



AI 周观察

数据专题研究
证券研究报告

国金证券研究所

分析师：刘道明（执业 S1130520020004） 分析师：黄晓军（执业 S1130525080005）
liudaoming@gjzq.com.cn huangxiaojun@gjzq.com.cn

英伟达一季度业绩优异，智谱推出高速版 GLM 5.1

摘要

- 本周海外 AI 应用中 Claude 活跃度持续攀升，ChatGPT、Gemini 及国内多数同类产品表现平稳；同时国内大厂在 AI 模型与系统级应用上迎来密集爆发：通义实验室发布新一代 Agent 基座大模型 Qwen3.7-Max，通过创新的解耦设计攻克长链路运行崩溃痛点，实现跨框架泛化与超长零中断自主推理；腾讯推出跨平台操作系统级 AI 助手“马维斯”，采用多智能体协同架构，打通系统文件与操作并支持纯本地断网的隐私模式；智谱则面向企业开放了 GLM-5.1 高速版 API，凭借底层深度的系统级重构，实现了刷新全球上限的 400 tokens/s 极速输出，为各类低延迟实时交互场景提供了强大的底层支持。
- 英伟达在 2026 年 1 季度（FY2027Q1）实现极为强劲的增长，总营收达到 820 亿美元，同比增长 85%，环比增长 20%，并实现连续第 14 个季度环比增长。数据中心业务收入达到 750 亿美元，同比增长 92%，环比增长 21%，其中数据中心计算收入 600 亿美元，同比增长 77%，数据中心网络收入 150 亿美元，同比增长近 3 倍。管理层表示 GB300 NVL72 需求强劲，前沿模型公司与超大规模云厂商已累计部署数十万颗 Blackwell GPU，成为公司历史上爬坡最快的产品。网络业务方面，公司强调 Spectrum-X 已经成为“规模超过所有其他以太网 AI 网络同行总和”的平台，而 InfiniBand 收入同比增长超过 4 倍，主要由 XDR 部署推动。盈利能力方面，公司 Q1 GAAP 毛利率为 74.9%，Non-GAAP 毛利率为 75%，环比基本持平。展望 Q2，公司预计营收 910 亿美元，上下浮动 2%，增长主要来自数据中心业务，同时对 2025 至 2027 自然年累计实现 1 万亿美元 Blackwell 与 Rubin 收入保持信心。

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
Qwen 3.7 Max 发布，智谱推出高速版 GLM 5.1.....	4
英伟达财报表现优异.....	5
风险提示.....	6



