

AI前沿发展日报 | 2026-09-30 (Asia/Shanghai)

日期：2026-09-30；覆盖窗口：2026-09-29 00:00—2026-09-30 07:10 (Asia/Shanghai)；美国当地2026-09-29的晚间发布计入本期；窗口外信息注明“近期补充”，仅有公告日期者不推定首发时刻。

今日要点

- 持续运行的个人代理开始交付。OpenAI 推出 dots，任务可以在两次对话之间继续推进；分批开放、地区限制和首月用量优惠，比演示中的“全天候助手”更影响实际采用。官方使用说明 (<https://help.openai.com/en/articles/20001530-getting-started-with-your-dot>)
- 模型调用成本与订阅权益出现不同方向的变化。GPT-6.1 Sol 降低了获得接近旗舰能力的 API 价格门槛；与此同时，Pro 200 恢复新订阅但降低包含用量。采购时需要分别计算两本账。模型发布 (<https://openai.com/index/introducing-gpt-6-1-sol/>)；订阅条款 (<https://help.openai.com/en/articles/9793128-about-chatgpt-pro-tiers>)
- 垂直应用仍能获得大额资本支持。EliseAI 宣布融资3.5亿美元、估值40亿美元，资金继续投向住房与医疗行政流程；值得追踪的是既有客户的业务扩展，而非单纯模型规模。公司原始公告 (<https://www.globenewswire.com/news-release/2026/09/29/3370681/0/en/eliseai-raises-350-million-at-4-billion-valuation-to-bring-ai-deeper-into-housing-and-healthcare-operations.html>)
- 美国最新政策信号落在企业内部与外部审查。白宫会议后披露的自愿协议，新增了审计报告交由董事会审阅的具体表述；不能据此认定已有新的强制监管制度。AP 原始报道的 WHDH 刊载页 (<https://whdh.com/news/trump-vows-to-never-stifle-ai-as-he-gathers-with-tech-executives-urging-caution/>)

今日结论

1. AI 预算应以完成一项业务的实际费用重算。低价模型、订阅额度和代理运行费用并不联动；先用相同任务测量成功率、重试及人工复核，再决定采购方式。
2. 行业应用的竞争优势更取决于能否接住完整业务。获客之后能否扩展到维护、续约、预约等后续环节，比新增一个聊天入口更能检验客户关系的深度；融资估值本身不能证明盈利。
3. 评估技术进展时，提高对结果可验证性的要求。模型榜单看测试条件，科研系统看实验验证，企业承诺看可交付的审查记录；三类证据不可互相替代。以上均为编辑判断。

AI 产品与应用

dots 开始分批开放：持续接手任务，但上线条件并不统一

当日增量 | 美国2026-09-29发布，对应上海2026-09-30凌晨。dots 是由 GPT-6 Astra 驱动的常驻代理，拥有独立云端电脑，能够使用用户连接的应用并持续处理任务。首发 Pro 开放范围排除欧洲经济区、瑞士和英国；Business Premium 覆盖所有受支持的 ChatGPT 地区；Enterprise 以管理员开启的测试版提供，默认关闭。官方使用与准入说明 (<https://help.openai.com/en/articles/20001530-getting-started-with-your-dot>)

为什么重要：用户不必每次重新发起工作，周期性资料整理、提醒和跟进可以变成持续服务。这是产品交付方式的变化，实际任务完成质量仍需验证。

商业影响：首批试点应选结果容易验收、需要反复执行的任务。官方仅承诺接下来一个月 dots 用量不计入符合条件的套餐额度，之后另行公布规则；不能按永久免费估算长期成本。个人套餐的地区范围也不能直接套用到企业套餐，账户开放可能需要数日。这里核验的是发布条件，未做账户实测。

模型与技术进展

GPT-6.1 Sol 把接近旗舰的任务表现带到更低价格档

当日增量 | 美国2026-09-29发布。官方称 GPT-6.1 Sol 在代码代理、电脑操作和专业工作上接近 Astra；在 DeepSWE v1.1 中，相比 GPT-6 Sol 的最好成绩提高6.4个百分点，且使用更低推理强度。这是厂商评测，不能直接外推到所有企业任务。发布与评测说明 (<https://openai.com/index/introducing-gpt-6-1-sol/>)

可核验的采购变化：标准短上下文 API 每百万输入 / 输出 token 分别为2 / 10美元，缓存输入为0.10美元。输入超过272K token 时，整次请求采用更高费率，不能把短上下文价格套用到全部长文档任务。官方模型价格与限制 (<https://developers.openai.com/api/docs/models/gpt-6.1-sol>)

