

2024年12月16日



华鑫证券
CHINA FORTUNE SECURITIES

豆包稳居 MAU 国内第一，谷歌 Gemini 2.0 聚焦智能体

—计算机行业周报

推荐(维持)

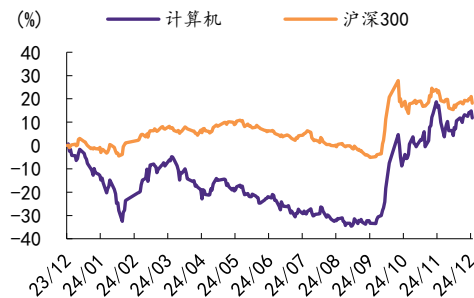
投资要点

分析师：宝幼琛 S1050521110002
baoyc@cfsc.com.cn

行业相对表现

表现	1M	3M	12M
计算机(申万)	1.7	67.6	10.0
沪深300	-0.9	24.5	17.7

市场表现



资料来源：Wind，华鑫证券研究

相关研究

- 《计算机行业点评报告：谷歌(GOOG.L)：净利润强劲增长，AI全面赋能业务》2024-12-10
- 《计算机行业点评报告：微软(MSFT.O)：营收利润稳健增长，资本开支继续扩张》2024-12-09
- 《计算机行业周报：智谱开展AgentOpenDay，展示Agent全新进展》2024-12-02

算力：算力租赁价格平稳，富士通 Monaka 处理器计划于 2027 财年亮相

本周恒源云 13 核+128G 价格环比上周持平。具体来看，显卡配置为 A100-40G 中，腾讯云 16 核+96G 价格为 28.64 元/时，阿里云 12 核+94GiB 价格为 31.58 元/时；显卡配置为 A100-80G 中，恒源云 13 核+128G 价格为 8.50 元/时；阿里云 16 核+125GiB 价格为 34.74 元/时；显卡配置为 A800-80G 中，恒源云 16+256G 价格为 9.00 元/时。

本周富士通展示了基于 Arm 的 2nm Monaka 处理器，该处理器采用博通推出的业内首个 3.5D XDSiP 技术平台。RIKEN 计算科学中心 (R-CCS) 主任、东京工业大学教授 Satoshi Matsuoka 公开了 Monaka 的更多技术细节：

- 在核心架构方面，Monaka 采用基于台积电的 CoWoS-L 封装技术的博通 3.5D XDSiP 技术平台，拥有 36 个计算小芯片。其中主要的 CPU 计算核心基于 Armv9 指令集，拥有 144 个 CPU 内核，采用台积电 2nm 制程制造，并使用混合铜键合 (HCB) 以面对面 (F2F) 方式堆叠在 SRAM tiles 上。
- 在内存接口方面，为应对广泛的数据中心工作处理需求，Monaka 不依赖于高带宽内存，而是预计使用主流的 DDR5 DRAM，有望在其 MR-DIMM 和 MCR-DIMM 实施过程中运用并提供足够的容量、降低数据中心处理器的成本。
- 在处理能力与安全性能方面，Monaka 处理器将使用基于 Armv9-A 指令集架构构建的内核，并集成 SVE2 指令（第二代可扩展矢量扩展指令）。鉴于其 A64FX 处理器具备支持高达 512 位向量的卓越能力，Monaka 处理器有望支持类似乃至更高位数的向量。同时，Monaka 将整合高级安全功能，包括 Armv9-A 的机密计算架构 (CCA)，提供增强的工作负载隔离和强大的保护。
- 在能效提升方面，富士通计划在 2026 年至 2027 年依赖空气冷却技术将处理器能效提高一倍。

AI 应用：豆包领跑国内 AI 应用市场，谷歌 Gemini 2.0 引领智能体时代

在国内 AI 模型领域，2024 年第四季度字节跳动旗下 AI 应用

产品豆包表现亮眼。在 10 月的国内月活跃人数排行榜上，豆包 APP 的 MAU（月活跃人数）达到 4717.47 万人次，远超前排名第二 Kimi 的 2028.65 万人次。字节跳动在 AI 应用领域的迅速崛起可以归功于以下因素：

1) 流量投放战略谋求“弯道超车”：相较于竞争对手，字节跳动在人工智能领域的起步时间相对延后。为了迅速缩小与行业的差距，字节跳动采用巨额投入的“烧钱”战略。截至 11 月，旗下豆包产品的投放金额已达 40363 万元，仅次于 Kimi 的投放金额 54214 万元，并且巨幅领先投放金额排名第三的星野。

2) 巨量引擎流量资源支持：在各厂商所采用的投放渠道之中，字节跳动旗下的巨量引擎占据着极为关键的地位。尽管豆包投放相对 Kimi 较少，但得益于字节跳动独有的投放经验和流量优势，豆包能够实现更加高效的获客，实现 10 月 MAU（月活跃人数）4717.47 万人次，远超前排名第二 Kimi 的 2028.65 万人次

当前，面临网络效应局限和付费订阅变现的难题，AI 对话产品可能仅为过渡形态，长期发展需要注重用户视觉化体验和较低的用户门槛。字节跳动预计将提升即梦产品优先级，探索打造 AI 时代的“抖音”。

在国外 AI 模型领域，12 月 12 日谷歌发布的 Gemini 2.0 Flash 实验版本，标志着 Gemini 2.0 系列的正式启航。相较于前一代产品，Gemini 2.0 Flash 在速度上实现了翻倍，支持图像、视频和音频等多模态输入。此外，2.0 Flash 版本新增了多模态输出功能，能够原生生成图像与文本的混合内容，并提供可控制的多语言文本转语音（TTS）音频。此次更新不仅是对语言模型的升级，而是构建了一个满足智能体基本需求的统一底层模型。

