
		2024/7/19	万兴喵影新增对 RTX Video HDR 的支持及 RTX 加速的 AI 功能	
		2024/8/16	万兴科技旗下亿图图示 V14 重磅焕新 首页模板智能推荐实现“千人千面”	
		2024/8/17	万兴科技：公司天幕多模态音视频大模型及相关算法技术已在 Filimora、喵影等产品中实现部分接入	
		2024/9/14	万兴科技焕新发布万兴 PDF V11 创新实现扫描二维码跨端传输 PDF 文档	
		2025/2/21	AI 赋能文本秒级图示化 万兴科技旗下亿图图示 PPT 插件全新发布 AI 赋能文本秒级图示化 万兴科技旗下亿图图示 PPT 插件全新发布。	
		2025/6/20	万兴天幕 2.0 灵相华为 HDC2025，公测通道正式开启	
		2025/7/17	万兴科技推出首个 AI Agent 产品“万兴超媒 Agent”	
		2025/8/6	万兴科技旗下万兴超媒 Agent 正式上线，打造全链路语音视频 AI 智能体	
浙文互联	600986.SH	2024/3/15	公司已成立杭州浙文智算公司，并逐步搭建起人员团队，持续推动相关业务落地	
		2024/4/15	公司已推出 AI 辅助写作工具“米文”	
		2024/4/29	公司围绕“AI+文化”、“AI+内容创意”，深入建设智慧内容生态平台，加速“AI+”应用类产品的研发迭代，并全面切入数字文化底座建设，目前已初步建立算力基础设施+应用落地的格局。	
		2024/5/17	公司打造的虚拟人 IP 已经初见成效，虚拟人“兰 LAN”成为奥迪 AudiSA “推荐官” 目前 AIGC 绘画工具“米画”、文生文工具“米文”、元宇宙场景生成工具“数字战士”、营销垂直行业文案小模型等应用，陆续推出进入商业应用 公司与循环智能签订了《战略合作协议》，作为 Kimi 的股东方及其 TOB 的战略合作伙伴，循环智能在打造行业大模型解决方案方面具有极强的优势	文心一言、Diffusion
		2024/8/16	浙文互联：旗下深瑞威行与巨量引擎在数字人共建项目中获突破性进展	
		2025/7/5	具身 AI 获数千万元战略投资 投后估值达 5 亿元	
神州泰岳	300002.SZ	2024/1/8	包含“泰岳小维”在内主要产品均已实现商业化落地并形成收入	
		2024/2/2	公司通过使用 StableDiffusion、Midjourney 等最新的 AIGC 工具，并持续的训练和迭代，生成具有壳木特质的数据资源，游戏画质精良，熟练驾驭多种画风	
		2024/3/4	公司在游戏业务中已经通过使用 Stable Diffusion、Midjourney 等 AIGC 工具，提高制作效率，降低制作成本	
		2024/3/20	公司在 NLP 领域积累了较为深厚的技术，已研发出语音机器人、文本机器人等产品	
		2024/4/19	公司旗下鼎富智能专注于“认知+”产业生态构建，运用自研 NLP（自然语言处理）技术	
		2024/5/23	获亚马逊云科技生成式 AI 能力认证	
		2024/6/3	共创 AIGC 定制化解决方案：神州泰岳携手亚马逊云科技助力出海企业打通“最后三公里”	
		2024/6/20	公司旗下 VCG.com 网站推出首个 AIGC 应用产品 AI 智能搜索（Beta 版），使用多模态、自然语言理解以及知识图谱等人工智能技术，可为新闻稿件、创意文案、新媒体推文、教师课件等内容高效配图	
		2024/6/27	神州泰岳：公司运用自研的 NLP 自然语言处理技术，推出了“泰岳智呼”平台，以智能催收和智能电销为重点，面向金融、零售等行业提供 SaaS 化的解决方案	
		2024/7/4	神州泰岳基于创新技术全新打造了‘1+N+X’的大模型产品	
		2024/8/16	神州泰岳 & AWS 推动生成式 AI 实践	
		2025/6/10	神州泰岳正式发布“泰岳灯塔”AI 大模型应用能力体系	
		2025/7/9	公司正式推出“泰岳灯塔”AI 大模型应用能力体系	
视觉中国	000681.SZ	2024/3/4	目前公司旗下网站 vcg.com 已推出 AI 智能搜索、创意工具、灵感绘图等功能，提高用户工作效率，提升创意体验	
		2024/3/27	公司投资控股国内领先的音视频素材交易平台光厂创意 (VJshi)，光厂创意拥有音视频创作者超过 7.6 万	
		2024/4/12	公司旗下网站 vcg.com 已推出 AI 智能搜索、创意工具、灵感绘图等功能	
		2024/4/26	视觉中国 2023 年度到 2024 第一季度营收净利润双增长，宣布将大力推进“AI+内容+场景”战略	
		2024/5/20	视觉中国持续加强音视频电商平台光厂创意 (VJshi.com) 和图片电商网站 veer.com 的建设	
		2024/6/5	打造以“AI”为核心驱动的智能服务平台	
		2024/7/29	视觉中国战略投资 CG 模型网，持续布局 3D 内容产业	
		2024/7/30	视觉中国与爱诗科技达成战略合作，共绘 AI 视觉生成技术发展新蓝图	
		2025/6/13	视觉中国加速布局 AI 领域，大象视觉获亿元增资	
		2025/8/6	视觉中国战略投资企业 CG 模型网引入全球顶级 3D 模型资源	
当虹科技	688039.SH	2024/4/3	ShowBiz AI 是公司正在开发的跨模态产品系列之一，但相关产品处于开发阶段	
		2024/4/23	当虹科技与广科院、华为、北京大学等 20 家企事业单位共同签署《基于大模型的 AIGC 产业应用战略合作协议》	

	HUAN RESEARCH	2024/4/28	当虹科技发布多模态视听大模型 BlackEye 当虹科技与国家广电总局广科院围绕超高清、AI 深入合作	
网达软件	603189.SH	2024/2/16	公司可对媒资内容进行视频结构化分析，实现自动打点、自动标注、自动快剪、自动人物集锦等功能，以减少了人工标注的工作量；实现了智能场景分析，智能抽图截图，智能实现情节分段，完成媒资的智能编辑工作	
		2024/6/9	网达软件旗下推出的元宇宙会议系统魔塔会议，基于空间计算技术，AI、3D 建模、XR、沉浸式交互、数字人等技术，构建沉浸式数字化交流平台	
		2024/2/20	网达软件申请 AI 边缘智能设备告警图片屏显专利，能够更加直观的了解告警内容	
		2024/3/26	网达软件申请 NeRF 专利，提高 3D 渲染的质量和效率	
		2024/4/30	公司全媒体智能创作平台可基于对视频内容的深层次理解，实现视频内容字幕及语音识别、后期音视频分离	
		2024/8/14	网达软件：公司以 AI 大模型赋能全媒体的内容生产、运营、分发为主要着力点，重构现有内容运营编辑系统、媒资管理系统，研发了系列智能化内容生产系统	
掌阅科技	603533.SH	2024/9/6	积极推动 AI 大模型在数字阅读垂直领域的应用，将国内领先 AI 大模型和掌阅 APP 进行深度融合。	
		2025/6/17	短剧《逃天》定档 6 月 20 日，或开 AI 与真人实拍深度融合先河	
		2025/7/28	掌阅科技上线 AI 讲书电台，打造“对话式”智能阅读新场景	
汤姆猫	300459.SZ	2024/1/18	在智能机器人硬件领域，公司于 2021 年下半年在海外市场推出了一款智能语音机器人 GameBud Talking Tom	
		2024/2/1	公司海外团队打造的首款 AI 手游《Talking Ben AI》正在研发过程中，该产品已在斯洛文尼亚、塞浦路斯、南非等国家开启首轮海外测试	
		2024/2/2	公司目前在研的 AI 应用正在优化产品、完善内容及玩法	
		2024/2/7	公司国内研发团队与西湖心辰合作的汤姆猫 AI 讲故事等产品，已初步完成主要功能的测试，公司正在进一步丰富汤姆猫 AI 讲故事的产品内容	
		2024/2/7	公司 AI 硬件团队正研发一款基于生成式人工智能技术的 AI 语音交互陪伴机器人，该智能机器人将聚焦在与人类的交互与陪伴领域	
		2024/3/26	汤姆猫：伴随着移动互联网新兴技术的发展与新型交互方式的涌现，公司大力开拓创新业务，相继探索了虚拟 IP 直播、体感游戏、车载游戏、AR 游戏、智能语音机器人等创新业态	
		2024/4/2	汤姆猫：公司目前正研发一款基于生成式人工智能技术的 AI 语音交互陪伴机器人	
		2024/5/13	汤姆猫：公司目前在研的 AI 应用国内主要接入的是西湖大模型，该模型有公司所投资的西湖心辰打造	ChatGPT、文心一言、Diffusion
		2024/5/16	汤姆猫：公司国内团队研发的 AI 应用目前主要接入的是西湖大模型	
		2024/6/20	目前公司研发团队与西湖心辰技术专家持续在共建情感陪伴领域具有鲜明汤姆猫性格特征的垂直模型，并完善主动聊天、情绪识别、性格设定、长期记忆等功能模块	
		2024/8/1	西湖心辰（杭州）科技有限公司（下称“西湖心辰”）研发国内首例端到端语音大模型“心辰 Lingo”，该模型将于 9 月 5 日在 2024 外滩大会上正式发布并开放试用申请。	
		2024/8/12	汤姆猫携手西湖心辰开启人机交互新纪元	
		2024/8/14	汤姆猫：公司建设于甘肃定西的集约式算力中心第一期项目已投产使用，该项目将为公司人工智能业务提供算力存储等基础服务	
		2024/9/3	西湖心辰（杭州）科技有限公司（下称“西湖心辰”）研发国内首例端到端语音大模型“心辰 Lingo”，该模型将于 9 月 5 日在 2024 外滩大会上正式发布并开放试用申请。	
		2024/11/20	汤姆猫：一代汤姆猫 AI 机器人产品计划于春节期间实现第一批产品的上市	
		2025/3/9	公司在 AI 业务上确立了“AI+IP+Agent”的业务发展方向	
		2025/5/11	AI 情感陪伴机器人 4 月销量环比增长 100%	
		2025/5/29	公司为 AI 机器人产品训练了专属的智商情商双高的“汤姆猫情感陪伴垂直模型”	
2) AI+游 戏	巨人网络	2024/1/29	已初步建立公司层面 AIGC 生产能力，实现“iImagine-AI 绘画云平台”、“CodeBrain-代码生成助手”、“百灵-AI 语音生成”等平台化基础设施能力建构	
		2024/2/2	巨人网络：巨人大模型 GiantGPT 完成备案，2024 年公司手游《太空杀》将推出全新 AI 玩法	
		2024/2/20	巨人网络：关注 Sora 突破性进展，积极布局多模态大模型领域	
		2024/3/7	公司的 AI 大模型 GiantGPT 主要服务于公司自研游戏产品	
		2025/4/22	公司已将自研 AI 视频风格迁移技术广泛应用于研运环节，赋能内容创意表达	
		2024/4/29	公司已正式完成游戏 AI 大模型 GiantGPT 的备案，是游戏行业首家对外确认实现了此成果的企业，标志着公司在人工智能领域的重要进展	
		2024/5/20	公司组建了 AI 创新实验室，招募 AI 精英人才，负责 AI 技术研发与应用能力建设，全面推动 AI 在各业务场景的落地普及。	
		2024/5/27	巨人网络与 Huawei 就鸿蒙生态达成合作，共同推动 HarmonyOS 智能化移动开放生态业务发展	文心一言
		2024/6/28	公司已正式完成游戏 AI 大模型 GiantGPT 的备案，是游戏行业首家对外确认实现了此成果的企业	
		2024/8/1	巨人网络发布多智能体大模型 AI 原生游戏玩法	
		2024/8/5	巨人网络发布“AI 残局挑战”，AI 持续推动游戏玩法实现突破	
		2024/11/22	巨人网络：公司“游戏+AI”战略最新落地成果，来自于推理派对手游《太空杀》近期推出的全新 AI 玩法“AI 残局挑战”	
		2024/12/12	巨人网络：公司 AI 实验室联合《太空杀》项目组研发的“AI 残局挑战”玩法，是行业内首个基于多智能体大模型（Multi-Agent LLM）打造的 AI 原生游戏玩法	
		2025/5/13	巨人网络与火山引擎开展 AI 合作，《太空杀》接入豆包大模型。	
掌趣科技	300315.SZ	2024/1/9	掌趣科技、悠米互娱智游英伟达拓展 AI 游戏创作平台技术生态	
		2024/1/23	公司与北京蓝亚盒子科技有限公司达成业务合作协议，并已合作推出最新版本 LayaAir 3.1 的测试版本，支持 AI 资源生成、AI 插件、编辑器 AI 控制与 AI 技术客服等功能	
		2024/3/11	公司与北京蓝亚盒子科技有限公司达成业务合作的小游戏引擎 LayaAir 3.1，为游戏开发者提供更加高效、智能的开发工具，更好地服务小程序游戏等游戏类型及相关内容开发生态	
		2024/4/3	LayaAir 3.1 测试版本支持 AI 资源生成等功能	
		2024/4/26	公司宣布围绕“AI+游戏”积极创新，通过“自研+外部合作+投资”的方式，推动“AI+UGC”战略。	
		2024/4/29	公司对数字图书内容进行编辑制作和聚合管理，利用“掌阅”、“得间”等数字阅读平台面向互联网用户发行数字阅读产品	
		2024/5/9	加大人工智能技术研发投入，推动 AIGC 在数字阅读平台的深化应用	
		2024/6/3	公司与北京蓝亚盒子科技有限公司达成业务合作，联合推出 LayaAir 3 游戏引擎，为广大游戏开发者提供更加智能、自动化的游戏引擎产品，推动游戏产业技术发展。	
		2024/7/3	“AI 游戏创作平台”获得了英伟达在 DLSS、Audio2Face 等 AI 技术的支持，未来双方将在 AI 游戏开发工具、AIGC、AI 应用场景等领域继续保持交流与合作。	
		2024/5/14	盛大网络：公司持续关注 AI 技术在应用端的进展，目前已接入月之暗面的文本生成模型，尚处于测试阶段，具体应用场景和产品化工作仍在开展	

