仅供内部参考，请勿外传

通信

光博会启示：看山还是山

“看山是山，看山不是山，看山还是山”——禅宗三重境界，恰是本次光博会传递出的产业认知迭代。1、本次参会人数再创新高，表明行业景气高涨；2、从产业到资本圈，关注度都高，还有预期差吗？3、光通信领域新技术、新概念层出不穷，令人目不暇接，“相信光”还会有新的催化。然而喧嚣之后回归本质，技术演进的方向终究要回到产业底层需求的牵引与工程化落地的约束上来。

【NPO：展会重点，确定性落地方案】

本届光博会最显著的信号之一，是 NPO（近封装光学）的展示密度和产业化成熟度超市场此前预期。

展示层面，主要光通信厂商几乎悉数携 NPO 产品亮相。中际旭创旗下智禾光通、新易盛、剑桥科技、光迅科技、华工正源、Coherent 等诸多光通信厂商均展出各自 NPO 方案、配套外置光源模块，3.2T、6.4T、7.2T 等多款 NPO 模组集中登场，内置光源方案亦在展会亮相。

国内头部云厂商亦对 NPO 积极拥抱。展期会议上国内主要云厂商纷纷展示各自 NPO 方案及部署计划，部分云厂商已与合作伙伴完成样品验证并预计开启试点部署。我们认为，NPO 方案在性能、运维与生态之间取得了务实平衡，是 scale-up 确定性最高、最成熟的落地方案，并正在切实走向工程化落地。

【相干：scale-across 需求与相干下沉 DCN】

Scale-across 驱动相干光模块需求。因算力需求激增、集群规模不断扩张，受单数据中心在电力、散热与物理空间上的容量约束，算力资源不得不分布于多座数据中心，开展协同训练，跨数据中心互联的规模化部署，正驱动 DCI 设备及供应链需求大幅度提升。

相干技术正以“Coherent Lite”形态向数据中心内部 scale-out 场景渗透。随着光模块单通道速率提升，IMDD 系统逐渐无法满足带宽和传输距离要求，在较长距离的互联情景中，400G/lane 速率下 PAM4 的链路预算和色散容忍度均已严重不足。因此，相干技术正在向数据中心内部下沉，Coherent Lite 在保留相干技术核心优势（高灵敏度、高频谱效率）的前提下，通过牺牲部分冗余功能来换取更低的功耗、成本和延迟，适合园区级应用。（图表及详细内容见正文）

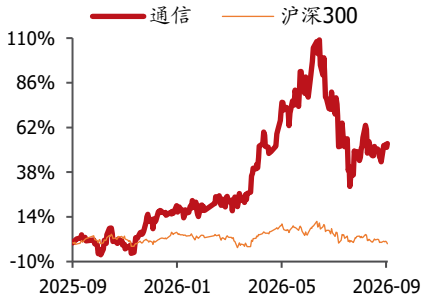
【OCS：参与方多元化与新技术路线】

OCS 参与方从谷歌单一玩家扩展为多极格局，硅光子技术路线潜力可观。

- **买方格局从谷歌单一驱动走向多极需求。**谷歌仍是 OCS 最成熟的用户，其规模化部署构成当前市场的基本盘。与此同时，其他 AI 巨头正加速入场：微软、Meta 已在新一代 AI 集群中测试并引入 OCS 技

增持（维持）

行业走势



作者

分析师 宋嘉吉
执业证书编号：S0680519010002
邮箱：songjiaji@gszq.com
分析师 黄瀚
执业证书编号：S0680519050002
邮箱：huanghan@gszq.com

相关研究

1、《通信：我们为什么选择无源和液冷赛道？》2026-09-06
2、《通信：液冷再出发：弹性、持续性与下一阶段》2026-09-01
3、《通信：预付款里的“发货密码”—从中报看光模块的交付储备》2026-08-30

术；OpenAI 与博通合作的自研推理 ASIC Jalapeño 引入 OCS 方案；英伟达在 OFC 2026 展示 Feynman 架构与 OCS 融合方案。

- **硅光技术路线正在重塑 OCS 的技术经济性。**本次 CIOE 期间多家公司推出了不同技术路径的 OCS 整机解决方案，其中包括硅光方案。传统 MEMS OCS 切换速度在毫秒级；硅光集成方案将光路切换推进至微秒甚至纳秒量级，且单端口成本更低，推动 OCS 从网络资源配置走向近计算侧动态调度。

【扩产与需求：光通信景气度的直观体现】

本周光博会的高频词是“缺货”和“扩产”。头部厂商订单排至明年，扩产力度空前，存货与预付款同步大幅攀升，其中无源器件厂商的品类和产能扩张强度也印证了我们之前的判断。产业端扩产与需求的双向旺盛，正是 AI 驱动下光通信行业景气周期的直观体现。看山还是山——短期筹码与情绪或有波动，长期发展趋势和景气度不变。

我们继续推荐算力产业链相关企业如光模块行业龙头中际旭创、新易盛等，同时建议关注光器件“一大五小”天孚通信+仕佳光子/太辰光/长芯博创/德科立/东田微，PIC 设计可川科技等，建议关注国产算力产业链，如其中的液冷环节英维克、东阳光等。

建议关注：

算力——

光通信：中际旭创、新易盛、天孚通信、源杰科技、太辰光、腾景科技、可川科技、光库科技、光迅科技、德科立、联特科技、华工科技、剑桥科技、铭普光磁、东田微、优迅股份、长光华芯、汇绿生态、联讯仪器。**铜连接：**沃尔核材、精达股份。**算力设备：**中兴通讯、紫光股份、锐捷网络、盛科通信、菲菱科思、工业富联、沪电股份、寒武纪、海光信息。**液冷：**英维克、东阳光、申菱环境、高澜股份。**边缘算力承载平台：**美格智能、广和通、移远通信。**卫星通信：**中国卫通、中国卫星、顺灏股份、海格通信。