在模型具体性能提升方面，Gemini 2.0 通过性能提升，增强了理解复杂指令、长期规划和组合函数调用能力，构成了其强大的规划能力基础。同时，Gemini 模型多模态能力同样引人注目，能够理解和生成多种信息形式，实现了与其他先进模型不同的突破，构成其感知能力的基础。此外，Gemini 2.0 在长上下文理解方面的优势、具备海量上下文窗口，以及原生调用 Google 搜索、执行代码和使用第三方功能的能力，进一步提升了其记忆能力和工具使用能力。

在模型应用方面，谷歌正将 Gemini 2.0 技术与其“智能体全家桶”相结合。基于 Gemini 2.0 架构之上，谷歌推出了三个新的 AI 智能体研究原型：Project Astra、Project Mariner 和 Jules，极大地增强用户与数字世界的交互体验感。

除了上述主要应用场景之外，谷歌还在游戏行业内进行了富有创意的探索。谷歌公司正与 Supercell 等顶尖游戏开发企业携手合作，研究人工智能代理在各类游戏中的运用，涵盖

了从《部落冲突》这类的策略性游戏到《卡通农场》这类的模拟经营类游戏。

Gemini 2.0 Flash 的正式版本预计将于 2025 年 1 月份推出。目前，谷歌正积极拓展 Gemini 2.0 应用范畴。

■ AI 融资动向：面壁智能/hohem 浩瀚获融资额前二，分别获数亿元/超亿元人民币融资

本期 AI 初创公司的融资中，面壁智能/hohem 浩瀚获融资额前二，分别获数亿元/超亿元人民币融资。今年上半年面壁智能发布的端侧 AI 模型面壁小钢炮 MiniCPM 凭借其以小博大、高效低成本的性能优势，接连带来端侧 GPT-4V 和 ChatGPT 时刻，并将无限长文本、超清 OCR 识图、实时视频理解等功能首次集成到端侧。面壁小钢炮 MiniCPM 系列自发布以来，累计下载量突破 300 万，成为端侧智能的标杆。更重要的是，端侧 AI 可以快速切入商业化。公司一边在主流消费电子硬件和新兴硬件融入端侧 AI，一边布局端云协同的未来范式，在端侧模型、算力、内存、模型算法和硬件调度等方面持续突破。当前，面壁智能正在与华为、联发科技、联想、英特尔、长城汽车、易来智能等企业紧密协作，业务覆盖 AI Phone、AI PC、智能座舱、智能家居与具身机器人等领域。

作为 AI 跟拍稳定器的开创者，hohem 将智能科技融入影像拍摄，旨在为用户提供更便捷、高效的拍摄体验。过去，手持稳定器多被视为专业影视制作工具，而 hohem 浩瀚通过创新，将这一专业设备推向更广泛的消费群体。随着智能手机摄影功能的不断提升，消费者对高质量视频内容的需求与日俱增。除了追求拍摄稳定性，hohem 浩瀚还洞察到用户在直播、短视频及旅行等场景下，对随行摄影师的需求，希望解放双手，记录生活中的美好瞬间。因此，hohem 浩瀚致力于开发能够满足这些需求的智能影像产品。公司不断投入技术研发，2021 年发布全球首款搭载端侧 AI 追踪模块的手机稳定器，无需蓝牙连接即可在任意 App 中实现人像追踪。今年 10 月，hohem 浩瀚再次推出全球首款 AI 触控手机稳定器 M7，集 AI 画面预览、可视化调参及远程触控功能于一体。M7 稳定器采用分体式触屏控制器，搭载 hohem 浩瀚自主研发的端侧 AI 追踪模块，通过 AI 可视化预览的触控交互系统，用户只需框选目标即可实现快速识别和智能跟拍，极大地提升了创作的灵活性和效率。hohem 浩瀚还推出了 X 系列和 V 系列稳定器，面向普通摄影爱好者，以轻巧便携的设计和专业的防抖性能，降低了高质量视频创作的门槛。

■ 投资建议

字节“大力出奇迹”弯道超车，关注其旗下的 AI 布局。相对竞品，字节跳动的 AI 起步相对较晚：国内百度、阿里巴巴、商汤、科大讯飞在去年 3 月开始都陆续推出了大模型新品和 AI 应用，而字节跳动的“豆包”则是在同年 8 月才发布。伽

马数据显示，10月国内头部AI应用广告投放金额超4亿元，近半年翻了5倍；环比上升24.51%，连续6个月攀升。广告投放金额居前的5家公司中，Kimi与腾讯元宝金额分别环比增长172%和326%，即梦、豆包、猫箱则全部为字节旗下产品。在各家的投放渠道中，基本都离不开字节的巨量引擎，而背靠字节的豆包将流量池的优势发挥到了极致，“大力出奇迹”模式验证可行，字节旗下AI布局有望迎来全面爆发。

AI 终端即将迎来大规模爆发。一方面，AI 陪伴或为 AI toC 赛道最快突破的方向，相较于其他品类更侧重于功能性需求，AI 陪伴玩具更像是“陪聊机器人”，主打提供情绪价值，目前大厂纷纷涉足 AI 玩具赛道。另一方面，以 AI 眼镜为代表的功能性 AI 端侧或以提升渗透率为路径，替代传统日常生活用品。随着 Ray-Ban Meta 验证了 AI 眼镜技术路径的可行性，11 月以来，AI 眼镜领域持续升温，百度推出小度 AI 眼镜，Looktech、谷东科技发布新品，闪极科技完成数千万元 A 轮融资。与此同时，雷鸟创新宣布新款 AI 拍摄眼镜即将发布，闪极科技 12 月份也将发布新品。

我们坚定持续看好 AI 软硬件机会，建议关注以 AI 为核心的龙头厂商科大讯飞（002230.SZ）、有望迎来需求爆发的 AI 应用金桥信息（603918.SH）、高速通信连接器业务或显著受益于 GB200 放量的鼎通科技（688668.SH）、已与 Rokid 等多家知名 AI 眼镜厂商建立紧密合作的亿道信息（001314.SZ）等。