	HUANAN RESEARCH	2025/5/30	公司接入 DeepSeek 模型研发的 AI 社交推理小游戏《密语炸弹》	
盛大 网络		2025/6/13	盛大网络：加速 AI 应用场景创新落地 构建“三维发展”格局	
		2025/7/11	公司目前正在开发一款 AI 社交出海产品，预计在今年正式进入海外市场	
华策 影视	300133.SZ	2024/2/8	华策影视：多模态分析平台已于去年四季度在内部上线使用，实现了多模态视频分析理解、多模态内容检索等功能，有利于帮助公司把拥有 5 万小时版权库变为视频元素数据库	
		2024/2/20	华策影视：公司去年组建了 AIGC 应用研究院，自研的编剧助手、剧本评估、视频检索、AI 剧照等功能已在内部应用或内测	
		2024/4/26	在公司内部形成“全员知 AI、会 AI、用 AI”的创新氛围，成为当前影视公司 AIGC 学习和应用的领跑者。	
		2024/5/15	华策影视：自研“有风”大模型通过国家备案	
		2024/5/23	公司重点在影视行业的 AIGC 应用，如项目评估助手、编剧助手等	
光线 传媒	300251.SZ	2024/4/26	在 GTC2024 大会上，参股公司七维科技主要展示了分镜故事板生成和 2.5D 场景环境生成这两项技术。	
		2024/5/31	公司与参股公司七维科技达成合作，委托其基于生成式人工智能和大语言模型技术（LLM），为公司定制开发在电影拍摄和制作过程中所需的生成式 AI 工具集（AI Studio），包括文字生成、语音生成、图像生成、视频生成、三维模型生成等。	
		2024/7/11	已经把动画制作的各个环节都跟 AI 进行了结合，特别是在动画的前期制作，比如形象设计、场景设计、分镜头脚本等方面，应用很广泛	
上海 电影	601595.SH	2024/3/1	上海电影官宣多项计划：全面入局 AI 新浪潮 探索中国动画学派大模型	
		2024/5/17	上海电影成立数智文创科技公司	
		2024/7/5	上海电影携手阶跃星辰亮相世界人工智能大会 推出两款“IP+AI”产品	
返利 科技	600228.SH	2024/2/3	返利科技 AI 导购 APP “如意”取得备案号告知书	
焦点 科技	002315.SZ	2024/2/23	AI 紫可目前系“图生视频”功能：利用供应商提供的图片生成视频，并利用 AI 技术帮助供应商添加字幕、语音、虚拟人形象等视频元素，帮助外贸企业解决视频制作费时费力的问题，提高产品展示效果	
		2024/4/30	宣布再次更新迭代，推出视频生成、智能排版、行业情报三大功能，目前已在图文、视频内容生成、智能营销、客户服务、大数据分析等方向进行了 AI 技术的探索与应用	
		2024/9/2	焦点科技：目前公司主要基于 AI 应用层进行研发，暂不涉及大模型的自研工作	
		2024/11/11	焦点科技 AI 发布 AI 紫可 4.0	
		2025/2/16	公司在人工智能代理（AI Agent）领域已推出两款产品，分别为 AI 紫可和 Mentarc	
吉宏 股份	002803.SZ	2024/4/30	公司跨境社交电商业务借助 AI 技术赋能包括选品、广告文案（图像与视频）设计与翻译、广告投放及客服等流程，并已实现业务流程和决策的自动化和智能化	
光云 科技	688365.SH	2024/4/16	公司的电商 SaaS 中的客服机器人、企业数字助理、自动生图等细分赛道都涉及人工智能技术	
		2024/4/26	公司与浙江云谷出资设立浙江光云科技有限公司，新公司拟建立数据中心、数据中心，更好地向市场提供大数据、大模型、物联网、人工智能与机器学习、服务器托管及网络接入等方面的服务。	
		2024/5/23	光云深绘 AI 美工机器人通过 AI 智能分析详情页、主图模板，实现智能化内容批量排版	
3) AI+电 商	300785.SZ	2024/1/25	什么值得买对话式机器人“小值”拟定于 2 月底上线	
		2024/1/31	公司的消费大模型是基于通用大模型的技术底座，结合公司积累的私有数据进行训练和调优，目前该模型已经在公司产品或者正在测试的产品中发挥作用	
		2024/2/1	值得买科技 CTO 王云峰出席未来直播电商科技峰会，畅谈 AI 赋能电商直播	
		2024/2/5	值得买：什么值得买“AI 问答机器人”拟定于 2024 年 2 月底上线	
		2024/2/23	值得买：公司消费内容社区“什么值得买”的“AI 问答机器人”新功能，目前已经正式通过了国家《生成式人工智能服务管理办法》备案	
		2024/3/5	值得买：值得买科技正在开展多模态的工作，具体包括“图生图”和“文生图”模型，短视频脚本的生成，短视频的自动生成，以及直播数字人等	
		2024/3/5	已经启动了值得买 APP 鸿蒙系统的开发	
		2024/4/26	公司宣布加大投入并推进相关业务场景与 AIGC 的融合，推动 AIGC 对相关业务场景商业生态的重塑。	
		2024/5/7	公司将脱敏后的消费内容和消费数据的 API 开放给 Kimi，帮助 Kimi 提升在消费类问答中的表现	
		2024/5/15	值得买：在成都设立了研发中心，进行出海架构开发，希望基于统一的底层架构在各个国家进行灵活的配置	
		2024/5/21	值得买：公司已经在消费内容社区“什么值得买”中部分商品详情页面接入了 Kimi，为用户提供商品提炼总结服务	
		2024/7/1	值得买科技加速共建 AI 生态，建成智能体成飞火星火大模型智能体中心	
		2024/7/12	值得买：VR 带货有望形成与用户全新的交互模式，从而带来全新的消费体验，我们认为具有巨大的潜力和商业化应用前景	
		2025/7/31	值得买：公司与大模型公司 Minimax 达成官方合作伙伴关系	
		2024/8/12	“什么值得买”正式完成鸿蒙原生应用开发	
		2024/9/6	值得买科技与腾讯云达成战略合作，双方将在 AI 产业共创、技术创新、营销生态、基础设施等方面展开全面合作	
		2024/9/9	值得买科技与腾讯云达成战略合作，结合值得买 AIGC 战略开展多维度合作	
		2024/11/18	值得买：基于“值得买消费大模型”研发的 AI 购物助手“小值”上线	
		2025/3/12	首次与 AIPG 端深入合作 值得买接入联想 AI 桌面助手“如意”	
		2025/5/28	首值得买科技发布海纳 MCP Server，开放四大核心接口	
		2025/6/28	值得买科技发布“什么值得买”GEN2 等多款 AI 产品	
		2025/6/11	值得买科技发布“海纳 MCP Server”：120 亿消费内容库驱动智能体生态加速发展	
易点 天下	301171.SZ	2024/2/4	易点天下：以 AI 化、数字化、场景化三大趋势为主要方向，以交付客户价值为中心，以产品技术为底座，持续打造标杆化和差异化的服务能力	
		2024/2/29	公司目前在 AI 上的研发投入主要聚焦在 AIGC 领域，用于提升广告创意的生产效率和运营效果	
		2024/3/27	易点天下：公司正在搭建的 AI 产品矩阵，帮助公司业务实现 AI 化，形成对行业上下游、广告主以及自身更有效的新质生产力	
		2024/3/28	易点天下与 AnyMind Group 达成深入合作 共同加速企业全球化商业成长	
		2024/4/23	易点天下受邀参与 2024AI 领袖者峰会，并携手阿里云共同发布了全新的“AI 营销+CI 云服务”企业出海数字化转型解决方案	
		2024/4/24	易点天下成功入选由虎嗅智库推出的「大鲸榜·AIGC 数字营销技术商竞争力评价（2024）」的 AIGC 数字营销技术商 TOP15	
		2024/5/16	公司已接入通义千问，除此之外还接入了 GPT-4、Anthropic、PalM、LLAMA、ChatGLM、StableDiffusion、Midjourney、文心一言、盘古等多种通用大模型	
		2024/5/30	CyberTargeter 携全新电商再营销解决方案亮相 eTail Asia Summit 2024	
		2024/6/17	公司近期发布了“AI+BI+CI”出海全链路解决方案，为更多的出海企业提供数字化营销推广的全链路解决方案	
		2024/6/23	易点天下与华为旗下鲸鸿动能达成深度合作	
		2024/7/19	易点天下与腾讯云达成合作，探索全球化智能营销创新模式	

