



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

GLM 5.3 Flash 发布，英伟达营收同比翻倍

摘要

- 本周腾讯开源混元 Hy4 preview，凭借 770B 总参数、1M 上下文及工程任务表现跻身开源第一梯队；智谱 GLM-5.3-Flash 进一步强化原生多模态与 Coding 能力，并以低价 API 和国产算力支撑突出性价比。海外方面，谷歌 Gemini Omni 1.1 Flash 将视频生成推进至 4K 成片阶段，而英伟达拟收购 Hugging Face，显示算力厂商正进一步向开源模型生态和平台入口延伸。
- 英伟达 2026 年第二季度（FY2027Q2）实现营收 960 亿美元，同比增长超过 100%，数据中心营收达到 890 亿美元，环比增长 18%。其中，超大规模业务营收 490 亿美元，环比增长 13%；ACIE 业务营收 400 亿美元，环比增长 25%、同比增长 138%。公司预计第三季度营收为 1,080 亿美元，上下浮动 2%，并初步预计 2028 财年营收同比增长约 70%。

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
GLM 5.3 Flash 发布, Nvidia 收购 Hugging Face	3
英伟达: Vera Rubin 加速放量, 强劲需求下供应仍是增长瓶颈	4
风险提示.....	5



海外市场行情回顾

图表1：截至8月28日海外AI相关个股行情

个股名称	股票代码	本周收盘价	周涨跌幅	产业链
Salesforce	CRM	256	22.39%	云服务
CrowdStrike	CRWD	218.4	13.78%	网络安全
Gitlab	GTLB	44.87	7.68%	云服务
微软	MSFT	513.53	6.27%	消费电子产品、云服务、软件
Meta	META	578.02	5.11%	消费电子产品、大模型、广告
Zeta Global	ZETA	30.54	5.06%	软件
Palo Alto Networks	PANW	371.59	3.83%	网络安全
Mongodb	MDB	446.62	3.67%	云服务
Palantir	PLTR	186.29	3.53%	云服务
苹果	AAPL	319.7	3.35%	消费电子产品
戴尔	DELL	456.24	3.2%	消费电子产品、服务器、交换机、网络安全
亚马逊	AMZN	266.43	3.02%	IaaS、电商平台、自营硬件、物流配送、广告变现
HP	HPQ.N	30.52	2.73%	消费电子产品
Cloudflare	NET	299.84	2.29%	网络安全
C3.AI	AI	10.53	2.23%	云服务
高通	QCOM	164.19	2.14%	芯片设计
Zscaler	ZS	184.23	1.37%	网络安全
英伟达	NVDA	217.55	1.32%	芯片设计、交换机
DataDog	DDOG	236.98	0.58%	芯片设计
谷歌	GOOGL	346.59	0.51%	云服务、大模型、广告
博通	AVGO	368.79	0.09%	芯片设计与制造、交换机
英特尔	INTC.O	89.47	-0.67%	芯片设计与制造
Qorvo	QRVO.O	94.73	-0.87%	芯片设计与制造
Snowflake	SNOW	328	-1.44%	云服务、软件
超威半导体	AMD	465.58	-1.62%	芯片设计
Skyworks	SKWS.O	65.79	-2.01%	芯片设计与制造
特斯拉	TSLA.O	348.75	-3.89%	整车、大模型
美满电子	MRVL	216.62	-8.61%	芯片设计与制造、交换机
Innodata	INOD	56.59	-11.87%	云服务

