

2024年12月22日



华鑫证券
CHINA FORTUNE SECURITIES

豆包大模型家族迎来全面升级，视觉理解模型迈入“厘时代”

—计算机行业周报

推荐(维持)

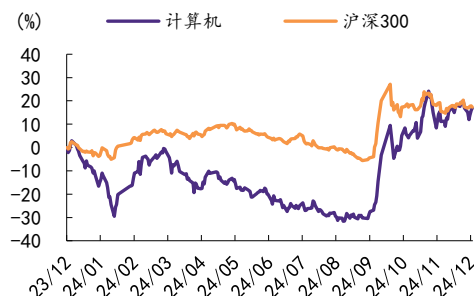
投资要点

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)	5.1	61.1	16.7
沪深300	1.6	22.7	17.7

市场表现



资料来源：Wind，华鑫证券研究

相关研究

- 《计算机行业周报：豆包稳居MAU国内第一，谷歌 Gemini2.0 聚焦智能体》2024-12-17
- 《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

算力：恒源云 13 核+128G 算力紧张，英伟达发布 AI 超级计算机 Jetson Orin Nano Super

本周恒源云 13 核+128G 算力紧张。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 算力紧张；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。

近日，英伟达推出了一款全新的尺寸小巧的生成式 AI 超级计算机 NVIDIA Jetson Orin Nano Super，该开发者套件只有手掌大小，能够为商业 AI 开发者、科技爱好者和学生等各类人群提供更强大的生成式 AI 功能与性能，目前建议零售价 2070 人民币。Jetson Orin Nano Super 凭借其卓越性能，无论是创建基于检索增强生成的 LLM 聊天机器人、构建视觉 AI 智能体，还是部署基于 AI 的机器人，均能得以有效实现。Jetson Orin Nano Super 亮点如下：

具备颠覆性的生成式 AI 模型性能。NVIDIA 最新软件更新将 Jetson Orin Nano 的 AI 性能从 40 TOPS 提升至 67 TOPS，并将内存带宽从 68 GB/s 提升至 102 GB/s，使开发者能够更高效地运行各种生成式 AI 模型。与上一代产品相比，NVIDIA Jetson Orin Nano Super 的生成式 AI 推理性能提升至 1.7 倍，性能提升 70%，达到 67 INT8 TOPS，内存带宽提高 50%，达到 102GB/s。NVIDIA Jetson Orin Nano Super 在各种热门大语言模型 (LLM)、视觉语言模型 (VLM) 和视觉 Transformer 上的性能相比前代产品均有显著提升。

助力解锁更多 AI 场景应用。NVIDIA Jetson Orin Nano Super 可提供出色的 AI 计算能力、快速的内存带宽和全面的软件堆栈，为新的生成式 AI 应用提供卓越的能效，它支持任何基于 Transformer 的生成式 AI 模型的快速推理。在机器人领域，从开发到仿真再到部署，借助 AI 加速机器人开发，实现移动、抓取和视觉等关键功能，实现为各行各业构建机器人。在视觉领域，其搭载的 NVIDIA Metropolis 集成了视觉数据和 AI，可以处理适用于顺畅零售、库存管理、智慧城市交通、工厂检查和医疗健康等应用的传感器数据。

目前，NVIDIA Jetson Orin 包含 Jetson AGX Orin、Jetson Orin NX 和 Jetson Orin Nano 三款系列产品，可为开发者提供不同性能和价位的加速计算能力，从而满足多种自主应用的需求。

■ AI 应用：豆包大模型家族全方位优化升级

2024 年 12 月 18 日火山引擎 FORCE 原动力大会上，火山引擎展示了其全新升级的豆包大模型家族，包含豆包视觉理解模型、豆包 3D 生成模型，以及全面更新的豆包通用模型 pro、音乐模型、文生图模型等，为企业提供多模态大模型能力。豆包大模型于 12 月实现日均 tokens 使用量超过 4 万亿，较 5 月发布时期增长超过 33 倍，稳居行业领先地位。

视觉理解模型重磅发布。豆包视觉理解模型具备以下三种能力：更强的内容识别能力，可识别出图像中的物体类别、形状等基本要素以及理解物体之间的关系、空间布局以及场景的整体含义；更强的理解和推理能力，能根据所识别的文字和图像信息进行复杂的逻辑计算；更细腻的视觉描述能力，可以基于图像信息，更细腻地描述图像呈现的内容以及多种文体创作。以上能力极大拓展了大模型的场景边界，在金融、医疗、建筑、地理、体育、物流等诸多行业有着广阔应用前景。除此之外，视觉理解的输入价格为每千 tokens 0.003 元，比行业平均价格降低 85%，相当于 1 元钱可以处理 284 张 720P 的图片，具备高性价比优势，视觉理解模型正式走进厘时代。同时火山引擎还将提供更高的初始流量，RPM 达到 15000 次，TPM 达到 120 万，让企业和开发者用好视觉理解模型，找到更多创新场景。