海外市场行情回顾

图表1: 截至5月22日海外AI相关个股行情

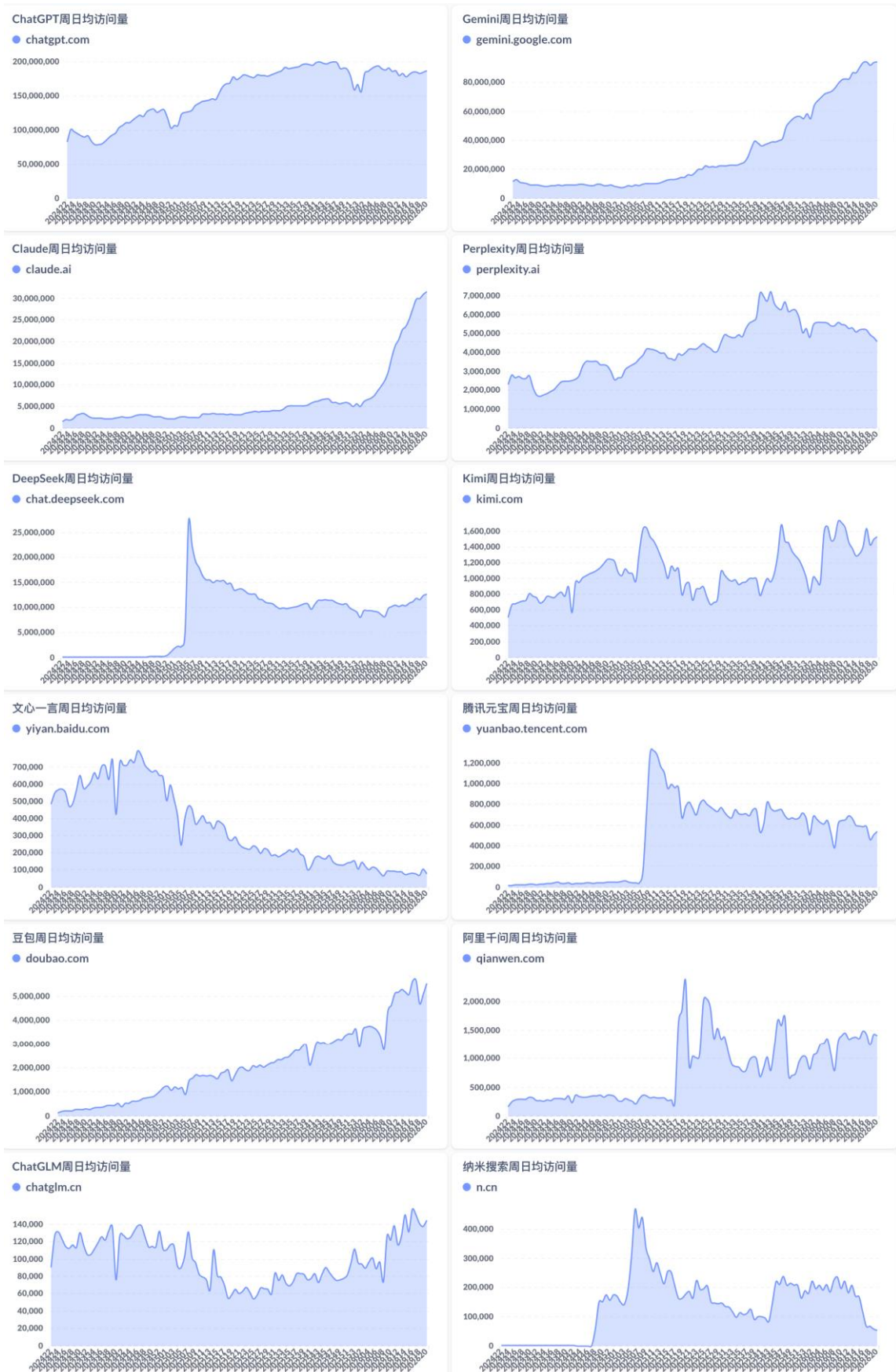
个股名称	本周收盘价	周涨跌幅	产业链
戴尔	295.19	21.98%	消费电子产品、服务器、交换机、网络安全
HP	25.24	21.29%	消费电子产品
Skyworks	82.42	20.27%	芯片设计与制造
高通	238.16	18.2%	芯片设计
Qorvo	106.43	15.37%	芯片设计与制造
Zscaler	182.37	13.24%	网络安全
Gitlab	26.73	12.98%	云服务
CrowdStrike	663.46	11.68%	网络安全
美满电子	196.33	10.99%	芯片设计与制造、交换机
超威半导体	467.51	10.24%	芯片设计
英特尔	119.84	10.18%	芯片设计与制造
Cloudflare	216.17	9.42%	网络安全
Snowflake	172.2	9.35%	云服务、软件
Zeta Global	18.79	9.31%	软件
C3.AI	9.29	7.4%	云服务
Palo Alto Networks	260.58	7.31%	网络安全
DataDog	222.32	6.89%	芯片设计
Mongodb	326.13	4.48%	云服务
Salesforce	180.07	3.78%	云服务
苹果	308.82	2.86%	消费电子产品
Palantir	136.88	2.16%	云服务
特斯拉	426.01	0.89%	整车、大模型
亚马逊	266.32	0.83%	IaaS、电商平台、自营硬件、物流配送、广告变现
Innodata	95.48	-0.65%	云服务
Meta	610.26	-0.65%	消费电子产品、大模型、广告
微软	418.57	-0.79%	消费电子产品、云服务、软件
博通	414.14	-2.6%	芯片设计与制造、交换机
谷歌	382.97	-3.48%	云服务、大模型、广告
英伟达	215.33	-4.43%	芯片设计、交换机

来源: Reuters、国金证券研究所



Qwen 3.7 Max 发布，智谱推出高速版 GLM 5.1

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, Claude 持续上升, ChatGPT 和 Gemini 保持平稳。国内多数聊天类应用网页端活跃度较为稳定。

通义实验室正式发布新一代 AI Agent 基座大模型 Qwen3.7-Max, 在多项权威评测中斩获国内第一, 旨在攻克智能体在长链路生产环境中易断片崩溃的痛点。该模型在未知硬件 ZW-M890LPPU 极限压测中, 凭借运行时反馈展现极强 In-Context 泛化力, 实现连续运行 35 小时、跨越 1158 次工具调用零中断的深度自主推理, 并斩获 10.0 倍几何平均加速。为解决模型对开发框架过拟合痼疾, 其首创任务-运行框架-验证器正交解耦设计, 将强化学习推向真实分布, 实现跨框架泛化与端到端交付复杂应用。

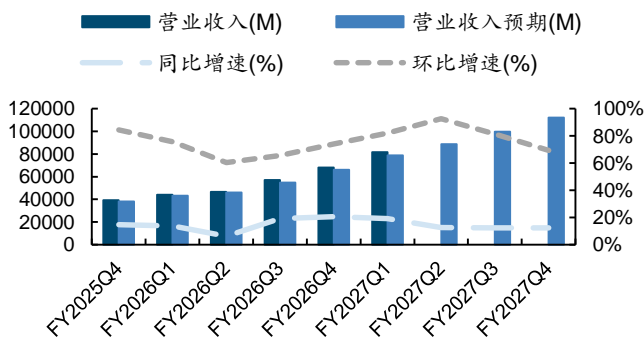
腾讯推出支持 Windows、Mac 及 Android 的操作系统级 AI 助手马维斯, 将终端系统、文件及算力整合至 AI 中间层, 出厂预设 6 个协同 Agent, 主 Agent 统筹调度, 专项 Agent 负责操作搜索等具体工作, 具备文档归类、图片处理与修复系统等功能。马维斯提供效率与隐私双模式, 隐私模式纯本地运行断网可用, 适配高敏感场景, 每人每日享 1000 万免费 Token。

