



算力产品更新，为AI应用发展提供有力保障

■ 走势比较



■ 子行业评级

计算机设备	无评级
计算机软件	中性
IT服务	无评级
云服务	无评级
产业互联网	无评级

相关研究报告

<<AI国产化趋势延续>>—2025-03-17

<<通用型AI Agent更进一步，AI应用蓄势待发>>—2025-03-10

<<数字虚拟商品综合服务商，发力云终端打造第二成长曲线>>—2025-03-10

证券分析师：曹佩

电话：

E-MAIL: caopeisz@tpyzq.com

分析师登记编号：S1190520080001

证券分析师：王景宜

电话：

E-MAIL: wangjy@tpyzq.com

分析师登记编号：S1190523090002

报告摘要

英伟达 GTC 2025 召开，发布多款软硬件新产品。硬件层面：1)

Blackwell Ultra: 具备 288GB HBM3e 内存和 15PF 稠密 FP4 算力，推理性能较前代提升 40 倍。2) **GB300 NVL72**: 机架级解决方案整合 72 颗 GPU 与 36 颗 Grace CPU，AI 性能提升 1.5 倍，预计 2025 年下半年上市。3) **Vera Rubin**: 2026-2027 年推出基于 Rubin 架构的下一代 GPU 以及更强的机柜 Vera Rubin NVL144 和 Vera Rubin NVL576。软件层面：1) **Dynamo**: 专为推理、训练和跨整个数据中心加速而构建的开源软件，旨在提高推理性能，确保更大程度地利用 GPU 资源。2) **GROOT N1**: 全球首个人形机器人开源基础模型 GROOT N1，采用“快思考”和“慢思考”双系统架构模拟人类思考模式。

英伟达对算力需求展望积极，**Agentic AI** 和推理带来百倍增长。黄仁勋提出 AI 技术三阶段：生成式 AI 到 Agentic AI（自主决策 AI）到 Physical AI（机器人时代），当前由于推理模型和 Agentic AI 的出现，计算量需求增长超过 100 倍。大会透露，北美四大云厂商 2025 年共计采购了 360 万颗 Blackwell 芯片，预计 2028 年数据中心资本支出规模突破 1 万亿美元。预计到 2028 年全球数据中心投资达 1 万亿美元。

华为合作伙伴大会召开，昇腾联合合作伙伴发布大模型应用一体机。

2025 年 3 月 20 日-21 日，华为携手广大合作伙伴召开主题为“因聚而生众智有为”的“华为中国合作伙伴大会 2025”。华为宣布推进全面智能化战略，包括打造算力底座，繁荣昇腾 AI 生态，用 AI 赋能华为产品。大会期间，昇腾联合合作伙伴，面向企业轻量化 AI 部署场景，发布开箱即用的一体机产品。现场，神州鲲泰、云从科技、科大讯飞等 22 家合作伙伴正式发布基于昇腾算力的大模型应用一体机，助力企业实现高效推理和快速部署。

HarmonyOS 5 全新升级，鸿蒙 PC 将于 5 月发布。3 月 20 日，华为推出首款全面搭载 HarmonyOS 5 的手机 HUAWEI Pura X。据华为披露，原生鸿蒙开启公测以来，截至 2025 年 3 月，已收获用户反馈的优化建议 400 多万条，完成系统更新迭代 30 多个版本，新增超过 150 项功能特性。在生态方面，微信、抖音、支付宝、高德地图、京东等 20 多个 APP 下载量超过 200 万次。基于 HarmonyOS 5，HUAWEI Pura X 等终端智能体验迎来全方位升级。包括融合了盘古、Deepseek 双模型，以及智能助手小艺，能协助用户处理复杂事务，具备如小艺帮接、AI 智能缓存、小艺时光机等功能。同时，华为宣布鸿蒙电脑将在 2025 年 5 月亮相。

