

2025 年 09 月 27 日



华鑫证券  
CHINA FORTUNE SECURITIES

# 英伟达 (NVDA.O)：重视 AI 推理计算，发布 Rubin CPX 芯片

—计算机行业点评报告

推荐(维持)

事件

分析师：任春阳 S1050521110006

✉ rency@cfsc.com.cn

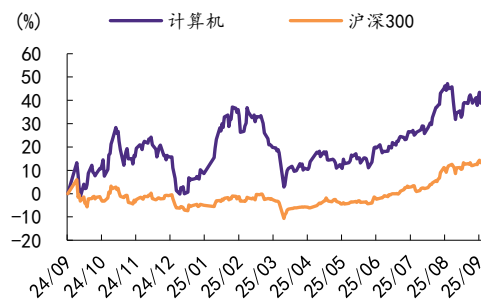
联系人：谢孟津 S1050123110012

✉ xiemj@cfsc.com.cn

## 行业相对表现

| 表现      | 1M   | 3M   | 12M  |
|---------|------|------|------|
| 计算机(申万) | -3.8 | 16.0 | 57.1 |
| 沪深 300  | 3.7  | 16.0 | 22.9 |

## 市场表现



资料来源：Wind，华鑫证券研究

## 相关研究

- 1、《计算机行业点评报告：谷歌 (GOOGL.O)：发布强大图像模型，巩固 AI 技术领先地位》2025-09-26
- 2、《计算机行业点评报告：微软 (MSFT.O)：发布整合性应用市场，强化云计算和 AI 应用生态》2025-09-26
- 3、《计算机行业点评报告：Figma (FIG.N)：AI 产品矩阵重塑增长路径，战略性投入着眼长期价值》2025-09-26

英伟达于 9 月 9 日宣布其 Blackwell Ultra 计算平台在 MLPerf Debut 测试中创下新的 AI 推理速度纪录。同日，英伟达发布了专门针对长上下文推理任务设计的 Rubin CPX 芯片。

## 投资要点

### ■ GB300 服务器创下 AI 推理速度记录，分离式推理成为重要方向

大模型发展对计算推理速度提出更高要求，英伟达算力系统具备领先表现。随着大语言模型参数量越来越大，且在推理过程中需要产生很多中间推理过程的 token，对算力硬件的推理能力提出了更高要求。近期英伟达对其最新的基于 Blackwell Ultra 架构的算力系统 (GB300 NVL72 服务器) 采用 MLPerf Inference v5.1 测试标准，对 DeepsSeek-R1、Llama 3.1、Whisper 等流行开源 AI 模型的推理速度进行了测试，结果显示，GB300 服务器在每 GPU 每秒生成 token 数量上创下新纪录，大幅领先公司此前的 DGX H200 和 GB200 服务器。

分离式推理成为重要方向，能够进一步增强 AI 推理速度。英伟达高度重视分离式推理的架构，在将 GB200 服务器中的 16 块 GPU 用于生成、56 块 GPU 用于上下文处理之后，能够进一步提升服务器的整体推理速度。该技术趋势有望在未来广泛应用并促进 AI 模型的推理效率，从而推动 AI 应用的部署。

### ■ 发布 Rubin CPX 芯片，加速长上下文推理性能和效率

AI 智能体系统需要多步推理、持久记忆和长上下文以处理复

杂任务，Rubin CPX 芯片针对长上下文处理和分离式推理进行优化。本次新发布的 Rubin CPX 芯片专门针对长上下文进行优化，并且对视频编码解码能力给予支持，在长上下文推理和视频模型推理中具备特殊效果。其设计本身反映了英伟达对强化 AI 推理计算效能和分离式推理架构的重视。我们认为未来该芯片的推广应用有望巩固英伟达在 AI 计算生态中的领先地位，并加速 AI 模型和应用的部署应用。

投资建议

英伟达始终处于 AI 计算领域的前沿，不断通过革新硬件和计算架构增强 AI 算力对 AI 模型计算的支持。建议投资者重点关注其 AI 算力产品更新、泛 AI 生态圈拓展延申等方面的进展。

风险提示

- (1) AI 技术投入与成本控制压力；(2) 市场竞争加剧；(3) 用户增长可持续性挑战；(4) 汇率波动与国际化风险。

重点关注公司及盈利预测

| 公司代码     | 名称   | 2025-09-27<br>股价 | EPS   |       |       | PE    |       |       | 投资评级 |
|----------|------|------------------|-------|-------|-------|-------|-------|-------|------|
|          |      |                  | 2024  | 2025E | 2026E | 2024  | 2025E | 2026E |      |
| MSFT. O  | 微软   | 511.46           | 11.86 | 13.46 | 15.75 | 43.12 | 38.00 | 32.47 | 未评级  |
| GOOGL. O | 谷歌 A | 246.54           | 8.13  | 9.77  | 10.6  | 30.32 | 25.23 | 23.26 | 未评级  |
| NVDA. O  | 英伟达  | 178.19           | 12.05 | 2.97  | 4.45  | 14.79 | 60.00 | 40.04 | 未评级  |

资料来源：Wind，华鑫证券研究（注：以上公司盈利预测皆取自万得一致预期，单位为美元）

■ 中小盘&主题&北交所、计算机&AI&互联网组介绍

**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。

**倪汇康：**金融学士，2025 年 8 月加盟华鑫证券研究所。

**谢孟津：**伦敦政治经济学院硕士，2023 年加入华鑫证券。

■ 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。

■ 免责条款

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。