模型家族迎来全面升级。火山引擎除了推出视觉理解模型之外，还发布、升级了多个其他模型，大模型家族成员更丰富，大模型能力再提升。大会发布豆包 3D 生成模型，其与火山引擎数字孪生平台 veOmniverse 结合使用，可以高效完成智能训练、数据合成和数字资产制作，成为一套支持 AIGC 创作的物理世界仿真模拟器。此外，豆包通用模型 pro 全面升级，对齐 GPT-4o，使用价格仅为后者的 1/8；音乐模型从生成 60 秒的简单结构，升级到生成 3 分钟的完整作品；文生图模型 2.1 版本在业界首次实现精准生成汉字和一句话 P 图的产品化能力，该模型已接入即梦 AI 和豆包 App。

AI 应用开发平台全新升级。火山引擎升级了火山方舟、扣子和 HiAgent 三款平台产品，帮助企业构建好自身的 AI 能力中心，高效开发 AI 应用。其中，火山方舟发布了大模型记忆方案，并推出 Prefix Cache 和 Session Cache API，降低延迟和成本。火山方舟还带来全域 AI 搜索，具备场景化搜索推荐一体化、企业私域信息整合等服务。

技术架构面向 AI 全面转型。云原生是过去十年最重要的计算范式，而火山引擎认为，未来十年计算范式应从云原生转变

为 AI 云原生。基于 AI 云原生的理念，火山引擎推出了新一代计算、网络、存储和和安全产品。在计算方面，火山引擎 GPU 实例，通过 vRDMA 网络，支持大规模并行计算和 P/D 分离推理架构，显著提升训练和推理效率，降低成本；在存储方面，新推出的 EIC 弹性极速缓存，能够实现 GPU 直连，使大模型推理时延降低至 1/50、成本降低 20%；在安全方面，火山将推出 PCC 私密云服务，构建大模型的可信应用体系。基于 PCC，企业能够实现用户数据在云上推理的端到端加密，性能卓越。

火山引擎持续致力于增强模型能力、完善模型服务，促进 AI 在各行各业的深度融合与创新应用，打造更强性能、更低价格、更易落地的豆包大模型，开启一个更加智能、高效、便捷的 AI 时代。

■ AI 融资动向:Databricks/智谱 AI 获本周 AI 融资市场估值前二，分获 100 亿美元/30 亿人民币融资

本周，Databricks/智谱 AI 获本周 AI 融资市场估值前二，分别获 100 亿美元/30 亿元人民币融资。Databricks 最初是美国加州大学伯克利 AMP 实验室的 Spark 大数据处理系统商业化项目，可以帮助企业以极快的速度分析其内部大数据，也凭此在获得了在硅谷立足之地。到 2020 年末，Databricks 推出了其数据仓库产品——Databricks SQL，凭借着对 AI 数据的理解，Databricks 先后打造了 Apache Spark、Delta Lake、MLflow 和 OneLakehousePlatform，开设了多条产品线。目前，Databricks 主要靠开发软件来提取、分析和构建人工智能应用程序，这些应用程序使用来自各种来源的复杂数据。商业模式方面，Databricks 根据客户每秒消耗的计算资源量收费，并打造了自家独有的 DBU 作为衡量单位。从官网客户资源库可以看到，Databricks 拥有移动通信公司 AT&T、电子商品公司惠普、语言 AI 工具 Grammarly 等超过 500 多家客户群体。

智谱 AI 或受益于清华系的科研底色，一直保持着较快的迭代速度。2020 年底，智谱 AI 研发 GLM 预训练架构，2021 年训练完成百亿参数模型 GLM-10B，同年利用 MoE 架构成功训练出收敛的万亿稀疏模型。2023 年，智谱 AI 推出了千亿基座的对话模型 ChatGLM，并开源了单卡版模型 ChatGLM-6B，使得研究者和个人开发者可以进行微调和部署。与此同时，智谱 AI 开放平台的价格从 2023 年年初到 2024 年下半年降低了 5000 倍。2024 年以来，智谱 AI 更是先后发布对标 Sora

的视频生成模型 CogVideoX，以及对标 GPT-4o 的端到端语音模型 GLM-4-Voice 和 GLM-4-VideoCall。

投资建议

大模型或从快速扩张迈入淘汰赛阶段，字节明年有望迎来爆款应用。今年以来，视觉理解模型在 GPT-4o、Google 的 Project Astra，以及 iPhone 16 和一众旗舰机型上都加速落地。豆包视觉理解模型不仅能精准识别视觉内容，还可根据图像信息进行复杂的逻辑计算，完成分析图表、处理代码、解答学科问题等任务。豆包视觉理解模型千 tokens 输入价格仅为 3 厘，大模型的价格体系实现了从分至厘的跨越性调整，以更低成本让多模态交互降低用户的使用门槛，辅助完成一系列复杂的任务。人类接收的信息超过 80% 来自视觉，视觉理解将极大地拓展大模型的能力边界，同时也会降低人们与大模型交互的门槛，为大模型解锁更丰富的应用场景。同时，随着爆款应用的诞生，算力需求预计大幅增加。

