

OpenAI网安双面出牌、DeepMind高层地震、阿里首开Max级权重

头部企业动态

- OpenAI 于 8 月 7 日披露,内部评估无法排除未发布模型 Astra 触及其《准备度框架》网络安全「关键」阈值,已暂停不满足强化安控要求的内部活动,系业界首例因网络安全风险主动放缓前沿模型。 **B+ · 89% 多源一致**
- 谷歌 8 月 5-6 日重组 DeepMind 领导层:哈萨比斯卸任 CEO,转任 DeepMind 董事长兼 Alphabet 首席科学家;CTO Koray Kavukcuoglu 接管日常运营;首席科学家 Jeff Dean 等离职创办公益公司 Discovery Loop,谷歌参投并任其云供应商。
B · 64% 待复核 · 链接探活失败
- Meta 于 8 月 10 日发布可在个人电脑本地运行的开源模型 Muse Glimmer(宽松许可),并称将向开发者开放更强的 Muse Spark 1.2;扎克伯格同步发表约 6500 字宣言《The Future is for Everyone》,反对超级智能被少数机构垄断。
B- · 65% 同源引述 · 链接探活失败
- Anthropic 于 8 月 7 日更新 Fable 5 的生物学安全分类器,重写其「宪法」规则集并重新训练,使良性生物医学问题的降级回退减少约 85%;病毒学、毒理学与分子设计等两用领域仍强制回退至 Opus 5。 **B+ · 91% 多源一致**

模型发布

- OpenAI 于 8 月 10 日推出基于 GPT-5.6 Sol 专门训练的网络安全模型 GPT-5.6-Cyber,并将 Daybreak 计划拆为 Blue(通用防御:事件响应、恶意代码分析、补丁验证)与 Red(漏洞研究与渗透测试)两档,仅向受信任伙伴开放。 **B+ · 92% 多源一致**
- 阿里于 8 月 3 日发布 Qwen3.8-Max:2.4 万亿参数稀疏 MoE、最高 100 万 token 上下文、原生多模态(文/图/视频输入),Terminal-Bench 2.1 得 86.6;官方称开源权重于次周(8 月 10 日当周)与 Qwen3.8-27B 一并放出,为 Max 级旗舰首次开放权重。 **B+ · 90% 多源一致**
- 谷歌 Gemini API 官方更新日志(最后更新 8 月 11 日)显示,8 月内无任何新模型条目,最近一次实质发布为 7 月 30 日 Gemini Robotics-ER 2 公开预览及 ER-1.6 弃用公告——即谷歌在本轮竞赛中确实处于发布空窗期。 **A · 96% 已核实**
- 字节跳动 Seed 团队于 8 月 5 日发布原生音视频全双工大模型 SeedRealtime,以统一架构整合音频、视频与文本,支撑豆包 App 的实时「看听说」交互;另于 8 月 8 日发布视频生成模型 Seedance 2.5,单次可生成 30 秒片段并支持多轮延长。
B · 76% 待复核

热点与趋势

- 欧盟《AI 法案》自 8 月 2 日进入执法期:第 50 条透明度义务(聊天机器人须自曝为 AI、合成内容须带机器可读标记、深度伪造须标注)可强制执行,AI Office 取得对通用模型的处罚权,罚则上限 1500 万欧元或全球营收 3%;高风险系统条款则推迟至 2027 年 12 月 2 日。 **B+ · 90% 多源一致**
- 「危险能力」正取代通用跑分成为前沿实验室的放行闸门:OpenAI 一周内先因 Astra 网安能力逼近关键阈值而刹车、再以分级授权放出攻击级 GPT-5.6-Cyber;Anthropic 同期精调生物护栏并预告可信研究者准入计划,行业从一刀切拒答转向分层可信访问。 **B · 79% 多源一致**
- 中国厂商把开放权重当作卡位手段:阿里 8 月首次承诺开放 Max 级旗舰权重并同步放出 Qwen3.8-27B,此前月之暗面于 7 月 16 日开源 2.8 万亿参数的 Kimi K3;闭源变现正让位于开源生态影响力的争夺。 **B- · 79% 同源引述**

