

# AI 半导体产业链深度研究报告

报告日期：2026.6.15 报告人：wang hua

## 从 GPU 短缺到系统部署瓶颈

本报告不构成任何买卖建议。文中公司和产业链节点仅用于研究框架、供需验证和风险识别。

## 执行摘要

2023 年以来，AI 基础设施的核心矛盾已经发生变化。早期市场最关心的是 GPU 是否短缺；到 2026 年，真正限制 AI 产能释放的，已经变成一整套基础设施能否同步扩张：电力能否并网，机柜能否供电，热量能否带走，HBM 和先进封装能否交付，光互连和网络能否支撑更大的集群，Neocloud 能否用高质量合同和资产负债表承接算力需求。

这意味着 AI 半导体研究不能继续停留在“GPU-HBM-光模块”的三件套框架里。GPU 仍然是价值中心，但 GPU 之外的瓶颈正在决定部署节奏。一个 AI 集群能不能上线，不只取决于芯片性能，还取决于园区电力、变压器、中压配电、液冷系统、CoWoS 产能、HBM 供应、800G/1.6T 光互连、交换网络、PCB、连接器、设备良率和融资结构。

本报告的核心判断是：AI 产业链正在从“芯片性能周期”进入“系统部署周期”。在这个阶段，真正值得跟踪的不是谁最会讲 AI 故事，而是谁处在难以速扩、客户认证严格、工程替代成本高、收入和毛利能够被验证的环节。

从当前证据看，高可信卡点集中在 HBM、先进封装、封装设备、高端光器件、InP/硅光材料、园区级电力、液冷部署和部分数据中心网络环节。中等可信但需要继续验证的方向包括 800VDC、GaN/SiC 功率器件、连接器、高速线缆、高端 PCB、被动元件和部分材料扩散链。反身性最高、最容易被市场过度外推的方向，则包括合同质量弱的 Neocloud、只有概念没有客户验证的小市值 CPO/液冷/电力映射标的，以及把具身智能零部件直接等同于当期 AI 数据中心资本开支的远期叙事。

历史上，最接近这一轮 AI 基础设施周期的参照物，是 1990 年代末到 2000 年代初的互联网基础设施扩张。早期互联网的流量增长是真实的，光纤、路由器、服务器、IDC 也是真实需求；但真实需求并不阻止供给过度建设、融资过度乐观和估值提前透支。AI 周期也类似：长期需求可能成立，但中短期胜负取决于哪一层是真卡点，哪一层只是扩散受益，哪一层已经被二级市场提前定价。

因此，本报告用 L0-L8 重新拆解 AI 半导体产业链：从电力和数据中心底座，到先进封装、光互连、网络、散热、PCB/连接器、材料、设备，再到具身智能和 Neocloud。这个框架的目的不是给出单一结论，而是把 AI 研究从“看概念”推进到“看部署、看订单、看良率、看回款、看资产负债表”。

## 一、核心判断

### 1. AI capex 的约束正在从 GPU 外溢到整套基础设施

Hyperscaler 的资本开支仍然是本轮周期的总闸门。Meta、Amazon、Microsoft、Alphabet 等平台的 AI 数据中心支出，决定了上游 GPU、HBM、封装、光模块、网络、电源和液冷的总需求。但当资本开支足够大之后，限制因素不再只是“有没有 GPU”，而是“GPU 能不能被部署成可用算力”。

部署一座 AI 数据中心，需要同时满足五类条件：芯片和 HBM 可获得，先进封装产能可用，机柜和园区电力可接入，散热和液冷可靠，网络和光互连能够支撑集群效率。任何一层失配，都会把订单从“纸面需求”拖慢为“延迟交付”。

这也是为什么 L0 电力、L1 封装/HBM、L2 光互连、L4 液冷和 Neocloud 合同质量，正在成为市场关注的主线。AI 的价值中心仍然在算力，但算力的兑现方式已经越来越像一个系统工程。