心一言
ChatGPT 等

HUAAN RESEARCH					
蓝色光标	300058.SZ	2024/2/23	雪花啤酒 X 蓝色光标全球首发 AI 定制多人场景化互动营销		
		2024/4/3	蓝色宇宙与 SOLANA 蓝色港湾举行战略合作签约仪式		
		2024/5/10	蓝色光标荣登 2024 中国上市公司品牌价值榜 TOP100		
		2024/7/4	蓝色光标与昆仑万维正式宣布达成战略合作，携手打造 AI 营销创新生态		
		2024/9/11	蓝色光标 Blue AI 携手快手可灵 AI，共同引领 AI 视频生成新时代		
		2025/7/7	蓝色光标接入文心 4.5 系列开源模型，加速建设 AI 营销新生态		
风语筑	603466.SH	2024/1/8	苹果头显 VisionPro 上市在即，风语筑成立 MR 研究院开拓新局		
		2024/2/1	公司积极拓展数字文旅业务，通过整合 VR/MR 相关产品和技术，结合各地历史文化传统和特色 IP，打造新型文旅消费场景，盘活存量空间资源		
		2024/2/23	风语筑，实时渲染、裸眼 3D、全息影像等前沿视觉技术手段广泛应用于公司各类数字化体验项目以及虚实结合的沉浸式数字文旅项目		
		2024/3/5	风语筑携手新华社 AIGC 绘中国，A13D 实景瞰代表委员履职		
		2024/3/29	公司已与百度签署战略合作协议共同探索并推进 AIGC 在文字、图片、虚拟空间模型生成等 3D 内容场景的模型训练、落地应用和模型优化		
		2024/4/29	风语筑将继续以高品质的数字创意、创新的设计、先进的技术，打造沉浸式的数字文娱体验空间。		
		2024/5/6	风语筑与中央美院达成合作，共同开发行进式 VR《消失的圆明园》		
		2025/3/20	风语筑成立“具身智能研究院”，推动 AI 人工智能与场景应用融合创新		
天地在线	002995.SZ	2024/2/1	天地在线推出 AI 智能办公产品“Aibase”		
		2024/2/2	“Aibase”新品发布会圆满举办，天地在线助力企业全方位智能服务		
		2024/3/14	推出“元启”“森森”等自有数字人形象		
		2024/3/28	天地在线：公司现已推出 AI 数字人 SaaS 产品“星河 AI 数字人系统”以及帮助企业提高工作效率和协作能力的 AI 智能应用产品“Albase”，满足客户在营销及提升企业经营效能的不同需求。		
		2024/4/16	公司全资子公司一天地太古与京东合作京小智智能营销客服、AI 智能外呼和 AI 虚拟直播		
		2024/6/3	公司已在虚拟数字内容、区块链、AI 等相关技术和业务方面布局，并已尝试为文创、旅游等行业领域客户提供数字藏品、数字场景、XR 直播等服务		
		2024/6/19	已推出 SaaS 产品“星河 AI 数字人系统”		
		2024/11/5	天地在线&大账房携手打造“财税有问必答”AI 应用，共启财税智能化新篇章		
		2025/7/9	天地在线等成立新公司，含 AI 软件开发业务		
利欧股份	002131.SZ	2024/5/9	利欧数字现已推出面向营销全行业的 AIGC 生态平台「LEO AIAD」以及营销领域大模型「利欧归一」	ChatGPT、文 心一言、 Diffusion、 DALL-E2 等	
		2024/9/10	利欧股份：推出 AIGC 生态平台和营销大模型，持有理想汽车折合 ADS 不到 1,000 万股，2024 年启动股份回购方案		
		2025/6/12	利欧数字主导广告行业首个 AIGC 标准落地，助力 AI 营销新基建		
4) 数字人	002354.SZ	2024/1/4	积极布局虚拟/混合现实内容业务，目前公司已在申请苹果 Vision Pro 开发套件		
		2024/1/23	公司自研企业级虚拟数字人大模型智者千问基于 Transformer 编解码模型框架，致力于为企业级虚拟数字人提供完整的智能化生成式行业大模型解决方案，为垂直行业同全面的支持和创新应用		
		2024/1/30	公司数字人“天妤”以收集壁画碎片为故事线索的系列短剧《千壁寻踪》一经推出就受到众多网友关注，天妤也在抖音收获粉丝近 350 万		
		2024/2/1	公司通过进一步整合抖音、快手等平台资源，依托“魔方 Mix-AIGC 短视频智能营销助手”，帮助客户更加精准地触达目标受众，优化广告创意并提高广告投放效率		
		2024/4/3	公司自研企业级虚拟数字人大模型智者千问基于 Transformer 编解码模型框架，致力于为企业级虚拟数字人提供完整的智能化生成式行业大模型解决方案，为垂直行业同全面的支持和创新应用	ChatGPT、文 心一言、 DALL-E2 等	
		2024/5/9	公司利用 AIGC 相关技术，通过文本、语音、动作等多模态交互方式，推出了用一张人像就能快速创建数字写真的 Crush 相机小程序，可以在多种场景提供情绪陪伴的电子相册“魔境”		
		2024/5/17	天娱数科：数字营销业务新增客户包括 Kimi 等 AI 社交、短剧、大模型领域企业		
		2024/5/21	天娱数科入选 36 强“2024 AI 应用标杆案例”		
		2024/6/7	大连数据流量谷正式开园，天娱数科企业级 AI 应用战略成型		
		2024/6/28	天娱数科成立数字科技新公司		
		2024/7/22	天娱数科“智者千问”大模型完成算法、模型双备案		
		2024/8/20	天娱数科空间计算与 3D 数据集技术，赋能游戏内容开发		
		2024/9/5	天娱数科打造 AI+文旅“种草”新模式，助力文旅智能化升级		
		2024/9/12	天娱数科：智境云创“智者千问”大模型已完成生成式人工智能服务备案，将支持公司各业务平台发展		
捷成股份	300182.SZ	2024/2/26	公司研发的 AIGC 一键成片系统，基于多模态大模型能力，能够自动化处理大量图片和视频素材	ChatGPT、文 心一言、世优 自研小模型	
		2024/3/15	捷成股份发布 AIGC 产品智能创作引擎 ChatPV		
		2024/3/27	捷成股份：ChatPV 是捷成股份探索 AI 技术在视频创作领域应用的一个阶段性成果		
		2024/7/12	捷成股份与华为云签署协议，共同建设视频大模型		
		2024/7/29	捷成股份：2024 年 3 月，公司自主研发的 AI 智能创作引擎 ChatPV 正式发布，并接入华为云盘古大模型的通用语言解析能力，服务于 AI 视频创作应用		
		2024/9/13	捷成股份：AI 智能创作引擎 ChatPV 于 7 月 29 日正式上线公测		
		2024/11/4	捷成股份：目前，chatPV 手机版本的开发工作进展顺利，并将于近期上线公测		
		2024/12/17	推出 AI 视频智能创作引擎 ChatPV，与华为签约共同研发视频垂直大模型。		
博瑞传媒	600880.SH	2024/9/6	博瑞传播：与成都青羊城建设发展有限公司签署合作协议，打造成都数字文创新地标		
5) AI 内容检测	新华网	603888.SH	2024/2/22	目前新华网 AIGC-Safe 平台已取得良好的阶段性运用效果，在内部已展开一些应用，考虑到投入运营成本较高，短期内暂不会产生很大经济效益	
			2024/4/29	公司 AIGC-Safe 平台亦在通过采集大量国内外生成式视频内容训练 AI 鉴别模型	
			2025/3/27	新华网 x 华弘数科联合发布 AI 新物种：全液冷智算一体机	
			2025/4/18	推出全场景可 AI 解决方案，赋能政企智能化升级	
6) AI+出版	荣信文化	301231.ZS	2024/2/22	2023 年公司与科大讯飞合作，利用其语音复刻技术为点读笔新增了“AI 声音定制”功能，帮助忙碌的爸妈与孩子建立更多的联结	
	中国高科	600730.SH	2024/9/6	中国高科：上半年自主研发 AI 科普机器人等产品	

世纪天鸿	300654.SZ	2024/2/7	公司积极响应教育数字化相关政策，持续投入技术研发，先后推出AI录播等创新产品方案，赋能教育数字化发展	投资的AI作家批改产品“笔神”采用基于Transformer的最新算法作为AI模型底层
		2024/2/20	目前公司已结合人工智能技术推出小鸿助教、笔神作文等相关教育产品	
		2024/3/22	小鸿助教相关产品目前上处于用户测试阶段，为产生业务收入	
		2024/4/15	公司的全资子公司奥威亚现有产品已融合多种人工智能技术	
		2024/11/19	世纪天鸿：公司以自身战略定位和资源禀赋为出发点，结合人工智能技术已推出了基于大语言模型研发的专注于服务老师的AI智能体“小鸿助教”	
		2025/2/10	公司已推出基于大语言模型的AI智能体“小鸿助教”	