投资建议：AI 方向我们继续重点推荐金山办公、科大讯飞、用友网

络、赛意信息、熵基科技、海光信息、联想集团等；华为鸿蒙方向我们推荐智微智能、法本信息等。

风险提示：AI 技术发展不及预期，算力需求不及预期，应用落地不及预期，市场竞争加剧。

一、本周行业热点事件

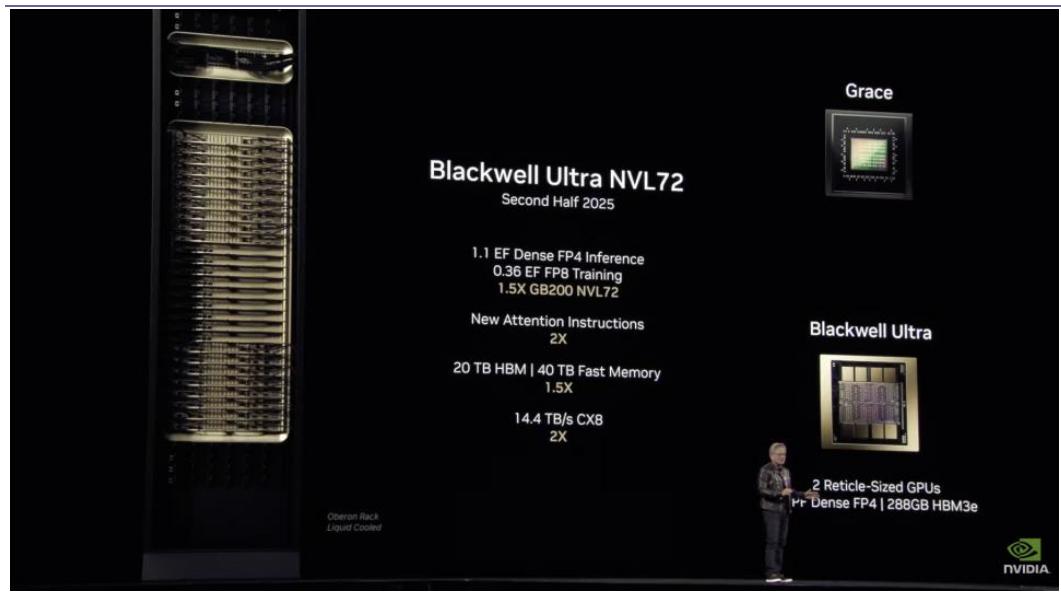
(一) 英伟达 GTC 2025 召开

2025 年 GTC 大会在 3 月 17 日-3 月 21 日于美国加州圣何塞举行。英伟达 CEO 黄仁勋于北京时间 2025 年 3 月 18 日举行主题演讲，介绍 AI 加速计算新时代、CPO 芯片及集成光子引擎、AI 技术在自动驾驶、机器人等前沿领域的应用。大会期间，英伟达推出多款产品。

➤ 硬件层面：

- **Blackwell Ultra**: FP4 计算性能方面预计较 B200 提升 50%，搭配了 12 层堆叠的 HBM3e 内存，显存提升至为 288GB，支持第五代 NVLink，可实现 1.8TB/s 的片间互联带宽。
- **Blackwell Ultra NVL72**: 推理专用机架，应用场景包括推理型 AI、Agent 以及物理 AI，相比前一代产品 GB200 NVL72 的 AI 性能提升了 1.5 倍。根据官方提供的信息，6710 亿参数 DeepSeek-R1 的推理，基于 H100 产品可实现每秒 100 tokens，而采用 Blackwell Ultra NVL72 方案，可以达到每秒 1000 tokens。
- 下一代超级芯片 **Vera Rubin**: 2026-2027 年推出基于 Rubin 架构的下一代 GPU 以及更强的机架 Vera Rubin NVL144 和 Vera Rubin NVL576。Rubin 将基于 TSMC 3nm 制程构造，提供 50 PFLOP 密集 FP4 计算，是 B300 的三倍多。