### 2. 真卡点通常具备四个特征

第一，扩产慢。HBM、CoWoS、InP 激光、先进封装设备、数据中心并网和液冷部署，都不是靠短期加班就能解决的供应。

第二，客户认证严格。AI 基础设施不能容忍高故障率，供应商必须经过客户导入、系统验证、可靠性测试和量产爬坡。

第三，替代成本高。真正的卡点往往嵌入客户架构，一旦进入平台路线，替代并不容易。

第四，财务数据能验证。订单、backlog、收入占比、毛利率、客户集中度、产能利用率和交付周期，必须能在财报或公告里找到对应证据。

如果一个方向只有“技术听起来重要”，但没有客户、订单、收入、良率或交付数据，它更适合被放进观察池，而不是直接归入真实卡点。

### 3. 反身性最高的不是芯片本身，而是承接算力需求的融资层

Neocloud、GPU 租赁、AI 数据中心运营商和部分重资产 hosting 公司，是这一轮周期里反身性最高的资产类型。它们一方面承接真实需求，另一方面也高度依赖融资、客户集中度、租约结构、上电时间和 GPU 残值。

这类公司的研究重点不应该只看 GPU 数量或合同金额，而要看合同质量：是否 take-or-pay，是否有预付款，最大客户占比多高，债务成本如何，机房是否已上电，收入是否已经回款，客户是否能持续续约。

在 AI 基础设施周期中，融资层可能放大上行，也可能最早暴露风险。它更像基础设施周期里的运营杠杆，而不是纯粹的半导体卡点。

## 二、L0-L8 产业链地图

层级	核心环节	解决的问题	主要验证指标	代表公司/方向	研究判断
L0	电网、变压器、UPS、母线、800VDC、园区电力	数据中心能否并网、上电、稳定供电	已获批 MW、已上电 MW、机柜 kW、PUE、变压器和中压设备交期	Schneider、Vertiv、Eaton、ABB、Infineon、Navitas、onsemi	真需求明确，但不是所有电力概念都等于卡点
L1	GPU、HBM、CoWoS、先进封装、TCB、Hybrid Bonding、ABF/基板	高算力芯片能否以足够带宽和良率交付	HBM sold-out 状态、CoWoS 产能、封装良率、设备订单、基板 load	NVIDIA、TSMC、SK hynix、Micron、Samsung、ASMPT、Besi、Ibiden	当前最高可信卡点之一
L2	光互连、800G/1.6T、CPO、硅光、InP 激光、SOI	大规模 GPU 集群如何降低网络功耗并提高带宽	800G/1.6T 出货、200G/lane 渗透、InP 产能、CPO 时间表	Broadcom、Marvell、Coherent、Soitec、AXT、IQE、AIXTRON、Lumentum	光互连是真趋势，但 CPO 兑现节奏不能前置
L3	数据中心网络、交换芯片、功率半导体、GaN/SiC	集群通信效率和高密度供电效率	51.2T/102.4T 交换芯片、800V pilot、PSU/IBC 效率、design win	Broadcom、Arista、Infineon、Navitas、onsemi	网络更接近硬瓶颈，功率器件属于扩散验证
L4	液冷、CDU、冷板、泵阀、热交换、机柜热管理	高功率机柜如何可靠散热	液冷 attach rate、CDU 出货、冷板覆盖率、故障率、运维标准	Vertiv、nVent、Schneider/Motivair、CoolIT、Supermicro、Boyd	工程刚需明确，但利润池会分散
L5	PCB、连接器、高速线缆、被动元件	板级和系统级信号、电流、连接可靠性	PCB 层数、连接器 ASP、224G/PCIe Gen7 导入、MLCC 用量	Amphenol、TE Connectivity、Murata、TTM、沪电股份	真实扩散受益，但不一定是独占卡点
L6	InP 衬底、epi wafer、SOI、特气、化合物半导体材料	光器件、功率器件和先进制造的材料底座	InP 产能、SOI revenue、epi lead time、材料 ASP	Soitec、AXT、IQE、Entegris	选择性卡点，必须追踪客户导入
L7	半导体设备、MOCVD、TCB、量测、检测、封装设备	扩产速度、良率和工艺窗口	book-to-bill、订单、交期、验机进度、服务收入	ASMPT、Besi、Onto、Camtek、K&S、AIXTRON	真受益但强周期，设备订单常领先终端收入
L8	具身智能硬件、执行器、传感器、磁材、机器人部件	AI 从数据中心走向物理世界	机器人出货、BOM 固化、客户导入、量产良率	Tesla、THK、TDK、Harmonic Drive 等	长期重要，但短期不能直接映射 AIDC capex
附加层	Neocloud、GPU 租赁、AI 数据中心运营、算力融资	谁拥有 GPU、谁融资、谁承担空置和电力风险	最大客户占比、合同年限、take-or-pay、预付款、债务和稀释	CoreWeave、Applied Digital、Core Scientific、Crusoe、Lambda	需求真实但反身性最高

