

AI 巨头极布局 上游， 弊热也 需注意风险

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

随着“内存墙”的带宽限制不断累积，光子元件、基板材料、核心设备、CPO 等符合产业趋势的 AI 环节确定性受益，字节跳动可能正通过海外布局，获取英伟达 AI 芯片，阿里千问郁博文加入字节跳动，火山引擎正式上线 ArkClaw，国家互联网应急中心发布关于 OpenClaw 安全应用的风险提示。

资本风向

英伟达与 Nebius 宣布达成战略合作伙伴关系，Meta 计划在 2027 年底前部署四款新一代自研 AI 芯片，以应对其快速扩张的 AI 计算需求，为应对芯片紧缺情况，马斯克计划建造芯片工厂，谷歌近日发布原生多模态嵌入模型 Gemini Embedding 2，Meta 最新 AI 模型推迟发布，性能表现不如同行，腾讯旗下全场景 AI 智能体 WorkBuddy 正式上线。

本周观点

近期 NV 多次增强对供应链的投资，供养产业链的决心非常坚定。国内腾讯、百度、阿里巴巴、字节跳动、美团、京东等互联网大厂，小米、华为、荣耀等手机大厂等科技大厂均发布了一键部署、接入 OpenClaw 的相关方式，国产算力需求加速向上。甲骨文第三季度剩余履约义务（RPO）达到 5530 亿美元，同比大增 325%，较上季度末再增 290 亿美元，公司将 2027 财年总营收指引上调至 900 亿美元，较此前预期大幅提高，传递出管理层对 AI 需求持续爆发的高度信心。继续看好 AI 产业链的高景气。

内容目录

一、产业前沿	2
AI 基础层: AI 发展推动上游新技术落地	2
AI 模型层: OpenAI 的 IPO 受到投资者关注	2
AI 应用层: 小龙虾火热, 安全应用值得重视	3
二、资本风向	4
AI 基础层: 英伟达、Meta、马斯克等布局上游, 光通信高景气	4
AI 模型层: Meta 最新 AI 模型推迟发布	7
AI 应用层: Oracle 业绩显示 AI 供不应求, 腾讯推出智能体	8
三、本周观点	10
风险提示	10

一、产业前沿

AI 基础层：AI 发展推动上游新技术落地

2026 年 3 月 12 日 CITRINI 发文称，随着“内存墙”的带宽限制不断累积，机遇正孕育于那些推动下代技术的企业——光子元件、基板材料、核心设备制造商，以及当共封装光学 (CPO) 成为必然趋势时将受益的相关企业。

根据 TrendForce 集邦咨询最新高速互连市场研究，NVIDIA(英伟达)下一代的 AI 算力柜架构显示，未来 GPU 设计重心将转向更高密度的芯片互连，以及更高速的数据传输，机柜内芯片互连 (Scale-Up) 及跨机柜的大规模互连 (Scale-Out) 将成为规划数据中心的核⻝课题。使用铜缆的传统电气传输方案，受物理限制无法应对超大规模的数据搬运需求，光学传输方案因此获得发展空间。

TikTok 母公司字节跳动正通过海外布局，大规模获取英伟达最尖端的 AI 芯片。据《华尔街日报》3 月 12 日报道，字节跳动正在中国境外组建价值超过 25 亿美元的 AI 计算系统，其中包含被美国政府禁止对华直接销售的 Blackwell 系列 (B200) 芯片。

AI 模型层:OpenAI 的 IPO 受到投资者关注

OpenAI 的 IPO 希望面临投资者群体的质疑。据一位收到接洽的市场人士透露，OpenAI 距离上市可能至少还需六个月，甚至更久。但随着该公司逐步推进首次公开募股计划，试图分一杯羹的投资银行正在联系公开市场投资者，以评估他们对 OpenAI 作为上市公司的前景预期。市场情绪喜忧参半。投资者预计该公司将在首次公开募股中筹集数百亿甚至数千亿美元，这意味着其规模可能轻松超越沙特阿

美 2019 年创纪录的 250 亿美元 IPO。但许多公开市场投资者-即便是那些看好 OpenAI 在人工智能竞争格局中有利地位的特保留态度。

此外，投资者还在质疑 OpenAI 通过首次公开募股筹集的资金是否足以支撑其实现盈利，或者未来是否需要再次融资，这可能稀释现有股东的权益。

阿里千问郁博文加入字节跳动。原阿里巴巴通义实验室 Qwen（千问）大模型后训练负责人郁博文已正式加入字节跳动，担任 Seed 团队视觉模型与多模态交互团队后训练负责人。这是继今年 3 月初郁博文从阿里离职后，其职业动向的首次明确。字节对此暂无回应。