图表1: Blackwell Ultra NVL72



资料来源：腾讯科技，太平洋证券整理

图表2：英伟达产品路线图

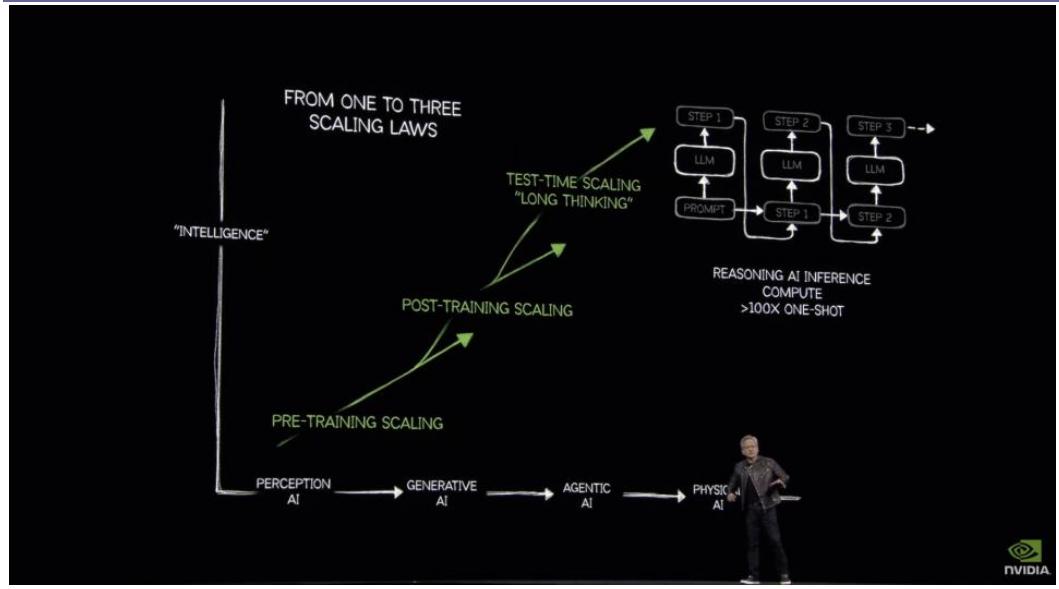
Nvidia Roadmap											
	2022	2023	2024	2025		2026	2027				
	Chip and Package Level										
Accelerator	Hopper		Blackwell		Rubin						
H100 (SXM)	H200	B200/ GB200	GB300 (Ultra)	B300 (single die, B300A)	VR200	VR300 (Ultra)					
GPU TDP (W)	700	700	700/1200	1,400	600	1,800	3,600				
Foundry Node	4N		4NP		N3P (3NP)						
Logic Die Configuration	1 x Reticle Sized GPU		2 x Reticle Sized GPU		2 x Reticle Sized GPU, 2x I/O chiplet	4 x Reticle Sized GPU, 2x I/O chiplet					
FP4 PFLOPs - Dense (per Package)	4*		10	15	4.6	50	100				
HBM	80GB HBM3	141GB HBM3E	192GB HBM3E	288GB HBM3E	144GB HBM3E	288GB HBM4	1024GB HBM4E				
HBM Stacks	5	6	8		4	8	16				
HBM Bandwidth	3.35TB/s	4.8TB/s	8TB/s		4TB/s	13TB/s	32TB/s				
Packaging	CoWoS-S		CoWoS-L		CoWoS-L						
SerDes speed (Gb/s uni-di)	112G		224G		224G	448G					
Nvidia CPU	Grace				Vera						
System Form Factor											
Maximum system density	NVL8		NVL72	144 compute chiplets 72 GPUs	NVL16	NVL144 144 compute chiplets 72 GPUs	NVL576 576 compute chiplets 144 GPUs				
Form Factor Supported	HGX		HGX,Oberon		HGX, Oberon, Kyber						
# of GPU Packages	8		72	72	16	72	144				
# of GPU dies	8		144	144	16	144	576				
Scale up links	UBB (PCB)		Copper Backplane	UBB (PCB)	Copper Backplane	PCB Backplane					
Aggregate FP4 PFLOPs (Dense)	32*		720	1,080	74	3,600	14,400				
Aggregate HBM capacity	14TB	14TB	14TB	21TB	64TB	21TB	147TB				
Aggregate HBM bandwidth	27TB/s	38TB/s	576TB/s	576TB/s	64TB/s	936TB/s	4,608TB/s				

