



信达证券
CINDA SECURITIES

Research and
Development Center

GTC2025 硕果良多，AI 创新进入快跑道

2025 年 3 月 21 日

证券研究报告

行业研究

行业专题研究

电子

投资评级 看好
上次评级 看好

莫文宇 电子行业首席分析师
执业编号: S1500522090001
邮箱: mowenyu@cindasc.com

杨宇轩 电子行业分析师
执业编号: S1500525010001
邮箱: yangyuxuan@cindasc.com

信达证券股份有限公司
CINDA SECURITIES CO., LTD
北京市西城区宣武门西大街甲 127 号金隅大厦 B 座
邮编: 100031

GTC2025 硕果良多, AI 创新进入快跑道

2025 年 3 月 21 日

本期内容提要:

- **GTC 大会十年磨剑, AI 进入快跑道。** GTC 大会是英伟达产品重磅发布的盛会, 见证了英伟达成长的历程。2014 年, 英伟达发布了 Pascal 架构。2016 年发布了基于 Pascal 架构的 P100 芯片。并于 2017 年发布 V100 芯片。2020 年, A100 芯片发布, 采用了台积电 7nm 制程, 超过 540 亿个晶体管。近几年英伟达硬件迭代加速, 性能提升也进入到新阶段。英伟达 2022 年发布了重磅 H 系列产品, 同步发售了 Quantum-X400 Infiniband 交换机。2024 年英伟达发布了 Blackwell 架构系列产品, 并在 2025 年进一步发布 Blackwell Ultra 系列产品更新。至于新一代 Rubin 架构系列产品则有望在 2026-2027 年推出。据 GTC2025 透露, 2028 年将会发布 Feynman 架构产品。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。此外, GTC 大会还有 1000 多场深具启发性的会议, 更有 400 多项展示、技术实战培训和大量独特的交流活动, 许多业界和学术界的翘楚也将参加会议。
- **GTC2025 发布多款软硬件产品, 建议关注相关优质个股。** 硬件层面, GTC2025 英伟达发布或预告了数款产品。(1) **Blackwell Ultra 架构系列产品:** Blackwell Ultra 搭配 12 层堆叠的 HBM3e 内存, 显存达到 288GB, 可实现 1.8TB/s 的片间互联带宽。GB300 NVL72 的 AI 性能是 NVIDIA GB200 NVL72 的 1.5 倍, 这使得建造 AI 工厂的收入机会与基于 NVIDIA Hopper™ 构建的系统相比增加了 50 倍。(2) **Rubin 架构系列产品:** 2026 年英伟达将推出 Rubin 架构, 并将机柜的节点进一步扩展, Rubin 将基于 TSMC 3nm 制程构造, 提供令人难以置信的 50 PFLOP 密集 FP4 计算, 是 B300 的三倍多。同时, 英伟达将扩大计算节点, 推出基于 Rubin 的 NVL144 及 NVL576。(3) **CPO 交换机:** GTC 英伟达推出了 Quantum-X 硅光共封芯片、Spectrum-X 硅光共封芯片以及衍生出来的三款交换机产品。NVIDIA 硅光交换机是全球领先的网络解决方案, 它们创新地集成了光器件, 减少了 4 倍的激光器数量, 与传统方法相比, 能源效率提高到 3.5 倍, 信号完整性提高到 63 倍, 大规模组网可靠性提高到 10 倍, 部署速度提高到 1.3 倍。**软件层面:** 英伟达发布了 Dynamo, 用于以最低的成本和最高的效率加速和扩展 AI 工厂中的 AI 推理模型。同时, 英伟达宣布开源全球首个人形机器人基础模型 GROOT N1。此外还有 Llama Nemotron 模型等等。我们认为, 当前 AI 发展明显加速, 推理成本的降低有望使得复杂的算力密集场景落地, 从而带来推理算力需求的增长。目前人形机器人、自动驾驶等赛道已经看到明显的 AI 赋能, 行业发展明显加速。其他 AI 端侧尤其是传统的消费电子终端如手机、PC 等有望被 AI 重新定义, 并衍生出新型的 AI 终端, 如智能管家、AI 眼镜、桌面机器人等等。今年各类端侧软件和硬件有望密集推出, 接入用户量的提升则有望推动云端算力成长。建议持续关注相关优质个股。
- **相关个股:** 【AI 云侧】工业富联/沪电股份/生益科技/深南电路/胜宏科技/寒武纪/海光信息; 【AI 端侧】蓝思科技/领益智造/鹏鼎控股/东山精密/乐鑫科技/瑞芯微/恒玄科技/全志科技/兆易创新/晶晨股份等。
- **风险提示:** 电子行业发展不及预期; 宏观经济波动风险; 地缘政治风险。