■ 风险提示

1) AI 底层技术迭代速度不及预期。2) 政策监管及版权风险。3) AI 应用落地效果不及预期。4) 推荐公司业绩不及预期风险。

重点关注公司及盈利预测

公司代码	名称	2024-12-15 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
001314.SZ	亿道信息	52.03	0.91	0.92	1.03	57.18	56.55	50.51	买入
002230.SZ	科大讯飞	50.97	0.28	0.40	0.56	182.04	127.43	91.02	买入
688256.SH	寒武纪-U	568.00	-2.04	-1.21	-0.50	-278.43	-469.42	-1136.00	买入
688668.SH	鼎通科技	38.49	0.67	1.04	1.41	57.45	37.01	27.30	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：算力租赁价格平稳，富士通 MONAKA 处理器计划于 2027 财年亮相	6
1.1、 数据跟踪：算力租赁价格平稳	6
1.2、 产业动态：富士通 Monaka 处理器计划于 2027 财年亮相	6
2、 AI 应用动态：KIMI 搜索访问量环比+4.17%，豆包稳居国内月活跃用户第二	8
2.1、 流量跟踪：Kimi 搜索访问量环比+4.17%	8
2.2、 产业动态：豆包稳居国内月活跃用户第二，谷歌 Gemini 2.0 引领智能体时代	8
3、 AI 融资动向：面壁智能/HOHEM 浩瀚获融资额前二，分别获数亿元/超亿元人民币融资	13
4、 行情复盘	15
5、 投资建议	17
6、 风险提示	17

图表目录

图表 1：本周算力租赁情况	6
图表 2：富士通 Monaka 处理器	7
图表 3：2024. 12. 9-2024. 12. 13 AI 相关网站流量	8
图表 4：AI 产品 MAU 国内总榜	9
图表 5：国内十款大模型产品今年以来（截至 11 月）投流金额	9
图表 6：近半年（截至 11 月）十款 AI 产品广告投放主要渠道及其占比	10
图表 7：Gemini 2.0 模型性能远超 Gemini 1.5	11
图表 8：Gemini 2.0 在游戏行业的运用	12
图表 9：本周 AI 初创公司的融资动态	14
图表 10：本周指数日涨跌幅	15
图表 11：本周 AI 算力指数内部涨跌幅度排名	15
图表 12：本周 AI 应用指数内部涨跌幅度排名	16
图表 13：重点关注公司及盈利预测	17

1、算力动态：算力租赁价格平稳，富士通 Monaka 处理器计划于 2027 财年亮相

1.1、数据跟踪：算力租赁价格平稳

本周恒源云 13 核+128G 价格环比上周持平。具体来看，显卡配置为 A100-40G 中，腾讯云 16 核+96G 价格为 28.64 元/时，阿里云 12 核+94GiB 价格为 31.58 元/时；显卡配置为 A100-80G 中，恒源云 13 核+128G 价格为 8.50 元/时；阿里云 16 核+125GiB 价格为 34.74 元/时；显卡配置为 A800-80G 中，恒源云 16+256G 价格为 9.00 元/时。

图表 1：本周算力租赁情况

显卡配置	CPU	内存	磁盘大小 (G)	平台名称	价格 (每小时)	价格环比上周
A100-40G	16	96	可自定, 额外收费	腾讯云	28.64/元	0.00%
	12 核	94G	可自定, 额外收费	阿里云	31.58/元	0.00%
A100-80G	13	128	系统盘: 20G 数据盘: 50GB	恒源云	8.50/元	0.00%
	16 核	125G	可自定, 额外收费	阿里云	34.74/元	0.00%
A800-80G	16	256	系统盘: 20G 数据盘: 50GB	恒源云	9.00/元	0.00%

资料来源：腾讯云，阿里云，恒源云，华鑫证券研究

1.2、产业动态：富士通 Monaka 处理器计划于 2027 财年亮相

本周富士通展示了用于数据中心的基于 Arm 的 144 核 Monaka 处理器的机械样品，该处理器采用博通推出的业内首个 3.5D XDSiP 技术平台，以达到高性能、低功耗和低成本目标。RIKEN 计算科学中心 (R-CCS) 主任、东京工业大学教授 Satoshi Matsuoka 公开了 Monaka 的更多技术细节：

1) 在核心架构方面，Monaka 采用基于台积电的 CoWoS-L 封装技术的博通 3.5D XDSiP 技术平台，拥有 36 个计算小芯片。其中主要的 CPU 计算核心基于 Armv9 指令集，拥有 144 个 CPU 内核，采用台积电 2nm 制程制造，并使用混合铜键合 (HCB) 以面对面 (F2F) 方式堆叠在 SRAM tiles 上。SRAM tiles 是基于台积电的 5nm 工艺制造的、计算和缓存堆栈伴随着生成的相对巨大的 I/O 芯片，该芯片集成了内存控制器、顶部带有 CXL 3.0 的 PCIe 6.0 通道以连接加速器和扩展器，以及公众期望从数据中心级 CPU 获得的其他接口。

2) 在内存接口方面，为应对广泛的数据中心工作处理需求，Monaka 不依赖于高带宽内存，而是预计使用主流的 DDR5 DRAM，有望在其 MR-DIMM 和 MCR-DIMM 实施过程中运用

并提供足够的容量、降低数据中心处理器的成本。

3) 在处理能力与安全性能方面，Monaka 处理器将使用基于 Armv9-A 指令集架构构建的内核，并集成 SVE2 指令（第二代可扩展矢量扩展指令）。但目前富士通尚未为该设计指定固定的矢量长度，范围为 128 至 2048 位。鉴于其 A64FX 处理器具备支持高达 512 位向量的卓越能力，Monaka 处理器有望支持类似乃至更高位数的向量。同时，Monaka 将整合高级安全功能，包括 Armv9-A 的机密计算架构（CCA），提供增强的工作负载隔离和强大的保护。