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。

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	2024-12-22 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
001314.SZ	亿道信息	52.71	0.91	0.92	1.03	57.92	57.29	51.17	买入
002230.SZ	科大讯飞	51.75	0.28	0.40	0.56	184.82	129.38	92.41	买入
688256.SH	寒武纪-U	675.95	-2.04	-1.21	-0.50	-331.35	-558.64	-1351.90	买入
688668.SH	鼎通科技	40.74	0.67	1.04	1.41	60.81	39.17	28.89	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：恒源云 13 核+128G 算力紧张，英伟达发布 AI 超级计算机 JETSON ORIN NANO SUPER	6
1.1、 数据跟踪：恒源云 13 核+128G 算力紧张	6
1.2、 产业动态：Jetson Orin Nano Super 蕴含超强 AI 性能，解锁多元场景应用新篇	6
2、 AI 应用动态：CHARACTER.AI 搜索访问量环比+7.07%， 豆包大模型家族全方位优化	9
2.1、 流量跟踪：Character.AI 搜索访问量环比+7.07%	9
2.2、 产业动态：豆包大模型家族全方位优化	9
3、 AI 融资动向：DATABRICKS/智谱 AI 获本周 AI 融资市场估值前二，分获 100 亿美元/30 亿人民币融资	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：Jetson Orin Nano Super 开发者套件	7
图表 3：Jetson Orin Nano Super 在性能上的提升	7
图表 4：2024.12.16-2024.12.20 AI 相关网站流量	9
图表 5：2024 年豆包大模型日均 tokens 使用量提升情况	10
图表 6：豆包视觉理解模型理解和推理能力展示	10
图表 7：豆包通用模型 pro 全面升级	11
图表 8：火山方舟发布 Prefix Cache 和 Session Cache API	111
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	177

1、算力动态：恒源云 13 核+128G 算力紧张，英伟达发布 AI 超级计算机 Jetson Orin Nano Super

1.1、数据跟踪：恒源云 13 核+128G 算力紧张

本周恒源云 13 核+128G 算力紧张。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 算力紧张；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	-	-
	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：Jetson Orin Nano Super 蕴含超强 AI 性能，解锁多元场景应用新篇

近日，英伟达推出了一款全新的尺寸小巧的生成式 AI 超级计算机 NVIDIA Jetson Orin Nano Super，该开发者套件只有手掌大小，能够为商业 AI 开发者、科技爱好者和学生等各类人群提供更强大的生成式 AI 功能与性能，目前建议零售价 2070 人民币。Jetson Orin Nano Super 凭借其卓越性能，无论是创建基于检索增强生成的 LLM 聊天机器人、构建视觉 AI 智能体，还是部署基于 AI 的机器人，均能得以有效实现。Jetson Orin Nano Super 亮点如下：

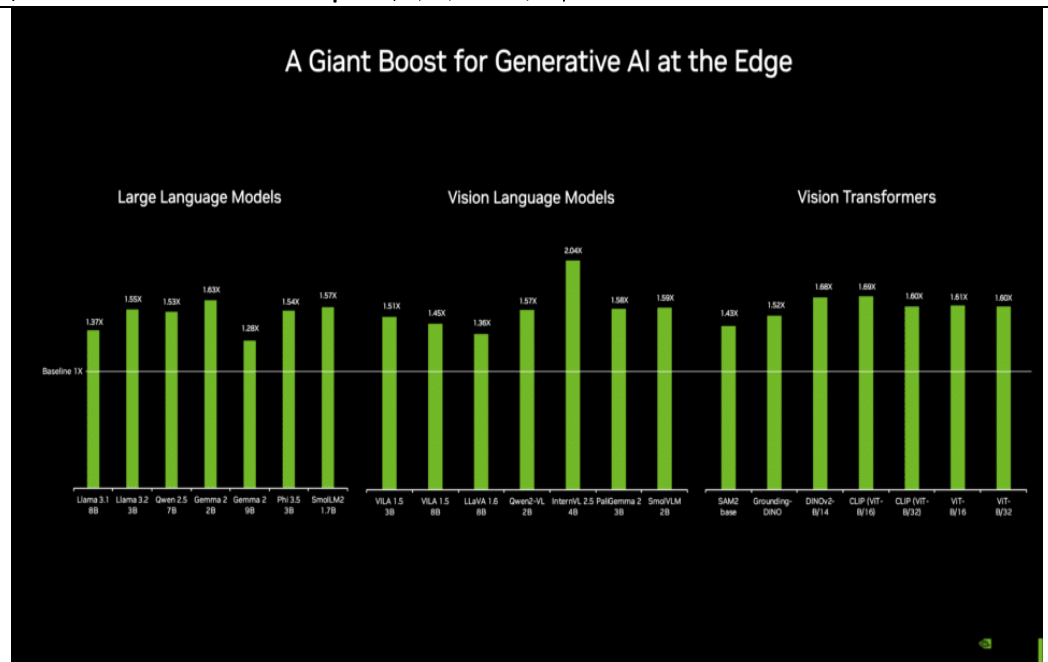
图表 2: Jetson Orin Nano Super 开发者套件



资料来源: NVIDIA, 华鑫证券研究

具备颠覆性的生成式 AI 模型性能。NVIDIA 最新软件更新将 Jetson Orin Nano 的 AI 性能从 40 TOPS 提升至 67 TOPS, 并将内存带宽从 68 GB/s 提升至 102 GB/s, 使开发者能够更高效地运行各种生成式 AI 模型。与上一代产品相比, NVIDIA Jetson Orin Nano Super 的生成式 AI 推理性能提升至 1.7 倍, 性能提升 70%, 达到 67 INT8 TOPS, 内存带宽提高了 50%, 达到 102GB/s。NVIDIA Jetson Orin Nano Super 在各种热门大语言模型 (LLM)、视觉语言模型 (VLM) 和视觉 Transformer 上的性能相比前代产品均有显著提升。