智谱面向企业定向开放 GLM-5.1 高速版 API, 输出达 400tokens/s 刷新全球大模型官方 API 速度上限, 支持 200K 上下文与 128K 最大单次输出, 打破旗舰级全尺寸与极致低延迟不可兼得的惯例。底层由 GLM 与 TileRT 团队深度联合重构系统级推理生态, 通过定制引擎重写核心路径、引入 DynamicBatching 及 KVCache 调度的智能合并以及集群协同硬件调优, 为 AI 编程、实时交互 3D 游戏、商业决策集群及无缝实时语音等场景实现降维打击般的极速响应。

英伟达财报表现优异

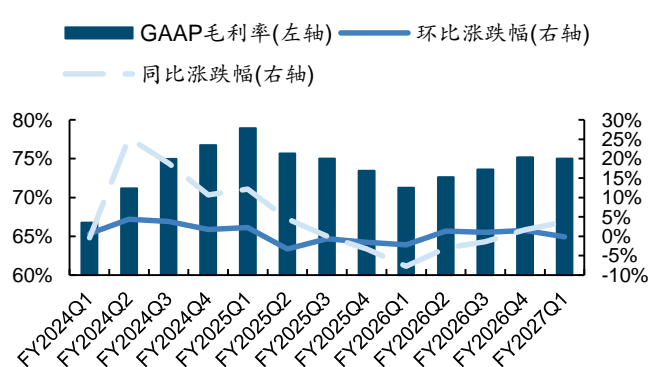
英伟达在 2026 年 1 季度 (FY2027Q1) 实现极为强劲的增长, 总营收达到 820 亿美元, 同比增长 85%, 环比增长 20%, 并实现连续第 14 个季度环比增长。数据中心业务收入达到 750 亿美元, 同比增长 92%, 环比增长 21%, 其中数据中心计算收入 600 亿美元, 同比增长 77%, 数据中心网络收入 150 亿美元, 同比增长近 3 倍。管理层表示 GB300 NVL72 需求强劲, 前沿模型公司与超大规模云厂商已累计部署数十万颗 Blackwell GPU, 成为公司历史上爬坡最快的产品。网络业务方面, 公司强调 Spectrum-X 已经成为规模超过所有其他以太网 AI 网络同行总和的平台, 而 InfiniBand 收入同比增长超过 4 倍, 主要由 XDR 部署推动。盈利能力方面, 公司 Q1 GAAP 毛利率为 74.9%, Non-GAAP 毛利率为 75%, 环比基本持平。展望 Q2, 公司预计营收 910 亿美元, 上下浮动 2%, 增长主要来自数据中心业务, 同时对 2025 至 2027 自然年累计实现 1 万亿美元 Blackwell 与 Rubin 收入保持信心。

图表3: 英伟达收入增速 (百万美元)



来源: IDC、国金证券研究所

图表4: 英伟达毛利率



来源: IDC、国金证券研究所

我们认为, 英伟达目前仍处于极强势阶段, 公司仍在持续扩大份额, 而不是仅仅跟随行业增长。本季度数据中心收入同比增长 92%, 同时黄仁勋表示公司增速应该快于 Hyperscaler 的资本开支。原因在于除了 Hyperscaler 之外, 英伟达也向全球围绕 AI 算力、AI 训练、AI 推理而建立的云服务商或算力平台提供算力支持。同时黄仁勋首次明确提出 Vera CPU 对应 2000 亿美元市场空间, 并预计今年 CPU 收入接近 200 亿美元, 而且特别强调这是独立 CPU 收入, 不包含 Rubin 系统内部 CPU。



未来几个季度应该继续关注 1) Blackwell & Rubin 出货节奏。这是未来两年最核心指标，尤其是 GB300 向 Vera Rubin 切换是否顺利。假设 26Q3 开始出货后，Q4 与明年 Q1 能否继续维持高速爬坡值得注意，如果 Rubin 再次实现超预期放量，意味着 AI 基础设施周期远未结束。2) Vera CPU 进展。如果未来 Vera CPU 销量快速增长，意味着英伟达将正式切入传统 CPU 市场，以及英伟达在服务器里路线的话语权进一步增强。3) 毛利率与自由现金流。未来如果 Rubin 大规模放量后毛利率仍稳定，说明英伟达在系统级 AI 基础设施中拥有极强定价权。

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。

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	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路1088号 紫竹国际大厦5楼	地址：北京市东城区建国内大街26号 新闻大厦8层南侧	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