



AI 周观察

数据专题研究

证券研究报告

分析师：刘道明（执业 S1130520020004）
 liudaoming@gjzq.com.cn

分析师：黄晓军（执业 S1130525080005）
 huangxiaojun@gjzq.com.cn

联系人：麦世学（执业 S1130123100111）
 maishixue@gjzq.com.cn

英特尔 18A 正式量产

摘要

- 本周 AI 聊天应用市场活跃度分化，海外 Gemini 下降而 Perplexity 与 Claude 上升，国内应用则因国庆假期普遍下滑。模型应用方面，OpenAI 正式上线迄今最强大的 GPT-5Pro API，凭借 40 万 Token 超大上下文窗口和多模态能力，瞄准科研、法律等高端市场，定价为每百万 Token15 美元。谷歌预览了能像人一样操作网页的 AI 代理模型 Gemini2.5 Computer Use，并升级视频模型 Veo 至 3.1 版本。此外，谷歌宣布其 AI 月处理代币数超 1300 万亿，主要反映了新模型导致后端计算复杂度与成本激增，而非用户活跃度的直接体现。
- Semianalysis 发布 InferenceMAX 推理基准框架，以“TCO 每百万 token”和“每 MW 产出”衡量 AI 推理系统经济性与能效，为 AMD 与 NVIDIA 竞争提供统一评估语境。结果显示，NVIDIA 在低至中交互性区间凭借 GB200/B200 与封闭软件栈保持优势，而 AMD 在中高交互性与 vLLM+MX4 组合下能效逼近、ROI 具竞争力。InferenceMAX 使比较从峰值性能转向单位经济性，揭示 AI 推理竞争将围绕 TCO 与能效展开。
- 英特尔 18A 节点正式量产，标志公司迈入 2nm 时代并实现 RibbonFET 与 PowerVia 双技术落地，提升能效但制造成本与复杂度同步上升。相较台积电 N2，18A 在性能与能效上具潜在领先，但晶体管密度与量产良率仍待验证。首批产能来自亚利桑那 Fab 52，出货时间延至 2025 年底。该节点是英特尔重建先进制程信誉的关键一步，但目前仍处“验证期”，能否成为真正领先节点取决于后续 Clearwater Forest 等产品表现。
- 2025 年 9 月，台湾存储产业显著回温。群联电子营收 65.15 亿元新台币，PCIe SSD 控制芯片出货量同比激增近 300%，受 AI 推理与企业级 SSD 需求带动。华邦电子第三季营收 217.7 亿元新台币，创 12 季高点，利基型 DRAM 合约价上扬与 DDR3 停产推升 DDR4 价量齐升。两家公司分别受益于 AI 存储需求与 DRAM 价格周期回升，显示台湾存储链在 AI 驱动下进入结构性复苏阶段。
- 2025 年 8 月，中国智能手机销量达到约 2080 万台，同比下滑 6.24%。小米、华为、vivo、OPPO、荣耀以 16.21%、15.08%、13.82%、13.79%、13.52% 的份额排名前五。具体型号方面，iPhone 16 Pro 销量排名第一，OPPO reno 14、iPhone 16 Pro Max 排名第二、第三。8 月国内 PC 市场销量基本同比持平，台式机销量约为 157 万台，同比下滑约 1.39%。笔电销量约为 232 万台，同比增长约 2.02%。

风险提示

芯片制程发展与良率不及预期

中美科技领域政策恶化

智能手机销量不及预期



内容目录

海外市场行情回顾.....	3
国内 AI 应用活跃度受假期影响下降，Google 发布 Computer Use.....	4
AMD vs. NVIDIA：在 InferenceMAX 框架下的性能与投资结论.....	5
英特尔 18A 节点点评：突破量产门槛，但领先仍在验证期.....	5
AI 需求驱动存储链回暖：群联控制器出货创新高，华邦 DRAM 价量齐升.....	6
消费电子动态.....	7
8 月国内智能手机市场销量同比下滑.....	7
8 月 PC 市场同比基本持平.....	8
风险提示.....	9



