

巨头 AI 投入 不减，积极发 力商业变现

分析师及联系人

孟灿

SAC:S1130522050001

mengcan@gjzq.com.cn

李忠宇

SAC:S1130524100002

lizhongyu01@gjzq.com.cn

核心要点

产业前沿

英伟达整合 Groq 强化能力，字节提升 26 年资本支出；OpenAI 强化商业变化，字节、智谱、MiniMax 等迭代大模型；头部 AI 应用竞争格局相对稳定，字节在 AI 软件和硬件应用端积极探索。

资本风向

上游流片、存储等多环节持续涨价，阿里升级模型版本，META AI 眼镜销量持续提升。

本周观点

英伟达战略性整合 Groq，并重组自身云服务团队，此举意在通过吸纳 Groq 的技术与人才，补强其在高能效推理领域的短板并强化其在 AI 推理芯片上的布局。英伟达能够通过组织变革持续优化其长期竞争能力，持续看好与 Groq 合作对英伟达高性能推理芯片研发的促进作用。三星电子、SK 海力士等存储供应商已上调明年 HBM3E 价格，涨幅接近 20%。我们认为此举是受 AI 加速器企业上调明年 HBM3E 订单量所致。OpenAI 算力利润率攀升，内部正讨论广告业务模式。这标志着该公司在探索广告业务方面已进入具体细节的规划阶段。字节明年资本支出高企，豆包 DAU 突破 1 亿大关我们看好国内大厂资本支出及大模型落地对 IDC 的提振作用。

内容目录

一、产业前沿	3
AI 基础层：英伟达整合 Groq，字节提升 26 年 AI 资本支出	3
AI 模型层：OpenAI 强化商业变现，头部厂商迭代模型	4
AI 应用层：头部应用格局相对稳定，字节 AI 软硬件应用探索积极	5
二、资本风向	7
AI 基础层：上游多环节涨价持续，头部厂商调整 AI 团队	7
AI 模型层：阿里推出新版本大模型	8
AI 应用层：Meta AI 眼镜销量持续提升	8
三、本周观点	9
四、产业链数据更新	10
11 月覆铜板拉货紧张程度继续升级，高景气度继续维持	10
风险提示	13

图表目录

图表 1：Meta Rayban 出货量持续上升	9
图表 2：2025Q3AI 眼镜份额	9
图表 3：中国台湾电子铜箔厂商月度营收同比增速	11
图表 4：中国台湾电子铜箔厂商月度营收环比增速	11
图表 5：中国台湾电子玻纤布厂商月度营收同比增速	11
图表 6：中国台湾电子玻纤布厂商月度营收环比增速	11

图表 7: 中国台湾覆铜板厂商月度营收同比增速	12
图表 8: 中国台湾覆铜板厂商月度营收环比增速	12
图表 9: 中国台湾 PCB 厂商月度营收同比增速	12
图表 10: 中国台湾 PCB 厂商月度营收同比增速	12

一、产业前沿

AI 基础层：英伟达整合 Groq，字节提升 26 年 AI 资本支出

美国政府 23 日宣布，将在 2027 年对中国芯片加征关税，结束了上届拜登政府发起的针对中国芯片的贸易调查。美媒分析称，尽管美国政府称中国在芯片产业中的做法“损害美国利益”，但最终决定至少在 18 个月内不对中国芯片加征额外关税。彭博社说，暂缓加征新关税是美国政府寻求巩固中美“休战”协议、稳定中美关系的最新信号。中国外交部发言人林剑 24 日在例行记者会上表示，中方坚决反对美方滥施关税，无理打压中国产业，并称如果美方一意孤行，中方必将坚决采取相应措施，维护自身正当权益。

