AI 技术每日分析：英伟达董事会恢复黄仁勋 CEO 职务，设出口管制委员会

中国高技术产业发展促进会新质生产力工作委员会

博雅云创 & 中科创新驱动

2026 年 11 月 24 日



摘要

英伟达董事会宣布恢复黄仁勋总裁、首席执行官及董事职务，消息推动公司股价在美国周一开盘时大幅上涨。亚马逊签约引入 Decart 的芯片及模型优化技术和研发团队，网络安全企业 CrowdStrike 则同意以约 439 亿美元企业价值收购 Zscaler。技术端，智谱公布采用循环 Transformer 的 GLM-5.4，腾讯微信 AI 开放面向本地多模态应用的 WePocketModel。算力生态竞争、企业智能体的治理架构和模型推理效率正在形成相互关联的三条产品路线。

一、英伟达董事会恢复黄仁勋职务，出口合规治理同步调整

11 月 23 日，英伟达投资者关系网站发布公告，董事会一致同意黄仁勋复任总裁、首席执行官和董事，即刻生效。SEC 披露的 8-K 文件进一步确认董事会决定作于 11 月 22 日，临时首席执行官 Colette Kress 继续担任首席财务官。美联社报道 11 月 23 日开盘股价上涨约 33%，显示资本市场对领导层变化的迅速定价。本期仅记录开盘行情，收盘指标不作为新闻依据。

公告同时说明，董事会此前设立的独立调查委员会将继续完成调查，并已另设由独立董事组成的国家安全与出口管制委员会，计划招聘首席合规官，后者向提名与公司治理委员会报告。先进计算芯片的交易和跨境数据中心服务涉及客户身份、最终用途、设备交付及监管规则，合规部门与技术销售体系之间的组织接口由此成为芯片公司的经营基础设施。行业更关注的是 CUDA 生态、系统产品交付节奏及客户服务能否保持连续性。董事会的新安排把这些商业关系背后的贸易治理明确纳入公司级监督。

二、亚马逊签约购买 Decart 技术，定制算力竞争延伸至优化软件

亚马逊 11 月 23 日正式宣布购买 Decart Optimization Stack（DOS）相关专有技术的协议，并计划引入三位创始人及其研究工程团队。双方自 2025 年已经合作优化 Trainium。Reuters 同日报道称交易金额未由公司公布，彭博援引消息人士给出约 75 亿美元的估计。交易签约已经发生，交割取决于相关条件。

DOS 关注从模型计算图到硬件执行的完整路径。对于同一模型，矩阵算子、注意力访问模式、低精度表示和设备间通信均会影响最终吞吐。

云厂商自主芯片提供了另一种硬件成本结构，软件优化决定计算资源能否转化为实际服务能力。亚马逊此次把芯片性能工程与模型推理优化人才纳入同一研发体系，反映出云平台竞争的焦点正在由购置加速卡延伸到编译、运行时、内存调度和生产级推理服务。对于企业用户，跨芯片迁移能否复用业务接口、模型版本和评测流程，将影响下一轮算力采购的可替代性。推理服务的实际成本还取决于在线请求的长度分布与并发水平。把同一套模型分别放在不同加速器上运行，可以通过单位有效输出、首字延迟、峰值服务能力及持续功耗建立可比较的采购口径，避免单纯比较芯片标称峰值。

三、CrowdStrike 拟以 439 亿美元收购 Zscaler，智能体连接成为安全平台的市场入口

两家公司 11 月 23 日公布全股票交易协议，企业价值约 439 亿美元，每股 Zscaler 股票拟交换 1.135 股 CrowdStrike 股票，基于 11 月 20 日收盘价计算溢价约 34%。双方预计在 CrowdStrike 2028 财年下半年完成交易，合并后的年经常性收入预计超过 100 亿美元。双方在交割前独立经营。

