由 Codex 自动生成

AI前沿发展日报 ｜ 2026-08-04（Asia)

日期：2026-08-04；覆盖窗口：2026-08-03 07:00 至 2026-08-04
ai），并纳入经官方资料复核、在本日到达实施节点的产品访问、计费管理与监管变化。

今日要点

- 今天的高信号不在通用模型榜单，而在三个生产节点：微软把多智能体网络防御开放预览，GitHub 将 Copilot 成本管理收回原生账单系统，白宫则称已完成先进模型的自愿评测安排但尚未公开关键规则。
- Project Perception 显示垂直 AI 的竞争开始由“单模型能力”转向信号、专用
、任务分工和执行接口的闭环；微软给出的 CyberGym 与成本数字仍是厂商自测，但已经把安全 agent 的采购问题带到效果与单位成本同一张表上。
- Google Docs 的 Gemini 新能力进入 Scheduled Release 域的
11 种语言开始获得跨 Drive、Gmail、Chat 与网页生成和改写文档的入口；语言扩展同时
放大了数据访问设置的实际影响。
- 当天未发现达到必看级、且能由官方或一级媒体确认的新大型 AI 融资。更重要的商业变量是 Copilot 用量制计费的管理入口成熟：预算、成本中心与个人级用量已从预览工具转入原生运营面板。

今日结论

1. 企业 AI 的下一轮价值验证会发生在高频、持续运行的垂直任务里；采购方应同时要求任务成功率、误动作率、人工接管率和单位任务成本，而不是只看模型基准。
2. 用量制 AI 已经进入 FinOps 常态。预算上限、成本归属和原始用量导出不再是采购后的补充工作，而是允许 agent 扩张之前的上线条件。
3. 美国前沿模型发布正在增加一个不透明的外部依赖：政府最多 30 天的预发布评测。模型商与企业客户都应把审查延迟、访问限制和版本替代写进交付计划。

AI 产品与应用

Project Perception 开放预览，网络防御从告警助手走向红、蓝、绿队
nt 闭环

微软确认 Project Perception 于 8 月 3 日进入全球公开预览。系统用红队 a
潜在入侵路径，蓝队 agent 结合上下文判断真实风险，绿队 agent 执行修复；首个场景
是软件漏洞管理，并把专用模型 MAI-Cyber-1-Flash 接入 MDASH 多模型 age
微软报告该配置在 CyberGym 达到 96%，比 Mythos 高 12 个百分点，同时较现有

配置节省近 50% 成本。微软官方说明（https://blogs.microsoft.com/
/rethinking-security-for-the-age-of-ai/）

产品意义不只是自动写一份漏洞摘要，而是把发现、判断和修复连接成可持续运行的控制回路。安全团队在试用时应把自动执行范围限制在可回滚动作，分别记录三类 agent 的证据链与责任人，并用自身资产、漏洞类型和误报成本复测；官方基准不能直接替代生产环境验收。

模型与技术进展

Gemini in Docs 的 11 种新增语言进入 Scheduled Rel

Google 的新版 Gemini in Docs 已于 8 月 1 日开始向 Schedule
，新增普通话、荷兰语、马来语、希伯来语、波兰语、土耳其语、捷克语、印尼语、瑞典语、丹麦语和挪威语。能力不止翻译：用户可从零生成带格式文档、跨全文改写、匹配既有写作风格与文档格式，并在启用 Workspace Intelligence 后调用 Drive、Gm
网页上下文。Google Workspace 官方更新（https://workspaceup
m/2026/07/expanded-language-support-for-gemini-i

多语言入口把“文档生成”变成跨系统上下文汇聚。对中文团队，真正的部署检查不是能否写中文，而是 Gemini for Workspace in Drive 与智能功能是否开启、哪些
进入上下文、建议改动由谁批准，以及跨语言生成后如何保留来源。渐进部署最长可达 15 天，单个租户在 8 月 4 日尚未看到功能不等于异常。

投融资与商业动态

GitHub 退役 Copilot Billing Preview，AI 用量成本
理

GitHub 于 8 月 3 日退役 Copilot Billing Preview 应用，原因是
盖现有账单设置中的个人级预算、成本中心和用量池分配。企业改用原生 AI usage 页面查看、分组、筛选和导出 AI Credits，用预算限制支出，并可通过用量报告与 Billing
PI 获取原始数据。GitHub 官方变更说明（https://github.blog/cha
-copilot-billing-preview-app-will-be-retired-on-a
ttps://github.blog/changelog/2026-06-01-updates-t
d-plans/）

这不是 Copilot 服务收缩，而是成本管理从过渡应用转成平台原生能力。自 6 月起所有 Copilot 计划已按 AI Credits 用量计费，代码审查还会消耗 GitHub Acti
程负责人因此需要同时管理 AI Credits 与 runner 成本，并把预算归属到团队或个人。当天未发现可信度与重要性更高的新大型融资交易。