4) 在能效提升方面，富士通计划在 2026 年至 2027 年依赖空气冷却技术将处理器能效提高一倍。凭借其显著的能源效率优势，Monaka 有望成为 AMD 的 EPYC 与英特尔的 Xeon 处理器的市场有力竞争者。考虑到 Monaka 是基于 Arm 架构的处理器，预计相较于 x86 架构处理器其能源消耗将更加经济。

Monaka 处理器计划于 2027 财年推向市场，具体时间范围为 2026 年 4 月 1 日至 2027 年 3 月 31 日。

图表 2：富士通 Monaka 处理器



资料来源：芯智讯，华鑫证券研究

2、AI 应用动态：Kimi 搜索访问量环比 +4.17%，豆包稳居国内月活跃用户第二

2.1、流量跟踪：Kimi 搜索访问量环比+4.17%

本期（2024.12.9-2024.12.13）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（907.3M）、Bing（413.2M）和 Canva（180.5M），访问量环比增速第一为 Kimi（4.17%）；平均停留时长前三位分别为 Character.AI（00:13:13）、Discord（00:12:12）和 DeepL（00:08:50）；平均停留时长环比增速第一为 Character.AI（0.89%）。

图表 3：2024.12.9-2024.12.13 AI 相关网站流量

应用	应用类型	归属公司	周平均访问量 (M)	访问量环比	平均停留时长	时长环比
ChatGPT	聊天机器人	OpenAI	907.3	1.84%	6:18	0.80%
Bing	搜索	微软	413.2	3.66%	6:42	-0.74%
Discord	游戏社区	微软	127.4	-0.55%	12:12	-0.41%
Canva	在线设计	Canva	180.5	-3.94%	7:50	-0.42%
Github	代码托管	微软	113.9	-0.35%	6:29	-0.51%
Gemini	聊天机器人	谷歌	62.86	1.09%	4:34	-0.72%
Character.AI	聊天机器人	Character.AI	43.26	-0.39%	13:13	0.89%
NotionAI	文本/笔记	Notion	33.97	-0.64%	8:43	0.00%
QuillBot	释义工具	QuillBot	17.09	-4.15%	4:06	0.00%
Kimi	聊天机器人	Moonshot AI	8.618	4.17%	3:30	0.48%
DeepL	翻译工具	DeepL	47.16	-1.46%	8:50	-0.38%
文心一言	聊天机器人	百度	4.683	-5.47%	2:56	-0.56%
Perplexity	AI 搜索	Perplexity	23.22	-1.53%	5:58	-0.28%

资料来源：similarweb, 华鑫证券研究

2.2、产业动态：豆包稳居国内月活跃用户第二，谷歌 Gemini 2.0 引领智能体时代

豆包稳居国内月活跃用户第二。在国内 AI 模型领域，2024 年第四季度字节跳动旗下

AI 应用产品豆包表现亮眼。豆包作为字节跳动旗下的热门 AI 智能助手产品，具备聊天机器人、写作助手等多种功能，其多功能性和便捷性深受用户青睐。当前，豆包产品以显著的优势领先于其他竞争者，成为最受市场欢迎的 AI 应用产品之一。在 10 月的国内月活跃人数排行榜上，豆包 APP 的 MAU（月活跃人数）达到 4717.47 万人次，远超前排名第二 Kimi 的 2028.65 万人次。字节跳动在 AI 应用领域的迅速崛起可以归功于以下因素：

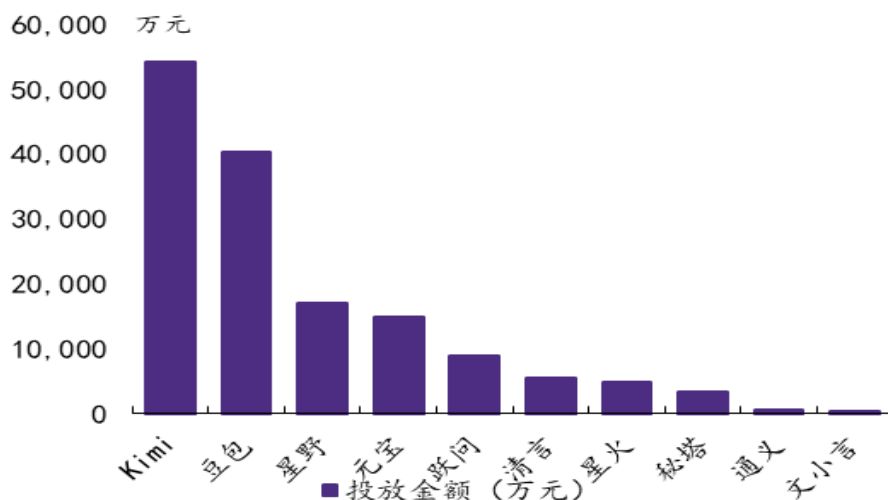
图表 4：AI 产品 MAU 国内总榜



资料来源：Xsignal AI&APP 全息数据洞察，华鑫证券研究

1) 流量投放战略谋求“弯道超车”：相较于竞争对手，字节跳动在人工智能领域的起步时间相对延后。自 2023 年 3 月开始，国内的百度、阿里巴巴、商汤科技以及科大讯飞等企业已相继推出了各自的大模型新品及人工智能应用。而字节跳动的 AI 产品“豆包”直至同年 8 月方才问世，在起步时间上稍显滞后。为了迅速缩小与行业的差距，字节跳动采用巨额投入的“烧钱”战略。截至 11 月，旗下豆包产品的投放金额已达 40363 万元，仅次于 Kimi 的投放金额 54214 万元，并且巨幅领先投放金额排名第三的星野。