12 月 27 日消息，英伟达与人工智能芯片初创公司 Groq 达成一项战略性技术整合协议，将支付约 200 亿美元现金获得人工智能芯片初创公司 Groq 的技术许可并聘请其核心工程团队。英伟达方面称，Groq 的低延迟芯片具备极高的输入响应速度，这项技术的融入将为英伟达的产品赋予全新功能，并帮助其开拓新的市场领域。Groq 表示，创始人乔纳森·罗斯、总裁桑尼·马德拉及其他员工将加入 Nvidia，而去年九月加入 Groq 担任首席财务官的西蒙·爱德华兹将出任新任 CEO。

12 月 25 日至 26 日，全国工业和信息化工作会议在京召开。会议指出，全年电信业务总量和软件业务收入预计将分别同比增长 9% 和 12% 左右，数字产业业务收入同比增长 9% 左右；1—11 月规模以上高技术制造业、装备制造业增加值同比分别增长 9.2% 和 9.3%，人工智能核心产业超过万亿元。全国工业 5G 专网项目超过 2 万个、5G 工厂超过 8000 家，截至 10 月底 5G 基站总数达到 475.8 万个，5G 融入 97 个国民经济大类中的 91 个。

12月23日消息,据金融时报报道,TikTok 母公司字节跳动已初步计划在2026年投入1600亿元人民币(230亿美元)用于资本支出,这一数字将高于该公司今年在人工智能基础设施领域投资的1500亿元人民币。知情人士透露,字节跳动已计划明年在人工智能处理器上花费850亿元人民币,并计划以测试订单的形式购买2万台H200,每台的价格可能约为2万美元,且如果字节跳动能够不受限制地购买更多H200设备,其2026年的资本支出可能会进一步大幅增加。

AI 模型层: OpenAI 强化商业变现, 头部厂商迭代模型

12月22日消息,据科技媒体 The Information 披露,在人工智能行业白热化竞争背景下,OpenAI 正通过优化成本结构强化市场优势。内部数据显示,截至今年10月,OpenAI 的算力利润率(即扣除模型服务成本后的营收占比)已攀升至70%,较2024年初的35%实现翻倍增长。这一关键指标的跃升,反映出公司在控制基础设施投入与扩大商业变现之间的平衡能力。

12月24日,据 The Information 援引知情人士消息,OpenAI 内部已在讨论调整其AI模型以优先展示赞助信息,并已着手制作广告在ChatGPT中呈现方式的内部模型。这标志着该公司在探索广告业务方面已进入具体细节的规划阶段。这些讨论中的方案包括,当用户进行相关查询时,在ChatGPT的回复中优先植入赞助内容。同时,近几周制作的内部模型展示了多种广告呈现方式,例如在主回复窗口的侧边栏显示赞助信息。此举旨在将ChatGPT庞大的用户流量转化为直接收入,从而挑战现有数字广告巨头的市场地位。

字节推出形式化数学推理专用模型 Seed Prover 1.5。字节跳动 Seed 团队宣布推出形式化数学推理专用模型 Seed Prover 1.5。据介绍,相比上一代模型,Seed Prover 1.5 在16.5小时内,针对IMO 2025的前5道题目生成了完整可编

译验证的 Lean 证明代码，换算成绩为 35/42，达到此前 IMO 评分标准的金牌分数线。Seed Prover 1.5 的技术报告已对外公开，后续将开放 API，邀请感兴趣的数学和 AI 研究者体验该模型。

国内厂商 MiniMax 正式开源 M2.1 编码与智能体专用大模型，采用 100 亿激活参数的 MoE 架构，在 SWE-Multilingual 与 VIBE-Bench 基准中超越 Gemini3Pro 及 Claude4.5Sonnet，实现多语言编程 SOTA，其稀疏架构在保证性能的同时大幅降低推理成本。

12 月 23 日，智谱 AI 正式上线并开源 GLM-4.7 模型。新版本主要针对编程场景进行优化，提升了代码生成、任务规划与工具调用能力。据悉，该模型在多项编程基准测试中取得了开源模型中的领先成绩。官方数据显示，其在 SWE-bench-Verified 测试中获得 73.8% 的得分，在 LiveCodeBench V6 中达到 84.9%。在工具调用评测 τ^2 -Bench 中，其成绩为 87.4 分。除编码能力外，GLM-4.7 在长程任务中支持保留式思考机制以优化性能，并在前端代码与文档生成方面提升了视觉美观度，例如 PPT 的 16:9 适配率有所提高。模型目前已通过 BigModel.cn 提供 API 服务。