IDC：东阳光，润泽科技、光环新网、奥飞数据、科华数据、润建股份。

母线：威腾电气等。

数据要素——

运营商：中国电信、中国移动、中国联通。**数据可视化：**浩瀚深度、恒为科技、中新赛克。

风险提示：AI 发展不及预期，算力需求不及预期，市场竞争风险。

重点标的

股票代码	股票名称	投资评级	EPS（元）				PE			
			2025A	2026E	2027E	2028E	2025A	2026E	2027E	2028E
300308.SZ	中际旭创	买入	9.17	29.35	55.61	74.57	93.70	31.55	16.65	12.42
300502.SZ	新易盛	买入	6.84	15.73	30.10	43.82	60.30	26.90	14.05	9.65
300394.SZ	天孚通信	买入	1.85	3.04	5.75	10.05	144.50	86.20	45.60	26.08
002837.SZ	英维克	买入	0.41	0.94	1.60	2.29	133.70	64.64	37.81	26.48
002463.SZ	沪电股份	买入	1.92	2.40	2.92	-	25.10	53.34	43.91	-

资料来源：Wind，国盛证券研究所

内容目录

1. 投资策略：光博会启示：看山还是山 4
2. 行情回顾：通信板块上涨，光通信表现相对较优 5
3. 光博会启示：看山还是山 6
4. OpenAI Agents API 开放公测：支持代码执行、工具调用和跨上下文任务运行，为开发者提供云端智能体基础设施 9
5. Arm 发布云和数据中心计算子系统 Neoverse CSS N4，支持 LPDDR6、PCIe Gen7 9
6. 海光信息首次推出“词元经营双芯加速方案”加速 AI 要素流转 10
7. LightCounting 曹丽：Scale-Across 重塑 AI 时代 DCI 需求 11
8. DeepSeek 发布最新模型 V4.1 Flash 12
9. 阿里云专家：自研 400G/800G/1.6T 硅光芯片已开启批量部署 13
10. OpenAI 扩大实时语音模型能力：GPT-Live-1 上线 API，支持打断处理、工具调用和电话语音智能体 14
11. 从可插拔到 NPO，AI 算力的“新基石” 14
风险提示 15

图表目录

图表 1：通信板块上涨，细分板块中光通信表现相对较优 5
图表 2：本周金安国纪领涨通信行业 5
图表 3：16QAM 调制格式 6
图表 4：相干光收发器的内部结构 7
图表 5：IMDD/coherent-lite/coherent 对比 8

1. 投资策略：光博会启示：看山还是山

本周建议关注：

算力——

光通信：中际旭创、新易盛、天孚通信、源杰科技、太辰光、腾景科技、可川科技、光库科技、光迅科技、德科立、联特科技、华工科技、剑桥科技、铭普光磁、东田微、优迅股份、长光华芯、汇绿生态、联讯仪器。

铜连接：沃尔核材、精达股份。

算力设备：中兴通讯、紫光股份、锐捷网络、盛科通信、菲菱科思、工业富联、沪电股份、寒武纪、海光信息。

液冷：英维克、东阳光、申菱环境、高澜股份。

边缘算力承载平台：美格智能、广和通、移远通信。

卫星通信：中国卫通、中国卫星、顺灏股份、海格通信。

IDC：东阳光，润泽科技、光环新网、奥飞数据、科华数据、润建股份。

母线：威腾电气等。

数据要素——

运营商：中国电信、中国移动、中国联通。

数据可视化：浩瀚深度、恒为科技、中新赛克。

本周观点变化：

本周海外算力板块整体走势分化。从整体市场看，纳斯达克指数累计下跌 0.9%，标普 500 指数下跌 1.2%。CSP 方面，谷歌本周股价累计下跌 1.1%，亚马逊股价累计下跌 0.8%，meta 股价累计上涨 6.1%，微软股价累计下跌 2.8%。光通信板块，lumentum 本周累计上涨 9.4%，coherent 本周累计上涨 15.5%。AI 芯片板块，英伟达本周累计下跌 4.3%。

我们继续推荐算力产业链相关企业如光模块行业龙头中际旭创、新易盛等，同时建议关注光器件“一大五小”天孚通信+仕佳光子/太辰光/长芯博创/德科立/东田微，PIC 设计可川科技等，建议关注国产算力产业链，如其中的液冷环节英维克、东阳光等。

2. 行情回顾：通信板块上涨，光通信表现相对较优

2026 年 09 月 07 日-2026 年 09 月 13 日上证综指收于 3888.11 点。各行情指标从强到弱依次为：中小板综>创业板综>沪深 300>万得全 A(除金融，石油石化)>万得全 A>上证综指。通信板块上涨，表现强于上证综指。

图表1：通信板块上涨，细分板块中光通信表现相对较优

指数	涨跌幅度
中小板综	-0.6%
创业板综	-0.6%
沪深 300	-0.8%
万得全 A(除金融，石油石化)	-1.0%
万得全 A	-1.1%
上证综指	-1.1%
国盛通信行业指数	4.8%
国盛光通信指数	9.6%
国盛量子通信指数	7.7%
国盛通信设备指数	3.5%
国盛卫星通信导航指数	0.8%
国盛移动互联指数	-0.1%
国盛运营商指数	-2.3%
国盛物联网指数	-3.6%
国盛云计算指数	-4.3%
国盛区块链指数	-8.4%

资料来源：Wind，国盛证券研究所

从细分行业指数看，光通信、量子通信、通信设备、卫星通信导航指数分别上涨 9.6%、7.7%、3.5%、0.8%；移动互联、运营商、物联网、云计算、区块链指数分别下跌 0.1%、2.3%、3.6%、4.3%、8.4%。

本周涨幅前五名中，受益于元件概念，金安国纪上涨 27.63%，领涨板块；受益于通信设备概念，长江通信上涨 17.24%；受益于通信设备概念，武汉凡谷上涨 15.06%；受益于军工电子概念，合众思壮上涨 14.76%；受益于通信设备概念，永鼎股份上涨 13.56%。