* Hopper doesn't have FP4 support, but downcast from other formats for sake of comparison.

资料来源: *SemiAnalysis*, 太平洋证券整理

英伟达认为推理模型和 Agentic AI 将带来算力需求持续扩张。英伟达判断，人工智能将逐步进入代理式人工智能（Agentic AI）阶段，随后发展到物理人工智能（Physical AI）阶段。当前由于推理模型和 Agentic AI 的出现，模型生成的 token 数是以前的 10 倍，计算量超过 100 倍。据大会公布数据，2024 年美国前四云厂总计采购 130 万颗 Hopper 架构芯片，2025 年共计采购了 360 万颗 Blackwell 芯片，预计 2028 年数据中心资本支出规模突破 1 万亿美元。

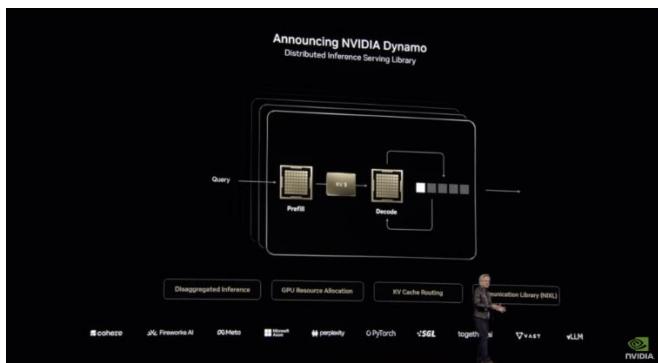
图表3：黄仁勋在 GTC 演讲中称推理让算力需求激增



资料来源：腾讯科技，太平洋证券整理

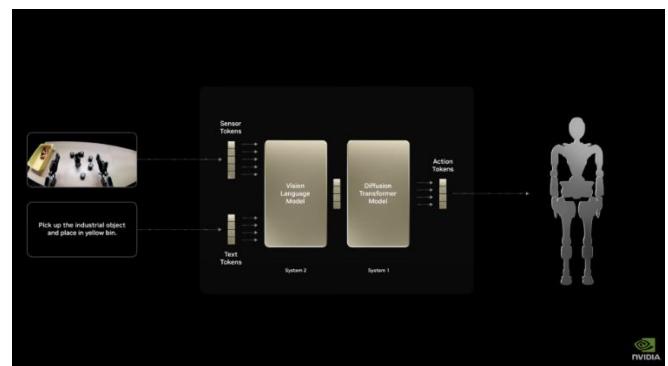
- 软件层面：发布并开源 **Dynamo** 推理软件和 **GROOT N1** 机器人模型
- **Dynamo**：专为推理、训练和跨整个数据中心加速而构建的开源软件，旨在提高推理性能，确保更大程度地利用 GPU 资源，为部署推理 AI 模型的 AI 工厂最大化其 token 收益。
- **GROOT N1**：全球首个人形机器人开源基础模型 **GROOT N1**，采用双系统架构模拟人类思考模式。“系统 1”是一个快速思考的动作模型，反映人类的本能反应或直觉。“系统 2”是慢思考模型，用于进行经过深度思考的决策制定。