目 录

GTC 2025 正在进行时，大咖云集探索未来	4
GTC2025 发布多项产品，软硬件共振推动 AI 发展	6
风险因素	9

表 目 录

表 1: 消费电子涨跌 TOP5.....	9
-----------------------	---

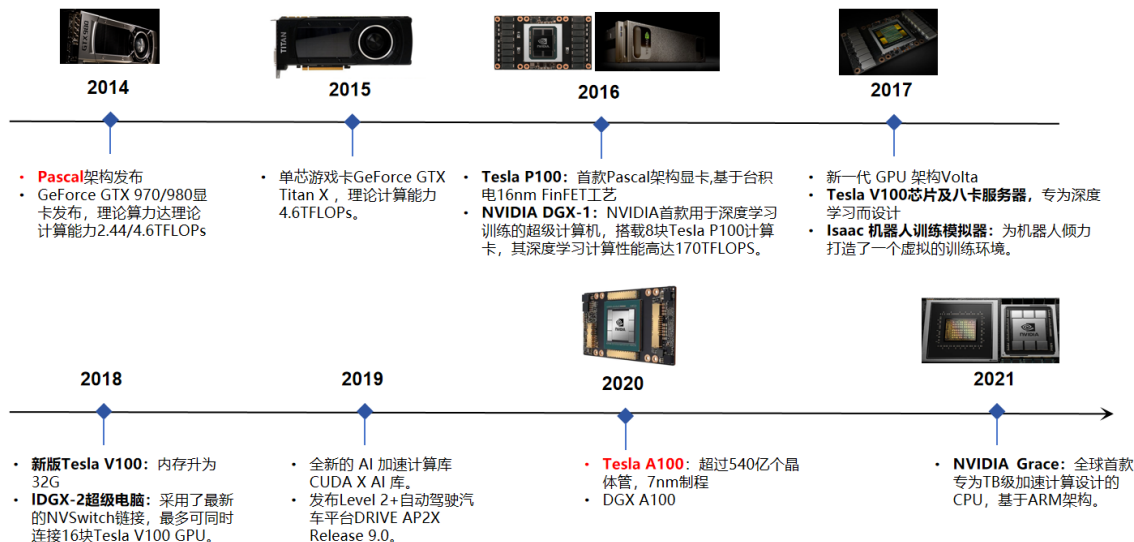
图 目 录

图 1: 历届 GTC 大会复盘	4
图 2: 英伟达 2022-2027 年产品规划	4
图 3: 2025 年 GTC 大会	5
图 4: 黄仁勋主题演讲	5
图 5: Blackwell Ultra NVL72	6
图 6: Rubin Ultra NVL576	6
图 7: Blackwell Ultra NVL144	6
图 8: CPO 交换机.....	6
图 9: GTC 大会公布的产品发布节奏	7
图 10: AI 将进入每一个行业	8

GTC 2025 正在进行时，大咖云集探索未来

GTC 大会十年磨剑，AI 进入快跑赛道。 GTC 大会是英伟达产品重磅发布的盛会，见证了英伟达成长的历程。2014 年，英伟达发布了 Pascal 架构。2016 年发布了基于 Pascal 架构的 P100 芯片。并于 2017 年发布 V100 芯片。2020 年，A100 芯片发布，采用了台积电 7nm 制程，超过 540 亿个晶体管。

