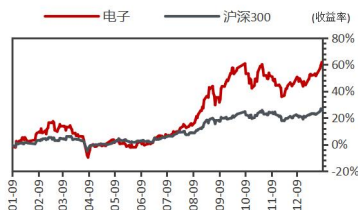


行业及产业

电子

强于大市

一年内行业指数与沪深 300 指数对比走势:



资料来源: 聚源数据, 爱建证券研究所

相关研究

《电子行业周报: 国产 DRAM 存储龙头 IPO 成功受理》2026-01-06

《爱建电子深度报告: 卫星通信加速进入发展元年》2026-01-05

《电子行业周报: AI Glasses 发展或迎来苹果动能》2025-12-30

《爱建电子深度报告: 半导体产业的发展复盘与方向探索》2025-12-26

《爱建电子专题报告: 存储芯片涨价将延续至 2026 年》2025-12-22

证券分析师

许亮
S0820525010002
0755-83562506
xuliang@ajzq.com

联系人

朱俊宇
S0820125040021
021-32229888-25520
zhujunyu@ajzq.com

CES 2026 AI 前沿信息汇总

——人工智能月度跟踪

投资要点:

- **事件:** 2026 年 1 月 6 日 CES 展会上, NVIDIA CEO 黄仁勋展示了 Vera Rubin AI 超级计算机平台, 并宣布该产品即将量产; AMD 同期推出了基于 MI 455X GPU 的 Helios 机柜级 AI 计算方案; Intel 则发布了其首款基于 18A 制程节点打造的计算平台。
- **作为全球知名芯片设计公司, NVIDIA 长期专注高性能 GPU 及系统芯片研发。**在架构早期阶段, 其通过 2010 年前的 Tesla、2010 年的 Fermi、2012 年的 Kepler 及 2014 年的 Maxwell 四大系列, 推动 GPU 从“图形加速专用硬件”成功转型为“通用并行计算引擎”, 期间持续优化可靠性、能效比与场景适配能力; 2016 年后, 伴随 AI 与高性能计算需求爆发, NVIDIA 加快迭代节奏, 先后推出 2016 年的 Volta (首推 Tensor Core 开启 AI 算力硬件加速)、2022 年的 Hopper (以 Transformer Engine 支撑大模型研发)、2024 年的 Blackwell (全栈协同设计提升 AI 推理与图形渲染性能) 等架构, 并计划通过下一代 Rubin、Feynman 架构突破算力密度与能效比极限, 服务超复杂场景。
- **NVIDIA 推出新一代 Vera Rubin AI 超级计算机平台, 该平台涵盖 Vera CPU、Rubin GPU 等 6 款新型芯片。**凭借跨芯片极致协同设计, 其不仅可将单 Token 推理成本降低 1/10, 还能把混合专家 (MoE) 模型训练所需的 GPU 数量缩减至 1/4。该平台的核心为 Vera Rubin Superchip, 其通过 NVLink-C2C 互联技术, 无缝整合两颗 Rubin GPU 与一颗 Vera CPU, 进而构建起统一的机架级计算执行域。1) Vera CPU 专为智能体推理打造, 在核心数量、内存带宽及算力等关键维度大幅超越 Grace CPU; 2) Rubin GPU 搭载第三代 Transformer 引擎, 在运算能力、内存带宽等核心指标上实现显著跃升。NVIDIA 宣布, Vera Rubin 2026 年进入全面量产阶段, AWS、Google Cloud、OCI 将同步部署; 同时, CoreWeave、Lambda、Nebius、Nscale 等 NVIDIA 云生态合作伙伴也将跟进部署。
- **AMD 与 Intel 分别在 CES 2026 推出算力新品。**AMD 发布搭载 2nm EPYC Venice Zen6 CPU 与 Instinct MI455X GPU 的 Helios AI 机架平台。该平台采用全液冷设计, 基础单元集成 DPU 与 AI 网卡, 完整机架可提供 2.9 Exaflops AI 算力等强劲性能。MI455X GPU 性能较前代提升 10 倍, EPYC Venice Zen6 CPU 性能与效率提升超 70%; Intel 推出基于 18A 工艺的 Intel Core™Ultra Series 3 Processors (开发代号 Panther Lake), 作为首款 AI 个人电脑平台, 该系列含 X9、X7 型号, 旗舰款最高配备 16 核 CPU、12 核 Xe 显卡及 50 TOPS NPU 算力。
- **投资建议:** NVIDIA 最新 Vera Rubin 平台的发布标志着高算力 GPU、CPU 和高速连接芯片仍然是未来 AI 平台的发展方向。建议关注国产高算力 GPU、CPU 和连接芯片的投资机会, 以及国内已经打入 NVIDIA 供应链的 PCB 等零部件供应商的投资机会。
- **风险提示:** 1) 国际贸易摩擦加剧 2) 下游需求不及预期 3) 技术升级进度滞后