这张表的价值，不是把公司越列越多，而是强制区分三类资产：第一类是真卡点，供给短期难以扩张；第二类是扩散受益，收入会增长但竞争也会增强；第三类是叙事映射，技术方向可能成立，但当期财务兑现不足。

## 三、各层深度拆解

### L0：电力基础设施从背景变量变成主约束

AI 数据中心的电力问题，不只是总耗电量上升，而是单机柜功率密度快速提高后，园区、楼宇、机柜和服务器内部电源架构都要重构。过去数据中心更像 IT 设施；未来 AI factory 更像电力、热管理和网络工程的综合体。

L0 的真实瓶颈通常在园区并网、变压器、中压配电、UPS、母线、土建和验收，而不是所有贴上“AI 电力”标签的零部件。800VDC 和更高功率密度机柜的方向成立，但规模化导入需要平台方、系统集成商、功率半导体和数据中心运营方共同验证。

研究 L0 时，最重要的是区分“签约 MW”和“已上电 MW”。前者说明项目意向，后者才说明算力能真正部署。若未来更多项目披露已 energize MW、机柜功率和液冷架构，L0 的景气度才算从叙事进入兑现。

### L1: HBM 与先进封装仍是最高可信卡点

HBM 和先进封装是当前 AI 半导体链条里最硬的瓶颈之一。GPU 性能提升越来越依赖高带宽内存和更复杂的 2.5D/3D 封装。CoWoS、TCB、Hybrid Bonding、ABF/高端基板和封装量测检测，共同决定高端 AI 芯片能否按时交付。

这一层的强度来自两个事实：一是需求端明确，GPU/ASIC 训练与推理都需要更高显存和更高 I/O；二是供给端扩张慢，产线、设备、材料、良率和客户认证都需要时间。

但 L1 也不是没有风险。先进封装和 HBM 一旦进入大规模扩产，未来也会经历周期性回落。真正需要跟踪的不是单次扩产公告，而是产能爬坡后毛利率、利用率、backlog 和客户订单能否维持。

### L2: 光互连是真趋势，CPO 不能被提前兑现

AI 集群越大，网络越重要。训练和推理都需要大量 GPU 之间的东西向通信，带宽、延迟、功耗和可靠性会直接影响算力效率。因此，800G/1.6T 光模块、InP 激光器、DSP、硅光 light engine、SOI 材料和测试封装，正在从外围供应链变成 AI 基础设施主线。

但 CPO 需要谨慎处理。CPO 的长期逻辑成立，因为把光更靠近交换芯片可以降低功耗、提高带宽密度；但这不等于所有 CPO 概念公司都已经进入收入兑现期。当前更接近现实出货主线的，仍然是 800G/1.6T pluggable optics 及其上游材料和器件。

L2 研究的关键是时间差。技术路线成立是一回事，客户量产、收入占比、ASP 和毛利兑现是另一回事。市场最容易犯的错误，是把“技术必要性”提前折现成“当期盈利确定性”。

### L3: 网络是吞吐瓶颈，功率器件是效率扩散

数据中心网络决定 AI job completion time。交换芯片、以太网/InfiniBand 互连、拥塞控制和网络软件，都会影响集群效率。随着 51.2T、102.4T 交换芯片和 800G/1.6T 端口推进，网络层的战略重要性继续提高。

功率半导体则解决另一个问题：高密度机柜如何减少电力转换损耗。GaN、SiC、800VDC、PSU 和中间母线转换器，都属于下一代高密度供电架构的一部分。

这两条线要分开研究。网络层更接近硬吞吐瓶颈，功率器件更像效率扩散链条。前者更容易形成平台级卡位，后者要看具体设计赢单、系统认证和量产份额。

### L4: 液冷是工程刚需，但利润池不会只属于单一部件

高功率 GPU 机柜把散热从“设施支持”推到了“部署前提”。液冷、CDU、冷板、泵阀、manifold、快接件、热交换和现场运维，都会影响 AI 集群能否稳定运行。