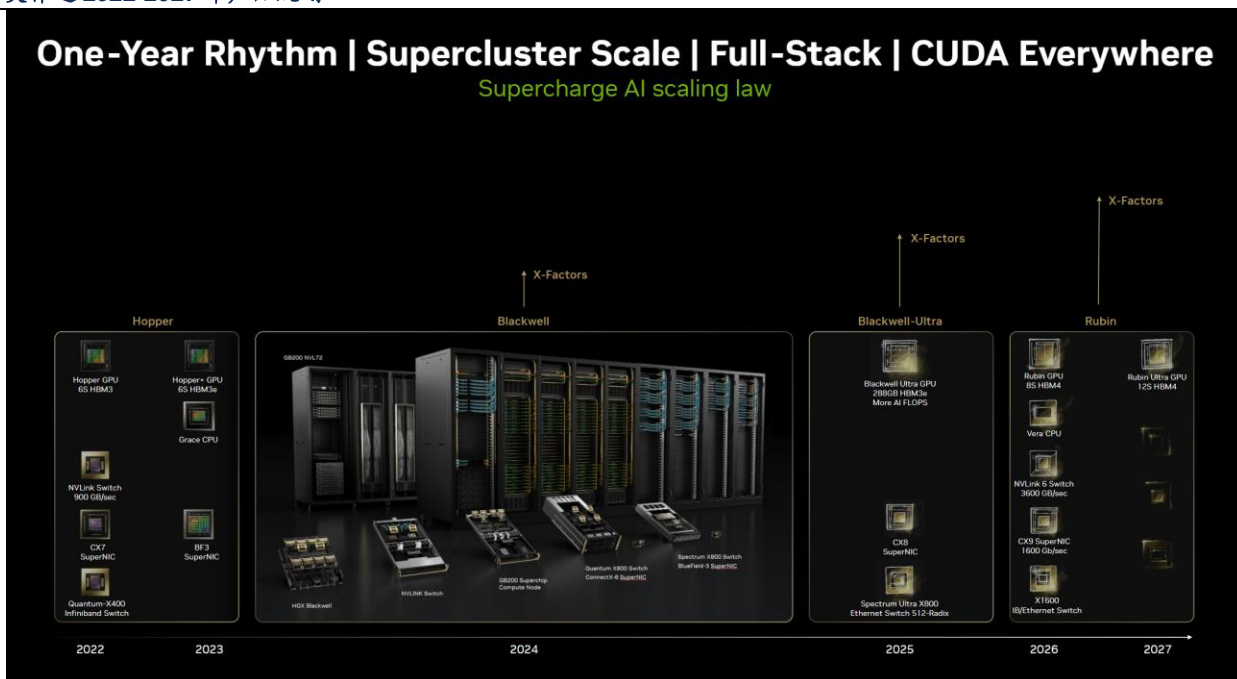
图 1：历届 GTC 大会复盘



资料来源：晶上世界，信达证券研发中心

英伟达硬件更新节奏趋紧，2022 年开始密集发布产品。硬件发布节奏看，英伟达 2022 年发布了重磅 H 系列产品，并且针对大模型做了硬件端的设计和优化，同步发售了 Quantum-X400 Infiniband 交换机。2024 年英伟达发布了 Blackwell 架构系列产品，并在 2025 年进一步发布 Blackwell Ultra 系列产品更新。至于新一代 Rubin 架构系列产品则有望在 2026-2027 年推出。

图 2：英伟达 2022-2027 年产品规划



资料来源：英伟达官网，信达证券研发中心

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。此外，GTC 大会还有 1000 多场深具启发性的会议，更有 400 多项展示、技术实战培训和大量独特的交流活动，许多业界和学术界的翘楚也将参加会议。