AI 应用层：头部应用格局相对稳定，字节 AI 软硬件应用探索积极

本周，海外聊天助手类应用活跃度中，受节假日影响，ChatGPT 访问量大幅下降，Gemini 活跃度保持稳定，Claude 和 Perplexity 小幅下降。国内应用中，多数应用活跃度保持稳定。

Quest Mobile 发布 AI 应用排名：豆包、DeepSeek、元宝、阿福位列前四。12 月 23 日，研究机构 Quest Mobile 发布的《2025 下半年 AI 应用交互革新与生态落地报告》显示，全市场 AI 原生 App 中，最新周活跃用户排名前四的是豆包、

DeepSeek、元宝、蚂蚁阿福，阿里千问位列第五，蚂蚁集团 11 月发布的通用 AI 助手灵光，进入前十。

12 月 25 日消息，据 36 氪报道，豆包的日均活跃用户数（DAU）已突破 1 亿大关，字节跳动 APP 王国里亿级 DAU 产品再添一子。据字节内部人士透露，豆包的 UG、市场推广费用，是字节历史上，所有破亿 DAU 产品中花费最低的。此外，豆包 Seedream 生图模型、Seedance 生视频模型，持续激活创作场景；三宫格等豆包 P 图玩法连续登上热搜，每天带动的自然下载量高达百万级。

12 月 24 日，蓝鲸财经报道中兴通讯相关人士首度回应与字节跳动的合作细节与战略考量。该人士提及，与字节跳动合作开发 AI 手机的原因首先是“战略同频”，即双方均将 AI 手机视为确定的未来趋势；其次是“优势互补”，中兴通讯具备深厚的硬件工程能力，而字节跳动则拥有豆包大模型及其“跨应用执行的智能体技术”，双方合作能够共同推动“硬件+AI”的融合创新；再次是豆包的技术能力与中兴的人工智能理念高度契合。同时，该人士表示目前中兴通讯已收到部分其他大模型厂商的合作邀约，相关沟通正在推进。这意味着与字节跳动的合作可能只是一个开端，中兴正逐步构建更加开放的 AI 生态体系。

12 月 22 日下午，华为正式推出定位中高端的 nova 15 系列手机。其中 nova 15 起售价为 2699 元，nova 15 Pro 起售价为 3499 元，nova 15 Ultra 起售价 4199 元。nova 15 Ultra 与 nova 15 Pro 均搭载麒麟 9010S 处理器、5000 万像素的前置红枫影像系统、6500mAh 电池与 100W 有线快充、北斗卫星消息和 2.4GHz 畅连无网通信；nova 15 Ultra 还支持星闪车钥匙、50W 无线快充，后置摄像头则搭载了全 5000 万像素全 RYYB 红枫影像系统。相比之下，nova 15 标准版搭载的是麒麟 8020 处理器，通过与鸿蒙操作系统深度协同，整机性能相比上一代提升 62%。

二、资本风向

AI 基础层：上游多环节涨价持续，头部厂商调整 AI 团队

据韩国《朝鲜日报》12月24日消息，三星电子、SK海力士等存储供应商已上调明年HBM3E价格，涨幅接近20%。报道援引行业消息称，三星电子和SK海力士等内存半导体企业将HBM3E的供应价格提高了近20%。考虑到通常在新一代HBM发布前价格会下调，这一涨幅被认为是异常情况。有观点认为，这是英伟达、谷歌、亚马逊等AI加速器企业上调明年HBM3E订单量所导致的。