海外市场行情回顾

图表1: 截至 10 月 10 日海外 AI 相关个股行情

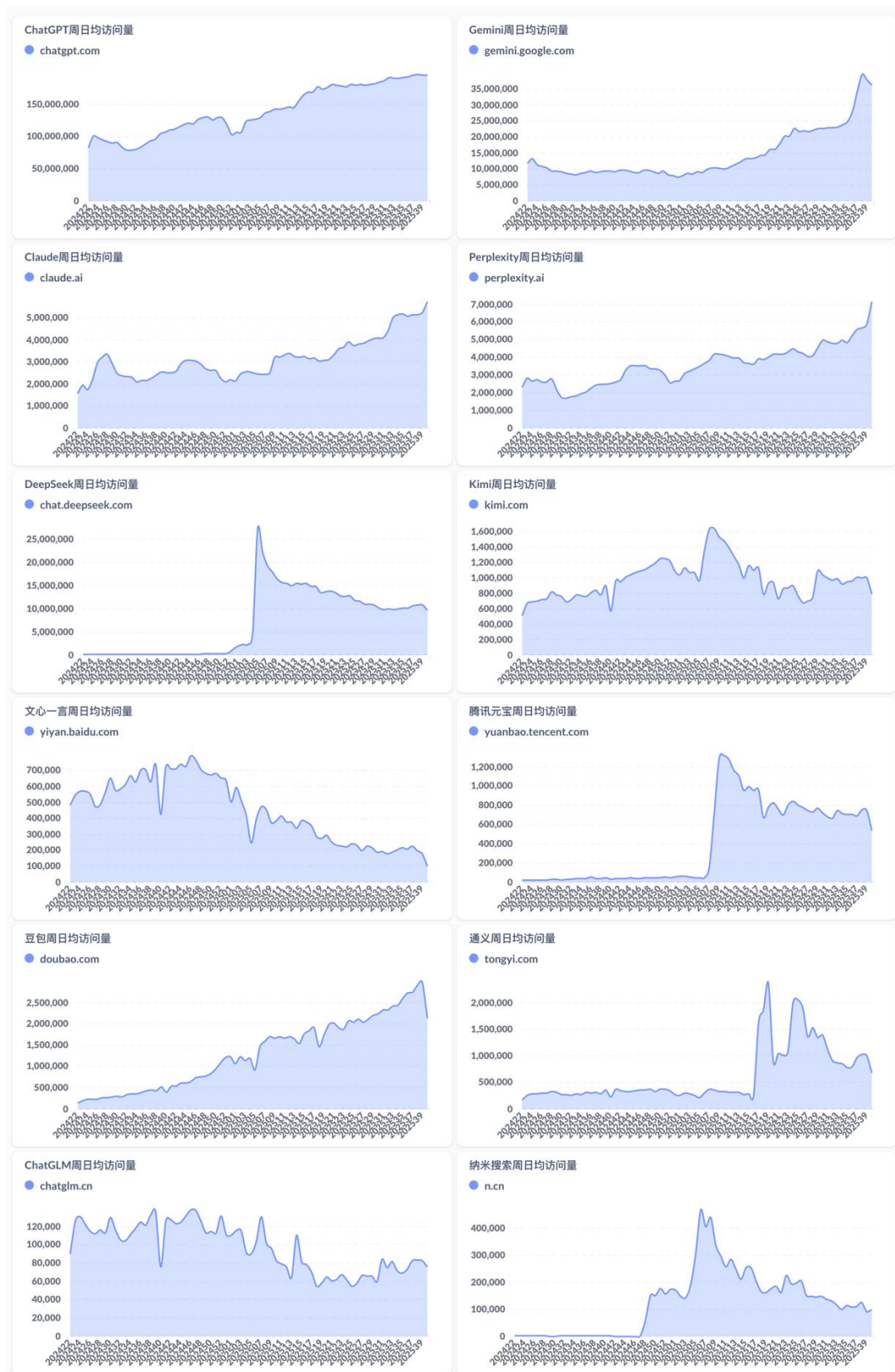
个股名称	本周收盘价	周涨跌幅	产业链
超威半导体	214.9	30.5%	芯片设计
戴尔	150.57	6.98%	消费电子产品、服务器、交换机、网络安全
DataDog	158.74	4.56%	芯片设计
Snowflake	242.17	3.01%	云服务、软件
Zscaler	309.88	1.46%	网络安全
Palantir	175.44	1.37%	云服务
CrowdStrike	493.66	0.77%	网络安全
Palo Alto Networks	208.55	0.66%	网络安全
Salesforce	241.68	0.55%	云服务
美满电子	85.61	-0.71%	芯片设计与制造、交换机
Meta	705.3	-0.74%	消费电子产品、大模型、广告
Innodata	83.28	-1.02%	云服务
微软	510.96	-1.24%	消费电子产品、云服务、软件
英特尔	36.37	-1.25%	芯片设计与制造
Cloudflare	213.82	-1.32%	网络安全
亚马逊	216.37	-1.43%	IaaS、电商平台、自营硬件、物流配送、广告变现
Mongodb	314.44	-2.11%	云服务
英伟达	183.16	-2.38%	芯片设计、交换机
谷歌	236.57	-3.58%	云服务、大模型、广告
特斯拉	413.49	-3.8%	整车、大模型
博通	324.63	-4.06%	芯片设计与制造、交换机
HP	25.55	-4.09%	消费电子产品
Zeta Global	18.4	-4.81%	软件
Gitlab	44.61	-4.9%	云服务
苹果	245.27	-4.94%	消费电子产品
C3.AI	17.91	-6.52%	云服务
高通	153.59	-9.22%	芯片设计
Skyworks	69.16	-10.3%	芯片设计与制造
Qorvo	83.24	-10.95%	芯片设计与制造