图表 3: Jetson Orin Nano Super 在性能上的提升



资料来源: NVIDIA, 华鑫证券研究

助力解锁更多 AI 场景应用。NVIDIA Jetson Orin Nano Super 可提供出色的 AI 计算能力、快速的内存带宽和全面的软件堆栈，为新的生成式 AI 应用提供卓越的能效，它支持任何基于 Transformer 的生成式 AI 模型的快速推理。在机器人领域，从开发到仿真再到部署，借助 AI 加速机器人开发，实现移动、抓取和视觉等关键功能，实现为各行各业构建机器人。在视觉领域，其搭载的 NVIDIA Metropolis 集成了视觉数据和 AI，可以处理适用于顺畅零售、库存管理、智慧城市交通、工厂检查和医疗健康等应用的传感器数据。

目前，NVIDIA Jetson Orin 包含 Jetson AGX Orin、Jetson Orin NX 和 Jetson Orin Nano 三款系列产品，可为开发者提供不同性能和价位的加速计算能力，从而满足多种自主应用的需求。

2、AI 应用动态：Character.AI 搜索访问量环比+7.07%，豆包大模型家族全方位优化

2.1、流量跟踪：Character.AI 搜索访问量环比+7.07%

本期（2024.12.16-2024.12.20）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（878.4M）、Bing（404.7M）和 Canva（171.6M），访问量环比增速第一为 Character.AI（7.07%）；平均停留时长前三位分别为 Character.AI（00:13:14）、Discord（00:12:12）和 DeepL（00:08:49）；平均停留时长环比增速第一为文心一言（2.27%）。

图表 4：2024.12.16-2024.12.20 AI 相关网站流量

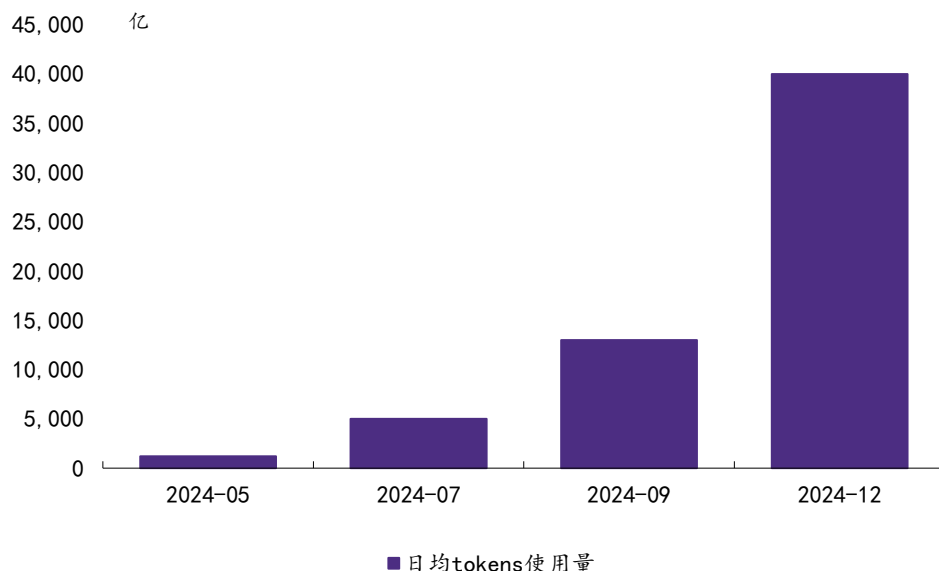
应用	应用类型	归属公司	周平均访问量 (M)	访问量环比	平均停留时长	时长环比
ChatGPT	聊天机器人	OpenAI	878.4	-3.19%	6:17	-0.26%
Bing	搜索	微软	404.7	-2.06%	6:37	-1.24%
Discord	游戏社区	微软	129.7	1.81%	12:12	0.00%
Canva	在线设计	Canva	171.6	-4.93%	7:50	0.00%
Github	代码托管	微软	111.2	-2.37%	6:27	-0.51%
Gemini	聊天机器人	谷歌	61.58	-2.04%	4:40	2.19%
Character.AI	聊天机器人	Character.AI	46.32	7.07%	13:14	0.13%
NotionAI	文本/笔记	Notion	32.68	-3.80%	8:46	0.57%
QuillBot	释义工具	QuillBot	14.67	-14.16%	4:05	-0.41%
Kimi	聊天机器人	Moonshot AI	8.612	-0.07%	3:28	-0.95%
DeepL	翻译工具	DeepL	46.22	-1.99%	8:49	-0.19%
文心一言	聊天机器人	百度	4.798	2.46%	3:00	2.27%
Perplexity	AI 搜索	Perplexity	21.92	-5.60%	5:56	-0.56%