图表 5：国内十款大模型产品今年以来（截至 11 月）投流金额



资料来源：硅谷AI见闻，华鑫证券研究

2) 巨量引擎的流量资源支持：对于多数 AI 应用，投放流量是前期用户增长的关键。在各厂商所采用的投放渠道之中，字节跳动旗下的巨量引擎占据着极为关键的地位。该平台汇聚了今日头条、抖音、西瓜视频等丰富的营销资源，成为众多厂商开展广告投放活动的重要依托。而豆包凭借与字节跳动的紧密关联，得以充分挖掘并最大化利用这一庞大流量池所蕴含的优势，在市场推广与用户触达方面展现出卓越成效。尽管豆包投放相对 Kimi 较少，但得益于字节跳动独有的投放经验和流量优势，豆包能够实现更加高效的获客，实现 10 月 MAU（月活跃人数）4717.47 万人次，远超前排名第二 Kimi 的 2028.65 万人次。

图表 6：近半年（截至 11 月）十款 AI 产品广告投放主要渠道及其占比

产品	第一大渠道 (占比)	第二大渠道 (占比)
Kimi	巨量 (77%)	腾讯 (20%)
豆包	巨量 (89%)	腾讯 (10%)
星野	巨量 (42%)	小米 / 腾讯 (23%)
元宝	腾讯 (98%)	巨量 (1%)
跃问	巨量 (55%)	腾讯 (42%)
清言	巨量 (62%)	小米 (29%)
星火	小米 (54%)	快手 (32%)
秘塔	腾讯 (99%)	快手 (1%)
通义	腾讯 (52%)	巨量 (46%)
文小言	腾讯 (47%)	小米 (35%)

资料来源：硅谷AI见闻，华鑫证券研究

当前，国内外 C 端 AI 产品多以聊天机器人为主，面临网络效应局限和付费订阅变现的难题，AI 对话产品可能仅为过渡形态，长期发展需要注重用户视觉化体验和较低的用户门槛。字节跳动预计将提升即梦产品优先级，探索打造 AI 时代的“抖音”。

谷歌 Gemini 2.0 发布，号称“为智能体时代设计”。在国外 AI 模型领域，12 月 12 日谷歌发布的 Gemini 2.0 Flash 实验版本，标志着 Gemini 2.0 系列的正式启航。该模型具

备卓越的反应速度（低延迟）和强大的性能，充分体现了谷歌 Gemini 技术的尖端水平。相较于前一代产品，Gemini 2.0 Flash 在速度上实现了翻倍，支持图像、视频和音频等多模态输入。此外，2.0 Flash 版本新增了多模态输出功能，能够原生生成图像与文本的混合内容，并提供可控制的多语言文本转语音（TTS）音频。此次更新不仅是对语言模型的升级，而是构建了一个满足智能体基本需求的统一底层模型。谷歌 CEO 桑达尔·皮查伊在公开信中提到，Gemini 1.0 旨在整理和理解信息，而 Gemini 2.0 则致力于使这些信息变得切实可用。

图表 7: Gemini 2.0 模型性能远超 Gemini 1.5

CAPABILITY	BENCHMARK	DESCRIPTION	Gemini 1.5 Flash 002	Gemini 1.5 Pro 002	Gemini 2.0 Flash Experimental
General	MMLU-Pro	Enhanced version of popular MMLU dataset with questions across multiple subjects with higher difficulty tasks	67.3%	75.8%	76.4%
Code	Natural2Code	Code generation across Python, Java, C++, JS, Go . Held out dataset HumanEval-like, not leaked on the web	79.8%	85.4%	92.9%
	Bird-SQL (Dev)	Benchmark evaluating converting natural language questions into executable SQL	45.6%	54.4%	56.9%
	LiveCodeBench (Code Generation)	Code generation in Python. Code Generation subset covering more recent examples: 06/01/2024 - 10/05/2024	30.0%	34.3%	35.1%
Factuality	FACTS Grounding	Ability to provide factuality correct responses given documents and diverse user requests. Held out internal dataset	82.9%	80.0%	83.6%
Math	MATH	Challenging math problems (incl. algebra, geometry, pre-calculus, and others)	77.9%	86.5%	89.7%
	HiddenMath	Competition-level math problems. Held out dataset AIME/AMC-like, crafted by experts and not leaked on the web	47.2%	52.0%	63.0%
Reasoning	GPQA (diamond)	Challenging dataset of questions written by domain experts in biology, physics, and chemistry	51.0%	59.1%	62.1%
Long context	MRCR (1M)	Novel, diagnostic long-context understanding evaluation	71.9%	82.6%	69.2%
Image	MMMU	Multi-discipline college-level multimodal understanding and reasoning problems	62.3%	65.9%	70.7%
	Vibe-Eval (Reka)	Visual understanding in chat models with challenging everyday examples. Evaluated with a Gemini Flash model as a rater	48.9%	53.9%	56.3%
Audio	CoVoST2 (21 lang)	Automatic speech translation (BLEU score)	37.4	40.1	39.2
Video	EgoSchema (test)	Video analysis across multiple domains	66.8%	71.2%	71.5%

资料来源：腾讯科技，华鑫证券研究

在模型具体性能提升方面，Gemini 2.0 Flash 模型是 Gemini 系列中除专为端侧设计的 Nano 模型外参数量最小的模型。根据谷歌 DeepMind CTO 格雷·卡武克奥卢的说法，Flash 模型相较于一年前的任何模型在性能上都有显著提升，同时成本大幅降低。Gemini 2.0 通过性能提升，增强了理解复杂指令、长期规划和组合函数调用能力，构成了其强大的规划能力基础。同时，Gemini 模型多模态能力同样引人注目，能够理解和生成多种信息形式，实现了与其他先进模型不同的突破，构成其感知能力的基础。此外，Gemini 2.0 在长上下文理解方面的优势、具备海量上下文窗口，以及原生调用 Google 搜索、执行代码和使用第三方功能的能力，进一步提升了其记忆能力和工具使用能力。综合来看，Gemini 2.0 在规划能力、感知能力、记忆能力与工具使用能力方面均有显著增强，将成为智能体时代的引领。