来源: Reuters、国金证券研究所



国内 AI 应用活跃度受假期影响下降，Google 发布 Computer Use

图表2：聊天助手类 AI 应用活跃度



来源：Similarweb、国金证券研究所

本周，海外聊天助手类应用活跃度中，Gemini 周均环比下降，ChatGPT 保持平稳，Perplexity 和 Claude 活跃度均有上升。国内应用中，受国庆假期影响，多数 AI 聊天类



应用活跃度下降。

OpenAI 正式上线其最新人工智能模型 GPT-5Pro API，作为其迄今最强大的模型向开发者全面开放。GPT-5Pro 拥有高达 40 万 Token 的上下文窗口，能处理更复杂的文本内容，适用于科研写作、法律分析等高要求任务，并支持文本与图像输入。该模型定价为每百万 Token 15 美元，虽比基础模型贵 15 倍，但反映了其在高端应用场景的价值。

谷歌预览了全新 Gemini AI 模型——Gemini2.5Computer Use，旨在赋予 AI 代理通过浏览器导航和交互网页的能力。该模型利用其视觉理解和推理能力，可像人类一样执行填写表单等复杂操作，主要应用于 UI 测试及为无 API 用户导航网页界面。此举正值 AI 代理功能竞争白热化，此前 OpenAI 发布了新的 ChatGPT 应用并持续关注 Agent 功能，Anthropic 也已发布带有“计算机使用”功能的 Claude 模型；谷歌发布 Veo 3.1 版本，为视频创作者带来视频时长增加、角色一致性增强、输出质量提升及新增的“多镜头多提示”功能，显著提升创作效率。此外，谷歌 CEO 桑达尔·皮查伊宣布，谷歌 AI 产品月处理代币数已超 1300 万亿。分析指出，该数字主要反映 AI 模型后端计算复杂度的上升，而非用户使用量的直接体现。代币消耗激增或由 Gemini2.5Flash 等新型推理模型驱动，其单次请求代币量为前代 17 倍，导致推理成本高出 150 倍，多模态功能亦计入总量。因此，该数字被视为衡量谷歌后端计算负载与基础设施扩展的指标。