资料来源：各公司互动平台，华安证券研究所整理

5. 海外主要大模型迭代情况

图表 9 海外主要大模型迭代情况（截至 2025-08-09）

模型名称	推出时间	简介
ChatGPT (OpenAI)		
ChatGPT 4	2023 年 3 月	不仅能够处理文本信息，还具备一定的图像理解能力。
ChatGPT 4 32k	2023 年 3 月	提供了更长的上下文记忆长度，增强了对长文本的处理能力。
GPT-4-32k (0613)	2023 年 6 月	增加了对大文本的理解能力，能够处理最多 32,000 个令牌的长文本。
ChatGPT 4-Turbo 预览版 (1106)	2023 年 11 月	支持 128k 的上下文，并且相较于之前的版本有较大幅度的价格降低。
GPT-4-vision-preview	2023 年 11 月	多模态处理；图像分析和理解；创意文本生成；跨语言翻译。
ChatGPT 4-Turbo 预览版 (0125)	2024 年 1 月	增强任务执行能力，更全面地完成代码生成等任务。
ChatGPT 4-Turbo	2024 年 4 月	自带读图能力，无需使用 4v 接口
GPT-4o	2024 年 5 月	可以实时对音频、视觉和文本进行推理，能够带着情感与用户对话。
GPT-4o mini	2024 年 7 月	最具成本效益的小型模型，学术基准测试中超越了 GPT-3.5 Turbo 和其他小型模型，在文本智能和多模态推理方面表现出色。
ChatGPT-o1	2024 年 9 月	深度思考能力强，但不具备浏览网页、处理文件和图像的能力。
ChatGPT-o1-mini	2024 年 9 月	适用于需要推理和无需广泛世界知识的任务
ChatGPT-o1-preview	2024 年 9 月	在复杂推理任务中表现出色，特别是在科学、编程、数学等领域。
GPT-4o with Canvas	2024 年 10 月	由 OpenAI 的 GPT-4o 来构建，需要在模型选择栏中单独 pick 出来使用——“GPT-4o with canvas”。允许用户对生成的内容进行二次创作和编辑。
o3&o3-mini	2024 年 12 月	o3 在编程领域，展示了媲美顶级程序员的实力；在科学推理方面，创造了多项新纪录；它的精简版 o3-mini 将在 1 月底向公众开放。
GPT-4.5	2025 年 2 月	GPT-4.5 在预训练规模上进一步 Scaling。在多项基准测试中，GPT-4.5 实力碾压 GPT-4o，尤其在数学能力上飙升 27%，编码能力提升 7%-10%。
o3-pro	2025 年 6 月	o3 推理模型的升级版。推理模型可逐步推理解答问题，在物理、数学和编程等领域表现得更可靠。在评估模型数学能力的 AIME 2024 测试中，o3-pro 的表现优于谷歌顶级模型 Gemini 2.5 Pro。在博士级水平的科学知识测试 GPQA Diamond 中，o3-pro 也击败了 Anthropic 最新发布的 Claude 4 Opus 模型。
GPT-5	2025 年 8 月	GPT-5 是 OpenAI 首个“一体化”的 AI 系统，整合了 GPT 系列（大语言模型）和 o 系列（推理模型），具备调度子模型的能力。作为 OpenAI 的最强模型，GPT-5 在编程能力、创意写作、健康领域等关键领域实现了显著提升。
Claude (Anthropic)		
Claude 初代	2023 年 3 月	在各种任务上有一定表现，但在编程、数学、推理等能力方面存在提升空间。
Claude 2	2023 年 7 月	输入长度增加：单次可输入长度达 10 万 token；能力提升：编程、数学、推理技能得到改善，在一些测试中的得分有所提高。
Claude 3	2024 年 3 月	Claude 3 Haiku：速度较快，能在短时间内理解和响应查询。 Claude 3 Sonnet：在智能水平和功能上有进一步提升。 Claude 3 Opus：是该系列中最智能、最强大的版本，在常见的 AI 基准测试中表现出色，在很多方面接近人类的理解和表达水平。
Claude 3.5	2024 年 6 月	翻译，文本创作，编程，视频处理，计算机使用。

Claude 3.5 加强版	2024 年 10 月	具备了像人类一样操作电脑的能力，能自动执行多个步骤的复杂任务，如在线填写表格、搜索数据、提交报告等，并且能够根据指令在电脑界面上自主导航，通过切换窗口、打开菜单、选择具体选项等操作完成复杂的多步骤任务。
Claude 4	2025 年 5 月	Claude Opus 4 和 Claude Sonnet 4。这两款模型均为混合模型，提供两种模式：即时回复和用于更深入推理的扩展思考（extended thinking）。最高可独立运行 7 小时，并推出了文件 API、提示词缓存等新功能。
Grok (xAI)		
Grok 1	2023 年 11 月	提供生成文本、代码、邮件、信息检索等功能。
Grok 1.5	2024 年 3 月	推理和问题解决能力提升；长文本理解能力增强；基础设施优化。
Grok 1.5V	2024 年 4 月	除了强大的文本功能外，还可以处理各种视觉信息，包括文档、图表、屏幕截图和照片。
Grok 2.0	2024 年 8 月	用户可以根据提示直接生成图像。创建表情包等创意内容。
Aurora	2024 年 12 月	X 公司为 Grok 推出了一个新的 AI 图像生成模型“Aurora”，该模型能够生成比 Grok 其他图像生成模型更加逼真的图像，并且对生成内容的限制也更少。
Grok 3.0	2025 年 2 月	引入了“思维链”（Chain of Thought）推理技术，支持文本、图像等多种数据类型的处理与生成。在数学推理、科学逻辑及代码生成等核心能力上全面超越 DeepSeekV3、GPT-4o 等头部模型，其推理性能较前代模型提升 16%-27%。
Grok 4	2025 年 7 月	Grok 4 系列包含 Grok 4 和 Grok 4 Heavy 两个版本，均为纯推理模型。Grok 4 是单代理版本，Grok 4 Heavy 是多代理版本，支持四个代理同时工作，上下文窗口最高支持 256k tokens。
Gemini (Google)		
Gemini 1.0	2023 年 12 月	原生多模态大模型，性能大致相当于 GPT-3.5 级别。
Gemini-Ultra		1.0 下最大规模参数版本，各项评测的结果都超过了 GPT-4V。
Gemini-Pro		比 Ultra 小一点版本的模型，注重自然语言理解、可视化、分析等方面。
Gemini-Nano		移动设备部署，两个版本：一个是 18 亿参数的 Gemini-Nano1，一个是 32.5 亿参数的 Gemini-Nano2。
Gemini 1.5	2024 年 2 月	能够同时处理文本、图像、音频、视频等信息。
Gemini 1.5 Pro	2024 年 5 月	处理的上下文从 12.8 万个 token 增加到 100 万个 token，首次提供了本地音频（语音）理解功能和全新的文件 API。
Gemini 1.5 Flash	2024 年 5 月	专为需要高频率、低延迟响应的任务优化，体积更小，保留了 100 万个 token 的上下文窗口。适用于特定场景下的快速响应需求。
Gemini-1.5-Pro-Exp-0801	2024 年 8 月	MMLU-Pro 基准测试中性能提升了约 7%，在 MATH 和 HiddenMath 基准中提高了约 20%，在视觉和代码用例中提高了约 2-7%。
Gemini 1.5 Pro Exp-0827	2024 年 8 月	在编程、复杂提示词处理等方面进行了增强。
Gemini 1.5 Flash Exp-0827	2024 年 8 月	Gemini 1.5 Flash 的升级版，在编程、复杂提示词处理等方面进行增强。
Gemini-1.5-Pro-002	2024 年 9 月	1.5 Pro 的价格降低了 50% 以上（针对提示词（prompt）<128K 的输入和输出），速率限制提升了 3 倍。
Gemini-1.5-Flash-002	2024 年 9 月	1.5 Flash 的速率限制提升了 2 倍。
Gemini-Exp-1114	2024 年 11 月	编码性能显著提高；提升推理能力，视觉理解能力。
Gemini-Exp-1121	2024 年 11 月	升级版实验模型，在上一代的基础上增强了代码能力、推理能力、多模态视觉能力。
Gemini 2.0	预计 2024 年 12 月	CoT（Chain of Thought）推理能力、多模态能力的增强、上下文窗口的扩展等。
Gemini 2.0 Flash Thinking	2024 年 12 月	只支持 32000 个 Token 的输入内容（大约相当于 50 至 60 页的文本量），并且每个输出响应能够生成 8000 个 Token。擅长“多模态理解、推理”以及“编码”任务。支持图像上传和分析；可通过下拉菜单向用户展示其逐步推理的过程。
Gemini 2.5 Pro	2025 年 3 月	具备统一的推理能力，以及用户所喜爱的 Gemini 的所有功能（长上下文、工具等）。
Gemini 2.5 Flash	2025 年 4 月	引入了革命性「思考预算」，可灵活控制推理深度，比肩 o4-mini。关闭思考模式成本直降 600%。具体来说，Gemini 2.5 Flash 关闭思考输出价格 0.6 美元/百万 token，开启思考输出价格 3.5 美元/百万 token。
Gemini 2.5 Pro Preview (I/O 版)	2025 年 5 月	可通过 Gemini API 以及谷歌的 Vertex AI 与 AI Studio 平台获取，价格则与升级前的 Gemini 2.5 Pro 模型相同。

Gemini 2.5 Flash-Lite	2025 年 6 月	Gemini 2.5 Pro、Gemini 2.5 Flash 从实验预览版变为正式版，可投入企业应用。Gemini-2.5-Flash-Lite 比 Gemini-2.5-Flash 便宜 30%-60%，输入价格为 0.1 美元（折合人民币约 0.7 元）/百万 tokens，输出价格为 0.4 美元（折合人民币约 2.9 元）/百万 tokens。
Gemini 2.5 Flash-Lite 稳定版	2025 年 7 月	能够原生支持高达 100 万 token 的上下文，每百万输入 token 的费用为 0.10 美元，而每百万输出 token 的费用为 0.40 美元，这与竞争对手 GPT-4.1Nano 的价格相当。
Gemini 2.5 Deep Think	2025 年 8 月	多个关键基准测试中超越了 OpenAI 的 o3 和 xAI 的 Grok 4。
Whisk (Google)		
Whisk	2024 年 12 月	用户无需使用典型的文本提示，只需上传一张图片，然后确定他们创作的主题、场景和风格，实现图片生成图片。
Mistral Large (Mistral)		
Mistral Large	2024 年初	流利地使用英语、法语、西班牙语、德语和意大利语等多种语言。
Mistral Large 2	2024 年 7 月	在代码生成、数学和推理方面比 Mistral Large 功能更强大，提供更强大的多语言支持和高级函数调用功能。
Mistral Small 3.1	2025 年 3 月	240 亿参数，128K 上下文，不仅能在本地运行，降低成本，而且部分性能超过 OpenAI 的 GPT-4o mini。
Mistral Medium 3	2025 年 5 月	同时具有前沿级别的性能和明显更低的成本，并具备多种企业功能。Medium 3 与 Anthropic 旗下的 Claude Sonnet 3.7 相比，在各项基准测试中的表现都达到或超过了后者的 90%，且成本仅为每百万 Token 输入 0.4 美元、每百万 Token 输出 2 美元。
推理模型 Magistral	2025 年 6 月	Magistral 能实现欧洲语言逻辑推理，突破美国模型仅限英语、中国模型仅限中文的局限。
Titan 系列 (Amazon)		
Titan Image Generator	2023 年 11 月	图像生成模型，能够根据用户输入的英文提示词句生成“专业等级”的图片。
Amazon Titan Text Express	2023 年 11 月	文字生成模型，提供最多 8000 token 输入，支持“检索增强的自然语言处理”、开放性文本生成、摘要重点、写代码、生成表格、文段改写、关联思考（chain of thought）、问答及聊天等。
Titan Text Lite	2023 年 11 月	文字生成模型，提供最多 4000 token 输入，为企业特定使用场景而设计，为摘要及文案写作（copywriting）做了重点优化。
Titan Image Generator V2	2024 年 8 月	生成图片之后，支持用户进一步细化，例修改和编辑特定区域等。
Llama 系列 (Meta)		
Llama 2	2023 年 7 月	模型架构与 Llama 1 基本保持不变，但用于训练基础模型的数据增加了 40%。新增功能：推出了针对对话进行微调的模型 LLaMA-2 Chat。
Llama 3	2024 年 4 月	最大版本的参数量飙升至超过 4000 亿。
Llama 3.1	2024 年 7 月	拥有 128k tokens 的大上下文长度、多语言能力和工具使用能力等。发布了两个新模型：Llama Guard 3 和 Prompt Guard，Llama Guard 3 是一个能够对 LLM 输入和生成进行分类的安全防护模型，Prompt Guard 是一个可以检测提示注入和越狱行为的小型分类器。
Llama 3.2	2024 年 9 月	11B 和 90B 版本加入了图像编码器，能够支持文档级的图表理解和视觉定位任务。
Llama 3.3	2024 年 12 月	共有 700 亿参数，性能方面媲美拥有 4050 亿参数的 Llama 3.1。Llama 3.3 模型效率更高、成本更低，可以在标准工作站上运行，降低运营成本的同时，提供高质量文本 AI 解决方案。Llama 3.3 模型重点优化了多语言支持，支持英语、德语、法语、意大利语、葡萄牙语、印地语、西班牙语和泰语 8 种语言。
DALL-E 系列 (OpenAI)		
DALL-E 3	2023 年 9 月	生成的图片还可以直接带上文字。
Midjourney (Leap Motion)		
V5 版本	2023 年 3 月	分辨率更高，达到了 1024 × 1024 像素；支持无限的宽高比。
V5.1 版本	2023 年 5 月	图像风格化方面更加强烈。
V5.2 版本	2023 年 6 月	图像质量进一步提升。
V6.1 版本	2024 年 8 月	图像的连贯性进一步增强；生成速度加快。
Stable Diffusion (Stability AI)		