目录

事件：CES 2026 于拉斯维加斯召开	4
1. NVIDIA 发布 Vera Rubin	4
1.1 NVIDIA 从 Tesla、Fermi 向 Vera Rubin 迭代升级	4
1.2 NVIDIA Vera Rubin 性能卓越	5
1.3 NVIDIA 宣布 Vera Rubin 2026 年进入全面量产阶段	7
2. AMD 发布“Helios”机架级平台	7
3. Intel 发布 Core Ultra Series 3 Processors	8
4. 风险提示	9

图表目录

图表 1 : NVIDIA 早期 GPU 技术蓝图	4
图表 2 : NVIDIA 2016-2026 年 GPU 技术路线持续迭代	5
图表 3 : NVIDIA Rubin 平台包含 6 款新型芯片	5
图表 4 : NVIDIA Vera Rubin Superchip 示意图	6
图表 5 : Vera CPU 较 Grace CPU 性能提升显著	6
图表 6 : NVIDIA Rubin GPU 示意图	7
图表 7 : AMD“Helios”搭载了 EPYC Venice Zen6CPU 与 Instinct MI455X GPU	8
图表 8 : MI445X GPU 及 EPYC Venice Zen 6 CPU 示意图	8
图表 9 : Intel Core™Ultra Series 3 Processors 示意图	9

事件：CES 2026 于拉斯维加斯召开

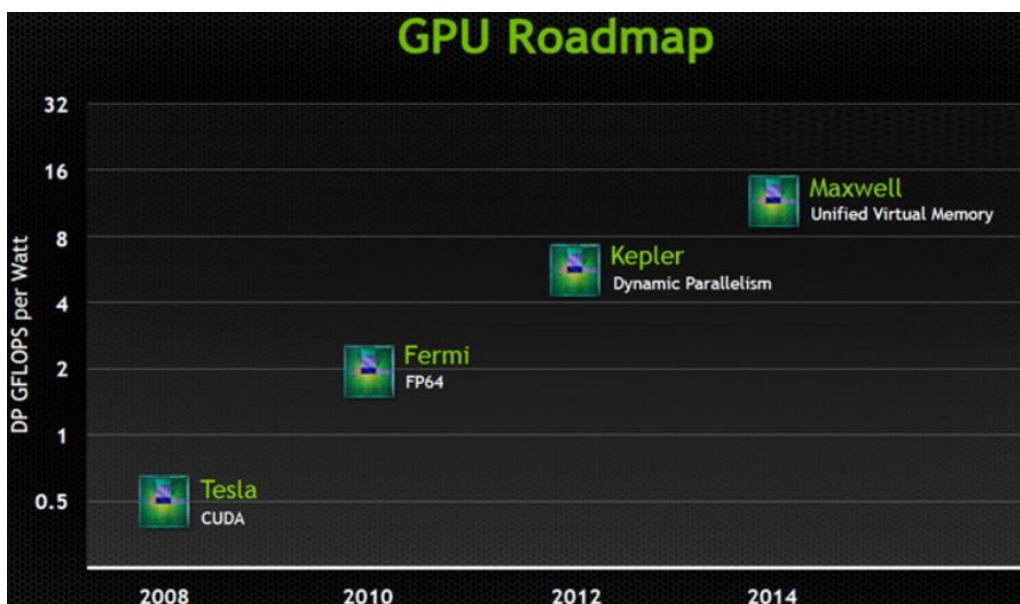
事件：2026 年 1 月 6 日 CES 拉斯维加斯展会上，NVIDIA CEO 黄仁勋展示了 Vera Rubin AI 超级计算机平台，并宣布该产品 2026 年进行量产；AMD 同期推出了基于 MI 455X GPU 的 Helios 机架级 AI 计算方案；Intel 则发布了其首款基于 18A 制程节点打造的计算平台。