AMD vs. NVIDIA : 在 InferenceMAX 框架下的性能与投资结论

10 月 10 日，Semianalysis 发布其 InferenceMAX 推理性能评估框架，InferenceMAX 是一个开源、自动化、按日更新的推理性能基准，核心在于以真实工作负载和可复现方法量化推理系统在“吞吐—交互性”之间的权衡。该框架以“TCO 每百万 token”与“每 MW 产出 (tokens per MW)”为核心指标，从经济性和能效两维刻画不同软硬件组合的性价比，并通过帕累托前沿展示在当前条件下的最优折中点。这一框架为比较不同 GPU 与推理栈提供了统一语境，使结论不依赖单点跑分，而锚定在实际交互区间与应用经济性。

在 InferenceMAX 的全域比较下，AMD 与 NVIDIA 并非单边压制关系。其优劣取决于模型类型、精度设定 (FP4/FP8/MX4)、推理引擎 (vLLM/SGLang/TRT-LLM) 以及目标交互性区间。在低至中交互性区间（如聊天类、agent 类服务），NVIDIA 通过 GB200/B200 及其封闭软件栈 (TRT-LLM、NVLink 并行策略) 展现出更强的机架级性能与单位经济性。其硬件与软件协同优化仍在快速演进，帕累托前沿持续下移，意味着其在单位 token 成本上的领先地位尚未固化，但总体仍保持优势。

在部分 vLLM+MX4/FP4 组合与特定中高交互性区间（如文档摘要、批量推理任务）中，AMD 在 TCO 归一化后具备显著竞争力。新一代 CDNA4 架构在能效提升上表现突出，使其在“每 MW 产出”维度上已逼近 NVIDIA 同代产品。虽然 Blackwell 在绝对性能上仍领先，但若考虑 TDP、整机定价与部署规模，AMD 在部分场景下的 ROI 接近或超越 NVIDIA。

正确的硬件选型与估值应以目标交互性为锚点，在该区间下比较经济性 (TCO/每百万 token) 与能效性 (tokens per MW)，而非单纯以峰值吞吐或极端并发结果做结论。InferenceMAX 通过夜间自动跑分，持续反映软件栈 (vLLM、SGLang、TRT-LLM) 在内核与并行策略上的迭代优化，使帕累托前沿动态更新、竞争格局持续重排。

短期内，性能公告需在 InferenceMAX 框架内解读，关注软件栈优化对 TCO 和交互性区间的边际影响。中期看，单位经济性与能效 (tokens per MW) 改善速度将决定盈利弹性，AMD 若能维持能效追赶并在生态上获得主流推理引擎支持，其竞争力将显著增强。长期来看，生态与机架级方案成熟度成为核心分水岭。NVIDIA 仍具先发与生态垄断优势，但 AMD 若通过开放生态与更优\$/TCO 路径获得规模化部署，将重塑 AI 推理市场的利润分布。

InferenceMAX 揭示了 AI 推理硬件竞争的真正衡量维度——不是峰值性能，而是单位 token 经济性与能效。NVIDIA 在机架级生态与分布式性能上领先，但 AMD 正凭借更优能效与开放栈快速追近。未来 AI 硬件竞争将围绕交互性区间的 TCO 而非单卡 TFLOPS 展开。

英特尔 18A 节点点评：突破量产门槛，但领先仍在验证期

英特尔宣布其 Core Ultra 3 “Panther Lake” 处理器已在 18A 节点上进入量产阶段，正式迈入 2nm 级制程时代。这一节点被视为英特尔重塑制造信誉与技术领导力的关键里程碑，既是公司内部 CPU 架构与制程技术能力的集中展示，也承载着吸引外部代工客户的战略目标。