图表4：Dynamo 的系统架构



资料来源：腾讯科技，太平洋证券整理

图表5：GROOT N1 的双系统架构



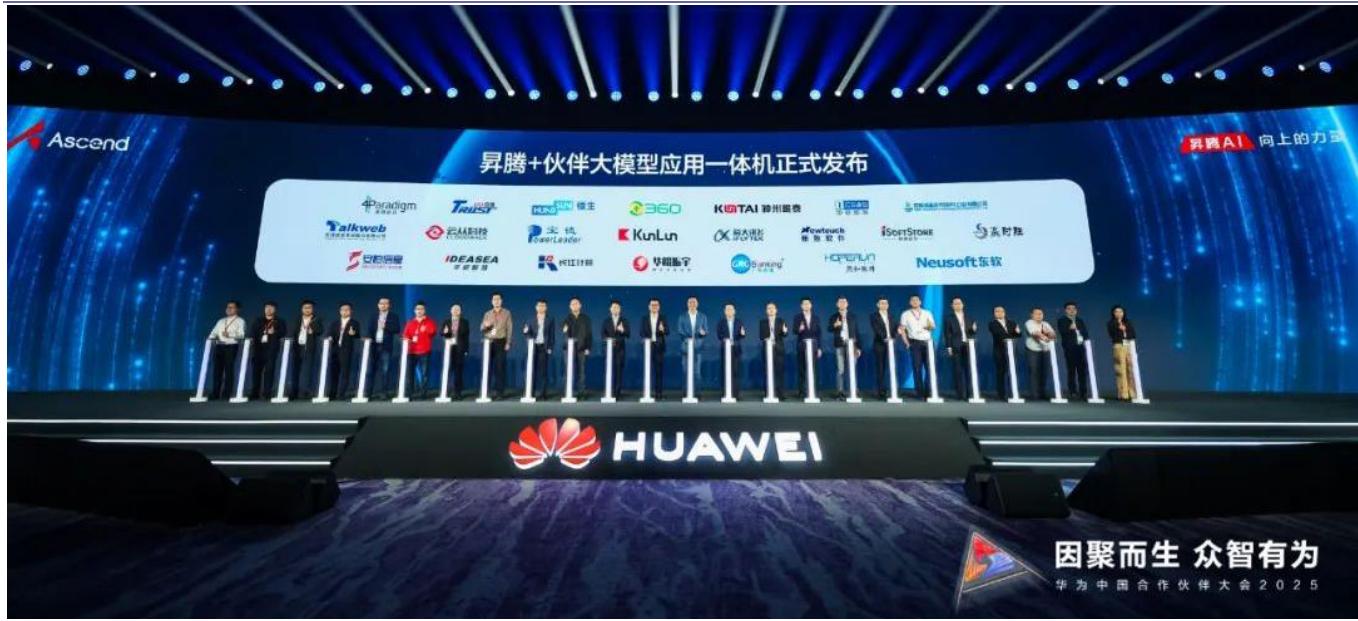
资料来源：腾讯科技，太平洋证券整理

(二) 华为合作伙伴大会 2025 召开

华为合作伙伴大会召开，推进全面智能化战略。2025年3月20日-21日，华为携手广大合作伙伴召开主题为“因聚而生 众智有为”的“华为中国合作伙伴大会 2025”。华为常务董事、ICT 基础设施管理委员会主任汪涛表示，AI 应用正重塑企业，大模型能力的跃迁和推理的爆发，为企业全面智能化铺平了道路，AI to B 的奇点时刻已经到来。为迎接通用人工智能时代的到来，华为推进全面智能化战略，包括打造算力底座，繁荣昇腾 AI 生态，用 AI 赋能华为产品。