1. NVIDIA 发布 Vera Rubin

1.1 NVIDIA 从 Tesla、Fermi 向 Vera Rubin 迭代升级

在 GPU 架构的早期演进中，NVIDIA 通过 Tesla、Fermi、Kepler、Maxwell 四大系列，实现了 GPU 从“图形加速专用硬件”到“通用并行计算引擎”的关键技术跃迁。2010 年前，NVIDIA 推出的 Tesla 架构，正式开启了 GPU 从图形加速向通用计算的跨越式转型；2010 年，Fermi 架构聚焦可靠性与通用性升级，不仅首次引入 ECC 错误校验码内存以保障数据计算准确性，还对 CUDA 核心进行优化，使其能支持更多编程语言；2012 年，Kepler 架构以“能效比革命”为核心，推出 SMX 流式多处理器以提升并行效率，同时首次支持 GPUDirect 技术，实现了 GPU 间及 GPU 与存储设备的直接数据传输；到了 2014 年，NVIDIA 又推出 Maxwell 架构，该架构采用台积电 28nm 工艺制程。彼时移动设备兴起，市场对低功耗、高性能 GPU 的需求大增，NVIDIA 也针对性优化了该架构在不同应用场景的适配能力。

图表 1：NVIDIA 早期 GPU 技术蓝图



资料来源：EXR Review，爱建证券研究所

随着人工智能与高性能计算需求的持续爆发，NVIDIA 自 2016 年后进一步加快架构迭代节奏，先后推出 Volta、Hopper、Blackwell 等一系列突破性架构，并计划通过下一代架构持续拓展算力边界。2016 年，Volta 架构凭借首创的 TensorCore 技术，开启 AI 算力硬件加速的商业化时代，为深度学习大规模落地提供关键算力支撑；2022 年，Hopper 架构聚焦大模型需求，以 Transformer Engine 夯实千亿参数模型研发

基础；2024 年发布的 Blackwell 架构，则通过“芯片-系统-软件”全栈协同设计，在 AI 推理效率与图形渲染性能上实现双重跃升。

图表 2：NVIDIA 2016-2026 年 GPU 技术路线持续迭代



资料来源：WCCFTECH，爱建证券研究所

而 NVIDIA 推出的下一代架构 Vera Rubin，进一步突破算力密度与能效比的极限，为超大规模 AI 集群、量子计算协同模拟等复杂场景提供底层技术支撑。

1.2 NVIDIA Vera Rubin 性能卓越

新一代 Vera Rubin AI 超级计算机平台包含 6 款新型芯片：Vera CPU、Rubin GPU、NVIDIA NVLink™ 6 交换器、NVIDIA ConnectX®-9 超级网卡、NVIDIA BlueField®-4 DPU 及 NVIDIA Spectrum™-6 以太网交换器。通过跨芯片的极致协同设计，Rubin 平台可将每个 Token 的推理成本降低 1/10，训练混合专家（MoE）模型所需的 GPU 数量减少至 1/4。