图表3: 18A 同时采用 RibbonFET 和 PowerVia



来源：英特尔、国金证券研究所

图表4: 18A D0 缺陷率持续下降



来源：英特尔、国金证券研究所

在技术架构上，18A 是英特尔首个在量产工艺中同时实现 RibbonFET（GAA）与 PowerVia 背面供电的节点。与传统前侧供电方式相比，背面供电可释放更多前侧空间用于逻辑互连，减少电阻与功耗，提升整体能效表现。这一并行创新体现出英特尔在制程结构上的进取姿态，但同时也显著提高了制造复杂度与成本，使其更适用于高端计算与 AI 加速等高毛利应用场景。

与台积电 N2 相比，18A 在性能与能效上有望取得领先，但在晶体管密度上仍略显不足。由于 PowerVia 将供电网络移至背面，英特尔的有效逻辑面积利用率更高，理论上可在同等面积下实现接近的计算密度。不过，实际晶圆成本更高，且能否在规模量产中保持良率与一致性仍待观察。

在产能布局方面，18A 首批量产来自亚利桑那州 Fab 52，而 Fab 62 仍在建设中。Panther Lake 的正式出货时间被推迟至 2025 年底至 2026 年初，反映出工艺爬坡速度慢于预期。虽然英特尔的缺陷密度指标持续改善，但性能与功耗良率的达标仍有不确定性。

从战略层面看，18A 的意义在于验证英特尔能否真正重返先进制造竞争行列。若未来 Clearwater Forest 等产品能顺利量产并展现性能优势，英特尔将重获代工信誉；反之，若产能爬坡与性能兑现再次受阻，18A 可能只是一次阶段性追赶。

总体而言，18A 标志着英特尔在过去十年最具技术含量的一次工艺跃迁。它实现了从 FinFET 向 GAA 的结构转型，也为未来 AI 时代的能效优化奠定了基础。然而，18A 目前仍处于“证明阶段”而非“统治阶段”：它展现了英特尔重新崛起的潜力，但距离真正与台积电并驾齐驱仍需时间与市场验证。

AI 需求驱动存储链回暖：群联控制器出货创新高，华邦 DRAM 价量齐升

2025 年 9 月，台湾存储产业呈现同步复苏态势，群联电子与华邦电子双双创下营收阶段高点，显示 AI 与企业级存储需求正成为产业核心驱动力。

群联电子 9 月营收达 65.15 亿元新台币，环比增长 9.8%，同比增长 47%，创历史单月次高；其中 PCIe SSD 控制芯片出货量同比暴增近 300%，反映 AI 推理带动的高性能存储需求爆发。公司明确指出，AI 应用对数据存取与训练缓存的扩张，正在推升企业级 SSD 采购节奏，并使 NAND 供需出现紧俏迹象。在此背景下，群联凭借控制器技术与客户黏性优势，正加速导入云端与服务器体系，企业级 SSD 市场份额有望持续提升。