资料来源：similarweb, 华鑫证券研究

2.2、产业动态：豆包大模型家族全方位优化

2024 年 12 月 18 日火山引擎 FORCE 原动力大会上，火山引擎展示了其全新升级的豆包大模型家族，包含豆包视觉理解模型、豆包 3D 生成模型，以及全面更新的豆包通用模型 pro、音乐模型、文生图模型等，为企业提供更多模态大模型能力。豆包大模型于 12 月实现日均 tokens 使用量超过 4 万亿，较 5 月发布时期增长超过 33 倍，稳居行业领先地位。

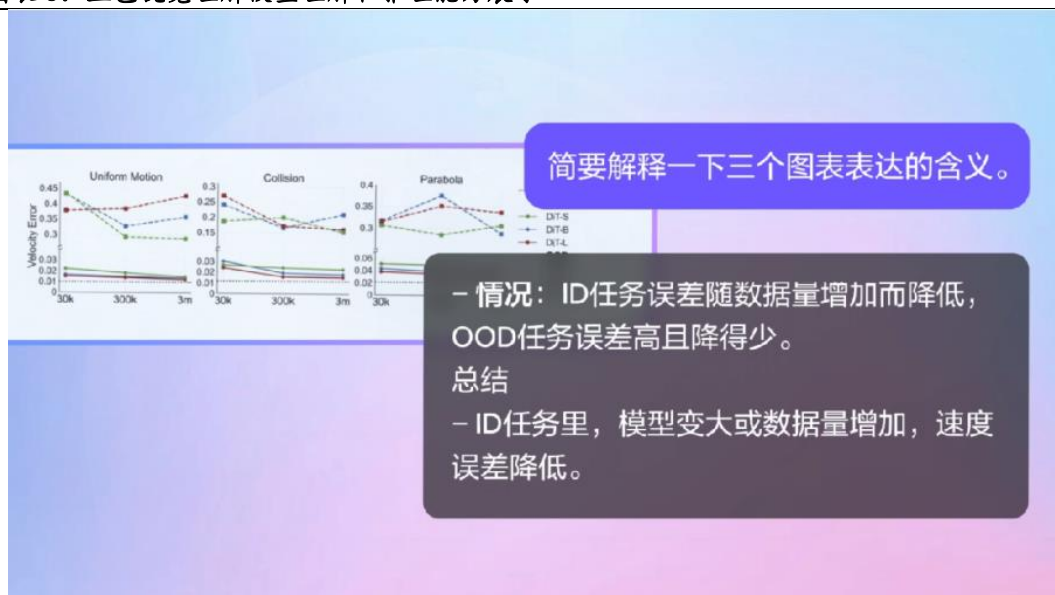
图表 5：2024 年豆包大模型日均 tokens 使用量提升情况



资料来源：火山引擎，华鑫证券研究

视觉理解模型重磅发布。视觉理解模型具备以下三种能力：更强的内容识别能力，可识别出图像中的物体类别、形状等基本要素以及理解物体之间的关系、空间布局以及场景的整体含义；更强的理解和推理能力，能根据所识别的文字和图像信息进行复杂的逻辑计算；更细腻的视觉描述能力，可以基于图像信息，更细腻地描述图像呈现的内容以及多种文体创作。以上能力极大拓展了大模型的场景边界，在金融、医疗、建筑、地理、体育、物流等诸多行业有着广阔应用前景。除此之外，视觉理解的输入价格为每千 tokens 0.003 元，比行业平均价格降低 85%，相当于 1 元钱可以处理 284 张 720P 的图片，具备高性价比优势，视觉理解模型正式走进厘时代。同时火山引擎还将提供更高的初始流量，RPM 达到 15000 次，TPM 达到 120 万，让企业和开发者用好视觉理解模型，找到更多创新场景