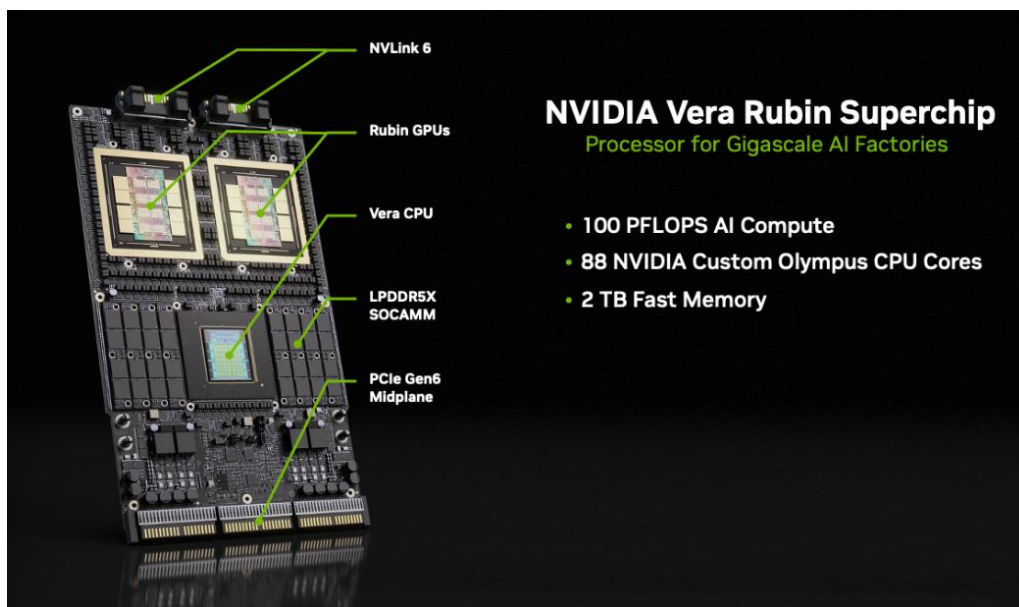
图表 3：NVIDIA Rubin 平台包含 6 款新型芯片



资料来源：NVIDIA DEVELOPER，爱建证券研究所

Rubin 平台的核心是 NVIDIA Vera Rubin Superchip，其 CPU 与 GPU 的性能跃升尤为值得关注。每颗超级芯片借助支持内存一致性的 NVLink-C2C 互联技术，无缝整合两颗 Rubin GPU 与一颗 Vera CPU，彻底打破传统 CPU 与 GPU 的架构壁垒，进而构建起统一的机架级计算执行域。

图表 4: NVIDIA Vera Rubin Superchip 示意图



资料来源: NVIDIA DEVELOPER, 爱建证券研究所

1) Vera Rubin CPU 专为智能体推理而设计, 是面向大规模 AI 工厂的能效最高的 CPU。这款芯片内置 88 个核心、176 条线程, 拥有 1.8 TB/s 的 NVLink-C2C 一致性内存互连、1.5TB 系统内存(相当于 Grace 的 3 倍)、1.2TB/s 的 SOCAMM LPDDR5X 内存带宽, 并支持机架级机密计算。整体算力、压缩和 CI/CD 性能均比 Grace 提升 2 倍。