液冷的确定性较高，因为高密度机柜继续上升时，纯风冷很难长期支撑。但液冷不是单一零部件故事，而是系统集成、现场交付和运维可靠性的故事。冷板或泵阀可能受益，但未必拥有长期高毛利。

研究 L4 时，应该优先看液冷 attach rate、命名客户、CDU 出货、整机平台适配、运维数据和故障率，而不是只看公司是否提到了“液冷”。

### L5: PCB、连接器和被动元件是扩散受益，不应被误读成独占卡点

AI 服务器和交换机升级，会提高 PCB 层数、连接器速率、高速线缆规格、电源完整性要求和被动元件用量。这一层的收入增长是真实的，尤其在 224G、PCIe Gen7、高端服务器 PCB、交换机背板和高速互连上。

但 L5 的问题在于竞争和替代。很多供应商可以受益，但不一定形成独占卡点。对这一层，应更关注 AI 收入占比、客户结构、ASP、毛利率和产能利用率，而不是单纯用“AI 服务器供应商”来定性。

## L6: 材料层的弹性来自小而关键的瓶颈

InP 衬底、epi wafer、SOI、特种气体和化合物半导体材料，往往不是市场最先关注的环节，但它们决定了光器件、功率器件和部分先进制造的上游供给。

材料层的研究难点在于间接性。需求可能来自 AI 数据中心，但收入兑现要经过器件厂、模块厂、系统客户和平台架构多层传导。因此，材料公司不能只看下游叙事，还要看客户导入、产能扩张、良率、价格和库存。

真正有价值的材料卡点，通常是市场小、大弹性、认证周期长、替代供应有限的环节。InP 和 Photonics-SOI 就属于需要重点跟踪的方向。

## L7: 设备是扩产的钥匙，但也最容易提前见顶

先进封装、HBM、InP 光器件、硅光、SiC/GaN 和高端检测，都需要设备。TCB、Hybrid Bonding、MOCVD、PVD、量测、检测和封装设备的订单，能提前反映扩产意愿。

但设备股的周期节奏通常早于终端收入。客户下单时，设备公司先受益；终端扩产完成后，如果需求没有继续上修，设备订单可能比终端收入更早回落。

因此，L7 的研究重点是 book-to-bill、backlog、客户验机、重复订单和服务收入。设备订单强并不自动等于终端长期短缺，必须同时看下游利用率和扩产后的价格环境。

## L8: 具身智能长期重要，但不能直接映射当前数据中心主链

具身智能、机器人、执行器、传感器、磁材和精密传动，是 AI 从数字世界走向物理世界的重要方向。但从当前产业链验证看，它和 AI 数据中心资本开支之间存在明显时间差。

L8 可以作为长期观察层，却不应被直接放进当前 AIDC 主链收入框架。机器人零部件要兑现，需要整机出货、BOM 固化、客户导入、工厂爬坡和可靠性验证。缺少这些证据时，它更像远期叙事，而不是当前 AI 半导体卡点。

## 附加层: Neocloud 是真实需求和金融反身性的交汇点

Neocloud 和 GPU 租赁平台解决的是“谁来拥有 GPU、谁来融资、谁来承担空置风险”的问题。在 hyperscaler 之外，企业 AI、模型公司和开发者需求，会推动一批 GPU 云和 AI hosting 公司扩张。

但这类资产最容易被误读。GPU 数量不是全部，合同金额也不是全部。研究重点应放在客户集中度、合同年限、take-or-pay、预付款、资产负债表、融资成本、上电节奏和真实利用率。

