

Anthropic发Fable 5.1、谷歌3.8 Flash三周再迭代、网安能力分分管控成主线

头部企业动态

- Anthropic 9月1日发布 Claude Fable 5.1 与 Mythos 5.1:同一底座、不同安全护栏,Fable 全面开放(API/Bedrock/Vertex/Azure),Mythos 仅限美国受信网安与生命科学机构;定价维持\$10/\$50,缓存读取降至\$0.25。 **A • 98% 已核实**
- OpenAI 9月1日称新模型 Astra 成为首个越过 Preparedness Framework 「Critical」 网络安全阈值的模型,测试中自主发现并串联两个零日漏洞;最强网安能力将先限于受信测试者,再经 Daybreak Blue 扩至防御用途。
B • 65% 待复核 • 链接探活失败
- Sony Music Publishing 与 Warner Chappell 8月29日在加州北区联邦法院起诉 Anthropic,并将 Dario Amodei、Benjamin Mann 列为个人被告,指控其盗版下载与抓取歌词训练 Claude,每首索赔最高15万美元;Anthropic 称将积极应诉。 **B+ • 91% 多源一致**
- 美国司法部 9月1日向曼哈顿联邦法院提交「利益声明」,支持 OpenAI 在《纽约时报》版权案中的合理使用抗辩,称训练大模型的公共收益远超竞争损害;副总检察长 Woodward 称之为「历史性声明」。 **B • 78% 待复核**

模型发布

- 谷歌 9月2日发布 Gemini 3.8 Flash 与 3.8 Flash Cyber,距 3.7 Flash 仅三周;年底前优惠价\$0.75/\$3.75 每百万 token,之后翻倍;Cyber 版仅经新设 Fairwind 计划向政府与关键基础设施「受信防御者」开放,20 种语言真实漏洞发现成功率超 70%。 **A • 98% 已核实**
- Meta Superintelligence Labs 9月1日推出首个实时音频感知模型 Muse Voice Transcribe:单模型同时做流式识别、20+说话人分离与端点检测,流式词错率3.1%、终稿延迟约0.16秒,支持25种语言,已上线 Meta Model API、Meta AI for Mac 与 Muse Code。 **A • 97% 已核实**
- 阿里通义 8月26日开源 Qwen3.8-Flash-Next:1250亿总参数、每 token 激活60亿,含510亿参数 N-gram 嵌入层,原生26 万上下文可扩至100万,定位 Qwen4 架构预览;生产版定价\$0.16/\$0.47,约为 Qwen3.8-Max 的十二分之一。
B • 78% 待复核
- DeepSeek 近期节奏:7月31日 V4-Flash 正式版 API 公测,8月13日 V4-Pro 正式版与官方 deepseek harness 上线,8月21 日再发 V4-Flash-Vision-Exp 视觉模型;V4 系列为1.6T总参/49B激活混合注意力架构,原生百万上下文。 **B • 72% 待复核**

热点与趋势

- 网安能力成为前沿模型分层发布的共同主线:OpenAI Astra 触发 Critical 阈值限流、谷歌 3.8 Flash Cyber 仅经 Fairwind 计划开放、Anthropic Mythos 5.1 仅限受信网安/生科机构,三家在同一周内同步走向「通用版公开+高危能力受控」双轨。
B+ • 90% 多源一致
- OpenAI 9月1日宣布 ChatGPT for Healthcare 接入 Epic 电子病历(覆盖超3.25亿患者),临床医生可在患者授权记录上只读问答、嵌入病历界面,并新增连接 PubMed、DailyMed、CMS 等九个公共医疗数据源的插件。 **B+ • 88% 多源一致**
- AI 算力需求持续外溢至定制芯片:博通 9月2日前后公布 FY26 Q3 营收296亿美元(+86%),AI 半导体收入同比+221%,Q4 指引 AI 收入217亿美元(+236%),并称已看到 FY27 1150亿、FY28 2300亿美元的 AI 半导体收入能见度。
B- • 65% 同源引述 • 链接探活失败

主动剔除(未核实,不外发)

- 月之暗面拟以500亿美元估值赴港 IPO、最快6个月递表——澎湃7月24日报道注明公司未置评,全部来自「据悉」,未采用。
- 「DeepSeek 9月2日更新 V3.2-Exp」——仅见于知乎汇总页,与官方 V4 时间线矛盾,疑为陈旧条目,未采用。

- 扎克伯格宣布下一代模型「Watermelon」与 Muse Spark 开源权重「即将推出」——仅 AI Weekly 转引 Techmeme,未找到原帖或官方页,未采用。
- 「Qwen3.8 27B 于9月2日由 Airside 发布」——llm-stats 与 LinkedIn 帖子来源互相矛盾,未核到通义官方公告,未采用。
- CrowdStrike 发布 SafeMind 红蓝双模型系统——仅 SiliconANGLE 单源,未全文复核,未采用。
- Meta 因 Superintelligence Labs 重组裁减600个 AI 岗位——仅 TradingView 转载 Invezz,日期与一手来源不明,未采用。

信息源清单(可审计)

条目	等级	可信度	探活	来源
Anthropic 9月1日发布 Claude	A	98%	✓活	anthropic.com/claude-fable-and-mythos-5-1...
OpenAI 9月1日称新模型 Astra 成为	B	65%	✗死	cnbc.com/2026/09/01/open-ai-astra-cyber-model...
Sony Music Publishing 与	B+	91%	✓活	techcrunch.com/2026/08/29/sony-music-warner-su...
美国司法部 9月1日向曼哈顿联邦法院提交「利益声	B	78%	✓活	theintercept.com/2026/09/02/trump-openai-copyr...
谷歌 9月2日发布 Gemini 3.8 Fla	A	98%	✓活	blog.google/innovation-and-ai/models-and-resea...
Meta Superintelligence L	A	97%	✓活	research.meta.ai/blog/introducing-muse-voice-t...
阿里通义 8月26日开源 Qwen3.8-Fla	B	78%	✓活	the-decoder.com/alibaba-releases-qwen3-8-flash...
DeepSeek 近期节奏:7月31日 V4-F	B	72%	✓活	zh.wikipedia.org/zh-cn/DeepSeek-V4...
网安能力成为前沿模型分层发布的共同主线:Open	B+	90%	✓活	blog.google/innovation-and-ai/models-and-resea...
OpenAI 9月1日宣布 ChatGPT fo	B+	88%	✓活	techcrunch.com/2026/09/01/chatgpt-health-adds-...
AI 算力需求持续外溢至定制芯片:博通 9月2日	B-	65%	✗死	investing.com/news/transcripts/earnings-call-t...

*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=22
WebFetch=20 成本≈\$2.4677672499999996。自动生成,源已分级核验。*