Stable Diffusion XL 2023 年 7 月 在 1024x1024 图像上进行训练。

Stable Diffusion XL Turbo	2023 年 11 月	使用对抗扩散蒸馏 (ADD) 技术进行训练，使得模型能够在一次迭代中生成逼真的图像。
Stable Diffusion 3	2024 年 6 月	将扩散 U-Net 替换为扩散变压器 (DIT)。
Stable Diffusion 3.5	2024 年 10 月	用户可通过修改提示词调取更多的图像风格。
SD 3.5 Large		80 亿参数，提供了卓越的图片质量和高度文本语义还原，是 Stable Diffusion 最强模型。
SD 3.5 Large Turbo		能在 4 步之内生成高质量的图像，并且对提示的遵循性非常好，速度比 Stable Diffusion 3.5 Large 快。
SD 3.5 Medium		25 亿参数法，能在消费级硬件上即开即用，可生成 0.25—200 万像素之间生成图像，兼顾了质量和易定制性。
sora (OpenAI)		
预览版	2024 年 2 月	文生视频。
独立推出	2024 年 12 月	Sora 成为独立产品，OpenAI 为其专门开设网站 Sora.com。

资料来源：各公司官方平台，华安证券研究所整理

6. 国内主要大模型迭代情况

图表 10 国内主要大模型迭代情况（截至 2025-08-09）

模型名称	推出时间	简介
文心一言（百度）		
ERNIE 1.0	2019 年 3 月	知识增强的文心大模型 ERNIE 1.0，文心一言的早期版本之一。
ERNIE 2.0	2019 年 7 月	学习大规模语料中的词法、语法、语义等知识。
ERNIE 3.0	2021 年 7 月	发布文心大模型 3.0 (ERNIE 3.0)，首次在千亿级预训练模型中引入大规模知识图谱。
文心一言 3.0	2023 年 2 月	官宣新一代大语言模型文心一言（英文名：ERNIE Bot）。
文心一言 3.5	2023 年 6 月	新增了插件机制，通过插件方式扩增了大模型的能力边界。
文心一言 4.0	2023 年 10 月	理解、生成、逻辑、记忆四大能力提升，重构了原来的搜索模式。
文心一言 4.0 Turbo	2024 年 6 月	上下文输入从 4.0 版的 2K tokens 升级到了 128K tokens；AI 生图分辨率提升，从 512x512 提升至 1024x1024；智能体技术，包括理解、规划、反思和进化。
文心一言 App 4.0.0 版本	2024 年 9 月	文心一言 App 升级为“文小言”，提供了问问题、陪聊天、写文章、画图片和下任务五大核心场景能力。
文心大模型 4.5	2025 年 3 月	首个原生多模态大模型，API 调用价格仅为 GPT4.5 的 1%。
文心大模型 X1	2025 年 3 月	性能上对标 DeepSeek-R1 的深度思考模型，同时还支持多模态、多工具调用能力，API 调用价格约为 R1 的一半。
文心大模型 4.5 Turbo	2025 年 4 月	文心 4.5 Turbo 的多项基准测试成绩显著优于 GPT 4o，平均分达到 77.68，超过 GPT 4o 的 72.76。而对比文心 4.5，文心 4.5 Turbo 速度更快、价格下降 80%，每百万 token 的输入价格仅为 0.8 元，输出价格 3.2 元，仅为 DeepSeek-V3 的 40%。
文心大模型 X1 Turbo	2025 年 4 月	文心 X1 Turbo 相较文心 X1，具备更长的思维链和更强的深度思考能力，同时进一步增强了多模态和工具调用能力，在性能提升的同时价格再降 50%，每百万 token 输入价格 1 元，输出价格 4 元。
通义千问（阿里巴巴）		
1.0 版本	2023 年 4 月	拥有多轮对话能力；文案创作与生成；逻辑推理与分析；多模态理解与交互，融合并理解图像、音频等多种信息源；多语言支持与跨语言沟通；知识问答与信息检索等。
Qwen1.5 版本	2024 年 2 月	开源了六个型号尺寸：0.5B、1.8B、4B、7B、14B 和 72B。支持 32K 上下文长度；开放了 Base + Chat 模型的 checkpoint；可与 Transformers 一起本地运行；发布了 GPTQ Int-4 / Int8、AWQ 和 GGUF 权重。
2.0 版本	2024 年 6 月	开源 5 款模型，分别是 0.5B、1.5B、7B、72B、MoE。新增图片理解、文档解析和语音聊天功能。行业应用模型包括，通义灵码：智能编码助手；通义智文：AI 阅读助手；通义听悟：工作学习 AI 助手；通义星尘：个性化角色创作平台；通义点金：智能投研助手；通义晓蜜：智能客服；通义仁心：个人专属健康助手；通义法睿：AI 法律顾问。

2.5 版本	2024 年 9 月	新增对文档的处理能力，可以处理长达 1000 万字、最多 100 个文档；新增音视频理解功能；训练参数量整体提升 18%+；支持高达 128KB 的上下文长度，可生成最多 8K 的内容；提供从 0.5B 到 72B 不同参数规模的模型。
Qwen-Agent	2025 年 1 月	Qwen-Agent 不仅仅是一个工具，更是一个全面的开发平台。开发者可以轻松开发出各种 Agent 应用，包括浏览器助手、代码解释器以及自定义助手等。赋予了智能体执行详细指令、使用工具及规划任务的能力，还整合了 RAG（检索增强生成）、代码解释器以及基于 Qwen2.5-Math 的数学推理等高级功能。
Qwen2.5-VL	2025 年 1 月	可以直接作为一个视觉 Agent 来自动化操作电脑、手机。此外，Qwen2.5-VL 还能理解超过 1 小时的长视频，并且能定位特定时间点发生的事件。
QwQ-32B	2025 年 3 月	320 亿参数的模型，其性能可与具备 6710 亿参数（其中 370 亿被激活）的 DeepSeek-R1 媲美。
QVQ-Max	2025 年 3 月	能够理解图片和视频内容，还能为上述信息提供分析并推理。还可以完成设计插图、生成短视频剧本等内容，甚至根据用户的需求创建角色扮演内容。
Qwen3 系列	2025 年 4 月	开源 Qwen3 系列模型，包含 2 个 MoE 模型、6 个稠密模型。发布 2 小时，Qwen3 模型在 GitHub 上的 star 数已超过 16.9k。其中旗舰模型 Qwen3-235B-A22B，在编程、数学、通用能力等基准评估中的表现优于 DeepSeek-R1、OpenAI o1、OpenAI o3-mini、Grok-3 和 Gemini-2.5-Pro 等业界知名模型。
Qwen3-Embedding	2025 年 6 月	专为文本表征、检索与排序任务设计，基于 Qwen3 基础模型进行训练。8B 参数规模的 Embedding 模型在 MTEB 多语言 Leaderboard 榜单中位列第一。Qwen3-Embedding 系列提供从 0.6B 到 8B 参数规模的 3 种模型配置，以满足不同场景下的性能与效率需求。
Qwen3-Coder	2025 年 7 月	编程能力登顶全球开源模型阵营，并超越 GPT-4.1 等闭源模型，比肩全球最强的编程模型 Claude 4。
Qwen-MT	2025 年 7 月	1.92 种语言互译。2. 高度可控性：提供术语干预、领域提示、记忆库等专业翻译功能，并支持用户自定义提示。3. 低延迟、低成本：采用轻量级 MoE 架构，在保证卓越性能的同时实现更快的响应速度和更低的 API 调用价格（每百万输出 token 低至 2 元）。
Qwen VLo	2025 年 6 月	一个多模态统一理解与生成模型。引入了一种新的生成机制：从上到下、从左到右逐步清晰的生成过程。适用于需要精细控制的长段落文字生成任务。现已可在 Qwen Chat 上直接体验新模型 Qwen VLo。
Qwen3-30B-A3B-Instruct-2507	2025 年 7 月	非思考模式的最新模型在技术架构上实现重要突破，长文本理解能力显著提升至 256K 规模。该模型采用创新的参数激活策略，仅需激活 3B 参数即可达到与 Gemini 2.5-Flash 和 GPT-4o 等主流闭源模型相当的性能水平。
混元（腾讯）		
混元大模型	2023 年 9 月	支持多轮对话、内容创作、逻辑推理、知识增强等功能；高级版每 1000 token 收费 0.14 元。
hunyuan-pro	2024 年 4 月	原“混元大模型高级版”；升级为 32K 长文大规模多专家模型，输入上限扩展为 30K tokens，输出上限 2K tokens；预训练数据量提升；支持文字生成图像能力；代码能力大幅提升 20%；正式对外开放“文生图”功能。输入：0.03 元/千 token；输出：0.1 元/千 token。
hunyuan-standard	2024 年 4 月	原“混元大模型基础版”，实现对长文本输入的处理。输入：0.0008 元/千 token；输出：0.002 元/千 token。
hunyuan-lite	2024 年 4 月	超长上下文 MOE 模型，最大输入支持 250K，最大输出支持 6K；在中文 NLP、英文 NLP、代码、数学等方向显著优于同等规模开源模型
hunyuan-standard-256k	2024 年 5 月	256K 超长上下文多专家模型，最大输入 250K，最大输出 6K。输入：0.0005 元/千 token；输出：0.002 元/千 token。
hunyuan-code	2024 年 7 月	混元代码专属模型，在代码对话生成场景具有更好的基础效果；Python、C++、JS、Java、Go 等代码语言效果评测优于市场同规模模型。
hunyuan-role	2024 年 7 月	角色扮演模型，基于混元模型结合角色扮演场景数据集进行在 IP 类角色、情感陪伴类角色设定与对话上具有更好的基础效果
hunyuan-functioncall	2024 年 9 月	支持并行多工具调用，支持多轮对话下的工具调用；支持 Workflow 形式的 System Prompt 设定；支持长输入意图、支持结构化输出。
hunyuan-turbo	2024 年 9 月	训练与推理方面的效率提高了一倍；支持高分辨率图片分析；AI 搜索功能增强；元宝 2.0 新增 AI 应用专属板块；融入腾讯生态产品
hunyuan-vision	2024 年 10 月	首个支持 7K 分辨率（最大 16:1 长宽比）；图片理解、图片问答等功能；输入输出：0.018 元/千 tokens。