12月23日消息，在英伟达首席执行官黄仁勋提出打造一家可与AWS竞争的云服务愿景两年多之后，他已经放弃了与这些公司正面竞争的想法，并在上周对公司的云团队进行了重组。根据一份内部备忘录，英伟达将规模达数百人的云团队并入工程与运营组织，该组织由负责软件工程的高级副总裁Dwight Diercks领导。这支被称为DGX Cloud的云团队今后将主要服务于英伟达工程师对英伟达芯片的需求，这些芯片用于开发开源AI模型，而不再把重点放在向外部企业客户销售云服务上。作为重组的一部分，英伟达较新的云服务DGX Cloud Lepton也将并入工程团队。该服务允许云服务商在英伟达运营的一个市场中，列出其未被使用的英伟达服务器算力。

12月22日消息，根据The Information报道，微软CEO萨提亚·纳德拉近期频繁现身公司内部一个由约100名核心技术人员组成的Teams频道，只要他认为AI产品表现不理想，便会直接在其中公开表达不满。报道称，过去几个月里，纳德拉逐渐转变为微软内部最具话语权的产品负责人。9月时，他曾告知员工自己将下放部分管理职责，把更多时间用于推动AI产品开发。纳德拉同时也加大了

对 AI 人才招聘的直接参与力度，不仅亲自联系候选人，还批准极具竞争力的薪酬方案，以从 OpenAI 和谷歌 DeepMind 等机构挖角顶尖研究人员。

12 月 23 日-24 日，多家媒体报道中芯国际已向下游客户发布涨价通知，且此次涨价主要集中于 8 英寸 BCD 工艺平台，涨价幅度在 10% 左右。有公司反映，预计涨价会很快执行。但由于之前存储产品价格过低，晶圆厂早已率先对其实施了涨价。有业内人士透露，此次涨价是对需求上涨的回应：“由于手机应用和 AI 需求持续增长，带动套片需求，从而带动了整体半导体产品需求的增长”；同时原材料涨价也是原因之一。

AI 模型层：阿里推出新版本大模型

阿里通义千问推出 Qwen Code v0.5.0 版本。阿里通义千问宣布，Qwen Code 推出 v0.5.0 版本。这次更新不仅包含了功能增强，更是 Qwen Code 从“命令行工具”向“开发生态”迈进的关键一步。（36 氪）

阿里升级新一代语音模型 Qwen3-TTS，可参考文字、声音生成拟人音色。12 月 24 日，阿里升级语音模型家族 Qwen3-TTS，发布音色创造 Qwen3-TTS-VD（VoiceDesign）和音色克隆 Qwen3-TTS-VC（VoiceClone）两款全新模型。

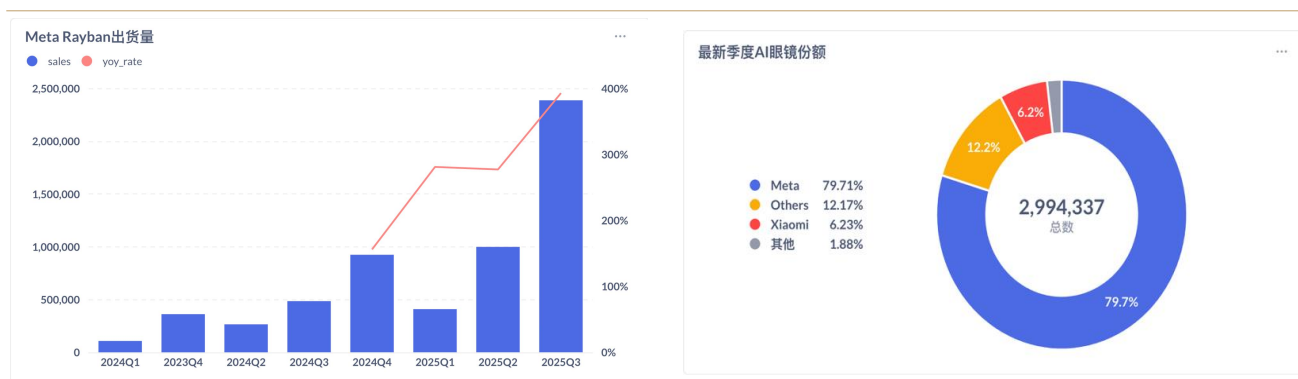
Qwen3-TTS 新模型可实现 DIY 声音设计和像素级音色模仿，甚至让动物“原生”开口说人话，音色自然、效果稳定、生成高效，可加速语音大模型在有声小说、AI 漫剧、影视配音等多专业领域落地。