AI 应用层：小龙虾火热，安全应用值得重视

本周，海外聊天助手类应用活跃度中，Gemini 和 Claude 持续快速上升，ChatGPT 小幅下降 Perplexity 保持稳定。国内多数聊天类应用网页端假期结束后活跃度回升后开始企稳。

办公自动化方面，Anthropic 对 Claude 的 Excel 与 PowerPoint 加载项进行重大升级，新增共享上下文和可复用技能功能，实现跨应用数据无缝协同及复杂 workflow 的一键调用。

国家互联网应急中心发布关于 OpenClaw 安全应用的风险提示。近期，OpenClaw（"小龙虾"，曾用名 Clawdbot、Moltbot）应用下载与使用情况火爆，国内主流云平台均提供了一键部署服务。此款智能体软件依据自然语言指令直接操控计算机完成相关操作。为实现"自主执行任务"的能力，该应用被授予了较高的系统权限，包括访问本地文件系统、读取环境变量、调用外部服务应用程序编程接口 (API) 以及安装扩展功能等。然而，由于其默认的安全配置极为脆弱，攻击

者一旦发现突破口，便能轻易获取系统的完全控制权。建议相关单位和个人用户在部署和应用 OpenClaw 时，采取一系列安全措施。

《科创板日报》9 日讯，今日，火山引擎正式上线 ArkClaw——开箱即用的云上 SaaS 版 OpenClaw。当前，“火山方舟 Coding Plan”用户可以抢先体验 ArkClaw，“Coding Plan Pro”用户登陆即可同步使用 ArkClaw，“Coding Plan Lite”用户可以免费体验 7 天。订阅后，ArkClaw “龙虾”可以畅享 Doubao-Seed-2.0 系列、Kimi2.5、MiniMax2.5 和 GLM 等主流模型，ArkClaw 协同 Doubao-Seed 2.0 Pro，复杂任务效果突出。

二、资本风向

AI 基础层：英伟达、Meta、马斯克等布局上游，光通信高景气

英伟达与 Nebius 宣布达成战略合作伙伴关系，共同开发和部署面向 AI 市场的下一代超大规模云服务，且将向 Nebius 投资 20 亿美元。近期 NV 多次增强对供应链的投资，供养产业链的决心非常坚定。

3 月 11 日，据彭博报道，计划在 2027 年底前部署四款新一代自研 AI 芯片，以应对其快速扩张的 AI 计算需求。目前，最新一代芯片已投入内容排序与推荐系统的训练任务；第二款芯片完成实验室测试，正推进部署；另有两款芯片定于 2027 年批量上线。

尽管自研芯片有助于降低对外部供应商的依赖并压缩长期成本，但 Meta 并未削减对外采购规模。该公司近期分别与英伟达和 AMD 签署价值数百亿美元的采购协议，锁定未来数年数吉瓦级的 AI 算力产能，形成“自研+外采”双轨并行的供给体系。

为应对芯片紧缺情况，马斯克计划建造芯片工厂，3月15日，马斯克发文称 TeraFab 项目将在“7 天后”启动，他希望实现每年 1000 亿~2000 亿颗芯片的产量，这一规模将使其成为全球最大芯片工厂之一。

3月11日晚，博通官网发布，其 3nm400G/通道光 PAM-4DSP——Taurus BCM83640 上市，该处理器针对 1.6T 收发器解决方案进行了优化，具备前所未有的带宽密度和效率。该设备配备 400G/通道串行光接口，使光收发器制造商能够以经济实惠的方式提供低功耗 1.6T 可插拔模块，以满足 AI 数据中心日益增长的带宽需求。

性能再升级，为更高性能 AI 网络互联奠定基础。采用 Taurus 的 1.6T 可插拔模块 BCM83640 使每条光通道带宽翻倍，有效实现 1RU 系统中的 102.4T 交换容量，以提升 AI 光互连的带宽密度。此外，采用 400G/通道光接口为最终部署 3.2T 可插拔光模块奠定基础，该模块解决方案配备 400G/通道电气接口，适用于 204.8T 交换机。

随着人工智能算力集群的加速扩张，数据中心内部高速互连需求正迎来爆发式增长。美国光模块制造商 Applied Optoelectronics Inc.（简称 AOI，纳斯达克代码：AAOI）于 3月9日宣布，已获得来自一家大型超大规模客户的首笔 1.6T 数据中心光收发器批量订单，订单总额超过 2 亿美元。