主动剔除(未核实,不外发)

- 「月之暗面以 500 亿美元投前估值洽谈 Pre-IPO 最后一轮融资、最快 6 个月内赴港上市」——仅中文财经媒体引述知情人士,无公司或港交所一手公告,且系 7 月 22 日旧闻;检索摘要中还出现 50 亿/500 亿美元自相矛盾的数字。
- 「智谱与 MiniMax 已于 2026 年初在港交所上市、估值均超千亿元」——仅见于一篇中文报道的背景段落,未找到交易所公告或招股书佐证。

- 「Gemini 3.6 Flash 稳定版、Gemini 3.5 Flash-Lite 已发布」——谷歌官方 Gemini API 更新日志(已打开核对)8 月区间无任何对应条目,无法证实。
- 「OpenAI 内部版 Astra 于 8 月 1 日攻克十个数学与理论计算机科学未解问题并在 GitHub 公开 Lean 形式化证明」——仅见于二手聚合站,未找到 OpenAI 一手页面或论文。
- 「OpenAI Doug 模型于 8 月 9 日亮相、将于 11 月发布」——单一聚合站说法,无一手源,模型名亦无旁证。
- 「Manus 分拆为独立公司、受影响用户须在 8 月 23 日前备份数据」——仅来自新闻聚合页摘要,未找到官方公告。
- 「欧盟 AI 法案高风险条款(风险管理、人工监督、合规评估)已于 8 月 2 日生效」——与已核实资料直接矛盾:高风险条款经 AI Omnibus 推迟至 2027 年 12 月 2 日,该说法系错误信息,主动剔除。
- 「ChatGPT 免费用户文字对话全面无限量、由 GPT-5.6 Luna 承载」——仅 TechCrunch 单条标题出现于检索结果,未打开正文,且模型代号无第二方佐证,本次不采用。

信息源清单(可审计)

条目	等级	可信度	探活	来源
OpenAI 于 8 月 7 日披露,内部评估无	B+	89%	√活	infosecurity-magazine.com/news/openai-pauses-d...
谷歌 8 月 5-6 日重组 DeepMind	B	64%	✗死	axios.com/2026/08/05/google-deepmind-demis-has...
Meta 于 8 月 10 日发布可在个人电脑本	B-	65%	✗死	usnews.com/news/business/articles/2026-08-10/z...
Anthropic 于 8 月 7 日更新 Fa	B+	91%	√活	the-decoder.com/anthropic-loosens-fable-5s-bio...
OpenAI 于 8 月 10 日推出基于 GP	B+	92%	√活	techcrunch.com/2026/08/10/as-ai-led-attacks-mu...
阿里于 8 月 3 日发布 Qwen3.8-Ma	B+	90%	√活	marktechpost.com/2026/08/03/alibaba-qwen-relea...
谷歌 Gemini API 官方更新日志(最后更	A	96%	√活	ai.google.dev/gemini-api/docs/changelog...
字节跳动 Seed 团队于 8 月 5 日发布原	B	76%	√活	stcn.com/article/detail/4059425.html...
欧盟《AI 法案》自 8 月 2 日进入执法期:	B+	90%	√活	helpnetsecurity.com/2026/08/04/eu-ai-act-enfor...
「危险能力」正取代通用跑分成为前沿实验室的放行 闸	B	79%	√活	forbes.com/sites/jonmarkman/2026/08/11/openai-...
中国厂商把开放权重当作卡位手段:阿里 8 月首次	B-	79%	√活	scmp.com/tech/article/3362738/alibabas-ai-mode...

评分:A 官方一手抓原文 / B+ ≥2独立源 / B 单家权威或未复核 / B- 同源引述 / C 二手(不采)。本期 WebSearch=18 WebFetch=16 成本≈\$1.7485905000000004。自动生成,源已分级核验。