AI 应用层：Meta AI 眼镜销量持续提升

2025 年第三季度，Meta Ray-Ban 的销量达到约 239 万台，同比增长 393%。整个 AI 眼镜市场中，Meta 占据了接近 80% 的份额，小米等公司凭借消费电子产品能力多年与手机生态的帮助开始进入这一市场。

AI 助手、拍照/视频记录、信息提醒等功能逐渐被用户接受，同时 Meta 的持续投入显著放大了销量弹性。若后续产品形态继续演进、应用场景进一步丰富，Meta Ray-Ban 将继续维持在 AI 眼镜市场的优势。

图表 1: Meta Rayban 出货量持续上升 图表 2: 2025Q3AI 眼镜份额



来源：IDC、国金证券研究所

来源：IDC、国金证券研究所

我们认为当前 AI 眼镜正从概念产品加速迈向规模化商业阶段：在低功耗 AI 处理器、微显示与传感器融合的推动下，产品功能从拍照/录音扩展到实时翻译、信息检索、导航与远程协作等高频场景，叠加工业、医疗、物流应用打开新的需求空间。与此同时，巨头与初创公司共同入局，软硬件一体化与生态内容成为竞争核心。不过，AI 眼镜也面临续航、发热、佩戴舒适度、人机交互与隐私合规等难题。但 AI 眼镜长期有望在消费与专业垂直场景双轮驱动下成长。

YouTube Gaming 上周宣布推出一个名为 Playables Builder 的开放测试项目。这一工具允许特定的 YouTube 创作者使用"基于 Gemini 3 构建的原型网络应用"来制作小型游戏，无需编程技能。YouTube 早在 2023 年就开始在桌面和移动平台上测试小型游戏的添加功能，去年还为 Playables 增加了多人游戏功能。

三、本周观点

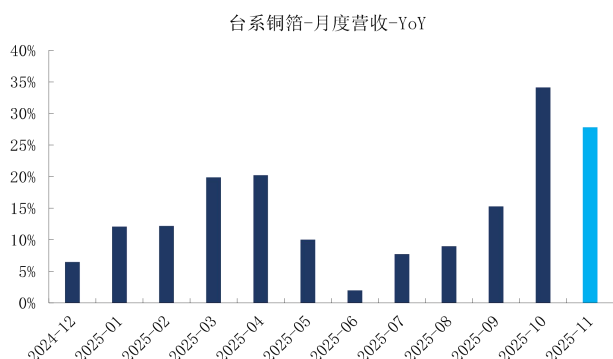
英伟达战略性整合 Groq，并重组自身云服务团队，此举意在通过吸纳 Groq 的技术与人才，补强其在高能效推理领域的短板并强化其在 AI 推理芯片上的布局。英伟达能够通过组织变革持续优化其长期竞争能力，持续看好与 Groq 合作对英伟达高性能推理芯片研发的促进作用。三星电子、SK 海力士等存储供应商已上调明年 HBM3E 价格，涨幅接近 20%。我们认为此举是受 AI 加速器企业上调明年 HBM3E 订单量所致。OpenAI 算力利润率攀升，内部正讨论广告业务模式。这标志着该公司在探索广告业务方面已进入具体细节的规划阶段。华为发布 nova 15 系列手机，均通过与鸿蒙系统协同实现整机性能提升。字节明年资本支出高企，豆包 DAU 突破 1 亿大关我们看好国内大厂资本支出及大模型落地对 IDC 的提振作用。