图表5: DRAM Wafer 周度涨跌幅

料号	2025年36周	2025年37周	2025年38周	2025年39周	2025年40周	2025年41周
DRAM:DDR3 2Gb 128Mx16 1600/1866	0.94%	0.93%	6.48%	6.96%	0.81%	0.00%
DRAM:DDR3 2Gb 256Mx8 1600/1866	0.00%	0.75%	5.97%	6.34%	1.99%	1.30%
DRAM:DDR3 4Gb 256Mx16 1600/1866	0.00%	0.93%	8.33%	8.55%	7.09%	0.74%
DRAM:DDR3 4Gb 512Mx8 eTT	2.00%	3.92%	3.77%	7.27%	5.08%	3.23%
DRAM:DDR4 16Gb (1Gx16) 3200	1.02%	2.15%	11.35%	8.91%	6.70%	4.22%
DRAM:DDR4 16Gb (2Gx8) 2666 Mbps	1.08%	2.68%	6.26%	8.35%	7.89%	6.72%
DRAM:DDR4 16Gb (2Gx8) 3200	2.99%	6.11%	8.59%	11.41%	12.74%	8.87%
DRAM:DDR4 16Gb (2Gx8) eTT Mbps	1.23%	2.13%	5.67%	8.19%	5.48%	3.47%
DRAM:DDR4 4Gb 512Mx8 eTT	1.03%	1.02%	5.05%	5.77%	3.64%	0.88%
DRAM:DDR4 4Gb (256Mx16) 2400/2666	-1.80%	1.83%	3.60%	2.17%	0.43%	1.27%
DRAM:DDR4 4Gb (512Mx8) 2400/2666	0.00%	1.57%	3.09%	2.25%	0.73%	0.73%
DRAM:DDR4 8Gb (1Gx8) 2666 Mbps	1.26%	2.66%	5.71%	6.55%	3.84%	2.81%
DRAM:DDR4 8Gb (1Gx8) 3200	2.26%	5.43%	9.92%	9.03%	3.50%	5.54%
DRAM:DDR4 8Gb (1Gx8) eTT	1.94%	0.63%	7.55%	9.94%	8.51%	3.43%
DRAM:DDR4 8Gb (512Mx16) 2666 Mbps	-0.66%	2.42%	8.37%	7.72%	5.88%	3.13%
DRAM:DDR4 8Gb (512Mx16) 3200	-0.51%	5.81%	6.95%	6.50%	5.11%	3.37%
DRAM:DDR5 16G (2Gx8) 4800/5600	0.50%	2.48%	10.32%	9.36%	8.69%	5.54%

来源: 中国闪存市场、dramexchange、国金证券研究所

华邦电子第三季营收 217.7 亿元新台币, 创近 12 季新高, 主因利基型 DRAM 合约价上扬与产品组合优化。公司预计第四季 DRAM 合约价涨幅将高于第三季, 显示需求复苏趋势稳固。受惠服务器端需求回升与原厂逐步停产 DDR3 造成的供给收缩, DDR4 价格自 9 月起明显走强, 带动价量双升。华邦以合约价销售为主, 价格波动滞后但稳定, 叠加 DDR3 补涨与工业/车载拉货动能, 第四季营收与毛利率均具进一步改善空间。

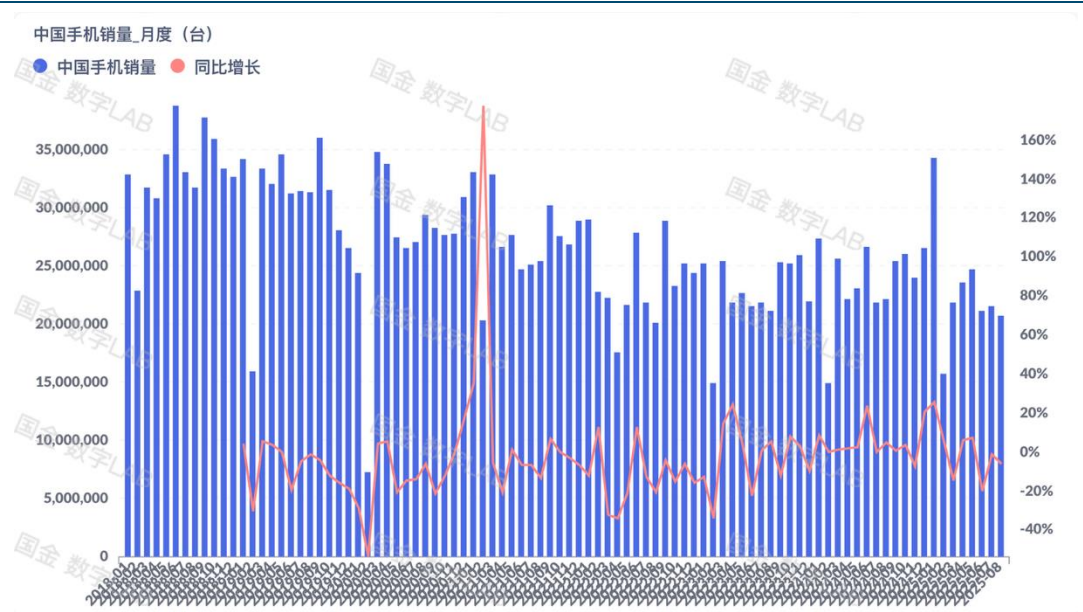
总体而言, AI 算力扩张正从上游 GPU 与 HBM 延伸至中下游存储链条, 推动控制器与利基型 DRAM 同步进入景气回升周期。群联受益于 AI 推理侧 SSD 放量, 华邦则凭特规 DRAM 与合约价机制实现稳定成长, 台湾存储产业链在 2025 年下半年有望维持结构性复苏。