图 3：2025 年 GTC 大会



资料来源：英伟达官网，信达证券研发中心

图 4：黄仁勋主题演讲



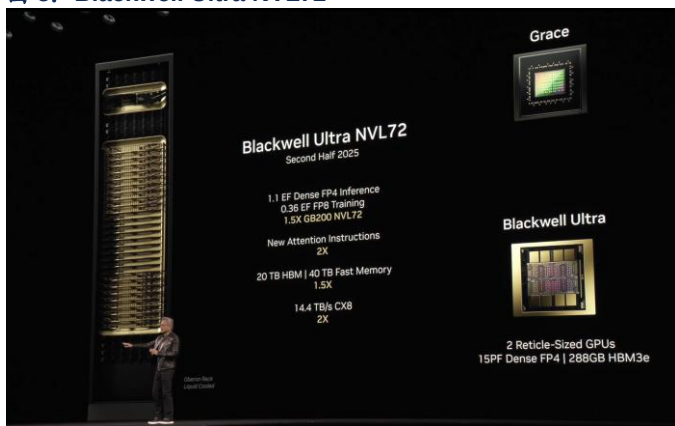
资料来源：英伟达官网，信达证券研发中心

GTC2025 发布多项产品，软硬件共振推动 AI 发展

硬件层面：

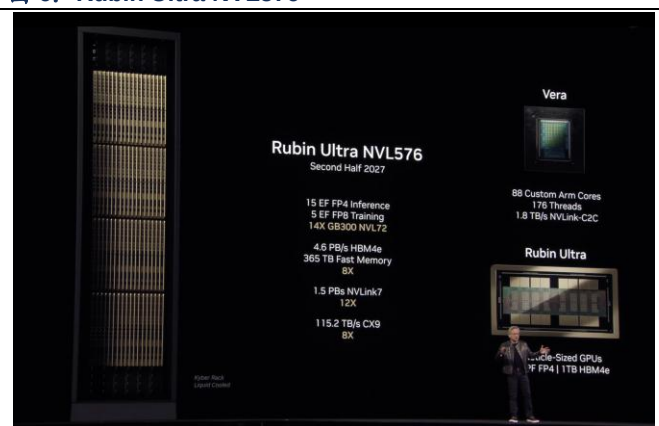
- **Blackwell Ultra 架构系列产品：**Blackwell Ultra 搭配 12 层堆叠的 HBM3e 内存，显存达到 288GB，可实现 1.8TB/s 的片间互联带宽。在此基础上，英伟达也将推出 Blackwell Ultra NVL72 机柜。GB300 NVL72 的 AI 性能是 NVIDIA GB200 NVL72 的 1.5 倍，这使得建造 AI 工厂的收入机会与基于 NVIDIA Hopper™ 构建的系统相比增加了 50 倍。合作伙伴预计将从 2025 年下半年开始推出基于 Blackwell Ultra 的产品。
- **Rubin 架构系列产品：**2026 年英伟达将推出 Rubin 架构，并将机柜的节点进一步扩展，推出 Vera Rubin NVL144 和 Vera Rubin NVL576。Rubin 将基于 TSMC 3nm 制程构造，提供令人难以置信的 50 PFLOP 密集 FP4 计算，是 B300 的三倍多。
- **CPO 交换机：**GTC 英伟达推出了 Quantum-X 硅光共封芯片、Spectrum-X 硅光共封芯片以及衍生出来的三款交换机产品：Quantum 3450-LD、Spectrum SN6810 和 Spectrum SN6800。NVIDIA 硅光交换机是全球领先的网络解决方案。它们创新地集成了光器件，减少了 4 倍的激光器数量，与传统方法相比，能源效率提高到 3.5 倍，信号完整性提高到 63 倍，大规模组网可靠性提高到 10 倍，部署速度提高到 1.3 倍。
- **产品节奏：**Blackwell Ultra 系列产品及 CX8 等将于 2025 年推出；Rubin/Rubin Ultra 及配套的 CX9 等产品将在 2026~2027 年推出。2028 年则将推出 Feynman 相关产品。