四、产业链数据更新

11 月覆铜板拉货紧张程度继续升级，高景气度继续维持

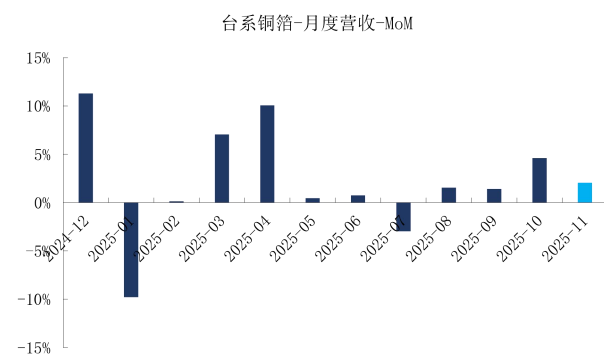
根据 11 月跟踪，整个产业链景气度判断为“保持高景气度”，主要原因来自于汽车、工控随着政策补贴加持同时 AI 开始大批量放量，并且从当前的能见度观察来看，四季度有望持续保持较高的景气度状态，中低端原材料和覆铜板在四季度又迎来一轮涨价，我们预计覆铜板涨价已经走到了斜率有望变高的重要时间点。

图表 3: 中国台湾电子铜箔厂商月度营收同比增速



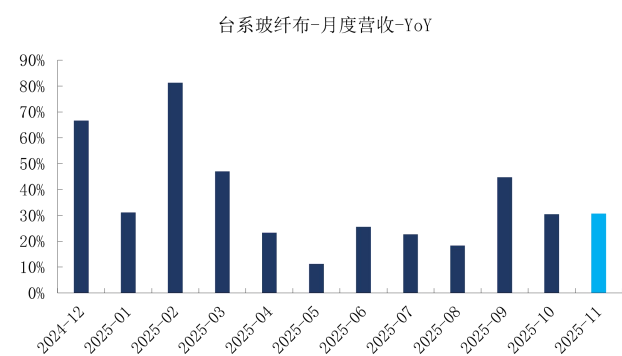
来源: Wind、国金证券研究所

图表 4: 中国台湾电子铜箔厂商月度营收环比增速



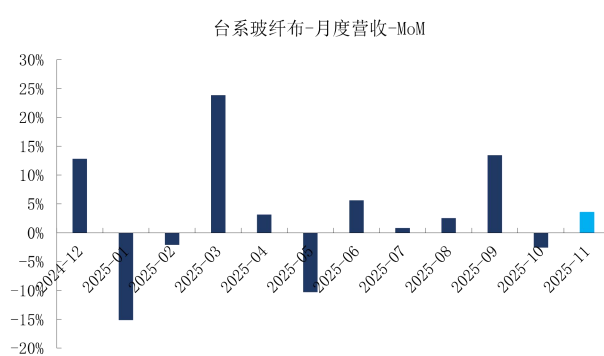
来源: Wind、国金证券研究所

图表 5: 中国台湾电子玻纤布厂商月度营收同比增速



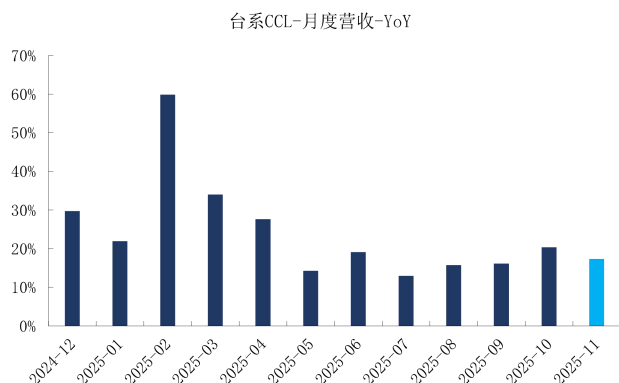
来源: Wind、国金证券研究所

图表 6: 中国台湾电子玻纤布厂商月度营收环比增速



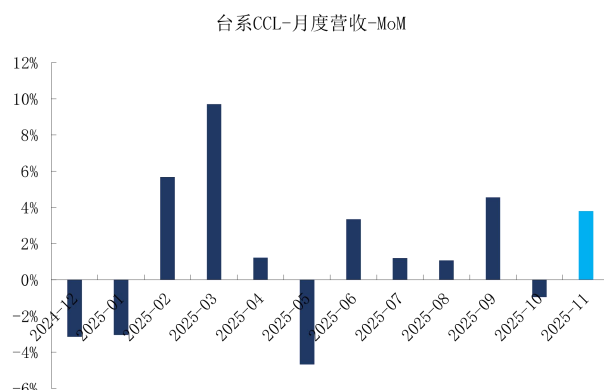
来源: Wind、国金证券研究所

图表 7: 中国台湾覆铜板厂商月度营收
同比增速



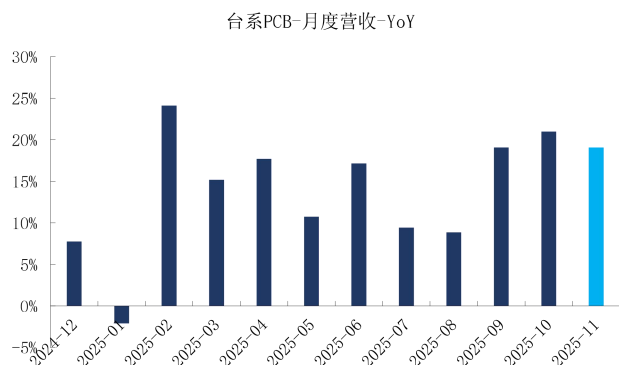
来源: Wind、国金证券研究所

图表 8: 中国台湾覆铜板厂商月度营
收环比增速