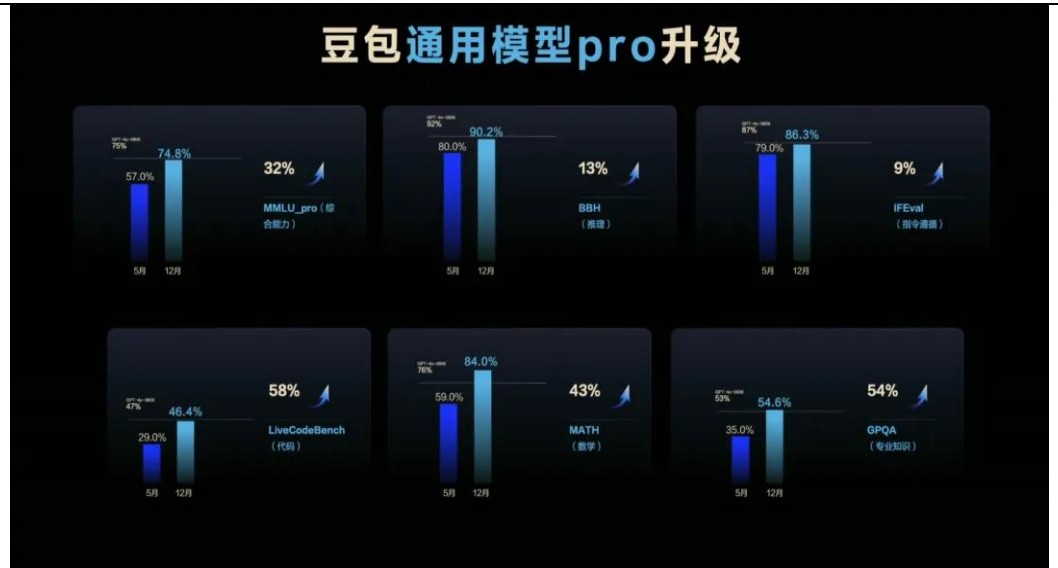
图表 6：豆包视觉理解模型理解和推理能力展示



资料来源：火山引擎，华鑫证券研究

模型家族迎来全面升级。火山引擎除了推出视觉理解模型之外，还发布、升级了多个其他模型，大模型家族成员更丰富，大模型能力再提升。大会发布豆包 3D 生成模型，其与火山引擎数字孪生平台 veOmniverse 结合使用，可以高效完成智能训练、数据合成和数字资产制作，成为一套支持 AIGC 创作的物理世界仿真模拟器。此外，豆包通用模型 pro 全面升级，对齐 GPT-4o，使用价格仅为后者的 1/8；音乐模型从生成 60 秒的简单结构，升级到生成 3 分钟的完整作品；文生图模型 2.1 版本在业界首次实现精准生成汉字和一句话 P 图的产品化能力，该模型已接入即梦 AI 和豆包 App。

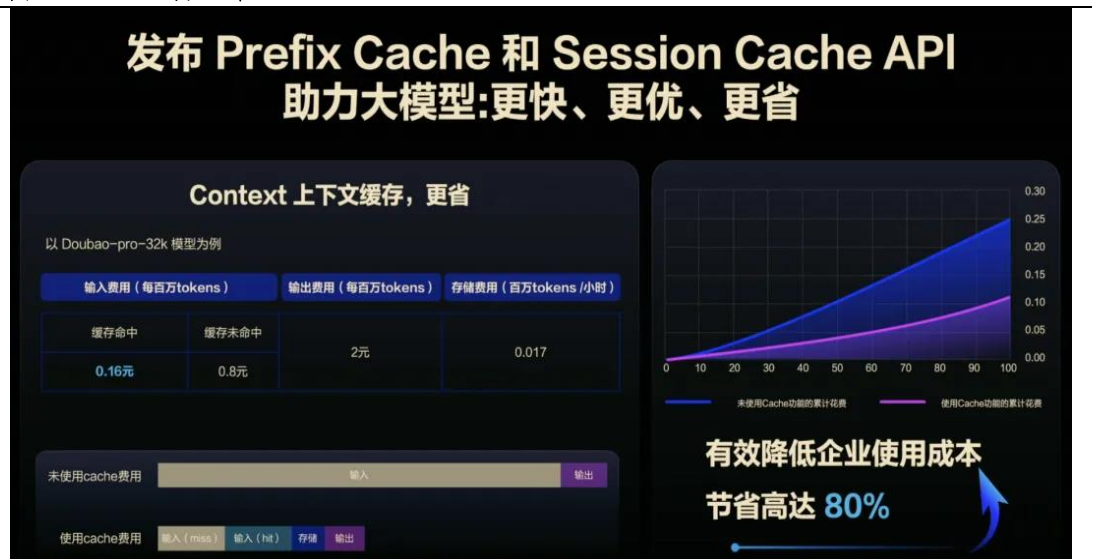
图表 7：豆包通用模型 pro 全面升级



资料来源：AI 前线，华鑫证券研究

AI 应用开发平台全新升级。火山引擎升级了火山方舟、扣子和 HiAgent 三款平台产品，帮助企业构建好自身的 AI 能力中心，高效开发 AI 应用。其中，火山方舟发布了大模型记忆方案，并推出 Prefix Cache 和 Session Cache API，降低延迟和成本。火山方舟还带来全域 AI 搜索，具备场景化搜索推荐一体化、企业私域信息整合等服务。

图表 8：火山方舟发布 Prefix Cache 和 Session Cache API



资料来源：AI 前线，华鑫证券研究

技术架构面向 AI 全面转型。云原生是过去十年最重要的计算范式，而火山引擎认为，未来十年计算范式应从云原生转变为 AI 云原生。基于 AI 云原生理念，火山引擎推出了新一代计算、网络、存储和和安全产品。在计算层面，火山引擎 GPU 实例，通过 vRDMA 网络，支持大规模并行计算和 P/D 分离推理架构，显著提升训练和推理效率，降低成本；在存储方面，新推出的 EIC 弹性极速缓存，能够实现 GPU 直连，使大模型推理时延降低至 1/50、成本降低 20%；在安全方面，火山将推出 PCC 私密云服务，构建大模型的可信应用体系。基于 PCC，企业能够实现用户数据在云上推理的端到端加密，性能卓越。

