

DeepSeek-V4 模型预览版正式上线并同步开源，海内外大模型厂商纷纷提质降价

——计算机行业跟踪报告

同步大市 (维持)

2026 年 04 月 27 日

行业核心观点:

上周，沪深 300 上涨 0.86%，创业板指下跌 0.29%，申万计算机行业指数下跌 0.27%，跑输沪深 300 指数 1.13 个百分点，跑赢创业板指 0.02 个百分点，在申万各一级行业中排名第 15 位。上周，DeepSeek-V4 模型预览版正式上线并同步开源，DeepSeek-V4 拥有百万字超长上下文，在 Agent 能力、世界知识和推理性能上均实现国内与开源领域的领先。上线后，华为云官方宣布为 DeepSeek-V4 提供首发适配，华为云 MaaS 模型即服务平台已为开发者提供免部署、一键调用 DeepSeek-V4-Flash API 的 Tokens 服务。DeepSeek 方面已经确认，受限于高端算力，目前 Pro 版本的服务吞吐较为有限，预计下半年昇腾 950 超节点批量上市后，Pro 的价格还将大幅下调。同时，英伟达、寒武纪、OpenClaw 等也纷纷表示和 DeepSeek-V4 完成了适配。我们认为此次 DeepSeek-V4 在国产算力深度参与下维持了领先的大模型能力，将进一步带动国产算力与更多大模型及 AI 应用端的适配需求，建议关注国产算力产业链相关的领先企业。此外，DeepSeek 宣布其全系 API 服务输入缓存命中价格降至原有价格的 1/10；OpenAI 推出的 GPT-5.5 模型相比前代系统百万 Token 成本降低至 35 分之一，每兆瓦 Token 输出量提升 50 倍；阶跃星辰发布的 StepAudio 2.5 ASR 自动语音识别模型推理速度提升 400%，定价骤减 90%。我们认为海内外大模型产品在向复杂化、专业化能力升级迭代的同时，持续呈现出提质降价的态势，大模型的整体竞争持续加剧，性价比和专用化能力的适配有望成为其核心竞争力之一。而在此趋势下我们认为 tokens 的使用量仍将保持高速增长态势，建议继续关注词元经济带来的投资机遇。中长期视角下，继续聚焦 AI 产业、数据产业两大投资主线。

投资要点:

产业动态: (1) **AI 大模型:** 4 月 24 日，DeepSeek-V4 模型预览版正式上线并同步开源。DeepSeek-V4 拥有百万字超长上下文，在 Agent 能力、世界知识和推理性能上均实现国内与开源领域的领先。模型按大小分为两个版本，分别是 deepseek-v4-pro 和 deepseek-v4-flash。(2) **AI 大模型:** 4 月 26 日 DeepSeek 官方宣布，即日起 DeepSeek 全系 API 服务，输入缓存命中价格降至原有价格的 1/10，Pro 模型还可在 5 月 5 日前叠加 2.5 折限时优惠。(3) **AI 大模型:** OpenAI 公司 4 月 24 日宣布推出 GPT-5.5 模型，是其迄今最智能、最直观的 AI 模型，在 Agentic Coding、计算机使用及科研领域表现卓越，相比较 GPT-5.4，在完成相同任务的情况下减少词元 (Token) 消耗。英伟达工程师通过 Codex 应用使用 GPT-5.5 已数周，基于 GB200 NVL72 系统，该模型实现了极具竞争力的经济效益：相比前代系统，百万 Token 成本降低至 35 分之一，每兆瓦 Token 输出量提升 50 倍。(4) **AI 大模型:** 阶跃星辰 4 月 24 日宣布推出新一代自动语音识别模型 StepAudio 2.5 ASR。实测数据显

行业相对沪深 300 指数表现



数据来源: 聚源, 万联证券研究所

相关研究

AI 大模型应用能力持续升级, 关注词元调用量增长及 AI 安全的部署需求

AI 大模型争相迭代, Anthropic 称 ARR 突破 300 亿美元

AI 产业应用商用化加速落地, 数据要素和数字人民币产业发展有望提速

分析师:

夏清莹

执业证书编号:

S0270520050001

电话:

(0755) 8322 3620

邮箱:

xiaqy1@wlzq.com.cn

示,模型推理速度提升 400%、时延降低 60%,推理峰值达 500 tokens/s,推理成本直降 80%。以 5 分钟左右的音视频为例,几乎可以实现即时转写。定价方面,StepAudio 2.5 ASR 仅为 0.15 元/小时,约为此前 Step ASR 2 的十分之一。

行业估值: 从估值情况来看,申万计算机行业 2026 年 4 月 24 日 PE-TTM 为 155.35 倍,处于较高水平,但接近 2023-2025 年历史 PE-TTM 的均值 158.26 倍。

风险提示: 中美科技摩擦;地缘政治风险;产业生态建设不及预期;应用落地不及预期,数据隐私安全风险,市场竞争加剧。

正文目录

1 行业周观点.....4

2 市场行情回顾.....4

3 产业动态.....5

 3.1 AI 大模型：DeepSeek-V4 模型预览版正式上线并同步开源.....5

 3.2 AI 大模型：降至原有价格 1/10，DeepSeek API 输入缓存降价.....6

 3.3 AI 大模型：OpenAI 最智能 AI 模型 GPT-5.5 登场，Token 成本降至 1/35、每兆瓦输出提升 50 倍.....6

 3.4 AI 大模型：阶跃星辰发布 StepAudio 2.5 ASR 自动语音识别模型：推理速度提升 400%，定价骤减 90%.....6

4 风险提示.....6

图表 1： 沪深 300、创业板指、申万一级行业指数涨跌幅（%）（上周）.....4

图表 2： 沪深 300、创业板指、申万一级行业指数涨跌幅（%）（2026 年初至 2026 年 4 月 26 日）.....5

图表 3： 申万计算机行业估值情况（PE-TTM，单位：倍）.....5

1 行业周观点

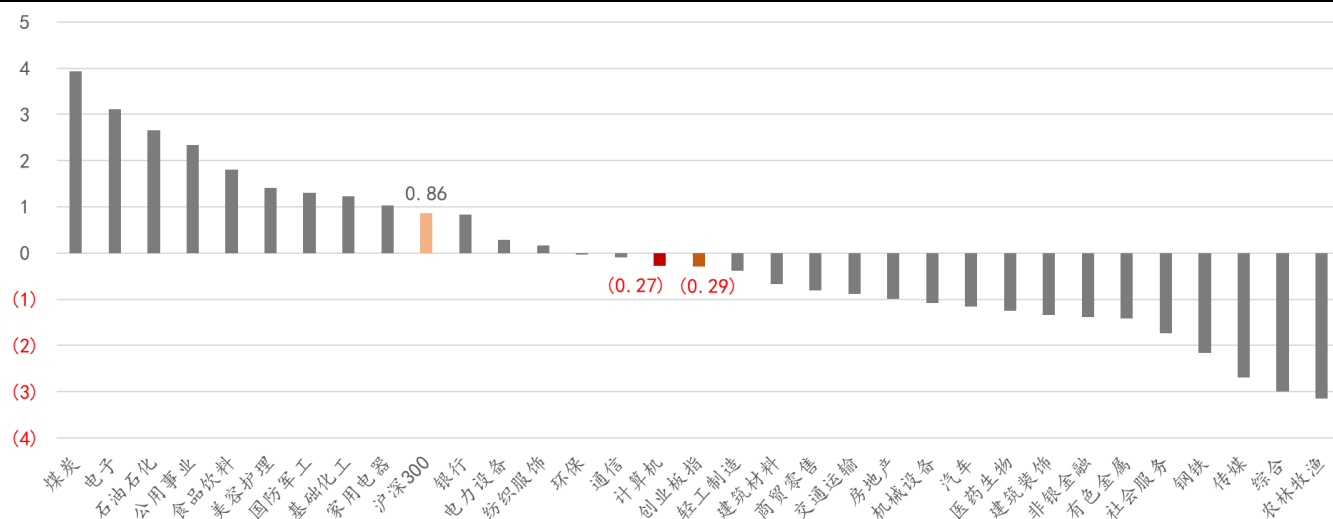
上周，DeepSeek-V4模型预览版正式上线并同步开源，DeepSeek-V4 拥有百万字超长上下文，在Agent能力、世界知识和推理性能上均实现国内与开源领域的领先。上线后，华为云官方宣布为DeepSeek-V4提供首发适配，华为云MaaS模型即服务平台已为开发者提供免部署、一键调用DeepSeek-V4-Flash API的Tokens服务。DeepSeek方面已经确认，受限于高端算力，目前Pro版本的服务吞吐较为有限，预计下半年昇腾950超节点批量上市后，Pro的价格还将大幅下调。同时，英伟达、寒武纪、OpenClaw等也纷纷表示和DeepSeek-V4完成了适配。我们认为此次DeepSeek-V4在国产算力深度参与下维持了领先的大模型能力，将进一步带动国产算力与更多大模型及AI应用端的适配需求，建议关注国产算力产业链相关的领先企业。此外，DeepSeek宣布其全系API服务输入缓存命中价格降至原有价格的1/10；OpenAI推出的GPT-5.5模型相比前代系统百万Token成本降低至35分之一，每兆瓦Token输出量提升50倍；阶跃星辰发布的StepAudio 2.5 ASR自动语音识别模型推理速度提升400%，定价骤减90%。我们认为海内外大模型产品在向复杂化、专业化能力升级迭代的同时，持续呈现出提质降价的态势，大模型的整体竞争持续加剧，性价比和专用化能力的适配有望成为其核心竞争力之一。而在此趋势下我们认为tokens的使用量仍将保持高增长态势，建议继续关注词元经济带来的投资机遇。