在模型应用方面，谷歌正将 Gemini 2.0 技术与其“智能体全家桶”相结合。基于 Gemini 2.0 架构之上，谷歌推出了三个新的 AI 智能体研究原型：Project Astra、Project Mariner 和 Jules，增大地增强用户与数字世界的交互体验感。其中，Project Astra 为系统级智能体，是于今年 5 月 I/O 大会上展示的 AI 助手，刚发布时尚不成熟，但搭载

Gemini 2.0 后性能获得了显著提升，成为当下表现亮眼的系统级智能体。Project Mariner 则是插件版智能体，作为一个实验性的 Chrome 扩展，它能够理解和分析浏览器屏幕上的所有信息，包括像素、文本、代码、图像和表单。此外，针对开发者群体，谷歌推出了实验性的 AI 智能体 Jules，同 Devin、Cursor 0.43 一样，其作为编码助手能够制定详尽的多步骤计划来解决问题，高效修改多个文件，甚至能够拉取请求、将修复直接提交回 GitHub。

除了上述主要应用场景之外，谷歌还在游戏行业内进行了富有创意的探索。谷歌公司正与 Supercell 等顶尖游戏开发企业携手合作，研究人工智能代理在各类游戏中的运用，涵盖了从《部落冲突》这类的策略性游戏到《卡通农场》这类的模拟经营类游戏。

图表 8: Gemini 2.0 在游戏行业的运用



资料来源：腾讯科技，华鑫证券研究

Gemini 2.0 Flash 的正式版本预计将于 2025 年 1 月份推出。目前，谷歌正积极拓展 Gemini 2.0 应用范畴，不仅致力于在虚拟领域探索智能体的能力边界，同时也积极尝试将 Gemini2.0 的空间推理能力应用于机器人技术，以期助力智能体在现实世界中发挥更大的作用，然而目前相关研发尚处于初始阶段。

3、AI 融资动向：面壁智能/hohem 浩瀚获融资额前二，分别获数亿元/超亿元人民币融资

本期 AI 初创公司的融资中，面壁智能/hohem 浩瀚获融资额前二，分别获数亿元/超亿元人民币融资。今年上半年面壁智能发布的端侧 AI 模型面壁小钢炮 MiniCPM 凭借其以小博大、高效低成本的性能优势，接连带来端侧 GPT-4V 和 ChatGPT 时刻，并将无限长文本、超清 OCR 识图、实时视频理解等功能首次集成到端侧。“端侧 AI”是指 AI 可以直接在移动设备上处理数据，无需连接到服务器或云端，能够在用户设备本地完成复杂的推理和决策。面壁小钢炮 MiniCPM 系列自发布以来，累计下载量突破 300 万，成为端侧智能的标杆。更重要的是，端侧 AI 可以快速切入商业化。公司一边在主流消费电子硬件和新兴硬件融入端侧 AI，一边布局端云协同的未来范式，在端侧模型、算力、内存、模型算法和硬件调度等方面持续突破。当前，面壁智能正在与华为、联发科技、联想、英特尔、长城汽车、易来智能等企业紧密协作，业务覆盖 AI Phone、AI PC、智能座舱、智能家居与具身机器人等领域。

成立于 2014 年的 hohem 浩瀚总部位于深圳，深耕 AI 智能影像设备市场。作为 AI 跟拍稳定器的开创者，hohem 浩瀚一直秉持用户至上的理念，将智能科技融入影像拍摄，旨在为用户提供更便捷、高效的拍摄体验。过去，手持稳定器多被视为专业影视制作工具，而 hohem 浩瀚通过创新，将这一专业设备推向更广泛的消费群体。随着智能手机摄影功能的不断提升，消费者对高质量视频内容的需求与日俱增。除了追求拍摄稳定性，hohem 浩瀚还洞察到用户在直播、短视频及旅行等场景下，对随行摄影师的需求，希望解放双手，记录生活中的美好瞬间。因此，hohem 浩瀚致力于开发能够满足这些需求的智能影像产品。公司不断投入技术研发，2021 年发布全球首款搭载端侧 AI 追踪模块的手机稳定器，无需蓝牙连接即可在任意 App 中实现人像追踪。今年 10 月，hohem 浩瀚再次推出全球首款 AI 触控手机稳定器 M7，集 AI 画面预览、可视化调参及远程触控功能于一体。M7 稳定器采用分体式触屏控制器，搭载 hohem 浩瀚自主研发的端侧 AI 追踪模块，通过 AI 可视化预览的触控交互系统，用户只需框选目标即可实现快速识别和智能跟拍，极大地提升了创作的灵活性和效率。hohem 浩瀚还推出了 X 系列和 V 系列稳定器，面向普通摄影爱好者，以轻巧便携的设计和专业的防抖性能，降低了高质量视频创作的门槛。