图表2：本周金安国纪领涨通信行业

涨跌幅前五名				涨跌幅后五名			
证券代码	证券名称	涨跌幅（%）	成交量（万手）	证券代码	证券名称	涨跌幅（%）	成交量（万手）
002636.SZ	金安国纪	27.627	337.18	300468.SZ	四方精创	-16.291	334.90
600345.SH	长江通信	17.235	81.76	002104.SZ	恒宝股份	-16.014	518.27
002194.SZ	武汉凡谷	15.060	343.52	300096.SZ	ST 易联众	-6.660	14.93
002383.SZ	合众思壮	14.763	225.31	002364.SZ	中恒电气	-5.307	156.24
600105.SH	永鼎股份	13.563	687.97	300050.SZ	世纪鼎利	-4.665	129.48

资料来源：Wind，国盛证券研究所

3. 光博会启示：看山还是山

“看山是山，看山不是山，看山还是山”——禅宗三重境界，恰是本次光博会传递出的产业认知迭代。1、本次参会人数再创新高，表明行业景气高涨；2、从产业到资本圈，关注度都高，还有预期差吗？3、光通信领域新技术、新概念层出不穷，令人目不暇接，“相信光”还会有新的催化。然而喧嚣之后回归本质，技术演进的方向终究要回到产业底层需求的牵引与工程化落地的约束上来。

【NPO：展会重点，确定性落地方案】

本届光博会最显著的信号之一，是 NPO（近封装光学）的展示密度和产业化成熟度超市场此前预期。

展示层面，主要光通信厂商几乎悉数携 NPO 产品亮相。中际旭创旗下智禾光通、新易盛、剑桥科技、光迅科技、华工正源、Coherent 等诸多光通信厂商均展出各自 NPO 方案、配套外置光源模块，3.2T、6.4T、7.2T 等多款 NPO 模组集中登场，内置光源方案亦在展会亮相。

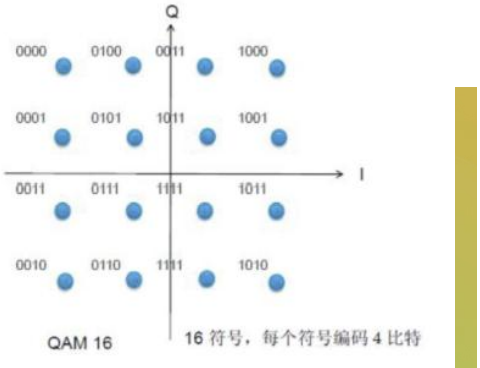
国内头部云厂商亦对 NPO 积极拥抱。展期会议上国内主要云厂商纷纷展示各自 NPO 方案及部署计划，部分云厂商已与合作伙伴完成样品验证并预计开启试点部署。我们认为，NPO 方案在性能、运维与生态之间取得了务实平衡，是 scale-up 确定性最高、最成熟的落地方案，并正在切实走向工程化落地。

【相干：scale-across 需求与相干下沉 DCN】

Scale-across 驱动相干光模块需求。因算力需求激增、集群规模不断扩张，受单数据中心在电力、散热与物理空间上的容量约束，算力资源不得不分布于多座数据中心，开展协同训练，跨数据中心互联的规模化部署，正驱动 DCI 设备及供应链需求大幅度提升。

相干光模块主要采用 DP-16QAM 调制方式，通过幅度和相位共同参与信息编码，并借助双偏振将信息容量翻倍，每符号可承载 8 bit 信息；而传统 IMDD 下的 PAM4 调制仅利用强度一个维度，每符号信息量仅为 2 bit。