中长期视角下，继续聚焦AI产业、数据产业两大投资主线。1)AI产业：主要关注国内AI大模型的技术升级和产品迭代、头部AI大模型厂商的生态建设、智能驾驶、具身智能等新型智能终端及智能体的应用落地进程，以及人工智能治理的生态建设。2)数据产业：主要关注数据基础设施建设、高质量数据集等数据要素资源的开发利用、数据要素的价值释放及应用落地以及数据隐私安全。

2 市场行情回顾

申万计算机行业跑输沪深300，跑赢创业板指，周涨跌幅位于申万31个一级行业的中游。上周，沪深300上涨0.86%，创业板指下跌0.29%，申万计算机行业指数下跌0.27%，跑输沪深300指数1.13个百分点，跑赢创业板指0.02个百分点，在申万各一级行业中排名第15位。

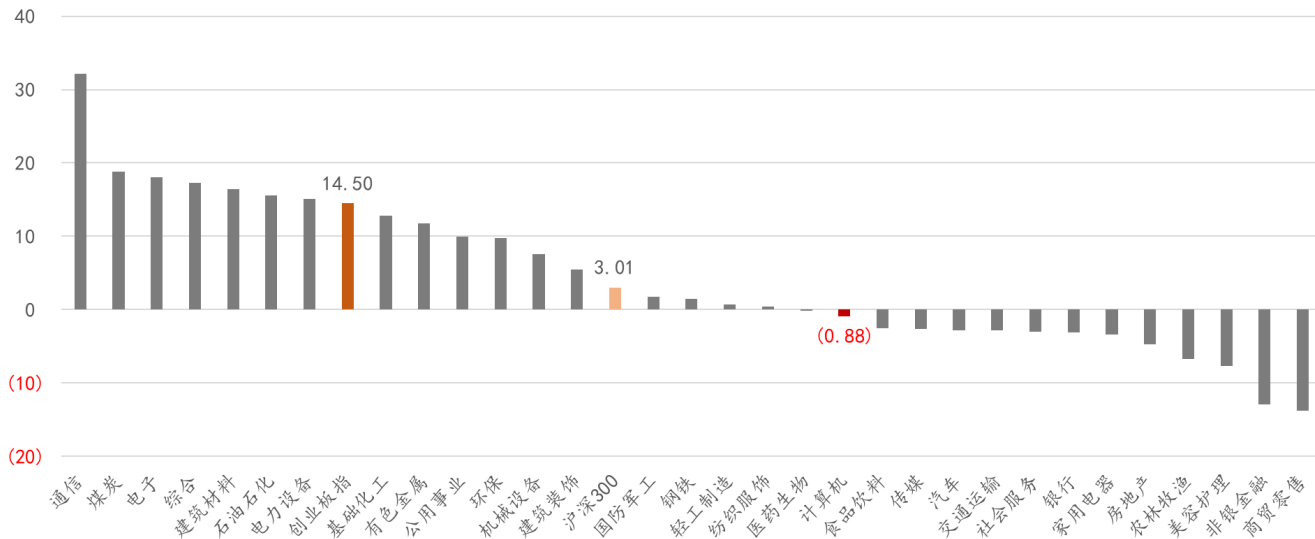
图表1: 沪深300、创业板指、申万一级行业指数涨跌幅(%) (上周)