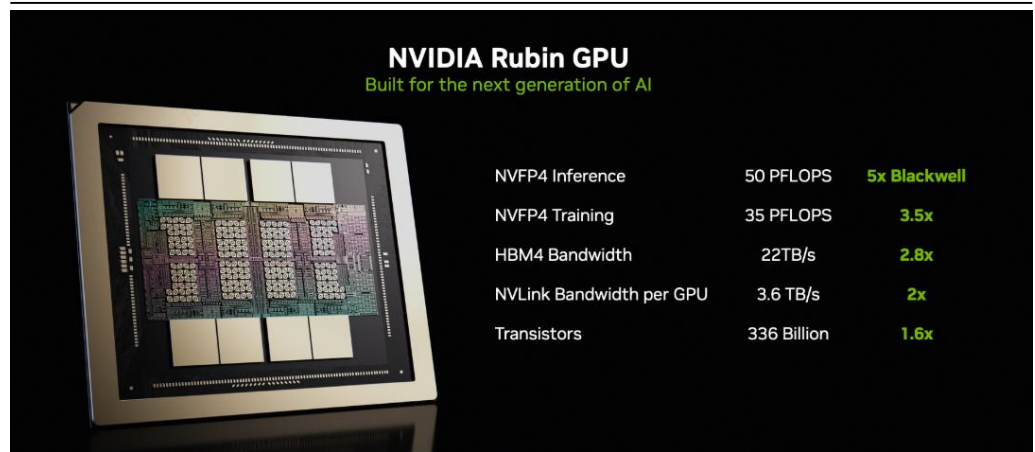
图表 5: Vera CPU 较 Grace CPU 性能提升显著

	Grace CPU	Vera CPU
核心	72 Neoverse V2 cores	88 NVIDIA Custom Olympus cores
线程	72	176 Spatial Multithreading
L2 Cache per core	1MB	2MB
Unified L3 Cache	114MB	162MB
内存带宽 (BW)	Up to 512GB/s	Up to 1.2TB/s
内存容量	Up to 480GB LPDDR5X	Up to 1.5TB LPDDR5X
SIMD	4x 128b SVE2	6x 128b SVE2 FP8
NVLINK-C2C	900GB/s	1.8TB/s
PCIe/CXL	Gen5	Gen6/CXL 3.1
机密计算	NA	支持

资料来源: NVIDIA DEVELOPER, 爱建证券研究所

2) Rubin GPU 采用第三代 Transformer 引擎与硬件加速自适应压缩技术, 可提供 50 petaflops 的 NVFP4 运算能力。其 HBM4 内存带宽提升至 22TB/s (相较前代提升 2.8X), 单 GPU 的 NVLink 互连带宽提升至 3.6TB/s (相较前代提升 2X)。

图表 6: NVIDIA Rubin GPU 示意图



资料来源: NVIDIA DEVELOPER, 爱建证券研究所

系统架构层面, Rubin 平台实现 100%全液冷覆盖, 采用第三代完全无缆化设计重构硬件形态, 旗舰产品 Vera Rubin NVL72 机架式系统采用无缆化、无风扇模块化设计, 构建“精准温控+高效换热”的整机散热生态。

1.3 NVIDIA 宣布 Vera Rubin 2026 年进入全面量产阶段

NVIDIA 宣布, Vera Rubin 于 2026 年正式进入全面量产阶段。

2026 年首批部署 Vera Rubin 的云厂商包括 AWS、Google Cloud、OCI 等; 同时, 英伟达的云生态合作伙伴 CoreWeave、Lambda、Nebius 与 Nscale 也将同步跟随部署。

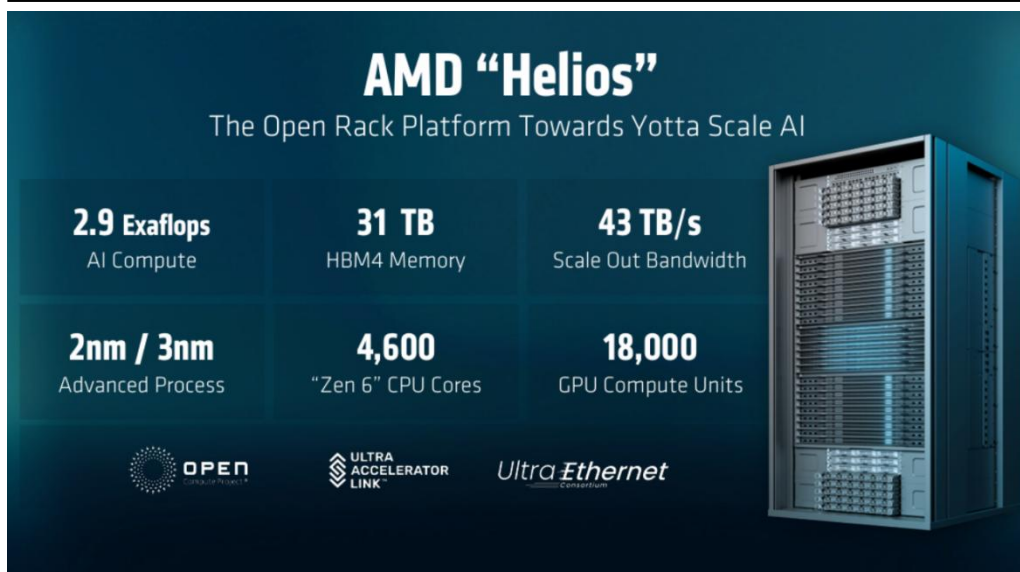
- 1) 微软将部署英伟达 Vera Rubin NVL72 机架级系统, 将其纳入新一代 AI 数据中心的核心基础设施, 并应用于未来的 Fairwater AI 超级工厂等场地。
- 2) CoreWeave 计划于 2026H2, 将英伟达 Rubin 系统全面整合至自有 AI 云平台。该平台具备多架构并行运行能力, 可支持客户将 Rubin 技术接入自有业务环境, 最大化释放训练、推理及代理型工作负载的效能。