火山引擎持续致力于增强模型能力、完善模型服务，促进 AI 在各行各业的深度融合与创新应用，打造更强性能、更低价格、更易落地的豆包大模型，有望开启一个更加智能、高效、便捷的 AI 时代。

3、AI 融资动向：Databricks/智谱 AI 获本周 AI 融资市场估值前二，分获 100 亿美元/30 亿人民币融资

本周，Databricks/智谱 AI 获本周 AI 融资市场估值前二，分别获 100 亿美元/30 亿元人民币融资。Databricks 最初是美国加州大学伯克利 AMP 实验室的 Spark 大数据处理系统商业化项目，可以帮助企业以极快的速度分析其内部大数据，也凭此在获得了在硅谷立足之地。到 2020 年末，Databricks 推出了其数据仓库产品——Databricks SQL，凭借着对 AI 数据的理解，Databricks 先后打造了 Apache Spark、Delta Lake、MLflow 和 OneLakehousePlatform，开设了多条产品线。目前，Databricks 主要靠开发软件来提取、分析和构建人工智能应用程序，这些应用程序使用来自各种来源的复杂数据。商业模式方面，Databricks 根据客户每秒消耗的计算资源量收费，并打造了自家独有的 DBU 作为衡量单位。从官网客户资源库可以看到，Databricks 拥有移动通信公司 AT&T、电子商品公司惠普、语言 AI 工具 Grammarly 等超过 500 多家客户群体。

智谱 AI 或受益于清华系的科研底色，一直保持着较快的迭代速度。2020 年底，智谱 AI 研发 GLM 预训练架构，2021 年训练完成百亿参数模型 GLM-10B，同年利用 MoE 架构成功训练出收敛的万亿稀疏模型。2023 年，智谱 AI 推出了千亿基座的对话模型 ChatGLM，并开源了单卡版模型 ChatGLM-6B，使得研究者和个人开发者可以进行微调和部署。与此同时，智谱 AI 开放平台的价格从 2023 年年初到 2024 年下半年降低了 5000 倍。2024 年以来，智谱 AI 更是先后发布对标 Sora 的视频生成模型 CogVideoX，以及对标 GPT-4o 的端到端语音模型 GLM-4-Voice 和 GLM-4-VideoCall。

图表 9: 本周 AI 初创公司的融资动态

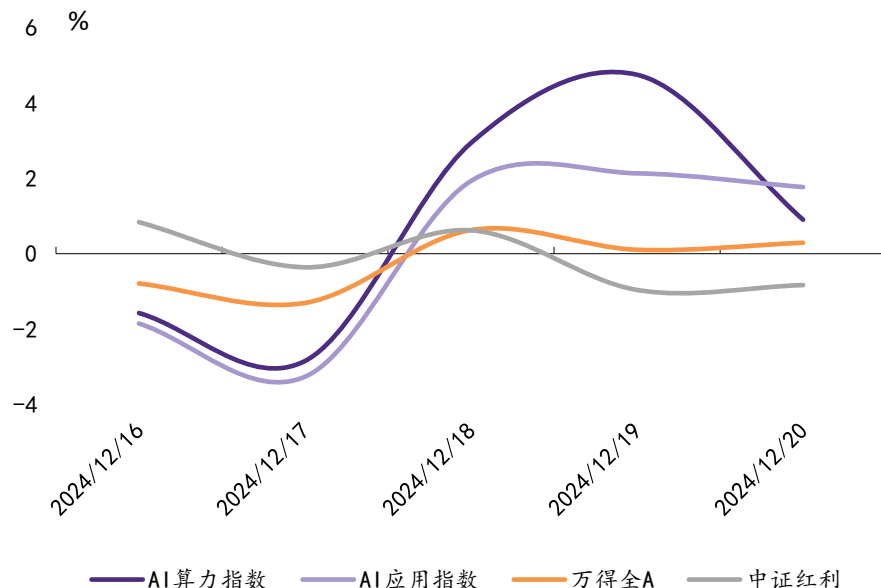
应用	应用类型	领投方	融资轮	融资额	目前累计融资额	目前估值
爱诗科技	AI 视频大模型	北京市人工智能产业投资基金	A+轮	3 亿元	——	——
Databricks	大型语言模型	Thrive Capital	J 轮	100 亿美元	——	620 亿美元
Perplexity AI	人工智能	Institutional Venture Partners	——	5 亿美元	——	90 亿美元
智谱 AI	AI 大模型	多家战投	——	30 亿元人民币	——	200 亿元