资料来源: iFinD, 万联证券研究所

申万计算机行业2026年初至4月26日累计下跌0.88%，位于申万31个一级行业的中下游。2026年初至4月26日，沪深300上涨3.01%，创业板指上涨14.50%，申万计算机行业指数下跌0.88%，分别跑输沪深300指数和创业板指3.89和15.39个百分点，在申万各一级行业中排名第19位。

图表2: 沪深300、创业板指、申万一级行业指数涨跌幅(%) (2026年初至2026年4月26日)



资料来源: iFinD, 万联证券研究所

行业估值水平较高但接近历史中枢水平。从估值情况来看，申万计算机行业2026年4月24日PE-TTM为155.35倍，处于较高水平，但接近2023-2025年历史PE-TTM的均值158.26倍。

图表3: 申万计算机行业估值情况 (PE-TTM, 单位: 倍)



资料来源: iFinD, 万联证券研究所

3 产业动态

3.1 AI大模型: DeepSeek-V4模型预览版正式上线并同步开源

4月24日, DeepSeek-V4模型预览版正式上线并同步开源。DeepSeek-V4 拥有百万字超长上下文, 在Agent能力、世界知识和推理性能上均实现国内与开源领域的领先。模型按大小分为两个版本, 分别是deepseek-v4-pro和deepseek-v4-flash。相比

DeepSeek-V4-Pro, DeepSeek-V4-Flash在世界知识储备方面稍逊一筹,但展现出了接近的推理能力。而由于模型参数和激活更小,相较之下V4-Flash能够提供更加快捷、经济的API服务。在Agent测评中,DeepSeek-V4-Flash在简单任务上与DeepSeek-V4-Pro旗鼓相当,但在高难度任务上仍有差距。DeepSeek-V4 开创了一种全新的注意力机制,在token维度进行压缩,结合DSA稀疏注意力(DeepSeek Sparse Attention),实现了全球领先的长上下文能力,并且相比于传统方法大幅降低了对计算和显存的需求。从现在开始,1M(一百万)上下文将是DeepSeek 所有官方服务的标配。DeepSeek-V4 针对Claude Code、OpenClaw、OpenCode、CodeBuddy等主流的Agent产品进行了适配和优化,在代码任务、文档生成任务等方面表现均有提升。(资料来源:IT之家)

3.2 AI大模型: 降至原有价格 1/10, DeepSeek API 输入缓存降价

4月26日DeepSeek官方宣布,即日起DeepSeek全系API服务,输入缓存命中价格降至原有价格的1/10, Pro模型还可在5月5日前叠加2.5折限时优惠。(资料来源: IT之家)