2. AMD 发布“Helios”机架级平台

CES 2026 大会上, AMD 发布了搭载 2nm EPYC Venice Zen6CPU 与 Instinct MI455X GPU 的 Helios AI 机架平台。

该机架采用全液冷设计, 基础单元配置为 1 颗 EPYC Venice Zen6 CPU 搭配 4 张 Instinct MI455X GPU, 同时集成 Pensando “Salina” 400 DPU 与 “Vulcano” 800 AI 网卡, 分别承担数据处理与高速互联任务; 扩展至完整机架形态后, 可提供 2.9 Exaflops AI 算力、31TB HBM4 内存与 43TB/s 扩展带宽, 最多集成 4600 个 CPU 核心与 18000 个 GPU 核心。

图表 7: AMD “Helios” 搭载了 EPYC Venice Zen 6 CPU 与 Instinct MI455X GPU

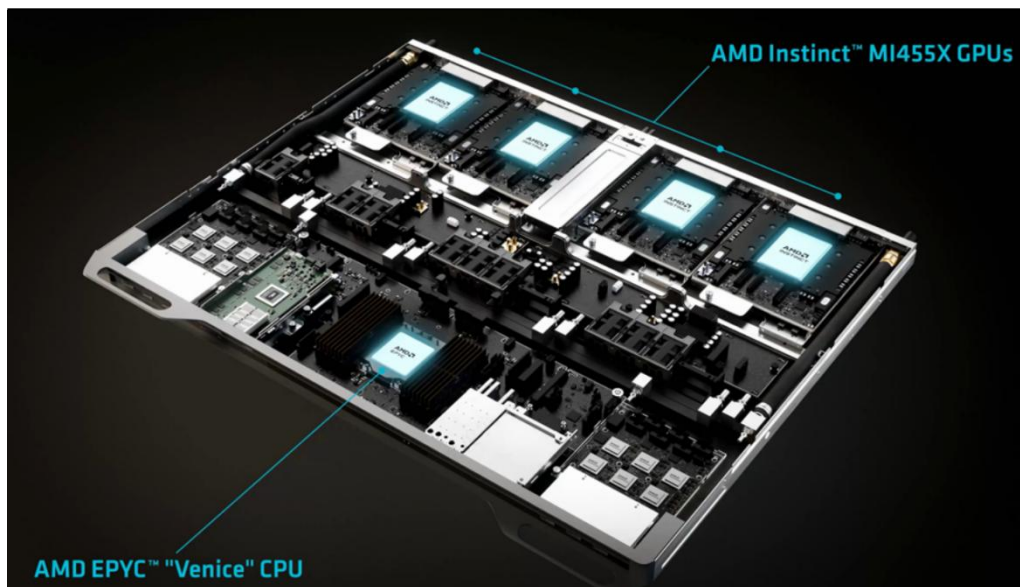


资料来源: AMD, 爱建证券研究所

同时, 苏姿丰展示了 MI445X GPU 与 EPYC Venice Zen 6 CPU。MI455X GPU 配备了两个巨大的 GCD (图形计算芯片)、两个 MCD (内存控制器芯片) 以及总共 16 个 HBM4 显存位点。MI455X 的性能比 MI355X 提高了 10 倍。