图表3：16QAM 调制格式

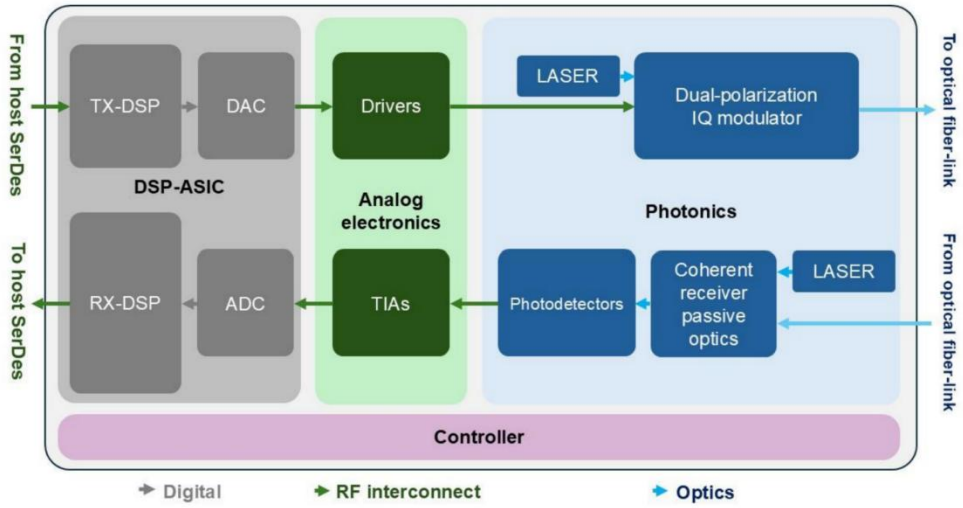


资料来源：《使用 DP-16QAM 调制格式实现 400G 通信》，卓越，国盛证券研究所绘制

相干光模块在接收端结构的复杂程度远高于普通光模块。

- **相干光模块的发射端：**1）光源：相干接收依赖相位信息，而激光器线宽直接决定相位噪声的大小，因此相干光模块需要窄线宽可调谐激光器（ITLA）作为光源。2）调制器：需要利用双偏振 IQ 调制器（DP-IQM）把信息同时加载到振幅和相位上，它由两个 IQ 调制器和一个偏振复用波导组成，IQ 调制器内部由多个马赫-曾德尔调制器（MZM）和移相器组成，将数据编码到光的幅度和相位上，而偏振复用波导将具有正交偏振态的两路调制信号合并。
- **相干光模块的接收端：**集成相干接收机（ICR），内部包含一个 90° 光混合器和四路平衡探测器。它需要一路本振光（通常来自模块内部的可调谐激光器，与发射端共用）与信号光在混合器中干涉。通过干涉，信号光的振幅、相位和偏振态被全部映射到四路电信号（X-I，X-Q，Y-I，Y-Q）上。
- **DSP：**相干光模块须配备强大的 DSP 芯片。接收端 ADC 采样后的四路高速数字信号，需要经过 DSP 依次完成：色散补偿、偏振解复用与均衡、频偏估计、相位恢复、判决和前向纠错（FEC）。正是 DSP 补偿了光纤链路的线性损伤，相干模块才能实现长距离传输，而无需色散补偿光纤。

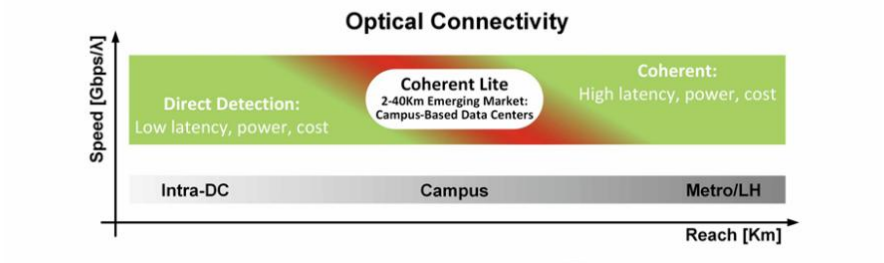
图表4：相干光收发器的内部结构



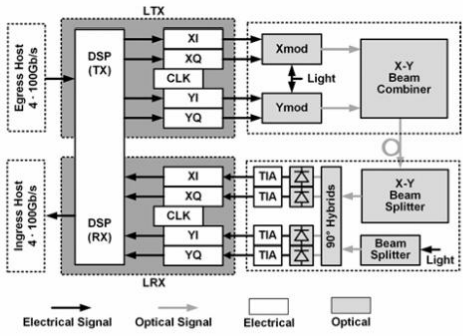
资料来源：《相干光收发器的测试与测量》，是德科技，国盛证券研究所

相干技术正以“Coherent Lite”形态向数据中心内部 Scale-out 场景渗透。随着光模块单通道速率提升，IMDD 系统逐渐无法满足带宽和传输距离要求，在较长距离的互联情景中，400G/lane 速率下 PAM4 的链路预算和色散容忍度均已严重不足。因此，相干技术正在向数据中心内部下沉，Coherent Lite 在保留相干技术核心优势（高灵敏度、高频谱效率）的前提下，通过牺牲部分冗余功能来换取更低的功耗、成本和延迟，适合园区级应用。

图表5：IMDD/coherent-lite/coherent 对比



	Direct Detection	Coherent-Lite	Coherent
Reach	<10Km	2-40Km	>80Km
Modulation	NRZ/PAM4	DP-QAM16	DP-QAM16/64
DSP complexity	Low	Med	High
Analog accuracy	Med	Med	High
Power efficiency	High	Med/High	Low



资料来源：光纤在线，国盛证券研究所

- **简化 DSP：** Coherent-Lite 架构采用 O-band 而非传统长距离 Coherent 系统所使用的 C-band。O-band 接近光纤的零色散波长，因此接收端 DSP 所需的色散补偿均衡器复杂度大幅降低。
- **降低对激光器的要求：**传统相干通常需要昂贵的窄线宽外腔激光器（线宽<500kHz）。Coherent Lite 因为链路短且色散小，容忍度放宽至 1MHz 左右，可以直接沿用 IMDD 系统中便宜成熟的 DFB 激光器，显著降低成本。

【OCS：参与方多元化与新技术路线】

OCS 参与方从谷歌单一玩家扩展为多极格局，硅光子技术路线潜力可观。

- 买方格局从谷歌单一驱动走向多极需求。谷歌仍是 OCS 最成熟的用户，其规模化部署构成当前市场的基本盘。与此同时，其他 AI 巨头正加速入场：微软、Meta 已在新一代 AI 集群中测试并引入 OCS 技术；OpenAI 与博通合作的自研推理 ASIC Jalapeño 引入 OCS 方案；英伟达在 OFC 2026 展示 Feynman 架构与 OCS 融合方案。

- 硅光技术路线正在重塑 OCS 的技术经济性。本次 CIOE 期间多家公司推出了不同技术路径的 OCS 整机解决方案，其中包括硅光方案。传统 MEMS OCS 切换速度在毫秒级；硅光集成方案将光路切换推进至微秒甚至纳秒量级，且单端口成本更低，推动 OCS 从网络资源配置走向近计算侧动态调度。

【扩产与需求：光通信景气度的直观体现】

本周光博会的高频词是“缺货”和“扩产”。头部厂商订单排至明年，扩产力度空前，存货与预付款同步大幅攀升，其中无源器件厂商的品类和产能扩张强度也印证了我们之前的判断。产业端扩产与需求的双向旺盛，正是 AI 驱动下光通信行业景气周期的直观体现。看山还是山——短期筹码与情绪或有波动，长期发展趋势和景气度不变。

我们继续推荐算力产业链相关企业如光模块行业龙头中际旭创、新易盛等，同时建议关注光器件“一大五小”天孚通信+仕佳光子/太辰光/长芯博创/德科立/东田微，PIC 设计可川科技等，建议关注国产算力产业链，如其中的液冷环节英维克、东阳光等。