政策、伦理与安全

白宫称先进模型自愿评测安排已完成，但关键规则仍未公开

白宫官员 8 月 3 日表示，依据 6 月 2 日行政命令制定的先进模型自愿评测安排已在截止日前完成，OpenAI、Anthropic 与 Google 曾对草案提供反馈，政府还在接触更多
。已知设计包括：开发者与政府判断在研模型是否适用；政府可在发布前最多 30 天访问模型；同时规定保密、网络安全、内部人员风险、知识产权保护和可信合作方早期访问要求。模型覆盖门槛与网络能力评测流程被列为机密。Axios 报道 （https://www.axios
2026/08/03/white-house-finalizes-ai-framework-beh
命令 （https://www.whitehouse.gov/presidential-acti
ced-artificial-intelligence-innovation-and-securi

当天新增事实是“已完成”，而不是“已可执行”：白宫没有公开文本、参与者清单、企业开始采用的日期或争议处理方式，并计划于 8 月 4 日与企业举行 staff-level 会议。对
采购方，最务实的动作是要求模型供应商披露政府审查是否影响发布日期、可用地区、客户白名单和版本回退，而不是把“自愿”理解为不会造成交付约束。

X 平台高信号

1. Project Perception 从发布预告转为可申请的公开预览

信号标题：微软多智能体网络防御于 8 月 3 日开放预览

类型：产品访问变化 / 网络安全

来源或链接：微软官方发布及 @MSFTSecurity 入口 （https://blogs.mic
g/2026/07/27/rethinking-security-for-the-age-of-a

核心观点：本日新增事实是公开预览节点到达；红队 agent 寻路、蓝队 agent 判险、绿队 agent 修复被接入同一持续防御系统，首个落地场景是漏洞管理。

为什么重要：安全 AI 开始从生成建议转向调用真实防护动作，评估单位也必须从回答质量改为端到端处置结果。

影响：试点应限定可回滚动作、保留每次判断的证据与请求者身份，并在自有漏洞集上复测 96% CyberGym 和近 50% 成本节省的厂商数字。

验证状态：开放预览日期、三类 agent 分工、基准与成本数据均由微软一手发布确认；效果数字尚无独立第三方复现。

2. Copilot Billing Preview 退役，成本控制迁入 GitHub 原生设置

信号标题：Copilot 过渡性账单应用于 8 月 3 日停止服务

类型：计费管理变化 / 开发者平台

来源或链接：GitHub 官方变更说明 （https://github.blog/changel
ilot-billing-preview-app-will-be-retired-on-augus

核心观点：退役的是预览应用，不是用量报表；AI usage 页面、个人级预算、成本中心、

用量池分配、导出与 Billing API 成为正式替代入口。

为什么重要：这表明 GitHub 已把 AI Credits 当作需要日常治理的云成本，而非席位订
阅里的模糊附加项。

影响：依赖旧应用导出或看板的组织应切换数据源，并把 Copilot 代码审查消耗的 AI Credits 与 Actions 分钟一起计入团队成本。

验证状态：退役日期、替代入口和预算能力由 GitHub 一手发布确认。

3. Gemini in Docs 的中文等 11 种语言进入受控发布域

信号标题：Scheduled Release 域开始接收多语言 Gemini 文档能力

类型：访问范围变化 / 企业生产力

来源或链接：Google Workspace 官方更新 （https://workspaceup
m/2026/07/expanded-language-support-for-gemini-i

核心观点：从 8 月 1 日开始的渐进部署正在进行，普通话等 11 种语言可获得生成完整初稿、全文改写、匹配风格与匹配格式等能力；Workspace Intelligence 可扩展至
e、Gmail、Chat 和网页上下文。

为什么重要：非英语团队首次更完整地获得跨文档和跨通信源生成入口，采用范围会明显扩大。

影响：管理员应复核 Drive 中 Gemini 与 Workspace 智能功能的开关、可访问数
和审批规则；租户可见性可能在 15 天部署期内分批到达。

验证状态：语言清单、能力范围、管理员前提、套餐与部署节奏均由 Google 一手发布确认。

4. HubSpot 撤回原定 8 月 4 日生效的数据条款，产品节点不能按旧通知理解

信号标题：Contact Discovery 的原始 8 月 4 日说明已被官方标记失效

类型：平台规则变化 / CRM 数据

来源或链接：HubSpot 官方撤回说明 （https://community.hubspot.c
-wrong-and-we-are-fixing-it/152063）；当前数据丰富化说明 （ht
bspot.com/records/how-data-enrichment-works-in-hu

核心观点：今天到达原计划节点，但 HubSpot 已在 7 月 5 日明确撤回 7 月 1 日公布
条款，并表示重新评估数据丰富化的清晰选择方式；当前知识库称该功能默认关闭，只有客户操作后才启用，且 AI 模型训练与数据丰富化是两个独立设置。

为什么重要：把旧邮件中的 8 月 4 日当作仍有效的自动数据共享生效日，会同时误判产品状态与合规责任。

影响：管理员应以当前租户设置和最新知识库为准，分别核查数据丰富化与 AI 模型训练；在官方重新确认前，不应依据已撤回通知假定 Contact Discovery 已按原方案上线。