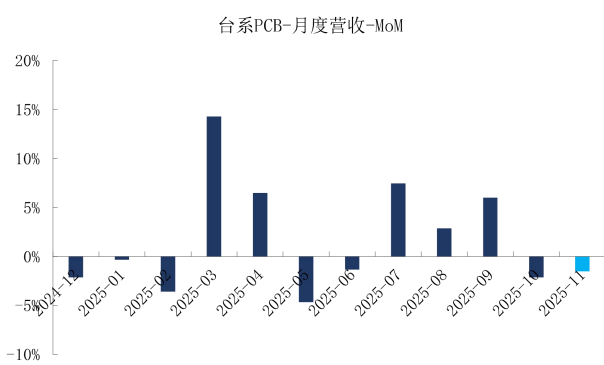
来源: Wind、国金证券研究所

图表 9: 中国台湾 PCB 厂商月度营收同
比增速



来源: Wind、国金证券研究所

图表 10: 中国台湾 PCB 厂商月度营收
同比增速



来源: Wind、国金证券研究所

风险提示

算力供不应求的风险：若训练或推理算力芯片不足，可能会推升现有芯片价格，降低 AI 应用的性价比，使得 AI 应用推广不及预期。

底层大模型迭代发展不及预期：若底层大模型迭代发展不及预期，可能会影响 AI 应用落地的深度，使其难以在诸多领域进行更加深入的应用。若底层大模型的蒸馏剪枝发展不及预期，可能会使其难以在边缘硬件上充分发挥性能。

国际关系风险：若出于国际关系原因，OpenAI 等海外大模型的调用或其他软硬件的进口受到影响，有可能使得国内 AI 应用的发展不及预期。

监管风险：AI 生成内容在版权归属、个人隐私、伦理等方面存在较大争议，面临一定监管风险，可能对技术及应用层也会带来一定影响。

并购整合不及预期的风险：部分公司进行并购扩大业务布局，如并购整合不能顺利完成，可能影响公司整体业绩表现。

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。

国金证券股份有限公司) 所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、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。