3.3 AI大模型: OpenAI 最智能 AI 模型 GPT-5.5 登场, Token 成本降至 1/35、每兆瓦输出提升 50 倍

OpenAI 公司4月24日宣布推出GPT-5.5模型,是其迄今最智能、最直观的 AI 模型,在Agentic Coding、计算机使用及科研领域表现卓越,相比较GPT-5.4,在完成相同任务的情况下减少词元(Token)消耗。OpenAI 总裁格雷格·布罗克曼(Greg Brockman)强调,GPT-5.5的核心突破,在于能够以更少的指导完成更多任务,最大的亮点在于在处理模糊问题时,展现出更强的自主性,能快速理解用户意图,自主规划并执行多步骤复杂任务,在代码编写调试、在线研究、数据分析及跨工具操作等场景表现出色。英伟达工程师通过Codex应用使用GPT-5.5已数周,基于GB200 NVL72系统,该模型实现了极具竞争力的经济效益: 相比前代系统,百万Token成本降低至35分之一,每兆瓦Token输出量提升50倍。(资料来源: IT之家)

3.4 AI大模型: 阶跃星辰发布 StepAudio 2.5 ASR 自动语音识别模型: 推理速度提升 400%, 定价骤减 90%

阶跃星辰4月24日宣布推出新一代自动语音识别模型StepAudio 2.5 ASR。该模型的核心突破在于率先将大语言模型的推理加速技术引入语音识别领域,在推理速度与转写精度两个维度均有显著提升,主要面向会议转写、语音交互、输入法、媒体内容处理、长音频识别等场景。传统语音识别模型受限于自回归生成机制,需要逐个Token依次输出,效率较低。StepAudio 2.5 ASR 采用 ASR+MTP-5深度融合架构,将此前应用在Step 3.5 Flash 大模型上的 MTP技术移植至语音识别领域。该技术使模型能够一次预测多个候选Token,并通过并行验证机制快速确认结果,打破了传统自回归机制逐个输出的效率瓶颈。实测数据显示,模型推理速度提升400%、时延降低60%,推理峰值达500 tokens/s,推理成本直降80%。以5分钟左右的音视频为例,几乎可以实现即时转写。定价方面,StepAudio 2.5 ASR仅为 0.15元/小时,约为此前Step ASR 2的十分之一。(资料来源: IT之家)

4 风险提示

中美科技摩擦; 地缘政治风险; 产业生态建设不及预期; 应用落地不及预期, 数据隐私安全风险, 市场竞争加剧。

行业投资评级

强于大市：未来6个月内行业指数相对大盘涨幅10%以上；

同步大市：未来6个月内行业指数相对大盘涨幅10%至-10%之间；

弱于大市：未来6个月内行业指数相对大盘跌幅10%以上。

公司投资评级

买入：未来6个月内公司相对大盘涨幅15%以上；

增持：未来6个月内公司相对大盘涨幅5%至15%；

观望：未来6个月内公司相对大盘涨幅-5%至5%；

卖出：未来6个月内公司相对大盘跌幅5%以上。

基准指数：沪深300指数

风险提示

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

证券分析师承诺

本人具有中国证券业协会授予的证券投资咨询执业资格并登记为证券分析师，以勤勉的执业态度，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。

免责声明

万联证券股份有限公司（以下简称“本公司”）是一家覆盖证券经纪、投资银行、投资管理和证券咨询等多项业务的全国性综合类证券公司。本公司具有中国证监会许可的证券投资咨询业务资格。

本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。本报告中的信息或所表述的意见并未考虑到个别投资者的具体投资目的、财务状况以及特定需求。客户应自主作出投资决策并自行承担投资风险。本公司不对任何人因使用本报告中的内容所导致的损失负任何责任。在法律许可情况下，本公司或其关联机构可能会持有报告中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、财务顾问或类似的金融服务。

市场有风险，投资需谨慎。本报告是基于本公司认为可靠且已公开的信息撰写，本公司力求但不保证这些信息的准确性及完整性，也不保证文中的观点或陈述不会发生任何变更。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。分析师任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本报告的版权仅为本公司所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登、发表和引用。未经我方许可而引用、刊发或转载的引起法律后果和造成我公司经济损失的概由对方承担，我公司保留追究的权利。

万联证券股份有限公司 研究所

上海浦东新区世纪大道1528号陆家嘴基金大厦

北京西城区平安里西大街28号中海国际中心

深圳福田区深南大道2007号金地中心

广州天河区珠江东路11号高德置地广场