来源：Reuters、国金证券研究所

GLM 5.3 Flash 发布，Nvidia 收购 Hugging Face

海外模型与生态方面，谷歌推出 Gemini Omni 1.1 Flash 视频生成模型，最高支持 4K 分

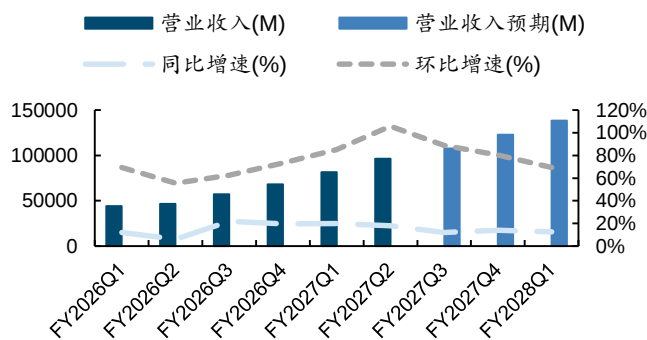


分辨率，并新增视频续生成、首尾帧控制与视频参考等能力，单条视频可按 10 秒步长扩展至最长 40 秒；360p 预览价格降至每秒 0.03 美元，速度较标准版最高提升 60%，720p、1080p 与 4K 价格分别为每秒 0.1、0.15、0.3 美元，进一步打通从低成本分镜预览到高质量成片的链路。同期，据 The Information 报道，英伟达将以 129 亿美元收购开源 AI 平台 Hugging Face，交易估值约为其 1.5 亿美元年收入的 80 倍；在 OpenAI、Anthropic 等闭源模型厂商加快自研芯片和多元化算力布局的背景下，英伟达有望借此强化开放权重模型生态与云服务入口。国内模型方面，腾讯发布并开源混元 Hy4 preview，总参数 770B、激活参数 49B，上下文长度达 163 名内部专家参与的 203 项工程任务盲测中获得 2.99 分，略高于 GPT-4 的 2.92 分与 Kimi K3 的 2.94 分，并已接入 Hugging Face、GitHub、ModelScope、Gitcode、腾讯云 TokenHub 及 OpenRouter。智谱 AI 同期推出 GLM-5.3-Flash，Artificial Analysis Intelligence 评分达 57 分，高于 DeepSeek V4 Pro 正式版的 53 分，每百万 Token 输入、输出价格分别仅 0.8 元和 2.8 元，同时原生集成视觉能力形成“生成-观察-修正”的 Coding 闭环，并依托国产芯片集群承载 OpenCode 与 OpenRouter 真实线上流量，在模型能力、推理成本与国产算力适配之间进一步取得平衡。

英伟达：Vera Rubin 加速放量，强劲需求下供应仍是增长瓶颈

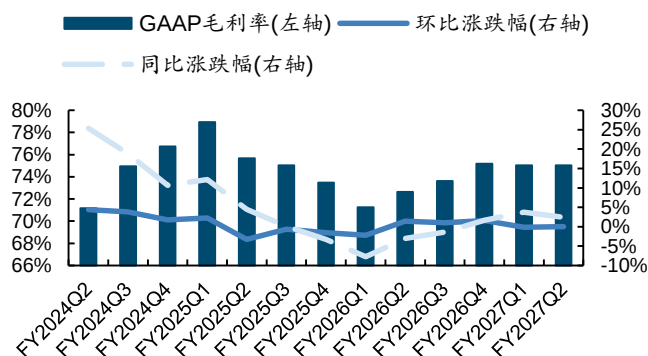
英伟达 2026 年第二季度（FY2027Q2）实现营收 960 亿美元，同比增长超过 100%，数据中心营收达到 890 亿美元，环比增长 18%。其中，超大规模业务营收 490 亿美元，环比增长 13%；ACIE 业务营收 400 亿美元，环比增长 25%、同比增长 138%。公司预计第三季度营收为 1,080 亿美元，上下浮动 2%，并初步预计 2028 财年营收同比增长约 70%。管理层表示，客户需求对应的增长接近 100%，但受制于整体供应链产能，公司目前只能有把握实现约 70% 的增长，供应瓶颈预计至少持续至 2028 财年末。

图表2：英伟达收入及增速（百万美元）



来源：路透、国金证券研究所

图表3：英伟达毛利率



来源：路透、国金证券研究所

Vera Rubin 将是未来几个季度最主要的产品驱动力。该产品已经开始生产交付，并获得所有主要超大规模云服务商、人工智能云服务商和系统 OEM 的采购订单。与 Grace Blackwell Ultra 相比，Vera Rubin 每兆瓦吞吐量提高 30 倍、每 Token 成本降低 35%，预计第三季度约占数据中心营收的 20%。随着供应逐步增加，超大规模业务预计从第四季度开始重新加速。英伟达每吉瓦对应的营收机会也由 Hopper 的约 180 亿美元、Grace Blackwell 的约 250 亿美元，提高至 Vera Rubin 的约 400 亿美元。

除 Rubin GPU 外，下一代 Vera CPU 已全面投产，目前已开始向 OCI 和 SpaceX AI 交付，并将从本季度开始向 AWS 交付。公司预计服务器 CPU 总需求约为 200 亿美元，2028 财年 CPU 营收有望增长超过一倍。首款机架级 LPU 系统 Groq 3 LPX 也已全面投产，预计本季度后期开始批量交付，由 Nebius 率先采用。管理层表示，全球绝大多数数据中心仍将主要采用 Vera Rubin 和 NVLink 72。

我们认为 Vera Rubin 快速放量将继续推动收入增长，但内存价格上涨和供应链满负荷运行会对短期盈利能力形成压力。未来需要重点关注 Vera Rubin 的实际交付速度、超大规模业务能否从第四季度开始重新加速，以及供应改善能否缩小需求增长与收入指引之间的差距。



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