消费电子动态

8 月国内智能手机市场销量同比下滑

2025 年 8 月, 中国智能手机销量达到约 2080 万台, 同比下滑 6.24%。小米、华为、vivo、OPPO、荣耀以 16.21%、15.08%、13.82%、13.79%、13.52% 的份额排名前五。具体型号方面, iPhone 16 Pro 销量排名第一, OPPO reno 14、iPhone 16 Pro Max 排名第二、第三。

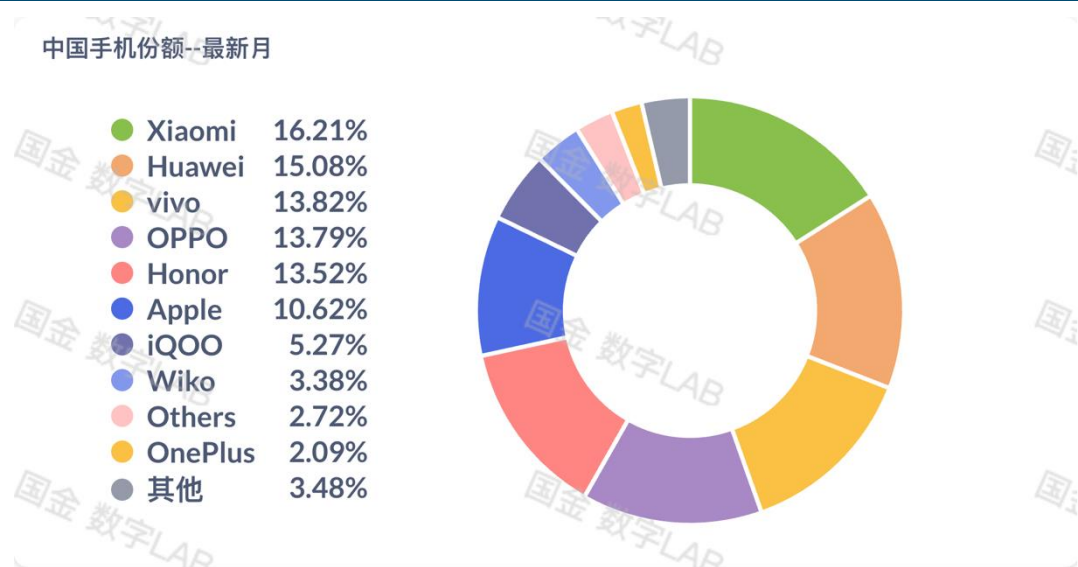
图表6: 中国智能手机销量(台)



来源: IDC、国金证券研究所



图表7: 2025 年 8 月中国智能手机份额



来源: IDC、国金证券研究所

图表8: 中国 8 月智能手机具体型号销量 (台)