图 5: Blackwell Ultra NVL72



资料来源：英伟达官网，信达证券研发中心

图 6: Rubin Ultra NVL576



资料来源：英伟达官网，信达证券研发中心

图 7: Blackwell Ultra NVL144



资料来源：英伟达官网，信达证券研发中心

图 8: CPO 交换机



资料来源：英伟达官网，信达证券研发中心

图 9：英伟达产品对比

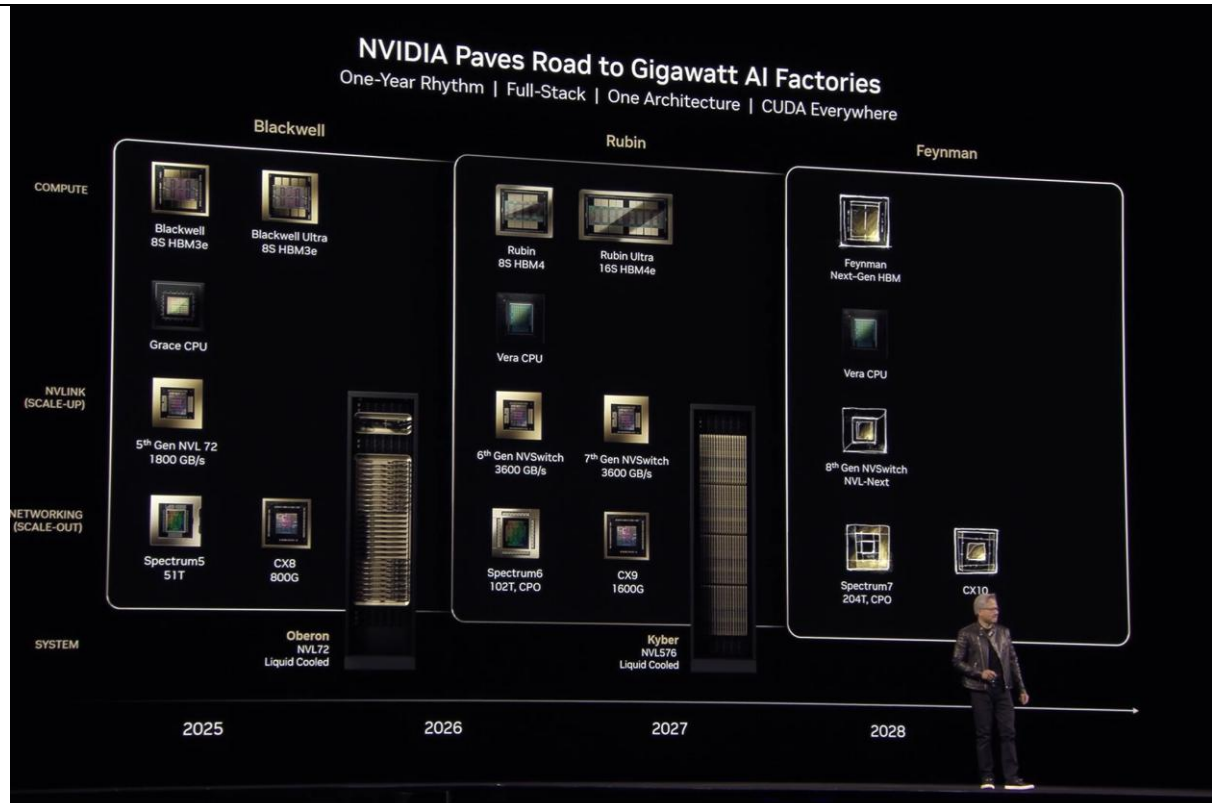
Nvidia Roadmap							
2022		2023	2024	2025		2026	2027
Chip and Package Level							
	Hopper		Blackwell			Rubin	
Accelerator	H100 (SXM)	H200	B200/ GB200	GB300 (Ultra)	B300 (single die, B300A)	VR200	VR300 (Ultra)
GPU TDP (W)	700	700	700/1200	1,400	600	1,800	3,600
Foundry Node	4N		4NP			N3P (3NP)	
Logic 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 comparision.