4. OpenAI Agents API 开放公测：支持代码执行、工具调用和跨上下文任务运行，为开发者提供云端智能体基础设施

据 IT 之家消息，OpenAI 于当地时间 9 月 10 日宣布推出 AgentsAPI 公测版，允许开发者通过 API 调用由 OpenAI 管理的云端 AI 智能体运行环境。

该服务复用了 Codex 背后的智能体执行框架和基础设施，开发者可以指定任务、模型、工具及运行环境，打造能够执行多步骤任务的 AI 应用。

OpenAI 表示，开发者只需一次 API 调用即可创建生产环境级别的智能体。智能体可以在云端运行代码、处理文件、保存中间结果，并持续执行跨越数小时甚至数天的任务。OpenAI 负责维护智能体执行框架，开发者则可以专注于自身应用所需的工具、知识和工作流程。

Agents API 支持 OpenAI 托管的沙盒环境，也支持部署在开发者自有基础设施或合作伙伴提供的沙盒中。OpenAI 已与 Blaxel、Cloudflare、Daytona、DigitalOcean、E2B、Modal、Oracle、Runloop 和 Vercel 等服务商合作，提供不同的 CPU、GPU、内存及存储配置。

IT 之家注意到，OpenAI 同时还推出了 OpenAI Hosted Sandbox 服务，开发者无需自行搭建沙盒基础设施，即可为智能体配置文件、软件包、技能和插件。该环境用于运行代码、处理文件和生成产物，并采用与 Codex 和 ChatGPT 相同的沙盒基础设施。

在长时间运行方面，Agents API 提供上下文管理能力，可在会话接近上下文限制时自动压缩早期信息，帮助智能体跨越多个上下文窗口继续执行任务。该 API 还支持工具搜索和程序化工具调用，智能体可以按需加载工具定义，并行运行调用、串联操作或筛选结果，以减少不必要的上下文占用。

多智能体支持是 Agents API 的另一项能力。开发者可以将复杂任务拆分为多个独立子任务，由子智能体并行处理，再由主智能体协调和汇总结果。OpenAI 提供的示例中，开发者可配置最多 3 个并发子智能体。

Agents API 基于开源 Codex 执行框架打造。OpenAI 表示，开发者可以查看公开代码库，了解协调模型调用、工具和上下文的核心逻辑，而 OpenAI 将负责维护该框架，并随模型更新持续改进智能体能力。

目前 Agents API 已向所有开发者开放公测，不额外收取 API 使用费用，用户仅需按照所使用的模型 Token 和工具付费。OpenAI 表示，公测期间将根据开发者反馈持续迭代，并逐步推进正式发布。

5. Arm 发布云和数据中心计算子系统 Neoverse CSS N4，支持 LPDDR6、PCIe Gen7

据 IT 之家 9 月 8 日消息，Arm 在 Arm Everywhere China 年度大会上发布了其面向云和数据中心的新款计算子系统 (Compute Subsystems) Neoverse CSS N4。

Neoverse CSS N4 面向智能体基础设施，是 Arm 迄今为止可配置性最高的 CSS，可帮助

合作伙伴以最快速度完成从 CSS 到芯片产品的落地。其也是首款支持 CPU 频率高达 3.8 GHz 的 N 系列 CSS、首款支持 LPDDR6 内存和 PCIe Gen 7 的 Neoverse CSS。

Neoverse CSS N4 基于 3nm 工艺，单颗裸片(Die)最高搭载 128 个原生支持 FP8 和 MMLA 的内核，支持 8000~12000MT/s 的 LPDDR6/DDR5/MRDIMM 内存与 128 通道 PCIe Gen7 互连。

与 Neoverse CSS N3 相比，其插槽性能最高可提升 100%，每瓦性能最高可达 1.25 倍，内存带宽最高可达 1.75 倍。

鸿钧微电子是首批采用 CSS N4 开发基于 Arm 架构的差异化基础设施芯片的合作伙伴之一。

6. 海光信息首次推出“词元经营双芯加速方案”加速 AI 要素流转

据 C114 讯，9 月 9 日，以“向智同行，连接算网新生态”为主题的中国联通合作伙伴大会正式开幕。海光信息首次推出“面向词元（Token）经营的双芯加速解决方案”，凭借 CPU 与 DCU 双芯协同及开放互连体系，将算力组网、大模型训练/推理、智能体 Agent 调度与运营纳入统一算力架构，为运营商词元经营提供从芯片到应用的全栈式算力新方案，助力 AI 要素高效流转与产业价值加速释放。

作为芯片厂商，海光信息以“降低百万 Token 成本”目标，将芯片与应用进行深度优化，并构建了涵盖驱动层、算子层、框架层、模型层及应用层的五维优化体系，全面提升系统性能并显著降低运营成本。

面对运营商加速算力网络布局、探索词元经营新模式的战略机遇，该方案紧扣词元“生产、调度、计量、成本核算、应用变现及复购扩容”的完整价值周期，精准匹配真实业务场景。

该架构确立了明确的分工协同机制：CPU 海光智能调度中枢，发挥拓扑感知优势，智能对齐 GPU 内存分配与 NUMA 节点，统筹编排、工具调度及记忆管理等复杂任务；以海光 DCU 为加速计算引擎，支撑并行计算加速、P/D 分离优化、批量解码、KV 缓存优化、预处理加速及长序列高效处理。双芯通过高效的数据协同与内存协同机制，实现业务流量的智能分流与动态调度，为词元规模生产、运营管理和行业落地提供系统级算力支撑。

针对词元经营过程中算力分布不均、成本波动、生态适配及行业需求差异化等挑战，海光方案在算力、互连、安全、软件栈四个维度全面开放，打破传统算力供给壁垒，为国产算力生态协同提供标准化对接框架。

在算力开放方面：支持双芯原生组合及异构芯片交叉适配，灵活满足不同并发规模与词元生产场景；在互连开放方面：基于自研 HSL 高速互连协议，突破带宽瓶颈，大幅提升芯片间数据流转效率；在安全开放方面：内置芯片级内生安全能力，为词元经营构筑全链路可信防护；在软件栈开放方面：全栈开发工具链适配百余种主流 AI 框架，显著降低