验证状态：条款撤回、重新评估承诺、当前默认关闭状态与设置分离均由 HubSpot 一手页面确认；未发现新的官方页面重新确认原方案在本日生效。

5. 白宫完成先进模型自愿评测安排，但对外仍不可见

信号标题：政府预发布评测从起草阶段转入企业沟通

类型：政府评测 / 模型发布条件

来源或链接：Axios 当日报道（https://www.axios.com/2026/08/0
lizes-ai-framework-behind-closed-doors）；白宫 6 月 2
whitehouse.gov/presidential-actions/2026/06/promo
telligence-innovation-and-security/）

核心观点：8 月 3 日的新增事实是白宫宣布已按期完成相关安排，并称正在与更多企业讨论下一步；已知设计允许政府在发布前最多 30 天访问适用模型，但覆盖门槛、正式文本和启动日仍未公开。

为什么重要：模型发布节奏可能受政府评测影响，而客户无法仅凭公开信息判断哪些模型会被纳入。

影响：采购与产品团队应要求供应商说明审查是否影响发布日期、地域可用性、客户白名单与回退版本，并给关键上线项目预留延期空间。

验证状态：完成状态、三家实验室反馈、最长 30 天访问期与 8 月 4 日企业会议来自 Axios 对白宫官员及知情人士的当日报道；行政命令由白宫一手文件确认，具体安排未公开，故其操作细节仍未完全验证。

6. Google Earth 回滚生成式卫星图像，水印不再被视为充分防线

信号标题：Nano Banana 2 图像生成功能因虚假灾害与冲突图像被暂停

类型：产品规则变化 / 合成内容安全

来源或链接：The Atlantic 对 Google 回滚声明的原始报道（https://ww
.com/technology/2026/07/google-earth-ai-images/68
https://www.techradar.com/ai-platforms-assistants
-from-google-earth-after-one-day-shows-how-bad-ou
-has-got）

核心观点：Google 在上线不足两天后撤回 Google Earth 内的生成图像入口；此前依赖
ynthID 和用户自行核验，随后改为暂停产品并补强防护。

为什么重要：当生成内容继承权威地图产品的视觉可信度时，事后水印不足以抵消传播中的误认风险。

影响：涉及新闻、灾害、保险、地产和 OSINT 的产品应把真实观测层与生成层硬隔离，并在导出图像上使用持续可见标识和来源链。

验证状态：回滚与 Google 声明由两家媒体直接引述确认；未找到 Google 独立公告页，故以一手回应经媒体交叉核验。

前沿研究速递

TokTier：为 coding agent 做“增量且完全一致”的分词

做了什么：TokTier（https://arxiv.org/abs/2607.29678）

工具结果、却重复提交超长历史的问题，在追加位置附近重做一个小窗口的分词，并通过稳定边界检查、窗口扩展和全量回退保证 token ID 与完整参考分词完全一致；无可复用前缀时则用 GPU 完成预分词和 BPE。

新在哪里：作者在 17 个 tokenizer 家族、150 亿次切分检查、12.4 TB 真实文
万多个 agent 步骤上报告零差异；增量修复处理 10 万至 300 万字符仅需 0.5—1.1
秒，接入 vLLM 后中位首 token 时间下降 16%—34%。

潜在应用方向：长会话 coding agent、工具调用循环、超长上下文 API 网关、提示缓存
前端与高并发推理服务。

一句话判断：当提示缓存命中率已经很高，分词会成为意外的首 token 瓶颈；结果亮眼，但目前仍是作者系统与流量条件下的论文数据。

ExtractBench：同时评估企业文档抽取的值、完整性、证据和成本

做了什么：ExtractBench (https://arxiv.org/abs/2607.296
、4,869 页、8 个业务域和 67 种文档类型，要求 agent 按用户 schema 输出字
给出可追溯到词和页的来源证据。

新在哪里：它不只看字段准确率，还衡量长列表是否被完整抽取、证据定位是否正确以及实际成本。论文发现商业 VLM 在短文档表现较好，却常在长文档中截断记录；coding agent
更准但成本明显更高。

潜在应用方向：发票与合同处理、尽调、保险理赔、监管报送、采购单解析和企业文档 agent 验收。

一句话判断：企业文档自动化的关键不是“抽到几个对的字段”，而是不能漏记录、必须能回到证据且成本可控；数据集与评测代码已公开，但榜单仍需外部复测。

Zero-Mem：把 agent 的记忆读写从额外 LLM 调用中拿掉

做了什么：Zero-Mem (https://arxiv.org/abs/2607.29377)
—上下文图和时间层级两种结构组织历史；查询时从两种视图检索并做确定性冲突校准，只有最终回答环节调用 LLM。

新在哪里：记忆操作本身不消耗 LLM 输入或输出 token，也不生成可能丢失细节的中间摘要。论文称在相同最终问答模型与上下文预算下，性能具竞争力，同时比最快对比基线减少 57.6% 的记忆操作时间成本。

潜在应用方向：长期个人助手、客服历史、项目型 agent、审计友好的企业记忆与低成本多轮交互。

一句话判断：结构化记忆未必需要另一个生成模型维护；公开实现承诺要等同行评审后提供

，当前可复现性仍有限。