如果合同质量强、客户稳定、融资成本可控、机房按时上电，Neocloud 能成为 AI 基础设施扩张的重要承接层。反之，它也可能成为周期里最脆弱的杠杆环节。

## 四、真实卡点、扩散受益与叙事映射

分类	代表环节	主要证据	典型风险	研究结论
高可信真实卡点	HBM、CoWoS、先进封装设备、InP 激光、高端光器件、园区电力、液冷系统	客户订单、产能紧张、sold-out、量产路线、设备订单、交付周期	扩产后供需缓解，毛利回落	优先纳入核心跟踪池
中等可信扩散受益	800VDC、GaN/SiC、连接器、高速线缆、PCB、被动元件、SOI/部分材料	技术路径成立，已有客户导入或收入受益	竞争加剧，收入占比不足，毛利被压缩	需要逐家公司验证份额和财务兑现
高反身性资产	Neocloud、GPU 租赁、重资产 hosting、部分小市值 AI 基建映射	合同金额、融资公告、客户合作、市场注意力	客户集中、融资收紧、稀释、上电延迟	不能只看增长，要看合同和资产负债表质量
远期叙事映射	具身智能、机器人零部件、部分遥远 AI 终端硬件	技术方向成立，但当期收入关联弱	商业化慢，估值先行	可观察，但不应混入 AIDC 主链卡点

这四类资产的研究方法完全不同。高可信卡点看产能、良率、客户和毛利；扩散受益看收入占比、份额和竞争；高反身性资产看合同、融资和回款；远期叙事看商业化时间表。把它们混在一起，是 AI 产业链研究最常见的错误。

## 五、公开 X 语料中的市场注意力地图

为了避免只凭主观判断划分产业链，本报告引入一组公开 X 语料观察作为辅助验证。该语料不是交易信号，也不代表任何单一账号观点；它只用于观察市场在不同时间段把哪些供应链节点和哪些标的放在一起讨论。

截至 2026 年 6 月 11 日，内部整理的 Serenity 公开语料共包含 5,894 条推文，命中 6 条供应链和 8 个核心节点。其分布显示，市场注意力已经从单一 GPU 叙事扩散到 Neocloud、光通信/CPO、GPU/ASIC/networking、HBM、800V 电力和 InP/epi 材料。

供应链/节点	命中数	代表性 ticker 共现	研究含义
Neocloud contract / financing	507	NBIS, IREN, MSFT, META, NVDA, CRWV	市场高度关注算力需求如何被合同和融资承接
GPU/ASIC/networking	415	NVDA, NBIS, LITE, IREN, MSFT, GOOGL	算力和网络仍是 AI 基础设施主轴
CPO / external light source	365	SIVE, LITE, NVDA, MRVL, AAOI, COHR	光互连从主题研究进入核心供应链研究
CW/DFB laser	337	SIVE, LITE, MRVL, COHR, AAOI, JBL	光器件上游的激光器和外置光源成为高频节点
Memory / HBM / NAND	327	MU, SNDK, NVDA, AXTI, LITE, TSM	存储从周期品转向 AI 产能约束讨论
Power / 800V DC	196	NBIS, NVDA, LITE, MSFT, IREN, GOOGL	电力和高压直流开始进入 AI 基础设施主线
InP substrate	109	AXTI, LITE, COHR, NVDA, GOOGL, MSFT	材料层被重新定价为光互连上游约束
Epiwafer	80	IQE, LITE, AXTI, COHR, SIVE, AAOI	外延片属于更上游、更窄但弹性更高的观察层

这组语料的价值在于帮助研究者发现“注意力迁移”。2025 年下半年，Neocloud、算力租赁、电力和融资结构是高频主题；2026 年一季度之后，CPO、外置光源、InP、epi wafer、HBM 和先进封装明显升温。换句话说，市场已经在沿着 AI capex 的真实传导路径寻找下一层瓶颈。

但必须强调：共现不等于供应关系，讨论频率不等于投资价值，历史关注不等于未来表现。正确用法是把它当成研究入口，再逐条回到公司公告、财报、客户验证和产业数据。

数据附注：本节来自 `data/processed/serenity/serenity_supply_chain_summary.md`、`data/processed/serenity/serenity_supply_chain_nodes.csv` 和 `data/processed/serenity/serenity_attention_shift.csv` 的汇总。该部分属于历史语料观察，不是 Serenity 原始买卖观点，也不是交易指令。