EPYC Venice Zen 6 CPU 配备 8 个大型 Zen 6C CCD (核心计算芯片)、两个 IOD (输入输出芯片) 及带管理控制器的微型芯片。AMD 称其性能与效率提升超 70%, 线程密度提升超 30%。该芯片还将推出标准的 192 核 Zen 6 版本, 配备 16 个 CCD, 每个 CCD 包含 12 个 Zen 6 核心, 以及 768MB 的 L3 缓存。

图表 8: MI445X GPU 及 EPYC Venice Zen 6 CPU 示意图



资料来源: AMD, 爱建证券研究所

3. Intel 发布 Core Ultra Series 3 Processors

在 CES 2026 上, Intel 发布了 Intel Core™ Ultra Series 3 Processors (开发代号

Panther Lake)，标志着首个基于英特尔 18A 工艺技术的 AI 个人电脑平台的首次亮相。

该系列含全新命名的 X9、X7 型号，集成高端锐炫显卡，旗舰款最高配备 16 核 CPU、12 核 Xe 显卡及 50 TOPS NPU 算力。其中，酷睿 Ultra X9 388H 在相近功耗下，多线程性能较 Lunar Lake 提升 60%，游戏性能提升 77%，电池续航最长达 27 小时。

图表 9：Intel Core™ Ultra Series 3 Processors 示意图



资料来源：Intel，爱建证券研究所

首批搭载 Intel Core™ Ultra Series 3 Processors 的消费级笔记本电脑，将于 2026 年 1 月 6 日启动预售，并自 2026 年 1 月 27 日起在全球范围内正式面市。

4. 风险提示

- 1) 国际贸易摩擦加剧
- 2) 下游需求不及预期
- 3) 技术升级进度滞后

爱建证券有限责任公司

上海市浦东新区前滩大道 199 弄 5 号

电话: 021-32229888

传真: 021-68728700

服务热线: 956021

邮政编码: 200124

邮箱: ajzq@ajzq.com

网址: <http://www.ajzq.com>

评级说明

投资建议的评级标准

报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 个月内的相对市场表现，也即以报告发布日后的 6 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场：沪深 300 指数（000300.SH）；新三板市场：三板成指（899001.CSI）（针对协议转让标的）或三板做市指数（899002.CSI）（针对做市转让标的）；北交所市场：北证 50 指数（899050.BJ）；香港市场：恒生指数（HIS.HI）；美国市场：标普 500 指数（SPX.GI）或纳斯达克指数（IXIC.GI）。

股票评级

买入	相对同期相关证券市场代表性指数涨幅大于 15%
增持	相对同期相关证券市场代表性指数涨幅在 5% ~ 15% 之间
持有	相对同期相关证券市场代表性指数涨幅在 -5% ~ 5% 之间
卖出	相对同期相关证券市场代表性指数涨幅小于 -5%

行业评级

强于大市	相对表现优于同期相关证券市场代表性指数
中性	相对表现与同期相关证券市场代表性指数持平
弱于大市	相对表现弱于同期相关证券市场代表性指数

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告采用信息和数据来自公开、合规渠道，所表述的观点均准确地反映了我们对标的证券和发行人的独立看法。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法可能存在局限性，请谨慎参考。

法律主体声明

本报告由爱建证券有限责任公司（以下统称为“爱建证券”）证券研究所制作，爱建证券具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管。

本报告是机密的，仅供我们的签约客户使用，爱建证券不因收件人收到本报告而视其为爱建证券的签约客户。本报告中的信息均来源于我们认为可靠的已公开资料，但爱建证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供签约客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见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。客户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、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，爱建证券及其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测后续可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，爱建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。

版权声明

本报告版权归爱建证券所有，未经爱建证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、转载、刊登和引用。否则由此造成的一切不良后果及法律责任由私自翻版、复制、转载、刊登和引用者承担。版权所有，违者必究。