hunyuan-large-longcontext	2024 年 11 月	擅长处理长文任务如文档摘要和文档问答等，同时也具备处理通用文本生成任务的能力；在长文本的分析和生成上表现优异。
hunyuan-translation-lite	2024 年 11 月	混元翻译模型支持自然语言对话式翻译。
hunyuan-turbo-latest	2024 年 11 月	hunyuan-turbo 模型的动态最新版本，与 C 端（腾讯元宝）保持一致。
hunyuan-turbo-vision	2024 年 11 月	在图文理解相关的基础识别、内容创作、知识问答、分析推理等能力上相比前一代模型全面提升。
hunyuan-large	2024 年 11 月	业界参数规模最大、效果最好的 transformer 结构的 MoE 模型；高质量合成数据；长文本处理能力等。
Turbo S	2025 年 2 月	Turbo S 在首字时延上降低了 44%，输入价格为 0.8 元/百万 tokens，输出价格为 2 元/百万 tokens，相比前代模型价格大幅下降。混元 Turbo S 还将成为腾讯混元系列衍生模型的核心基座。
图生视频模型	2025 年 3 月	模型总参数量保持 130 亿，模型适用于多种类型的角色和场景，包括写实视频制作、动漫角色甚至 CGI 角色制作的生成。
混元 T1	2025 年 3 月	输入价格为每百万 token 1 元，输出价格为每百万 token 4 元。速度方面，混元 T1 正式版在可比参数量下的解码性能高于行业 2 倍，首字秒出，吐字速度每秒 60 至 80token，具备优秀的长文处理能力。
混元图像 2.0	2025 年 5 月	基于模型架构创新，在行业内率先实现实时生图。腾讯混元图像 2.0 模型还发布了实时绘画板功能，基于模型的实时生图能力，用户在绘制线稿或调整参数时，预览区同步生成上色效果，突破了传统“绘制-等待-修改”的线性流程，可助力专业设计师的创作。
混元 3D 2.1	2025 年 6 月	首个全链路开源的工业级 3D 生成大模型。优化了细节建模，使网格精度更高，具备更好的拓扑一致性，为后续纹理映射提供了基础；还可生成基础颜色、金属度、法线、粗糙度等贴图，支持皮革、木质、金属、陶瓷等多种复杂材质的高质量渲染。混元 3D 2.1 模型适配于消费级显卡。
Hunyuan-A13B	2025 年 6 月	首个混合推理 MoE 模型 Hunyuan-A13B，总参数 80B，激活参数仅 13B，效果比肩同等架构领先开源模型，推理速度更快，性价比更高。已经在 Github 和 Huggingface 等开源社区上线，同时模型 API 也在腾讯云官网正式上线，支持快速接入部署。
Hunyuan3D-PolyGen	2025 年 7 月	业界首个美术级 3D 生成大模型 Hunyuan3D-PolyGen。结合自研高压缩率表征 BPT 技术，可生成面数达上万面的复杂几何模型，建模效率提升超 70%。
混元-0	预计 2025 年	基于腾讯自研大模型混元的首个端到端语音通话模型 Hunyuan-Voice，最快今年 6 月上线腾讯元宝 App。而且腾讯已规划多模态和全模态路线，最快今年将推出全球首个“全模态模型”，代号为混元-0。

盘古（华为）

3.0 版本	2023 年 7 月	包含五大基础大模型：盘古 3.0 包含 5 个基础大模型，分别是 NLP 大模型、CV 大模型、多模态大模型、预测大模型和科学计算大模型。提供数据工程套件、模型开发套件、应用开发套件；NLP 大模型功能：包括会议助手、文案生成、多轮对话、实体抽取、翻译、知识问答等能力；CV 大模型功能：多模态大模型功能：支持图像分类、物体检测、姿态估计等；图像生成、理解、视频生成等；预测大模型功能。
5.0 版本	2024 年 6 月	包括十亿级参数的 Pangu E 系列、百亿级参数的 Pangu P 系列、千亿级参数的 Pangu U 系列以及万亿级参数的 Pangu S 系列超级大模型。图片和视频识别方面，可支持 10K 超高分辨率；内容生成方面，采用业界首创的 STCG 技术。
盘古 Pro MoE	2025 年 7 月	盘古 Pro MoE 模型基于分组混合专家模型（Mixture of Grouped Experts, MoGE）架构构建，总参数量为 720 亿、激活参数量达 160 亿，并针对昇腾 3001 Duo 和 8001 A2 平台进行系统优化。

星火（科大讯飞）

星火大模型 V1.0	2023 年 5 月	具备文本生成、语言理解、知识问答、逻辑推理、数学能力、代码能力、多模交互七大核心能力。
星火大模型 V1.5	2023 年 6 月	提升了开放式知识问答、多轮对话、逻辑和数学能力。
星火大模型 V2.0	2023 年 8 月	提升了代码能力、多模态能力。
星火大模型 V3.0	2023 年 10 月	提升了七大核心能力，在中文语境下超越 ChatGPT。
星火大模型 V3.5	2024 年 1 月	提升了逻辑推理、语言理解、文本生成、数学答题、代码、多模态能力，语言理解、数学能力超越 GPT4-Turbo。数学、语言理解、语音交互能力超 GPT-4 Turbo，并首次发布星火语音大模型，首批 37 个主流语种语音识别效果超过 OpenAI Whisper V3。

星火大模型 V4.0	2024 年 6 月	V4.0 在文本生成、语言理解、知识问答、逻辑推理和数学五大能力方面完成了对 GPT-4 Turbo 的整体超越，并进一步缩小了在代码、多模态能力方面的差距，多模态能力达到了后者 97% 的水平。
星火大模型 4.0 Turbo	2024 年 10 月	七大核心能力全面超过 GPT-4 Turbo，数学和代码能力超越 GPT-4o，效率相对提升 50%，在国内外中英文 14 项主流测试集中实现了 9 项第一。
星火大模型 X1	2025 年 1 月	我国首个基于国产算力训练的具备深度思考和推理能力的大模型。与通用大模型相比，其解题过程更接近人类的“慢思考”方式。
GLM (智谱 AI)		
GLM-130B	2022 年 7 月	具备双语支持；自回归掩码语言建模；旋转位置编码 (RoPE)；深度规范化 (DeepNorm) 等核心能力。
ChatGLM-130B	2023 年 3 月	具备双语支持，高精度表现，可复现性，跨平台支持，无需后期训练的情况下达到 INT4 量化。
ChatGLM-6B	2023 年 3 月	更小的版本 ChatGLM-6B 开源。
ChatGLM2	2023 年 6 月	ChatGLM2-6B 的上下文长度扩展到了 32K，并在对话阶段使用 8K 的上下文长度训练，允许更多轮次的对话；推理速度相比初代提升了 42%；对学术研究完全开放；发布了 ChatGLM2-6B-32K 模型（提升对长文本的理解能力）和 CodeGeeX2 模型（加入代码预训练实现）。
ChatGLM3	2023 年 10 月	引入了多模态理解能力；首次加入了代码识别模块 Code Interpreter；支持网络搜索；推理框架基于最新的高效动态推理和显存优化技术，推理速度提升了 2-3 倍，推理成本降低了一倍；推出了端侧模型；支持在多种国产硬件生态上进行大规模预训练和推理。
GLM-4	2024 年 1 月	性能全面提升 60%；支持长达 128K 的上下文处理能力；更高精度的文生图性能和更丰富语义的图片理解；GLMs 个性化智能体定制能力上线；能够处理 128K token 的长文本；推出了 ChatGLM-Math；开发了 Self-Contrast，一种无需反馈的对齐策略；开发了 AgentTuning 框架。
GLM-4-9B	2024 年 6 月	支持多轮对话交互；预训练数据量提升了 3 倍；训练效率提高了 3.5 倍；模型规模提升至 9B，预训练计算量增加了 5 倍；GLM-4-9B 集成了 All Tools 功能。 包括多个版本： 基础版本：GLM-4-9B (8K) 对话版本：GLM-4-9B-Chat (128K) 超长上下文版本：GLM-4-9B-Chat-1M (1M) 多模态版本：GLM-4V-9B-Chat (8K)
GLM-4-PLUS	2024 年 8 月	模型在理解、指令遵循和长文本处理等方面表现与 GPT-4 和拥有 405B 参数的 Llama3.1 不相上下。GLM-4-Plus 通过精准的长短文本数据混合策略，显著提升了长文本推理效果。
AutoGLM 沉思	2025 年 3 月	全新智能体不仅具备深度研究能力 (Deep Research)，还能实现实际操作 (Operator)，真正推动 AI Agent 进入“边想边干”的阶段。
CogView4	2025 年 3 月	首个支持中英双语提示词的开源文生图 ai 模型，也是全球首个支持生成汉字的开源文生图模型，尤其擅长理解和遵循中文提示词，能在画面中生成汉字。
GLM-Z1-AirX	2025 年 4 月	GLM-Z1-AirX (极速版)：国内最快推理模型，推理速度高达 200 tokens/s，8 倍于常规速度； GLM-Z1-Air (高性价比版)：价格仅为 DeepSeek-R1 的 1/30，适合高频调用场景； GLM-Z1-Flash (免费版)：支持免费使用，进一步降低模型使用门槛。
GLM-4.1V-Thinking	2025 年 7 月	轻量版 GLM-4.1V-9B-Thinking 模型参数控制在 10B 级别，在兼顾部署效率的同时实现性能突破。
豆包 (字节跳动)		
2024 年 1 月迭代版本	2024 年 1 月	首页改版进一步强化智能体的露出，智能体置顶到首页顶部，支持滑动展示；开始对智能体做进一步的分类；将搜索框固定到首页顶部，强化搜索功能。