## 六、与早期互联网基础设施周期的类比

这一轮 AI 基础设施周期最值得类比的，是早期互联网基础设施扩容。

早期互联网时期，浏览器、门户、搜索、电商和企业上网推动流量增长，市场提前为光纤、路由器、服务器、IDC 和网络设备定价。需求是真实的，但供给建设也是真实过度的。最后留下来的，不一定是当年最热门的概念公司，而是掌握核心网络设备、标准、客户关系、工程交付和资产负债表韧性的公司。

AI 周期也有相似结构。模型能力和推理需求是真实的，hyperscaler capex 也是真实的，GPU、HBM、先进封装、光互连、电力和液冷都是真实基础设施。但真实需求不意味着所有环节都能长期享受超额利润，也不意味着所有小市值映射都能兑现成收入。

互联网周期给 AI 研究的启发有三点。

第一，基础设施先于应用。应用爆发之前，市场会先为网络、服务器、数据中心和设备链条定价。AI 也一样，当前更容易被验证的是基础设施，而不是所有终端应用。

第二，瓶颈会迁移。互联网早期从带宽、骨干网、路由器、IDC 逐步转向应用和平台；AI 也会从 GPU、HBM、封装，逐步外溢到电力、散热、网络、光互连、运营和推理成本。

第三，真实产业趋势和资本市场泡沫可以同时存在。互联网不是假趋势，但大量公司失败了；AI 也不是假趋势，但并不意味着所有 AI 产业链资产都会持续重估。

因此，本轮 AI 产业链研究最重要的不是判断“AI 是否重要”，而是判断“哪一层当前最稀缺，哪一层已经被充分定价，哪一层只是远期叙事，哪一层的财务质量可能先出问题”。

## 七、公司与产业链位置映射

下表不是投资建议，而是产业链定位表。它的用途是帮助研究者把公司放回正确层级，避免把供应链位置、收入兑现和二级市场叙事混为一谈。

公司	代码	层级	产业链位置	主要验证点	风险提示
NVIDIA	NVDA	L1/L2/L3	GPU 平台、网络、CPO/硅光生态牵引方	GPU 平台代际、networking 增长、生态绑定	估值和 capex 预期敏感
TSMC	TSM	L1/L7	先进制程与 CoWoS 核心产能	CoWoS 扩产、良率、客户需求	地缘和扩产周期
SK hynix	000660.KS	L1	HBM 核心供应商	HBM3E/HBM4 出货、ASP、客户结构	HBM 扩产后价格风险
Micron	MU	L1	HBM 与高端存储	HBM 收入占比、gross margin、产能分配	存储周期波动
Samsung Electronics	005930.KS	L1	HBM、存储、先进封装潜在恢复	HBM 客户认证、良率改善	执行落后风险
ASMPT	0522.HK	L1/L7	TCB 与先进封装设备	AI/HBM 相关订单、book-to-bill	设备订单周期
Besi	BESLAS	L1/L7	Hybrid Bonding 与封装设备	先进封装订单、客户导入	订单波动
Broadcom	AVGO	L2/L3	交换 ASIC、定制 ASIC、CPO/硅光	AI networking、ASIC 客户、CPO 路线	大客户集中和估值
Marvell	MRVL	L2/L3	DSP、光互连、定制芯片	AI 定制芯片收入、光互连产品	客户集中和兑现节奏
Arista	ANET	L3	AI 以太网交换网络	hyperscaler 订单、端口升级	竞争和客户集中
Coherent	COHR	L2/L6	InP 激光、PD/PIC、800G/1.6T	InP 扩产、AI datacom 收入	扩产后价格回落
Soitec	SOLPA	L2/L6	Photonics-SOI	硅光客户导入、SOI revenue	渗透节奏
AXT	AXTI	L6	InP 衬底	InP 需求、许可、产能	地缘和客户集中
IQE	IQE.L	L6	InP epi wafer	数据中心光器件需求、产能利用	经营杠杆和盈利波动
AIXTRON	AIXA.DE	L6/L7	MOCVD 设备	光电子订单、客户 capex	订单周期
Infineon	IFX.DE	L0/L3	800VDC、PSU、功率半导体	design win、AI 电源收入	份额兑现
Navitas	NVTS	L3	GaN/SiC 800V 架构	量产范围、客户导入	盈利和稀释风险
Vertiv	VRT	L0/L4	电源、热管理、液冷系统	backlog、液冷订单、交付毛利	项目执行和估值
Schneider Electric	SU.PA	L0/L4	电力系统、800VDC、液冷	AI 数据中心项目、Motivair 整合	规模化速度
nVent	NVT	L4	CDU、manifold、液冷基础设施	AI 液冷订单、客户结构	竞争和毛利
Amphenol	APH	L5	高速连接器和线缆	IT datacom 增速、AI 占比	不一定形成独占卡点
TE Connectivity	TEL	L5	高速与电源连接器	224G/PCIe Gen7、数据中心客户	收入占比披露有限
TTM Technologies	TTMI	L5	高端 PCB	数据中心/网络 PCB 增长	行业扩产和价格竞争
沪电股份	002463.SZ	L5	AI 服务器/HPC PCB	数据通信收入、客户结构	客户名和份额透明度
天孚通信	300394.SZ	L2	光器件和封装相关器件	高速光模块链条份额	客户和收入结构需核验
英维克	002837.SZ	L4	液冷和温控系统	AIDC 客户、海外收入、CDU/冷板	竞争和交付
Onto Innovation	ONTO	L7	先进封装量测	AI packaging 收入、订单	周期和客户波动
Camtek	CAMT	L7	先进封装检测	订单、客户导入	周期性
CoreWeave	CRWV	附加层	GPU 云和 AI 基础设施运营	客户集中度、合同、债务、回款	高反身性和融资风险
Applied Digital	APLD	附加层	HPC hosting 和 AI 机房	租户质量、融资结构、上电	稀释和执行风险
Core Scientific	CORZ	附加层	HPC hosting 和电力基础设施	CoreWeave 合同、MW 上电	客户集中
Crusoe	private	附加层	AI 工厂和能源一体化运营	私有合同、园区上电	财务透明度不足
Lambda	private	附加层	GPU 云和 AI 开发者平台	客户、融资、利用率	私有公司披露不足
Tesla	TSLA	L8	Optimus 与实体 AI	机器人出货、产线进度	当前与 AIDC 主链关联弱
THK	6481.T	L8	机器人运动部件	客户导入、机器人量产	远期叙事