昇腾联合合作伙伴发布大模型应用一体机。华为计算产品线总裁张熙伟表示，DeepSeek 将带来三波机遇，分别为大模型应用一体机、大规模专家并行(EP)集群推理方案、“微调+强化学习”来打造行业专属的 R1 模型。华为预计在第二季度推出昇腾多机大 EP 方案。大会期间，昇腾联合合作伙伴，面向企业轻量化 AI 部署场景，发布开箱即用的一体机产品。现场，神州鲲泰、云从科技、科大讯飞等 22 家合作伙伴正式发布基于昇腾算力的大模型应用一体机，助力企业实现高效推理和快速部署。根据财联社报道，华为今年还将新增大模型一体机伙伴，计划发展十几家大模型伙伴。今年面向伙伴投入将达到 100 亿元，同比提升 25%。

图表6：昇腾+伙伴大模型应用一体机正式发布



资料来源：华为计算微信公众号，太平洋证券整理

(三) HarmonyOS 5 全新升级，鸿蒙 PC 将于 5 月发布

3月20日，华为推出首款全面搭载 HarmonyOS 5 的手机 HUAWEI Pura X。据华为披露，原生鸿蒙开启公测以来，截至 2025 年 3 月，已收获用户反馈的优化建议 400 多万条，完成系统更新迭代 30 多个版本，新增超过 150 项功能特性。在生态方面，微信、抖音、支付宝、高德地图、京东等 20 多个 APP 下载量超过 200 万次。基于 HarmonyOS 5，HUAWEI Pura X 等终端智能体验迎来全方位升级。包括融合了盘古、Deepseek 双模型，以及智能助手小艺，能协助用户处理复杂事务，具备如小艺帮接、AI 智能缓存、小艺时光机等功能。同时，华为宣布，鸿蒙电脑将在 2025 年 5 月亮相，带来全场景互联、更流畅、更智能、更安全、更精致的崭新体验。

图表7：HarmonyOS 5 升级计划



资料来源：华为，太平洋证券整理

二、 投资建议

AI 方向我们继续重点推荐金山办公、科大讯飞、用友网络、赛意信息、熵基科技、海光信息、联想集团等；华为鸿蒙方向我们推荐智微智能、法本信息等。

三、 风险提示

AI 技术发展不及预期，算力需求不及预期，应用落地不及预期，市场竞争加剧。

投资评级说明

1、行业评级

看好：预计未来 6 个月内，行业整体回报高于沪深 300 指数 5%以上；
中性：预计未来 6 个月内，行业整体回报介于沪深 300 指数-5%与 5%之间；
看淡：预计未来 6 个月内，行业整体回报低于沪深 300 指数 5%以下。

2、公司评级

买入：预计未来 6 个月内，个股相对沪深 300 指数涨幅在 15%以上；
增持：预计未来 6 个月内，个股相对沪深 300 指数涨幅介于 5%与 15%之间；
持有：预计未来 6 个月内，个股相对沪深 300 指数涨幅介于-5%与 5%之间；
减持：预计未来 6 个月内，个股相对沪深 300 指数涨幅介于-5%与-15%之间；
卖出：预计未来 6 个月内，个股相对沪深 300 指数涨幅低于-15%以下。

太平洋证券股份有限公司

云南省昆明市盘龙区北京路 926 号同德广场写字楼 31 楼



太平洋证券
PACIFIC SECURITIES



研究院

中国北京 100044

北京市西城区北展北街九号

华远·企业号 D 座

投诉电话： 95397

投诉邮箱： kefu@tpyzq.com

免责声明

太平洋证券股份有限公司（以下简称“我公司”或“太平洋证券”）具备中国证券监督管理委员会核准的证券投资咨询业务资格。

本报告仅向与太平洋证券签署服务协议的签约客户发布，为太平洋证券签约客户的专属研究产品，若您并非太平洋证券签约客户，请取消接收、订阅或使用本报告中的任何信息；太平洋证券不会因接收人收到、阅读或关注媒体推送本报告中的内容而视其为太平洋证券的客户。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机构和个人的投资建议，投资者应自主作出投资决策并自行承担投资风险，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。

本报告信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映研究人员的个人观点。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我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业务服务。本报告版权归太平洋证券股份有限公司所有，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。任何人使用本报告，视为同意以上声明。