图表 9: 本周 AI 初创公司的融资动态

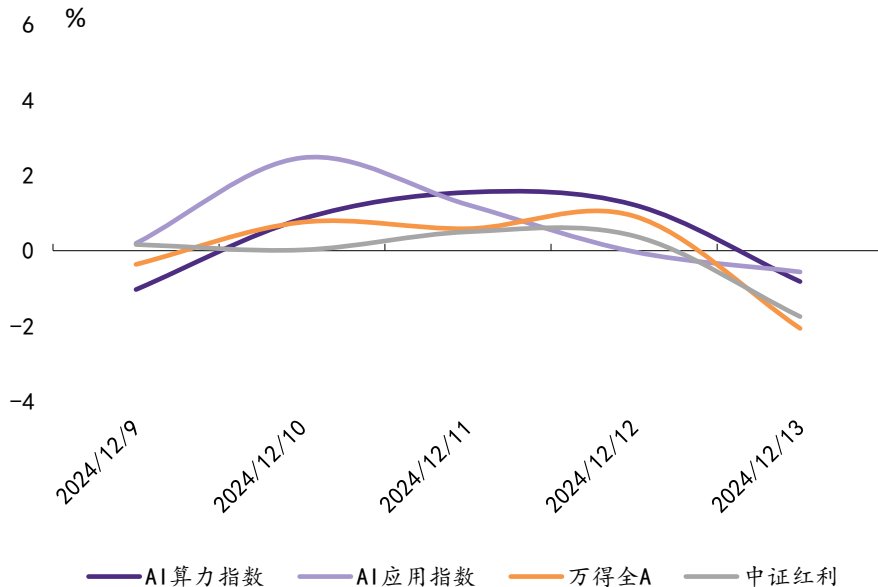
应用	应用类型	领投方	融资轮	融资额	目前累计融资额	目前估值
数聚链	AI 中台	汇孚资本、道生资本	A 轮	数千万	——	——
hohem 浩瀚	AI 智能影像设备	顺为资本、君联资本	B 轮	超亿元	——	——
面壁智能	AI 大模型	中关村科学城基金、赛富投资	——	数亿元	——	——
北斗智影	AI 内容变现平台	南京文投	——	近亿元	——	——

资料来源: MSN, 新浪新闻, 腾讯网, 36Kr, 华鑫证券研究

4、行情复盘

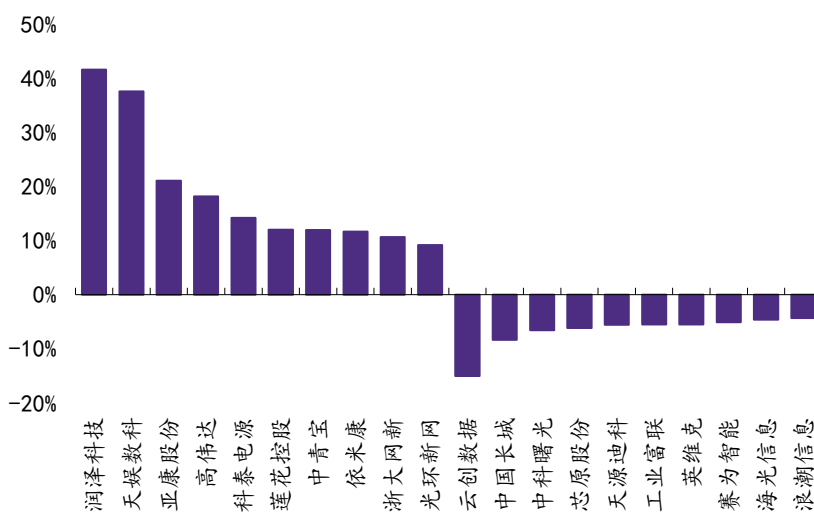
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利日涨幅最大值分别为 1.55%/2.47%/0.92%/0.50%，日跌幅最大值分别为-1.03%/-0.56%/-2.06%/-1.75%。AI 算力指数内部，润泽科技以+41.54%录得本周最大涨幅，云创数据以-15.03%录得本周最大跌幅。AI 应用指数内部，乐鑫科技以+50.20%得本周最大涨幅，易点天下以-11.39%录得本周最大跌幅。

图表 10：本周指数日涨跌幅



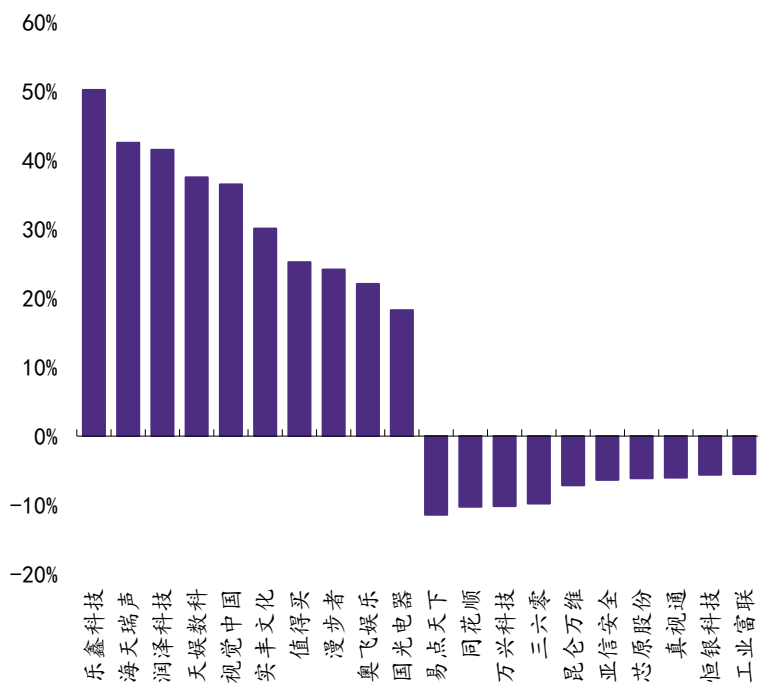
资料来源：wind, 华鑫证券研究

图表 11：本周 AI 算力指数内部涨跌幅度排名



资料来源：wind, 华鑫证券研究

图表 12：本周 AI 应用指数内部涨跌幅度排名



资料来源：wind, 华鑫证券研究

5、投资建议

字节“大力出奇迹”弯道超车，关注其旗下的 AI 布局。相对竞品，字节跳动的 AI 起步相对较晚：国内百度、阿里巴巴、商汤、科大讯飞在去年 3 月开始都陆续推出了大模型新品和 AI 应用，而字节跳动的“豆包”则是在同年 8 月才发布。字节在 2023 年通过 AI 布局、内部竞争、迅速补充短板、全力进攻和投入大量资金，一年完成弯道超车，站到了中国乃至全球的领先地位。伽马数据显示，10 月国内头部 AI 应用广告投放金额超 4 亿元，近半年翻了 5 倍；环比上升 24.51%，连续 6 个月攀升。广告投放金额居前的 5 家公司中，Kimi 与腾讯元宝金额分别环比增长 172%和 326%，即梦、豆包、猫箱则全部为字节旗下产品。在各家的投放渠道中，基本都离不开字节的巨量引擎，而背靠字节的豆包将流量池的优势发挥到了极致，“大力出奇迹”模式验证可行，字节旗下 AI 布局有望迎来全面爆发。