2024 年 2 月迭代版本	2024 年 2 月	虚拟聊天智能体增加聊天背景封面优化聊天氛围；国内版智能体开发平台“扣子”上线，和豆包打通，支持发布豆包；豆包 AI 拓展字节内部其他产品，包括抖音应用内上线 AI 搜索功能，今日头条 APP 首页右上角上线豆包 AI 功能。
2024 年 3 月迭代版本	2024 年 3 月	智能体开始灰度支持定制个人的音色功能；增加文档对话分析能力，支持上传多种格式文件；接入 bing 搜索，联网搜索拓展全网搜索；上线豆包 PC 客户端和浏览器插件。
2024 年 5 月迭代版本	2024 年 5 月	豆包大模型在火山引擎原动力大会上正式发布，包含通用模型 pro、通用模型 lite、语音识别模型、语音合成模型、文生图模型等 9 款模型。同时，豆包大模型的定价大幅低于行业价格，主力模型 pro-32k 版的价格比行业低 99%。
2024 年 6 月迭代版本	2024 年 6 月	豆包大模型推出桌面客户端，支持 Windows 与 macOS。
2024 年 8 月迭代版本	2024 年 8 月	豆包大模型支持实时语音通话。
2024 年 9 月迭代版本	2024 年 9 月	豆包两款视频生成模型发布，多个垂类模型上新，豆包通用模型 pro 和文生图模型、语音合成模型等垂类模型升级。
豆包视频生成-PixelDance		多镜头语言能力；支持生成具有连贯动作的视频；能够根据单一文本提示生成具有起承转合的完整故事短片；支持多种艺术风格。
豆包视频生成-Seaweed		支持从文本或图像生成视频；原生支持多种分辨率的视频生成；高逼真度；专业级视觉美感；多分辨率生成；默认输出为 720p 分辨率、24fps、时长 5 秒，可动态延长至 20-30 秒。
豆包音乐模型		采用独特的技术方案，实现了音乐生成的通用架构，从词曲唱三个方面来生成高质量的音乐作品。
豆包同声传译模型		采用端到端的模型架构，让翻译高精准、高质量，时延更低，还支持音色克隆，媲美真人同传效果。
视觉理解模型	2024 年 12 月	为企业提供极具性价比的多模态大模型能力。豆包视觉理解模型千 tokens 输入价格仅为 3 厘，一元钱就可处理 284 张 720P 的图片。
1.5 Pro	2025 年 1 月	采用了创新的 Speech2Speech 端到端框架，能够实现更自然的语言表达，包括情感化的语音输出，甚至可以说方言和唱歌。
1.5 • 深度思考模型	2025 年 4 月	包含两个版本，第一个主线版本是大语言模型 Doubao-1.5-thinking-pro，第二个是具备多模态能力的视觉版 Doubao-1.5-thinking-pro-vision。前者推理能力更强，后者支持视觉推理，开发者可以按需使用。模型采用 MoE 架构，总参数量为 200B，激活参数仅 20B，在多项基准测试中达到或接近全球第一梯队水平。
文生图模型 3.0	2025 年 4 月	Seedream3.0 支持 2K 分辨率直出，生图结构准确性、小字生成与排版、美感、真实度等方面具有优势。
1.6	2025 年 6 月	豆包 1.6 系列模型支持多模态理解和图形界面操作，能够理解和处理真实世界问题。图形界面操作能力则让豆包 1.6 进一步具备“行动力”，演示案例显示，豆包 1.6 可自动操作浏览器完成酒店预定，识别购物小票并整理成 Excel 表格等任务。
视频生成模型 Seedance 1.0	2025 年 6 月	视频生成模型 Seedance1.0 pro 支持文字与图片输入，可生成多镜头无缝切换的 1080P 高品质视频，主体运动稳定性与画面自然度较高。在国际知名评测榜单 Artificial Analysis 上，Seedance 在文生视频、图生视频两项任务上均排名首位。
图像编辑模型 3.0	2025 年 7 月	能够通过自然语言指令，实现对图像的精细化编辑，包括消除多余内容、改变光影效果、替换文字等操作。该模型还具备对风格、结构与语义的精准控制力，能够理解复杂的指令，并进行推理，从而解锁更多创新的修图场景。
同声传译模型 2.0	2025 年 7 月	采用了全双工语音理解和生成框架，将传统机器同传的语音延迟从 8-10 秒降低到 2-3 秒，实现了文本与语音的同步生成。
可灵（快手）		
初始版本	2023 年 11 月	基于快手自研大模型可灵和可图，提供视频及图像生成与编辑能力，提供视频生成、图像生成、虚拟试穿等接口能力。
V1.0 版本	2024 年 6 月	可灵 1.0 模型在高品质模式下可生成 720p 视频。
V1.5 版本	2024 年 9 月	视频生成新增可灵 1.5 模型，在画质质量、动态质量、美学表现、运动合理性以及语义理解等方面均有显著提升。引入了全新的“运动笔刷”功能，进一步提升视频生成的精准控制能力。

API 服务升级	2024 年 11 月	对应用程序接口 (API) 进行重大升级，引入了 v1.5 高品质模型和视频延长功能，为企业级用户提供了更加强大的视频及图像生成与编辑服务。视频延长功能允许用户对生成的视频进行 4 到 5 秒的续写，并且支持多次续写，最长达 3 分钟的总时长。
V2.0 版本	2025 年 4 月	发布可灵 2.0 视频生成模型及可图 2.0 图像生成模型。两款模型在团队内部的多指标横向评测中，均稳居业内第一。在图生视频领域，可灵 2.0 对比谷歌 Veo2 的胜负比为 182%，对比 Runway 的 Gen-4 的胜负比达 178%，在语义响应、画面质量、动态质量等维度上显著超越对手。
V2.1 版本	2025 年 5 月	高品质模式 (1080p) 下生成 5s 视频仅需不到 1 分钟。其在标准模式 (720p) 下生成 5s 视频仅需 20 灵感值，高品质模式 (1080p) 下也只需 35 灵感值，成本下降 65%。
天工 AI (昆仑万维)		
天工 1.0	2023 年 4 月	主要功能包括文案创作、知识问答、代码编程、逻辑推演、数理推算、AI 搜索产品、多轮次深度对话、图文答案。
天工 2.0	2024 年 2 月	采用 MoE 混合模型架构，响应速度和效率提升；支持图文对话、文生图等多模态应用；新增 AI 绘画功能；新增数据分析功能；新增 AI 伴侣功能；新增 AI 算命功能；新增热梗百科功能。
天工 3.0	2024 年 4 月	新增了搜索增强功能，提升复杂对话和深度内容创作的表现；引入了研究模式，自动生成研究大纲、图谱、实践总结、思维导图等；新增了调用代码及绘制图表的能力；能够进行多次联网搜索；针对性地训练了模型的 Agent 能力；新增了图表对比生成功能。
天工 4.0 01 版	2024 年 11 月测试	具备自我反思和思考链技术，能模仿人类的思维过程。
Skywork 01 Open		Llama 3.1 8B 的开源模型，解锁了许多轻量级模型无法解决的复杂数学任务，为推理模型在轻量级设备上的部署提供了可能性。
Skywork 01 Lite		具备完整的思考能力，特别优化了中文支持，并显著提升了推理与思考速度。
Skywork 01 Preview		完整版推理模型，展现了更多样且深入的思考过程，提供了更为完善和高质量的推理结果。
天工超级智能体	2024 年 5 月	建了一个由“5 个专家智能体”+“1 个通用智能体”组成的垂直专业系统。具体来看，5 个专家智能体分别专注于专业文档 (doc)、PPT、表格 (excel)、播客和网页的生成。其中，文档、PPT、表格是办公人群最常用的需求，也是 Microsoft Office、Google Workspace 的核心功能。
Kimi (月之暗面)		
V1.1	2023 年 11 月	主要功能包括：多文件处理能力及解析能力增强，搜索能力；多语言支持；多终端同步输出，各个终端可以同步输出内容。
V1.2	2023 年 12 月	新增了扫描件识别功能；推出了小程序版的 Kimi 智能助手。
V1.3	2024 年 1 月	支持 Kimi 常用语；新增语音输入/播报功能；支持搜索引用溯源。
功能更新	2024 年 7 月	开放平台上下文缓存 (Context Caching) 功能开启公测，可为开发者降低最高 90% 的长文本旗舰大模型使用成本。
Kimi 浏览器插件	2024 年 7 月	上线浏览器插件，提供点问笔和总结器两个功能。
Kimi 探索版	2024 年 10 月	月之暗面正式上线具备 AI 自主搜索能力的 Kimi 探索版，搜索量是普通版的 10 倍，一次搜索即可精读 500 个页面。
Kimi 数学版 (K0-Math)	2024 年 11 月	Kimi k0-math 的数学能力可对标 OpenAI o1 系列可公开使用的两个模型：o1-mini 和 o1-preview。
视觉思考模型 k1	2024 年 12 月	k1 模型基于强化学习技术打造，原生支持端到端图像理解和思维链技术，并将能力扩展到数学之外的更多基础科学领域。在数学、物理、化学等基础学科的基准能力测试中，初代 k1 模型的表现超过了全球标杆模型 OpenAI o1、GPT-4o 以及 Claude 3.5 Sonnet。
kimi - latest	2025 年 2 月	上下文长度能达到 128k，用户可以根据实际需求选择不同的模型计费，如 8k、32k 或 128k。新增了视觉模型支持，能理解和分析图片，具备自动上下文缓存功能，缓存中的 Tokens 费用仅为 ¥1/MTokens。在功能方面，kimi - latest 与原有的 moonshot - v1 系列保持一致，包括 ToolCalls、JSONMode、PartialMode 以及联网搜索功能等。
Kimi-Researcher	2025 年 6 月	首个 Agent 产品。基于端到端自主强化学习技术训练的新一代 Agent 模型，在 HLE 测试中，超过 Claude 4 Opus、Gemini 2.5 Pro 及 OpenAI Deep Research，与 Gemini-Pro 的 Deep Research Agent 打平。

资料来源：各公司官方平台，华安证券研究所整理

2.2 互联网

1. 本地生活

● 美团宣布向中小商户发放助力金 单店最高 5 万元

8 月 5 日，美团宣布，为坚定落实“反内卷”，促进行业健康发展，即日起正式启动中小商户发展扶持计划，向中小商户发放助力金，支持更多餐饮小店实现稳定增长。（电商派）

● 抖音生活服务上线“随心团”独立入口

8 月 6 日，抖音生活服务“随心团”业务近日已在拥有一个新的独立入口。用户在抖音内搜索“随心团”即可看到该入口。页面内除了显示本周推荐的活动商品，还有下午茶、夜宵等分类下的商品，部分商品支持免费配送。入口内还显示有优惠补贴活动，活动贯穿整个 8 月，这也将进一步强化价格吸引力，这一动作也将为随心团进行引流。（电商派）

2. 综合电商

● 淘宝上线大会员体系：打通阿里全系资源

淘宝上线全新大会员体系，打通饿了么、飞猪、盒马等阿里系资源，包含购物、外卖、旅行、出行等权益，全面覆盖吃喝玩乐衣食住行，88VIP 用户权益也将全面升级。（电商派）

● 飞猪提升至淘宝一级入口

淘宝 APP 日前改版，飞猪被提升至一级入口，与今年重点发力的“淘宝闪购”并列。长期位于二级频道的旅行业务首次进入淘宝前台核心流量区，意味着飞猪在淘宝体系中的战略地位显著提升，将与即时零售共同承担新的增长任务。（电商派）

3. 海外消费

● Shopify 公布 2025 年第二季度财务业绩

8 月 7 日，Shopify 公布 2025 年第二季度财务业绩。财报显示，Shopify 二季度营收 26.8 亿美元，同比增长 30.7%；商品交易总额（GMV）达到 878 亿美元，同比增长 29%；运营利润为 2.91 亿美元，净利润从去年同期的 1.71 亿美元跃升至 9.06 亿美元；调整后每股收益为 0.35 美元；自由现金流达 4.22 亿美元，自由现金流利润率达到 16%，连续八个季度保持两位数水平。（电商派）