终端监测覆盖程序、进程和工作负载，零信任网络则负责建立经过授权的连接。企业智能体不断调用外部 API、访问数据库、运行命令，传统以用户登录为中心的安全边界开始向具体任务、服务身份与连接生命周期扩展。Reuters 同日另报道部分美国银行正收紧内部智能体的访问权限、运行时长和凭证管理，说明该并购所指向的企业需求已有现实背景。企业软件提供商可围绕独立服务身份、工具访问策略、日志留存和业务连续性建立产品能力。

四、智谱 GLM-5.4 引入循环 Transformer，编码能力延伸至工程设计

智谱 11 月 23 日发布 GLM-5.4 技术博客，将已有模型层重复计算的路径用于提高有效计算深度，并展示 FreeCAD 机械臂设计、互动三维场景、视频及游戏生成等应用。公司将其称为广义编码，目前通过 Coding Plan、ZCode 及聊天产品提供服务，按量 API 和权重开放列为后续安排。循环结构有机会在保持参数复用的同时增加推理计算量，其工程价值与缓存策略、运行时延迟和实际任务完成质量密切相关。工业软件可把代码生成能力连接到几何内核、约束求解器和仿真工具，并以可执行模型与测试结果作为验收对象。

五、腾讯 WePocketModel 开放 2B 和 4B 权重，本地多模态模型走向统一底座

腾讯微信 AI 的 GitHub 项目明确记录 11 月 23 日开放 WePM 1.0，提供 2B、4B 两个尺寸及 Transformers、GGUF 格式，统一处理向量表示、内容生成与结构化判断。研发团队报告 4B 模型的 BF16 共享权重约 8GB，三类能力共享视觉语言骨干，避免分别常驻多个模型。模型支持图片检索、文字识别及候选项排序等操作。在企业工作站和终端设备上，本地文件检索、图片理解和业务动作路由可以复用同一推理进程，降低重复模型的内存占用。端侧部署的采购和运维评价也将更多关注真实任务组合下的常驻内存、响应速度与能耗。

参考资料

1. NVIDIA ｜董事会公告

https://investor.nvidia.com/news/press-release-details/202

6/NVIDIA-Board-Reinstates-Jensen-Huang-as-President-CEO-and-Board-Director/default.aspx

2. SEC | SEC 8-K
https://www.sec.gov/Archives/edgar/data/1045810/000104581026000090/nvda-20261122.htm

3. AP | AP 报道
https://apnews.com/article/39d4b34ffaa867b4e669893b5fd5028f

4. Amazon | 官方公告
https://www.aboutamazon.com/news/company-news/amazon-decart-ai-optimization

5. Reuters | Reuters 报道
https://www.reuters.com/business/retail-consumer/amazon-buy-decarts-ai-chip-optimization-tech-latest-ai-talent-grab-2026-11-23/

6. CrowdStrike | 交易公告
https://ir.crowdstrike.com/news-releases/news-release-details/crowdstrike-acquire-zscaler-bringing-network-security-and-zero

7. CrowdStrike | 交易条款
https://www.crowdstrike.com/en-us/press-releases/crowdstrike-to-acquire-zscaler-bringing-network-security-and-zero-trust-to-the-falcon-platform/

8. Reuters | Reuters 银行报道

https://www.reuters.com/legal/litigation/some-us-banks-tighten-ai-agent-controls-concern-about-models-escaping-safeguards-2026-11-23/

9. Z.ai | GLM-5.4 技术博客
   https://z.ai/blog/glm-5.4
10. The Next Web | The Next Web 报道
    https://thenextweb.com/news/zai-glm-5-4-loop-transformer-coding-model
11. GitHub | WePocketModel 代码和模型
    https://github.com/Tencent/WePocketModel

机构介绍与关注

新质生产力工作委员会

中国高技术产业发展促进会新质生产力工作委员会，专注于推动工业人工智能、智能制造、数字化转型等前沿技术发展，为企业提供政策解读、技术咨询和产业对接服务。

工业智能算网

专注于工业人工智能、新质生产力、工业软件 CAE、智能制造等前沿技术。提供每日动态分析、技术趋势解读、解决方案分享，推动工业智能化转型。

关注高促会新质生产力工委会公众号

关注工业智能算网平台

https://gyznsw.cn