## 八、验证框架：如何区分真 AI 资产、AI 壳叙事和纯远期映射

### 真 AI 基础设施资产

真 AI 基础设施资产通常能够回答五个问题：客户是谁，产品处在哪个系统位置，收入何时确认，供给为什么难扩，毛利为什么能维持。

这类公司需要有财报、订单、产能、客户导入、技术路线或供应链文件支撑。典型指标包括 AI 收入占比、backlog、book-to-bill、客户认证、产能利用率、毛利率、lead time、上电 MW、合同年限和预付款。

### AI 壳叙事

AI 壳叙事通常有技术词汇，但缺少财务和客户证据。常见表现包括：公司只说“服务 AI 数据中心”，但没有披露客户、订单、收入占比；只讲 CPO、液冷、800V、机器人，但没有量产时间表；只展示合作新闻，却没有后续回款。

这类资产不是一定没有机会，但应放在观察层，等待硬数据验证。

### 纯远期映射

纯远期映射常见于具身智能、机器人零部件、部分材料或小市值跨界公司。它们的长期方向可能成立，但和当前 AI 数据中心资本开支没有直接收入链条。

这类资产研究的重点不是“AI 是否重要”，而是“公司什么时候进入客户 BOM、什么时候量产、收入占比多少、毛利率能否改善”。没有这些证据，不能把它们和 HBM、CoWoS、电力、液冷等当前主链卡点放在同一置信层级。

## 九、监控仪表盘

模块	指标	观察频率	主要用途
需求侧	Hyperscaler capex、AI revenue run-rate、GPU/ASIC 平台代际、推理 token 和调用量	季度/月度	判断总需求是否继续扩张
数据中心	签约 MW、已上电 MW、并网进度、机柜功率密度、PUE	月度/季度	验证 L0 和 Neocloud 的真实部署
封装与存储	CoWoS 产能、HBM sold-out、封装良率、基板 load	季度	验证 L1 卡点是否延续
光互连	800G/1.6T 出货、200G/lane 渗透、InP 产能、CPO 量产时间	季度	判断 L2 从叙事到收入的进度
网络与电源	51.2T/102.4T 端口、800V pilot、PSU 效率、功率器件 design win	季度	区分网络硬瓶颈和功率扩散链
液冷	液冷 attach rate、CDU 出货、冷板覆盖率、故障率	季度	验证 L4 工程部署
财务质量	收入增速、毛利率、backlog、存货、应收账款、客户集中度	季度	区分真实兑现和叙事升温
融资质量	债务成本、增发、warrant、租约结构、prepayment、take-or-pay	发生时/季度	识别 Neocloud 和重资产运营商风险
市场状态	EV/Sales、EV/EBITDA、成交量、short interest、机构持仓	周度/月度	判断拥挤度和估值透支