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

资料来源：Semianalysis，信达证券研发中心

图 10：GTC 大会公布的产品发布节奏

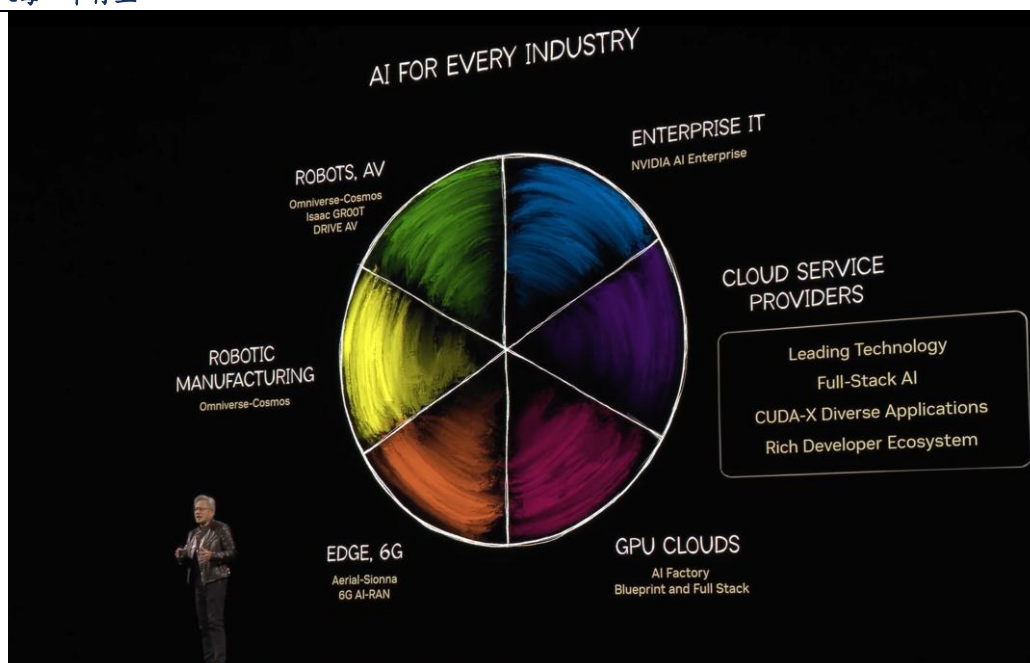


资料来源：英伟达官网，信达证券研发中心

软件层面：

- **Dynamo:** 这是一款开源推理软件，用于以最低的成本和最高的效率加速和扩展 AI 工厂中的 AI 推理模型。在 GB200 NVL72 机架的大型集群上运行 DeepSeek-R1 模型时，Dynamo 的智能推理优化还将每个 GPU 生成的 token 数量提高了 40 倍以上。
- **GROOT N1:** 英伟达宣布开源全球首个人形机器人基础模型 GROOT N1。GROOT N1 基础模型采用双系统架构，灵感来自人类认知原理。“系统 1”是一种快速思考的行动模型，反映了人类的反应或直觉。“系统 2”是一种慢速思考的模型，用于深思熟虑、有条不紊的决策。
- **Llama Nemotron:** 由 Llama 系列模型衍生而来，相较于 Llama 本体，这款模型仅有 48B。模型系列分为三档：入门级的 Nano、中端的 Super 和旗舰 Ultra，每一款都针对不同规模的企业需求。

图 11：AI 将进入每一个行业



资料来源：英伟达官网，信达证券研发中心

投资建议：我们认为，当前 AI 发展明显加速，推理成本的降低有望使得复杂的算力密集场景落地，从而带来推理算力需求的增长。目前人形机器人、自动驾驶等赛道已经看到明显的 AI 赋能，行业发展明显加速。其他 AI 端侧尤其是传统的消费电子终端如手机、PC 等有望被 AI 重新定义，并衍生出新型的 AI 终端，如智能管家、AI 眼镜、桌面机器人等等。今年各类端侧软件和硬件有望密集推出，接入用户量的提升则有望推动云端算力成长。建议持续关注相关优质个股。

表 1：消费电子涨跌 TOP5

板块	股票代码	股票简称	总市值（亿元）	净利润（亿元）			PE		
				2024E	2025E	2026E	2024E	2025E	2026E
AI 云侧	601138.SH	工业富联	4,140.51	241.78	307.60	356.79	17.13	13.46	11.60
	002463.SZ	沪电股份	663.84	25.61	34.14	41.32	25.92	19.44	16.06
	600183.SH	生益科技	686.05	18.16	24.53	30.45	37.78	27.97	22.53
	002916.SZ	深南电路	679.46	25.46	30.90	36.16	26.69	21.99	18.79
	300476.SZ	胜宏科技	730.70	11.73	25.05	32.71	62.31	29.17	22.34
	688256.SH	寒武纪	2,937.35	-4.58	0.68	5.23	-	-	-
	688041.SH	海光信息	3,525.09	29.54	41.84	59.29	119.32	84.25	59.45
AI 端侧	300433.SZ	蓝思科技	1,302.52	39.15	55.63	69.55	33.27	23.41	18.73
	002600.SZ	领益智造	673.49	20.11	29.04	37.61	33.49	23.19	17.91
	002938.SZ	鹏鼎控股	849.75	36.19	45.16	51.12	23.48	18.82	16.62
	002384.SZ	东山精密	587.52	18.05	28.53	36.28	32.55	20.59	16.19
	688018.SH	乐鑫科技	260.77	3.45	4.60	6.05	75.62	56.63	43.07
	603893.SH	瑞芯微	709.20	5.37	7.78	10.60	132.05	91.10	66.88
	688608.SH	恒玄科技	446.76	4.07	6.10	8.33	109.78	73.28	53.64
	300458.SZ	全志科技	363.25	2.19	3.59	5.02	165.57	101.13	72.30
	688521.SH	芯原股份	488.53	-5.22	0.12	1.32	-	#####	370.66
	603986.SH	兆易创新	852.65	11.19	16.67	21.51	76.19	51.14	39.64
	688099.SH	晶晨股份	347.29	8.20	11.16	14.26	42.34	31.11	24.35