前十大手机型号销量 (台) --最新月度数据

month	brand	model_name	sales
2025-08	Apple	iPhone 16 Pro	960,491
2025-08	OPPO	Reno 14	839,736
2025-08	Apple	iPhone 16 Pro Max	829,483
2025-08	Honor	X70	822,348
2025-08	vivo	S30	640,347
2025-08	Others	Others	566,643
2025-08	OPPO	A5 Energy Edition	522,930
2025-08	Honor	400	494,438
2025-08	Huawei	nova 14	491,171
2025-08	Xiaomi	15	433,934

来源: IDC、国金证券研究所

8 月 PC 市场同比基本持平

8 月国内 PC 市场销量基本同比持平, 台式机销量约为 157 万台, 同比下滑约 1.39%。笔电销量约为 232 万台, 同比增长约 2.02%。

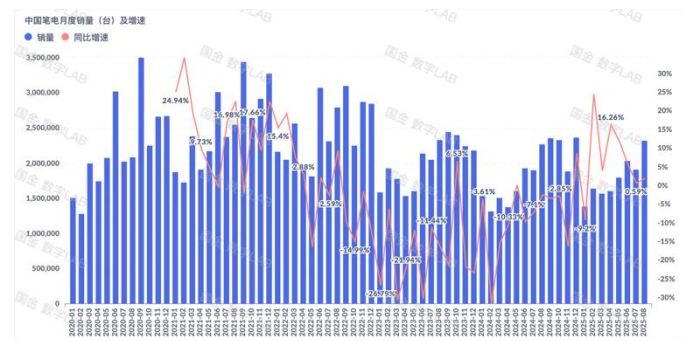


图表9：中国台式机月度销量及增速



来源：IDC、国金证券研究所

图表10：中国笔电月度销量及增速



来源：IDC、国金证券研究所

风险提示

1. 芯片制程发展与良率不及预期：半导体工艺的发展面临诸多挑战，主要包括技术瓶颈、良率提升难度、研发成本高企以及供应链不确定性等问题。随着工艺节点微缩变得愈发复杂，先进制程的实现难度和成本不断攀升，可能导致量产延迟，甚至影响产品性能和成本控制。此外，地缘政治风险和出口管制可能扰乱供应链，进一步拖累产能扩张。
2. 中美科技领域政策恶化：中美在 AI 领域竞争激烈，美国限制先进芯片和半导体对中国的出口，随着竞争的加剧，未来可能会推出更严格的限制政策，限制国内 AI 模型的发展。
3. 智能手机销量不及预期：智能手机销量与产品本身质量关系紧密，若产品本身有缺陷则智能手机销量可能受到影响。同时宏观经济变化也有可能导致消费者消费意愿发生变化从而影响智能手机销量。



特别声明:

国金证券股份有限公司经中国证券监督管理委员会批准,已具备证券投资咨询业务资格。

形式的复制、转发、转载、引用、修改、仿制、刊发,或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发,需注明出处为“国金证券股份有限公司”,且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料,但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法,故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致,国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断,在不作事先通知的情况下,可能会随时调整,亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用,在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险,可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突,而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品,使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况,以及(若有必要)咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议,国金证券不就报告中的内容对最终操作建议做出任何担保,在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下,国金证券的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,并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密,只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》,本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级(含 C3 级)的投资者使用;本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要,不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具,本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资,遭受任何损失,国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告,则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议,国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有,保留一切权利。

上海

电话: 021-80234211

邮箱: researchsh@gjzq.com.cn

邮编: 201204

地址: 上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话: 010-85950438

邮箱: researchbj@gjzq.com.cn

邮编: 100005

地址: 北京市东城区建国门内大街 26 号

新闻大厦 8 层南侧

深圳

电话: 0755-86695353

邮箱: researchsz@gjzq.com.cn

邮编: 518000

地址: 深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究