一个方向只有在硬数据和软确认同时改善时，才算 thesis 被验证。硬数据包括收入、毛利、backlog、MW、客户集中度和现金流；软确认包括客户导入、产品从 pilot 到量产、液冷或 800V 进入真实部署。只有叙事热度而没有硬数据，仍然是未验证状态。

## 十、风险框架

风险	传导机制	早期信号	应对方式
Capex 放缓	hyperscaler 下修支出，上游订单同步放慢	财报指引下修、园区延期、GPU 订单节奏放缓	跟踪四大云厂商 capex 和项目进度
产能过剩	卡点扩产后供需缓解，价格和毛利回落	lead time 缩短、backlog 下滑、存货上升	把扩产公告和利用率放在一起看
技术路线变化	CPO、800V、液冷、封装路线变化影响供应商份额	客户延后导入、替代方案发布	跟踪平台路线和 design win
毛利率压缩	扩散受益环节竞争加剧	收入增长但毛利下降	区分卡点供应商和普通供应商
客户集中	单一 hyperscaler 或 anchor tenant 占比过高	最大客户占比上升、续约不确定	审计合同年限和回款质量
融资和稀释	重资产公司靠债务和增发扩张	高息债、warrant、频繁股权融资	分析资产负债表而非只看项目规模
上电延迟	电力、土建、液冷或设备交付拖慢部署	signed MW 多、energized MW 少	区分签约、建设、上电、收入四个阶段
地缘和出口管制	材料、设备、先进封装、HBM 或光器件受限	许可证延迟、地区收入突变	看地区收入和供应链替代
市场叙事退潮	估值先于基本面回落	成交缩量、EV/Sales 压缩、short interest 变化	不把热度当作唯一信号
数据误导	社交媒体或二手资料夸大客户和订单	与财报、公告、客户文件不一致	以一手文件为准

当前最现实的风险不是 AI 需求突然消失，而是三类错配：第一，资本开支与工程部署错配；第二，真实收入增长与估值提前透支错配；第三，合同金额与现金流质量错配。AI 产业链研究必须同时看技术、供需、财务和市场结构。

## 十一、结论

AI 半导体产业链的主线，已经从“谁拥有最强芯片”扩展为“谁能让 AI 基础设施真正部署”。在系统部署周期里，GPU 仍然重要，但仅看 GPU 不够；HBM、先进封装、光互连、电力、液冷、网络、材料、设备和 Neocloud 合同质量，都可能决定下一阶段的供需节奏。

本轮周期最像早期互联网基础设施扩张：长期方向真实，资本开支前置，基础设施先行，瓶颈不断迁移，同时伴随局部泡沫和过度外推。最后真正留下来的，不一定是会讲 AI 故事的公司，而是那些在客户认证、工程交付、材料和设备壁垒、系统集成能力、收入质量和资产负债表上经得起验证的公司。

因此，研究 AI 产业链时，应坚持三条原则。

第一，先看产业位置，再看公司故事。公司必须被放回 L0-L8 的具体层级里判断。

第二，先看硬证据，再看叙事热度。订单、收入、毛利、客户、良率、MW 和回款，比市场口号更重要。

第三，先区分卡点、扩散和映射，再讨论估值。真卡点、扩散受益、高反身性资产和远期叙事，不应使用同一套判断标准。

AI 基础设施不是一个单点行情，而是一张由电力、芯片、封装、网络、光、热、材料、设备和融资共同构成的产业网络。能否把这张网络拆清楚，决定了研究者能否在热闹的 AI 叙事里找到真正可验证的产业信号。