AI 终端即将迎来大规模爆发。一方面，AI 陪伴或为 AI toC 赛道最快突破的方向，相较于其他品类更注重于功能性需求，AI 陪伴玩具更像是“陪聊机器人”，主打提供情绪价值，目前大厂纷纷涉足 AI 玩具赛道。另一方面，以 AI 眼镜为代表的功能性 AI 端侧或以提升渗透率为路径，替代传统日常生活用品。随着 Ray-Ban Meta 验证了 AI 眼镜技术路径的可行性，11 月以来，AI 眼镜领域持续升温，百度推出小度 AI 眼镜，Looktech、谷东科技发布新品，闪极科技完成数千万元 A 轮融资。与此同时，雷鸟创新宣布新款 AI 拍摄眼镜即将发布，闪极科技 12 月份也将发布新品。

我们坚定持续看好 AI 软硬件机会，建议关注以 AI 为核心的龙头厂商科大讯飞（002230.SZ）、芯片技术有望创新突破的寒武纪（688256.SH）、高速通信连接器业务或显著受益于 GB200 放量的鼎通科技（688668.SH）、已与 Rokid 等多家知名 AI 眼镜厂商建立紧密合作的亿道信息（001314.SZ）等。

图表 13：重点关注公司及盈利预测

公司代码	名称	2024-12-15 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
001314.SZ	亿道信息	52.03	0.91	0.92	1.03	57.18	56.55	50.51	买入
002230.SZ	科大讯飞	50.97	0.28	0.40	0.56	182.04	127.43	91.02	买入
688256.SH	寒武纪-U	568.00	-2.04	-1.21	-0.50	-278.43	-469.42	-1136.00	买入
688668.SH	鼎通科技	38.49	0.67	1.04	1.41	57.45	37.01	27.30	买入

资料来源：wind, 华鑫证券研究

6、风险提示

1) AI 底层技术迭代速度不及预期。2) 政策监管及版权风险。3) AI 应用落地效

果不及预期。4) 推荐公司业绩不及预期风险。

■ 计算机&中小盘组介绍

宝幼琛：本硕毕业于上海交通大学，多次新财富、水晶球最佳分析师团队成员，7年证券从业经验，2021年11月加盟华鑫证券研究所，目前主要负责计算机与中小盘行业上市公司研究。擅长领域包括：云计算、网络安全、人工智能、区块链等。

任春阳：华东师范大学经济学硕士，6年证券行业经验，2021年11月加盟华鑫证券研究所，从事计算机与中小盘行业上市公司研究

周文龙：澳大利亚莫纳什大学金融硕士

陶欣怡：毕业于上海交通大学，于2023年10月加入团队。

■ 证券分析师承诺

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并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，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收到任何形式的补偿。

■ 证券投资评级说明

股票投资评级说明：

	投资建议	预测个股相对同期证券市场代表性指数涨幅
1	买入	> 20%
2	增持	10% — 20%
3	中性	-10% — 10%
4	卖出	< -10%

行业投资评级说明：

	投资建议	行业指数相对同期证券市场代表性指数涨幅
1	推荐	> 10%
2	中性	-10% — 10%
3	回避	< -10%

以报告日后的12个月内，预测个股或行业指数相对于相关证券市场主要指数的涨跌幅为标准。

相关证券市场代表性指数说明：A股市场以沪深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；美国市场以道琼斯指数为基准。

■ 免责声明

华鑫证券有限责任公司（以下简称“华鑫证券”）具有中国证监会核准的证券投资咨询业务资格。本报告由华鑫证券制作，仅供华鑫证券的客户使用。本公司不会因接收人收到本报告而视其为客户。

本报告中的信息均来源于公开资料，华鑫证券研究部门及相关研究人员力求准确可靠，但对这些信息的准确性及完整性不作任何保证。我们已力求报告内容客观、公正，但报告中的信息与所表达的观点不构成所述证券买卖的出价或询价的依据，该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。投资者应当对本报告中的信息和意见进行独立评估，并应同时结合各自的投资目的、财务状况和特定需求，必要时就财务、法律、商业、税收等方面咨询专业顾问的意见。对依据或者使用本报告所造成的一切后果，华鑫证券及/或其关联人员均不承担任何法律责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露。

本报告中的资料、意见、预测均只反映报告初次发布时的判断，可能会随时调整。该等意见、评估及预测无需通知即可随时更改。在不同时期，华鑫证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。华鑫证券没有将此意见及建议向报告所有接收者进行更新的义务。

本报告版权仅为华鑫证券所有，未经华鑫证券书面授权，任何机构和个人不得以任何形式刊载、翻版、复制、发布、转发或引用本报告的任何部分。若华鑫证券以外的机构向其客户发放本报告，则由该机构独自为此发送行为负责，华鑫证券对此等行为不承担任何责任。本报告同时不构成华鑫证券向发送本报告的机构之客户提供的投资建议。如未经华鑫证券授权，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。华鑫证券将保留随时追究其法律责任的权利。请投资者慎重使用未经授权刊载或者转发的华鑫证券研究报告。