4. 产业互联网

● 二季度小米手机超越苹果，成为欧洲市场份额排名第二的厂商

根据全球技术市场分析公司 Canalys 的最新数据，今年二季度，小米手机成功超越苹果，成为欧洲市场份额排名第二的厂商。（电商派）

2.3 传媒

1. 影视

● 影视票房

上周（2025/07/28-08/03），全国电影市场实现票房收入 17.5 亿元，较上上周 11.88 亿元增长显著，票房整体呈显著上升状态，今夏暑期档观影热潮持续升温。由申奥执导的抗战题材影片《南京照相馆》以单周 11 亿票房产出遥遥领先，稳居票房榜首。据猫眼专业版数据，截至 8 月 3 日，2025 年暑期档总票房（含预售）破 70 亿，《南京照相馆》《长安的荔枝》《侏罗纪世界：重生》位列暑期档票房榜三甲。（腾讯网，中国电影报）

● 《浪浪山小妖怪》累计票房破 3 亿

8 月 7 日，电影《浪浪山小妖怪》官方微博发文宣布，累计票房破 3 亿。
(IT之家)

2. 游戏

● 2025年上半年的全球手游收入排行榜

外媒 Mobilegamer.biz 发布了 2025 上半年的全球手游收入排行榜。《王者荣耀》仍稳居榜首，今年上半年收入高达 10.05 亿美元。两款 4X 策略游戏《LastWar》和《Whiteout Survival》（无尽冬日）的持续崛起，《LastWar》的收入从去年上半年的 4.38 亿美元猛增至今年上半年的 8.95 亿美元，而《无尽冬日》的 IAP（应用内购买）收入同样实现了翻倍以上的增长，从 3.69 亿美元增至 8.34 亿美元。（腾讯网）

图表 11 国内游戏畅销榜 top 10 (2025-08-08)

#	应用	总榜	游戏榜	分类排名	关键词覆盖数/TOP3	评分统计
1	和平精英 Shenzhen Tencent Tianyou Technology Ltd	1 总榜 ▲ 3	1 游戏 ▲ 2	1 动作 ▲ 2	33714/2088	4.3 1,501万个评分
2	王者荣耀 Shenzhen Tencent Tianyou Technology Ltd	2 总榜 ▼ 1	2 游戏 ▼ 1	2 动作 ▼ 1	17812/2002	3.3 1,425万个评分
3	英雄联盟手游 Shenzhen Tencent Tianyou Technology Ltd	3 总榜 ▼ 1	3 游戏 ▼ 1	3 动作 ▼ 1	11941/1353	4.5 378万个评分
4	无尽冬日 AD Shanghai Qingdi Technology Company Limited	7 总榜 ▲ 1	4 游戏 ▲ 1	1 模拟	12830/325	4.8 45.7万个评分
5	金铲铲之战 Shenzhen Tencent Tianyou Technology Ltd	8 总榜 ▼ 1	5 游戏 ▼ 1	1 卡牌	17195/963	4.7 767万个评分
6	三角洲行动 Shenzhen Tencent Tianyou Technology Ltd	11 总榜	6 游戏	4 动作	9866/399	4.7 109万个评分
7	地下城与勇士：... AD Tencent Mobile Games	12 总榜	7 游戏	1 角色扮演	14530/783	3.3 55.2万个评分
8	蛋仔派对 网易移动游戏	13 总榜 ▲ 5	8 游戏 ▲ 2	1 家庭聚会	33684/2163	4.3 196万个评分
9	穿越火线-枪战王者 Shenzhen Tencent Tianyou Technology Ltd	16 总榜 ▼ 1	9 游戏	6 策略	12642/1231	3.8 89.1万个评分
10	火影忍者 AD Shenzhen Tencent Tianyou Technology Ltd	19 总榜	10 游戏 ▲ 1	8 动作	14507/965	4.6 339万个评分

资料来源：七麦数据，华安证券研究所

图表 12 国内游戏畅销榜 top 10 (2025-08-02)

#	应用	总榜	游戏榜	分类排名	关键词覆盖数/TOP3	评分统计
1	王者荣耀 Shenzhen Tencent Tianyou Technology Ltd	1 总榜 ▲ 7天	1 游戏	1 动作	20179/2195	3.3 1,423万个评分
2	和平精英 Shenzhen Tencent Tianyou Technology Ltd	3 总榜	2 游戏	2 动作	34049/2155	4.3 1,500万个评分
3	穿越火线-枪战王者 Shenzhen Tencent Tianyou Technology Ltd	4 总榜 ▲ 40	3 游戏 ▲ 24	3 策略 ▲ 5	12109/1278	3.8 89万个评分
4	三角洲行动 Shenzhen Tencent Tianyou Technology Ltd	7 总榜 ▼ 1	4 游戏 ▼ 1	4 动作 ▼ 1	10314/449	4.7 104万个评分
5	地下城与勇士：... AD Tencent Mobile Games	8 总榜 ▼ 1	5 游戏 ▼ 1	1 角色扮演	13861/706	3.3 55.2万个评分
6	无尽冬日 AD Shanghai Qingdi Technology Company Limited	9 总榜	6 游戏 ▼ 1	1 模拟	12356/324	4.8 45.2万个评分
7	梦幻西游 AD 网易移动游戏	15 总榜 ▼ 3	7 游戏 ▼ 1	6 动作 ▼ 1	39887/2087	4.3 24.3万个评分
8	英雄联盟手游 Shenzhen Tencent Tianyou Technology Ltd	17 总榜 ▲ 1	8 游戏 ▲ 1	7 动作	11473/1321	4.5 378万个评分
9	FC足球世界-2025 AD Tencent Mobile Games	18 总榜 ▲ 2	9 游戏 ▲ 3	1 体育	11126/1211	4.6 91.6万个评分
10	QQ飞车 Shenzhen Tencent Tianyou Technology Ltd	20 总榜 ▲ 33	10 游戏 ▲ 23	8 动作 ▲ 5	14124/747	4.1 90.4万个评分

资料来源：七麦数据，华安证券研究所

3 公司公告

图表 13 公司本周公告（2025-08-04 至 2025-08-08）

发布时间	证券代码	证券名称	公告标题
2025-08-08 19:10	002131.SZ	利欧股份	利欧股份：关于公司为全资子公司提供担保的公告
2025-08-08 19:08	002131.SZ	利欧股份	利欧股份：关于进行证券投资的公告
2025-08-08 16:56	603466.SH	风语筑	风语筑：上海风语筑文化科技股份有限公司大股东减持股份结果公告
2025-08-08 16:35	000681.SZ	视觉中国	视觉中国：2024 年年度权益分派实施公告
2025-08-08 19:01	1896.HK	猫眼娱乐	盈利警告
2025-08-08 16:20	300792.SZ	壹网壹创	壹网壹创：关于控股股东股份解除质押的公告
2025-08-07 18:17	600633.SH	浙数文化	浙数文化：浙数文化 2025 年半年度报告
2025-08-07 16:37	600373.SH	中文传媒	中文传媒：中文传媒关于为下属子公司江西新华发行集团有限公司提供担保的实施公告
2025-08-07 18:22	300182.SZ	捷成股份	捷成股份：关于控股股东部分股份质押和解除质押的公告
2025-08-07 11:44	300058.SZ	蓝色光标	蓝色光标：对外担保的进展公告
2025-08-05 20:17	300866.SZ	安克创新	安克创新：关于可转换公司债券转股价格调整的公告
2025-08-05 18:44	603888.SH	新华网	新华网：新华网股份有限公司关于使用闲置募集资金进行现金管理赎回并继续进行现金管理的进展公告
2025-08-05 18:05	601928.SH	凤凰传媒	凤凰传媒：凤凰传媒第六届董事会第二次会议决议公告
2025-08-05 18:05	601928.SH	凤凰传媒	凤凰传媒：凤凰传媒关于聘任高级管理人员的公告
2025-08-05 11:21	000156.SZ	华数传媒	华数传媒控股股份有限公司 2025 年一季度合并及母公司财务报表(终稿)
2025-08-04 21:15	002803.SZ	吉宏股份	吉宏股份：关于公司董事长减持股份的预披露公告
2025-08-04 17:49	002607.SZ	中公教育	中公教育：关于股东部分股份解除冻结及质押的公告
2025-08-04 17:37	002739.SZ	万达电影	万达电影：关于持股 5%以上股东权益变动触及 1%的公告
2025-08-04 16:11	002291.SZ	遥望科技	遥望科技：关于持股 5%以上股东部分股份质押的公告
2025-08-04 15:34	600037.SH	歌华有线	歌华有线：歌华有线关于高级管理人员变动的公告
2025-08-04 20:14	2400.HK	心动公司	正面盈利预告
2025-08-04 18:20	300770.SZ	新媒股份	新媒股份：关于回购公司股份的进展公告
2025-08-04 17:46	300343.SZ	ST 联创	ST 联创：关于回购公司股份的进展公告

资料来源：iFinD，华安证券研究所

风险提示

消费恢复不及预期风险、用户增长不及预期风险、行业竞争格局剧变风险、宏观经济波动风险、政策及监管趋严风险、科技创新落地进展不及预期风险。

分析师与研究助理简介

分析师：金荣，香港中文大学经济学硕士，天津大学数学与应用数学学士，曾就职于申万宏源证券研究所及头部互联网公司，金融及产业复合背景，善于结合产业及投资视角进行卖方研究。2015年水晶球第三名及2017年新财富第四名核心成员。执业证书编号：S0010521080002

分析师：刘京松，曼彻斯特大学经济学硕士，主要覆盖美股半导体领域，2022年加入华安证券研究所

分析师：赵亮，悉尼大学金融学硕士，主要覆盖美股互联网、港股互联网、港股新消费等细分领域，2023年加入华安证券研究所

重要声明

分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

免责声明

华安证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告由华安证券股份有限公司在中华人民共和国（不包括香港、澳门、台湾）提供。本报告中的信息均来源于合规渠道，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或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。华安证券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并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经华安证券研究所书面授权，本研究报告的任何部分均不得以任何形式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。如欲引用或转载本文内容，务必联络华安证券研究所并获得许可，并需注明出处为华安证券研究所，且不得对本文进行有悖原意的引用和删改。如未经授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

投资评级说明

以本报告发布之日起6个月内，证券（或行业指数）相对于同期相关证券市场代表性指数的涨跌幅作为基准，A股以沪深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；香港市场以恒生指数为基准；美国市场以纳斯达克指数或标普500指数为基准。定义如下：

行业评级体系

增持—未来6个月的投资收益率领先市场基准指数5%以上；

中性—未来6个月的投资收益率与市场基准指数的变动幅度相差-5%至5%；

减持—未来6个月的投资收益率落后市场基准指数5%以上；

公司评级体系

买入—未来6-12个月的投资收益率领先市场基准指数15%以上；

增持—未来6-12个月的投资收益率领先市场基准指数5%至15%；

中性—未来6-12个月的投资收益率与市场基准指数的变动幅度相差-5%至5%；

减持—未来6-12个月的投资收益率落后市场基准指数5%至15%；

卖出—未来6-12个月的投资收益率落后市场基准指数15%以上；

无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。