这笔订单主要用于提升面向 AI 工作负载的网络带宽。AOI 表示，随着产品认证完成，预计将于 2026 年第三季度开始出货，并在第四季度完成全部交付。公司预计，该订单将使这一客户重新成为其收入贡献占比达 10% 或以上的核心客户。分析人士指出，当前 AI 大模型训练对算力集群内部的通信带宽提出了极高要求，

传统的 400G 正加速向 800G 乃至 1.6G 过渡。此次大额订单的落地，标志着 1.6T 光模块已从技术验证走向规模化商用，为高速光模块产业链释放积极信号。

3 月 11 日，腾景科技在互动平台回答投资者提问时表示，公司具备高密度连接、低损耗耦合、高精度对准等技术能力，自研 CPO（共封装光学）光连接器产品已在开发验证中。

2026 年 3 月 12 日，全球高速光互联解决方案领导者和创新者，新易盛宣布，将在 2026 年 OFC 展会上展示其新一代基于 IMDD（强度调制直接检测）400G/λ 技术的 1.6T DR4 OSFP 光模块。

新易盛此次展示的 1.6T DR4 OSFP 光模块，集成了最先进的 (8:4) PAM4 数字信号处理器 (DSP)，可将主机侧 8×200G 电气接口转换为 4×400G 光学接口。其中，GAUI-8 电气接口的采用，使得该模块可直接适配当前主流的 102.4T 交换机，具备极强的兼容性和实用性。

与上一代基于 8×200G 方案的 1.6T DR8 模块相比，该解决方案凭借 4×400G 光学接口设计，可帮助客户减少 50% 的光纤使用量，显著降低数据中心布线复杂度和总拥有成本。同时，该模块还为现有 102.4T 交换机与未来采用 400G 每通道传输接口的平台之间，提供了平滑的迁移路径，助力客户实现基础设施的高效升级。

“该产品成功验证了 400G 每通道光学接口的技术可行性，”新易盛杰出工程师德克·卢茨 (Dirk Lutz) 表示，“这款业界首创的模块让我们能够对 400G 每通道传输进行测试和特性分析。除了评估光学性能外，我们还可在现有交换机中开展功能测试，从而更深入地了解下一代光学接口的系统”

AI 模型层：Meta 最新 AI 模型推迟发布

谷歌近日发布原生多模态嵌入模型 Gemini Embedding 2，可将文本、图像、视频、音频及 PDF 文档统一映射至同一语义向量空间，支持最长 8192 个文本 Token 及原生音频处理，并在多项跨模态检索基准中领先。

Meta 最新 AI 模型推迟发布，性能表现不如同行。Meta 从去年 12 月研发一款代号为“牛油果 (Avocado)”的文本人工智能模型，并计划在第一季度推出。然而最新消息称，Meta 已将“牛油果”人工智能模型代码的发布时间从 3 月至今推迟至 5 月。据知情人士透露，Meta 耗时数月研发了这款 AI 模型，但其在内部推理、编码和写作测试中，性能表现却不及谷歌、OpenAI 和 Anthropic 等竞争对手的领先人工智能模型。

阿里 Qwen 管理调整出炉：周靖人问模型一号位，刘大一恒管理更多团队。在上周林俊暘突然离职后，阿里在 3 月 9 日下午有了新的管理安排：Qwen 模型一号位由阿里云 CTO 和通义实验室负责人周靖人代管，他会深入了解模型发展需要的资源，提升各环节协作效率，确保模型高效迭代。负责 Qwen 预训练的刘大一恒，则将同时代管后训练和 Coding 团队。刘大一恒和 Qwen 模型团队的其他 leader 向周靖人汇报。

OpenRouter 平台匿名上线 Hunter Alpha 与 Healer Alpha 两款神秘新模型，前者拥有 1T 参数与 100 万 Token 超长上下文，专攻智能体工作流，后者支持 262K 上下文且具备全模态能力，社区普遍推测两者可能为智谱 AI 的新一代旗舰大模型。

AI 应用层：Oracle 业绩显示 AI 供不应求，腾讯推出智能体

Oracle: 关注未来产能能否按期交付、RPO 能否高质量转收入。Oracle 2025Q4 (FY2026Q3) 总收入 172 亿美元，同比增长 22%；按固定汇率计算，云应用同比增长 11%，达到 161 亿美元。其中 Fusion ERP 增长 14%，Fusion SCM 增长 15%，Fusion HCM 增长 15%，Fusion CX 增长 6%，NetSuite 增长 11%，而面向酒店、建筑、零售、银行等行业的行业 SaaS 整体增长 19%。这说明 Oracle 传统 SaaS 业务依然稳健，并且随着 AI 能力逐步嵌入应用系统，其产品竞争力仍在增强，但在当前，它更多扮演的是稳定现金流与客户基础的角色，而非最核心的增长引擎。