资料来源: MSN, 新浪财经, 21 经济网, 华鑫证券研究

4、行情复盘

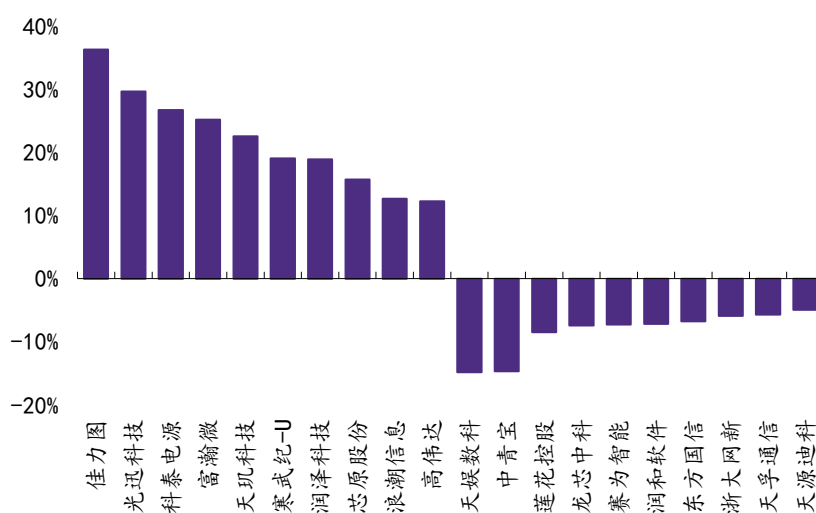
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利日涨幅最大值分别为 4.75%/2.13%/0.63%/0.84%，日跌幅最大值分别为-2.85%/-3.27%/-1.31%/-0.96%。AI 算力指数内部，佳力图以+36.27%录得本周最大涨幅，天娱数科以-14.82%录得本周最大跌幅。AI 应用指数内部，光云科技以+35.48%得本周最大涨幅，实丰文化以-22.16%录得本周最大跌幅。

图表 10：本周指数日涨跌幅



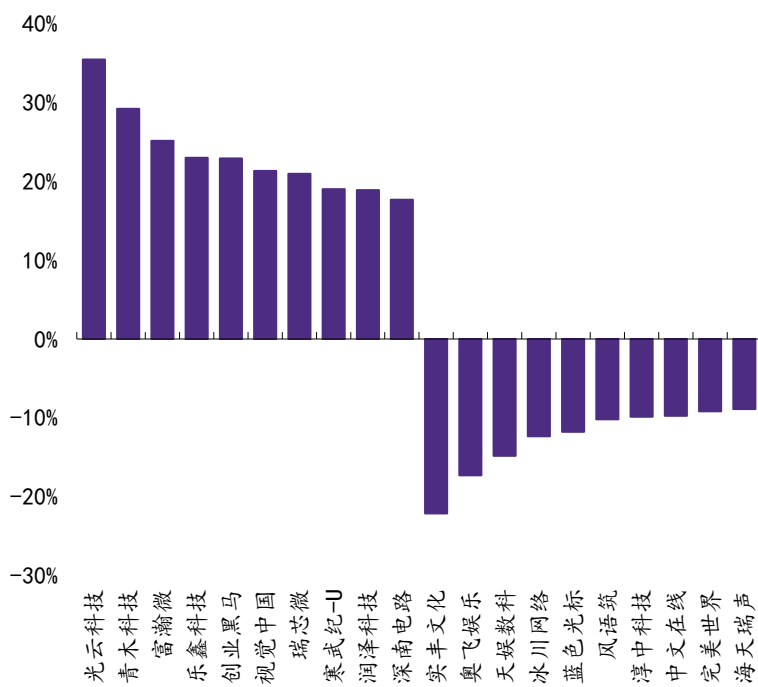
资料来源：wind, 华鑫证券研究

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

大模型或从快速扩张迈入淘汰赛阶段，字节明年有望迎来爆款应用。今年以来，视觉理解模型在 GPT-4o、Google 的 Project Astra，以及 iPhone 16 和众多旗舰机型上都加速落地。豆包视觉理解模型不仅能精准识别视觉内容，还可根据图像信息进行复杂的逻辑计算，完成分析图表、处理代码、解答学科问题等任务。豆包视觉理解模型千 tokens 输入价格仅为 3 厘，大模型的价格体系实现了从分至厘的跨越性调整，以更低成本让多模态交互降低用户的使用门槛，辅助完成一系列复杂的任务。人类接收的信息超过 80% 来自视觉，视觉理解将极大地拓展大模型的能力边界，同时也会降低人们与大模型交互的门槛，为大模型解锁更丰富的应用场景。同时，随着爆款应用的诞生，算力需求预计大幅增加。

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	2024-12-22 股价	EPS			PE			投资评级 2023
			2023	2024E	2025E	2023	2024E	2025E	
001314.SZ	亿道信息	52.71	0.91	0.92	1.03	001314.SZ	亿道信息	52.71	0.91
002230.SZ	科大讯飞	51.75	0.28	0.40	0.56	002230.SZ	科大讯飞	51.75	0.28
688256.SH	寒武纪-U	675.95	-2.04	-1.21	-0.50	688256.SH	寒武纪-U	675.95	-2.04
688668.SH	鼎通科技	40.74	0.67	1.04	1.41	688668.SH	鼎通科技	40.74	0.67

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。