迁移与开发门槛，加速词元价值转化。

用户可依据业务并发规模、任务复杂度、词元生成量级动态调度算力资源，助力词元价值向千行百业实体场景高效落地。

为解决框架适配“最后一公里”的兼容性挑战，海光还组建了专门的优化团队，在指令集架构及应用开发阶段即与合作伙伴展开深度交流，并在大模型发布前完成适配工作，确保基于大模型构建的应用在发布后即可直接落地部署。

未来，海光将依托光合生态，持续联合整机、基础软件、行业应用等上下游伙伴深化产品适配与联合创新，推动国产 CPU、DCU 及各类异构加速硬件实现统一调度与协同运行。行业用户可充分利用现有硬件资源，结合业务发展节奏实现分阶段平滑扩容，让国产算力以更高质量、更高效率赋能实体经济数字化转型。

7. LightCounting 曹丽：Scale-Across 重塑 AI 时代 DCI 需求

据 C114 9 月 9 日消息，第二十七届中国国际光电博览会（CIOE 中国光博会）期间举办的“AI 时代高速光传输技术创新论坛”上，LightCounting 高级分析师曹丽发表《AI 数据中心互联市场分析》的主题演讲，系统阐述了数据中心互联（DCI）的需求。

曹丽表示，目前，一类全新的 DCI 需求正在兴起，旨在实现跨数据中心的算力协同。在数据中心与终端用户之间建立连接，通过数据复制和同步，支撑容灾备份与多活业务部署。

主要是因为，随着模型参数规模增长，算力需求激增、集群规模不断扩张，受单栋数据中心在电力、散热与物理空间上的容量约束，算力资源不得不分布于多座数据中心，开展协同训练。由此引出两个关键问题：Scale-across 扩展能延伸多远？需要多大的网络带宽？

目前，分布式模型训练具有多种并行方案。TP（张量并行）、CP（上下文并行）与 EP（专家并行）对带宽和时延要求极高，因此被限制在单个 AI 区域之内；PP（流水线并行）与 DP（数据并行）则可跨楼部署。

将数据并行还是流水线并行置于最外层（DP-out 与 PP-out），取决于模型密度与跨楼链路的带宽收敛比（oversubscription）：DP-out 使用跨楼链路的频率更低，因而更适合稠密模型；而对 MoE（混合专家）模型，随着专家数量增加，DP 数据量显著增长，PP-out 反而成为更优选择。

FSDP（Fully Sharded Data Parallel，完全分片数据并行）并不苛求极高带宽，但对实时时延的要求极为严苛：网络拥塞可能直接击穿计算-通信重叠（overlap）的时间阈值。相比之下，100km 光纤带来约 1ms 的往返传播时延，其开销微不足道。

曹丽介绍，可靠性制约最大传输距离，包括信号劣化：光功率衰减、色散与放大器噪声累积导致光纤链路信噪比（SNR）下降，距离越长，误码率（BER）越高；中间网元增多：额外的硬件引入更多故障点，瞬态丢包概率随之升高；多跳排队抖动：每台设备排队时延的不确定性逐跳累积，放大长距离路径上的时延波动；拥塞丢包加剧：跨楼共享

链路易受突发流量冲击，引发缓冲区溢出与拥塞丢包；环境扰动：光纤振动、温度漂移与机械应力会造成短时 SNR 波动。

从 Meta 发布的论文来看，距离越远，丢包率与时延波动越大。实测数据显示，数据中心跨区域通信的丢包率比楼内通信高出 20 倍以上；地理分布式环境会进一步加剧该问题，丢包率介于 0.1%-10%之间。

对于 AIDC 间的异步数据传输场景，核心诉求是面向大规模跨站点数据迁移的高且可持续的带宽吞吐；此类场景可容忍中等且固定的传播时延，对超低时延和抖动抑制并无严格要求。

目前，头部互联网企业正在打造吉瓦级（GW）数据中心集群，其设计理念是优先打造紧凑型数据中心集群，从而利用短距离灰光（gray optics）模块实现互联。短期内，首要任务有两项：针对 10km 与 100km 两个量级优化光传输架构；让吉瓦级集群的部署能力走向成熟。

曹丽表示，DCI 的长期目标是构建灵活、透明的底层承载网络（underlay），实现无约束的跨区域数据中心算力调度。

目前，Meta 定义了四个典型距离阈值：＜3km：FR4 光模块的典型覆盖范围，对应市内（intra-city）超短距 DCI；＜10km：LR4 光模块的典型传输距离，对应市内/近城域（intra-city/near-metro）场景；约 600km：城际（inter-city）区域 DCI，是 ZR/ZR+光模块的主要工作范围；约 1200km：区域（regional）互联的实际上限，更远距离归入骨干网范畴。

对于 DCI 的新技术，有高密度线缆、增加新路由，谷歌认为的光纤即新波长，IM/DD 色散补偿技术，Coherent lite 和相干模块。目前正在从 400G 向 800G 升级。

在 DCI 的市场预测方面，LightCounting 数据显示 DWDM 光模块的销售额在 2025 年超过 20 亿美元，2026 年将超过 30 亿美元，到 2030 年将突破 110 亿美元。

8. DeepSeek 发布最新模型 V4.1 Flash

据 C114 网报道，9 月 10 日，DeepSeek V4.1 Flash 模型正式发布。

DeepSeek 方面表示，这是该公司全新模型结构系列中的最小尺寸的模型，具备原生多模态视觉理解能力。新模型结构的设计初衷是：能力上限更高、推理速度更快、吞吐更大、可扩展到更大参数模型。

据悉，DeepSeek V4.1 Flash 为 552B 参数的 MoE 模型，采用了全新的 Causal-Encoder-Decoder 结构，输入和输出不对称，输入激活只有 8B，输出激活 16B，成本显著低于已知的同尺寸模型。同时，V4.1 Flash 采用了新的预训练方式、经过了更大规模的强化学习后训练，在基准测试中，成功超越了包括 DeepSeek V4 Pro 在内的一众旗舰模型的智能水平。