多云数据库收入同比增长 531%，AI 基建收入同比增长 243%。这两个指标显示 Oracle 在 AI 算力基础设施与跨云数据库部署方面的需求正在快速爆发。与此同时，公司当前拥有 5530 亿美元的 RPO，并明确指出无论是 GPU 还是 CPU 仍然持续超过供给。这意味着 Oracle 目前面临的核心问题已经不再是需求不足，而是如何将庞大的订单池逐步转化为可交付的计算能力并最终转化为收入。

在 AI 基础设施业务层面，Oracle 表示未来三年通过合作伙伴已经锁定超过 10GW 的数据中心和电力容量，其中超过 90% 的容量已经由合作伙伴完成资金支持。过去一年公司制造基地数量增加至原来的三倍，机架产量提升四倍，同时从机架交付到产生收入的时间已经缩短 60%。在实际交付方面，公司在 2025Q4 向客户交付了超过 400MW 的算力容量，且 90% 的承诺容量按计划或提前完成交付。这些数据表明 Oracle 的 AI 基础设施业务已经从订单签约阶段进入到工程化放量阶段。

我们认为仍需关注 Oracle 在 Stargate 项目上的进展情况，尽管 2025Q4 完成了超过 400MW，但在 3 月 6 日，Oracle 与 OpenAI 宣布其在 Abilene 的

扩建项目缩减提高了未来数据中心交付不及预期的风险。同时 Oracle 自由现金流情况不佳，也为未来资本开支能否顺利实现打上问号。

谷歌宣布基于 Gemini 的任务自动化功能进入 Beta 测试，AI 可直接模拟人类 UI 操作完成打车或订餐等复杂流程，并设有强制人工复核机制保障安全。

腾讯：旗下全场景 AI 智能体 WorkBuddy 正式上线。3 月 9 日，腾讯宣布，旗下全场景 AI 智能体 WorkBuddy 正式上线，该产品能力跟目前火热的“小龙虾” OpenClaw 类似，无需复杂的部署环节，从下载安装到连接企微，最快 1 分钟，用户就可以通过手机“遥控”腾讯版小龙虾干活。

腾讯：目前已初步形成覆盖个人、开发者及企业级部署的智能体“养虾”矩阵。继 3 月 9 日腾讯版“小龙虾”（WorkBuddy）发布引发广泛关注后，腾讯目前已初步形成覆盖个人、开发者及企业级部署的智能体“养虾”矩阵。在提供核心部署方案的同时，腾讯同步完善了国民级应用场景中的 AI 生态支持。在办公协同方面，企业微信官方已正式降低小龙虾接入标准，开发者仅需三步即可通过 API“长连接”模式将智能体无缝接入企微，并结合 Webhook 技术实现数据向企业“智能表格”的自动流转。

腾讯秘密启动微信 AI 智能体项目。3 月 10 日，据外媒腾讯正秘密为其广受欢迎的微信通讯应用打造一款新型人工智能代理，以期在中国本土人工智能市场的竞争中超越阿里巴巴集团和字节跳动等对手，该项目可能极大拓展人工智能代理在中国的应用范围——这类代理能代替人类执行任务，而微信早已深度融入人们的日常生活。这款新型智能体将连接微信内运行的数百万个提供各类服务的小程序，涵盖从预约出租车到订购杂货等众多领域。

三、本周观点

英伟达与 Nebius 宣布达成战略合作伙伴关系，共同开发和部署面向 AI 市场的下一代超大规模云服务，且将向 Nebius 投资 20 亿美元。近期 NV 多次增强对供应链的投资，供养产业链的决心非常坚定。国内腾讯、百度、阿里巴巴、字节跳动、美团、京东等互联网大厂，小米、华为、荣耀等手机大厂等科技大厂均发布了一键部署、接入 OpenClaw 的相关方式，国产算力需求加速向上。甲骨文第三季度剩余履约义务 (RPO) 达到 5530 亿美元，同比大增 325%，较上季度末再增 290 亿美元，公司将 2027 财年总营收指引上调至 900 亿美元，较此前预期大幅提高，传递出管理层对 AI 需求持续爆发的高度信心。

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。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）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。