资料来源:ifind，信达证券研发中心（表中盈利预测为 ifind 一致预期，截至 2025 年 3 月 20 日）

风险因素

- (1) 电子行业发展不及预期；
- (2) 宏观经济波动风险；
- (3) 地缘政治风险。

研究团队简介

莫文宇，电子行业分析师，S1500522090001。毕业于美国佛罗里达大学，电子工程硕士，2012-2022 年就职于长江证券研究所，2022 年入职信达证券研发中心，任副所长、电子行业首席分析师。

郭一江，电子行业研究员。本科兰州大学，研究生就读于北京大学化学专业。2020 年 8 月入职华创证券电子组，后于 2022 年 11 月加入信达证券电子组，研究方向为光学、消费电子、汽车电子等。

杨宇轩，电子行业分析师，华北电力大学本科，清华大学硕士，曾就职于东方证券、首创证券、赛迪智库，2025 年 1 月加入信达证券电子组，研究方向为半导体等。

王义夫，电子行业研究员。西南财经大学金融学士，复旦大学金融硕士，2023 年加入信达证券电子组，研究方向为存储芯片、模拟芯片等。

李星全，电子行业研究员。哈尔滨工业大学学士，北京大学硕士。2023 年加入信达证券电子组，研究方向为服务器、PCB、消费电子等。

分析师声明

负责本报告全部或部分内容的每一位分析师在此申明，本人具有证券投资咨询执业资格，并在中国证券业协会注册登记为证券分析师，以勤勉的职业态度，独立、客观地出具本报告；本报告所表述的所有观点准确反映了分析师本人的研究观点；本人薪酬的任何组成部分不曾与，不与，也将不会与本报告中的具体分析意见或观点直接或间接相关。

免责声明

信达证券股份有限公司（以下简称“信达证券”）具有中国证监会批复的证券投资咨询业务资格。本报告由信达证券制作并发布。

本报告是针对与信达证券签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本报告仅提供给上述特定客户，并不面向公众发布。信达证券不会因接收人收到本报告而视其为本公司的当然客户。客户应当认识到有关本报告的电话、短信、邮件提示仅为研究观点的简要沟通，对本报告的参考使用须以本报告的完整版本为准。

本报告是基于信达证券认为可靠的已公开信息编制，但信达证券不保证所载信息的准确性和完整性。本报告所载的意见、评估及预测仅为本报告最初出具日的观点和判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会出现不同程度的波动，涉及证券或投资标的的历史表现不应作为日后表现的保证。在不同时期，或因使用不同假设和标准，采用不同观点和分析方法，致使信达证券发出与本报告所载意见、评估及预测不一致的研究报告，对此信达证券可不发出特别通知。

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，也没有考虑到客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测仅供参考，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人做出邀请。

在法律允许的情况下，信达证券或其关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能会为这些公司正在提供或争取提供投资银行业务服务。

本报告版权仅为信达证券所有。未经信达证券书面同意，任何机构和个人不得以任何形式翻版、复制、发布、转发或引用本报告的任何部分。若信达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，信达证券对此等行为不承担任何责任。本报告同时不构成信达证券向发送本报告的机构之客户提供的投资建议。

如未经信达证券授权，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。信达证券将保留随时追究其法律责任的权利。

评级说明

投资建议的比较标准	股票投资评级	行业投资评级
本报告采用的基准指数：沪深 300 指数（以下简称基准）； 时间段：报告发布之日起 6 个月内。	买入： 股价相对强于基准 15% 以上；	看好： 行业指数超越基准；
	增持： 股价相对强于基准 5%~15%；	中性： 行业指数与基准基本持平；
	持有： 股价相对基准波动在 $\pm 5\%$ 之间；	看淡： 行业指数弱于基准。
	卖出： 股价相对弱于基准 5% 以下。	

风险提示

证券市场是一个风险无时不在的市场。投资者在进行证券交易时存在赢利的可能，也存在亏损的风险。建议投资者应当充分深入地了解证券市场蕴含的各项风险并谨慎行事。

本报告中所述证券不一定能在所有的国家和地区向所有类型的投资者销售，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专业顾问的意见。在任何情况下，信达证券不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者需自行承担风险。