同时，新一代模型大幅减少了 KV Cache 缓存的大小，与上一代模型相比，对 HBM 的需求减少到 1/4，对 SSD 的需求减少到 1/8。在 Agent 使用场景中，缓存命中的费用往往占比较高，对 KV Cache 的压缩大幅降低了 Agent 类任务的使用成本。

伴随新模型的发布，DeepSeek 同步调整了 API 定价。该公司称，DeepSeek V4.1 Flash 能够以更低的成本服务更多的用户，因此相应下调了 V4.1 Flash 的定价——依然采用峰谷定价，闲时价格为高峰时段价格的一半，鼓励用户根据实际使用情况调整任务时间。

腾讯（WorkBuddy、CodeBuddy）和 OpenCode 作为官方合作伙伴，现已全量接入 DeepSeek V4.1Flash。

9. 阿里云专家：自研 400G/800G/1.6T 硅光芯片已开启批量部署

据 C114 9 月 10 日消息，光电融合作为后摩尔时代的核心突破方向，正以万亿级市场规模的强劲势头重塑产业格局，推动光通信行业全面迈入架构革命新时代。2026 年光通信行业的技术迭代主线已清晰可见，硅光技术加速替代传统 EML 技术，尤其是在 800G、1.6T 等高端产品中，硅光技术的应用比例大幅提升。

9 月 10 日，在 CIOE 中国光博会同期举办的“光电融合技术与产业发展论坛”上，阿里云光网络技术专家尹延龙发表题为《数据中心高速高密光互联技术演进趋势》的主旨演讲，围绕硅光芯片、Scale Up 光网络与 NPO 等热点，分享了阿里云在高速高密光互联领域的技术实践与判断。

AI 算力需求旺盛，正推动数据中心智算光网络带宽和速率的快速提升，光互联技术更新迭代，从 400G、800G 迈向 1.6T、3.2T 甚至 6.4T、12.8T。“阿里云数据中心已批量部署百万级 400G、800G 光互联模块，1.6T 光模块预计下半年进入量产，同步开启 3.2T/6.4T NPO 研发验证。”尹延龙说。

尹延龙介绍，随着 AI 模型的发展，算力规模的扩大，催生 scale up 光网络快速落地，呈现“近端铜、远端光”，“铜进光退”多种趋势发展。在光互联技术路线上，目前硅光产品满足了跨数据中心的相干需求，在数据中心内部，已覆盖单波 100G 和单波 200G 的短距光互联产品，同时在 NPO、CPO 以及 Scale Up 中都已启动新产品开发。“硅光技术相比于 VCSEL 以及 EML 技术，具备成本、供应、国产化等优势，在国内需求占比显著提升。”

阿里云全栈自研 100G/Lane 硅光互联芯片，首款单波 4X100G 硅光芯片性能优异，支持 FRO/LPO 等产品开发，并实现量产，率先完成规模部署。基于全国产硅光工艺线，阿里云已经启动单波 200G 产品硅光互联芯片的研发及量产，调制带宽>65GHz，满足数据中心光模块的指标需求。

尹延龙指出，面向单波 400G，纯硅光平台、铌酸锂平台、InP 平台等均在加速探索。在 400G-IMDD 时代，InP 平台仍表现出更为优异的带宽性能。

在 Scale up 高速高密光互联的技术选择上，尹延龙表示，NPO 开放解耦的技术优势，相比 CPO 具备更快的落地节奏；相比于可插拔，具备更好的高速链路传输性能，在 scale Out 和 Scale Up 光网络中成为重点关注的光互联技术。他同时指出，NPO 落地应用中，光引擎仍面临高密耦合、高速封装等技术难点，生态完备性、方案标准化需要时间。

10. OpenAI 扩大实时语音模型能力：GPT-Live-1 上线 API，支持打断处理、工具调用和电话语音智能体

据 IT 之家 9 月 11 日消息，OpenAI 宣布将 GPT-Live-1 引入 API，开发者可据此打造支持实时语音交互的应用和业务流程。该模型支持全双工对话，能够同时听取和输出语音，并在对话过程中处理打断、停顿及背景噪声。

GPT-Live-1 还支持电话场景，可用于预订餐厅、客户服务等全双工语音智能体。OpenAI 表示，开发者可以将其与 Codex 等工具连接，也可以通过 OpenAI Presence 开发能够查询企业系统、执行获批操作并在必要时转交人工的语音工作。

据介绍，GPT-Live-1 将语音理解与语音输出整合在同一模型中，减少传统“语音转文字—大语言模型—文字转语音”的多次衔接，从而降低延迟。模型还支持将复杂推理和工具调用交由后端文本模型处理。

开发者可以通过系统提示词调整其语气、语速和对话风格，也可以配置后端模型、工具及智能体框架。

OpenAI 表示，GPT-Live-1 支持原生语音识别转写和回复文本，并具备字母数字理解、关键词偏置及轮次检测能力。

在早期评估中，语言学习平台 Speak 使用 GPT-Live-1 后，学习者思考停顿期间的误打断次数较此前基于轮次的系统减少近 80%。医疗服务平台的联合创始人兼首席技术官 Tony Stoyanov 表示，其团队将代码量减少了 80%，删除约 2.3 万行代码。

OpenAI 公布的评测结果显示，GPT-Live-1 在 Full Duplex Bench 测试中的表现比 GPT-Realtime-2.1 高出 30 个百分点，在搭配 GPT-6 Astra 中等推理强度时，Tau3 端到端语音智能体任务测试排名第一。

目前，GPT-Live-1 已在 API 中提供，前端语音层价格为每分钟 0.05 美元。按每小时 60 分钟计算，单纯前端语音层的费用约为 3 美元（IT 之家注：现汇率约合 20.2 元人民币），后端模型和智能体工具的费用另行计算。

OpenAI 同时扩大了实时语音选项，新增 Quartz、Ripple、Vesper、Willow、Stone、Gleam、Meridian、Bossa、Tempo、Beacon、Delta 和 Cinder 等声音，并计划在未来数月继续增加语音和语言支持。

11. 从可插拔到 NPO，AI 算力的“新基石”

据 C114 网报道，大模型的迭代正把 AI 算力集群从千卡推向万卡乃至十万卡规模。当算力成为新的生产力，"如何连接"正在成为与"如何计算"同等重要的命题——作为算力"血管系统"的光互连，也从网络幕后走向算力舞台的中央。

9 月 7 日至 8 日，IFOC 2026 · 第 24 届讯石光通信大会在深圳洲际酒店举行。光迅科技受邀参与本届大会：营销市场部副总经理蒋波出席圆桌论坛，市场总监高郢发表主题演讲。两场议程，一条主线——用光互连，把 AI 算力真正连接起来。

9 月 8 日，圆桌论坛在深圳洲际酒店 2F 伊敦大宴会厅如期举行。光迅科技营销市场部副总经理蒋波与产业链各方嘉宾同台，围绕 AI 算力浪潮下光互连的路线之争、产业协同与市场格局展开面对面交锋，共话光互连的现在与未来。

光通信从“连接底座”逐步走向“算力中枢”，光模块在整个产业链中的角色和定位也随之发生了变化，光模块不仅仅是一个连接器件，更是成为了整个算力网络中的关键节点，行业竞争也从单一的功能成本竞争，演化为芯片、架构、系统的一体化竞争。

Scale Up 中光的渗透率越来越高，CPO、NPO、OIO 等相关的新技术新架构也层出不穷。目前来看，NPO 在功耗、密度、可维护性、供应链成熟度之间可以取得最大公约数，是现阶段工程落地综合最优解；CPO 则在理论上功耗最低、带宽密度最高，非常适合超高带宽集群，但是产业链配套还不成熟，大规模部署仍有大量性能可靠性、可制造性、可维护性等现实问题需要解决；OIO 则是把光 I/O 做到计算芯片内部，技术想象空间最大，但需要芯片架构、封装、光器件等全链条重构，属于远期技术储备。

光迅科技作为国内领先的光器件/模块供应商，坚持多技术路线并行，致力推动“芯片-器件-模块-标准”的全产业链开放协同，兼顾技术先进性、工程可落地性、供应链可靠性与运维现实性。

当光互连离计算芯片"越来越近"，NPO 正从技术幕后走向产业话语的中心。本届大会上，光迅科技市场总监高郢发表主题演讲《NPO 构筑 AI 光互联的新基石》，系统回应这一命题。

AI 集群的互连带宽与功耗正在迅速攀升，传统可插拔光模块在交换机面板密度与整机功耗上渐近极限；CPO 以极致功耗与密度给出远期答案，却也同时重置了运维模式与产业生态。在这两者之间，NPO（近封装光学）把光引擎推向交换芯片与计算芯片的近旁，在带宽密度、系统功耗与可维护性之间取得平衡，正成为 scale-up 与 scale-out 中兼具现实性与前瞻性的选择。

演讲中，高郢系统梳理了 AI 集群光互连架构的演进脉络，解析 NPO 的技术定位与工程化关键点，并结合光迅科技在 NPO 方向的产品布局与实践，给出了 AI 光互连"新基石"的一条可落地路径。

9 月 9 日至 11 日，第 27 届中国国际光电博览会（CIOE）在深圳国际会展中心（宝安）开幕，光迅科技展位号 11B53。2026 年恰逢光迅科技成立五十周年，从器件起步到全栈光互连，我们期待与业界新老朋友在深圳相聚——聊技术、聊产业、聊下一个五十年的光。

风险提示

AI 发展不及预期，算力需求不及预期，市场竞争风险。

免责声明

国盛证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司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。

本报告的信息均来源于本公司认为可信的公开资料，但本公司及其研究人员对该等信息的准确性及完整性不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，可能会随时调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司力求报告内容客观、公正，但本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。

本报告版权归“国盛证券股份有限公司”所有。未经事先本公司书面授权，任何机构或个人不得对本报告进行任何形式的发布、复制。任何机构或个人如引用、刊发本报告，需注明出处为“国盛证券研究所”，且不得对本报告进行有悖原意的删节或修改。

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，本报告所表述的任何观点均精准地反映了我们对标的证券和发行人的个人看法，结论不受任何第三方的授意或影响。我们所得报酬的任何部分无论是在过去、现在及将来均不会与本报告中的具体投资建议或观点有直接或间接联系。

投资评级说明

投资建议的评级标准		评级	说明
评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准。	股票评级	买入	相对同期基准指数涨幅在 15%以上
		增持	相对同期基准指数涨幅在 5%~15%之间
		持有	相对同期基准指数涨幅在-5%~+5%之间
		减持	相对同期基准指数跌幅在 5%以上
	行业评级	增持	相对同期基准指数涨幅在 10%以上
		中性	相对同期基准指数涨幅在-10%~+10%之间
		减持	相对同期基准指数跌幅在 10%以上

国盛证券研究所

北京
地址：北京市丰台区丽泽路 22 号院新华社国家金融信息大厦西塔 B 座 38 层
邮编：100077
邮箱：gsresearch@gszq.com

上海
地址：上海市浦东新区南洋泾路 555 号陆家嘴金融街区 22 栋
邮编：200120
电话：021-38124100
邮箱：gsresearch@gszq.com

南昌
地址：南昌市红谷滩新区凤凰中大道 1115 号北京银行大厦
邮编：330038
传真：0791-86281485
邮箱：gsresearch@gszq.com

深圳
地址：深圳市福田区福华三路 100 号鼎和大厦 24 楼
邮编：518033
邮箱：gsresearch@gszq.com