为什么重要与商业影响：对需要多轮工具调用的应用，较低模型单价有利于扩大可经济执行的任务范围，但最终节省取决于调用次数、缓存命中与重试。现已进入 API、ChatGPT Work 和 Codex，尚未进入普通 Chat；模型可用性与订阅额度属于不同条件。本条讨论能力和 API 成本，订阅权益变化集中见下方 X 信号第一条。

投融资与商业动态

EliseAI 融资3.5亿美元，继续向行业内部扩展

发布：2026-09-29 19:01 (Asia/Shanghai)。EliseAI 宣布以40亿美元估值完成3.5亿美元融资，由 a16z 和 Bessemer 领投。资金将支持产品开发及工程、部署、销售团队扩张，并在旧金山建立第二个工程中心。公司经 GlobeNewswire 发布的公告 (<https://www.globenewswire.com/news-release/2026/09/29/3370681/0/en/eliseai-raises-350-million-at-4-billion-valuation-to-bring-ai-deeper-into-housing-and-healthcare-operations.html>)；Reuters 报道的 Boursorama 刊载页 (<https://www.boursorama.com/bourse/actualites/la-societe-d-ia-eliseai-valorisee-a-4-milliards-de-dollars-lors-de-son-dernier-tour-de-table-32491fad79368469ff471f4d4ed35cb8>)

为什么重要：公司从租赁沟通延伸到住户服务、维修与续约，医疗业务则覆盖预约、转诊及保险核验等行政环节。新资金押注的是进入日常操作后的扩展空间。

商业影响：对中国行业软件团队，更值得借鉴的是围绕已有客户打通前后流程，而非复制美国市场估值。公告重述公司在2026-06已披露 ARR 超过2亿美元；这不是本期新增收入，也不是利润指标。融资与运营数据以公司披露为主，原公告正文直连失败后，以实时检索返回的官方全文及 Reuters 报道交叉核验；未获得审计财报。

政策、伦理与安全

白宫自愿协议新增“审计报告交董事会审阅”的责任安排

当日增量 | 美国2026-09-29会议。AP 报道，特朗普称已与科技企业领导人签署涵盖内部及外部审查的自愿协议。扎克伯格表示，企业董事会将独立审阅审计报告；特朗普另提出设立约十人的监督委员会，并将在之后确定负责人。AP 原始报道的 WHDH 刊载页 (<https://whdh.com/news/trump-vows-to-never-stifle-ai-as-he-gathers-with-tech-executives-urging-caution/>)

新增事实：本期关注企业审查结果向董事会汇报的安排，与此前政府间事故沟通渠道不同。不重复展开国际协调叙事。

为什么重要与商业影响：若落实，模型风险报告将进入企业最高治理层的决策材料。采购方可关注供应商能否提供具体评估与整改证据，而不能把“签署协议”当作产品通过认证。现有报道不足以确认统一审计标准、公开披露期限或处罚机制；委员会也不能写成已经运行。协议性质及有关表态依据现场原始报道，完整文本未独立核验。

X 平台高信号

本期可交叉确认的 X 原帖信号偏少：第一条有原帖入口与官方条款互证，其余四条明确采用官方渠道补充，不冒充已核验的 X 发言。以下影响均为编辑分析。

1. Pro 200 恢复新订阅，原有额度只保留到指定日期

信号标题：同为每月200美元，新旧 Pro 200 用户的额度安排不同

类型：高管 X 信号 / 订阅价格与权益变更

来源或链接：Thibault Sottiaux 原帖 (<https://x.com/thstottiaux/status/2104823812042940713>)；官方 Pro 套餐说明 (<https://help.openai.com/en/articles/9793128-about-chatgpt-pro-tiers>)

核心观点：2026-09-29更新确认 Pro 200 恢复新订阅，不符合过渡资格的新用户采用较低额度；符合资格的既有订阅保持旧额度至2026-10-29，之后降低，月费仍为200美元。

为什么重要：订阅标价没有变化，包含的使用权益却改变了。

影响：团队应按新额度评估续费和额外用量，不能用 API 降价推断订阅可完成的工作必然增加；官方帮助页未量化缩减比例，本条不采用网传倍数。

验证状态：X 原帖两次访问均受限，发言由第三方索引定位；额度调整、过渡日期和价格已直接核验官方帮助页。X 原文未完全验证，事实采用官方条款。

2. MTurk 到达官方公布的关闭日期

信号标题：人工标注与复核流程需要迁出 Mechanical Turk

类型：官方渠道补充 / 平台关闭与服务可用性

来源或链接：AWS 服务关闭说明 (<https://docs.aws.amazon.com/mturk/>)；SageMaker 人力选项说明 (<https://docs.aws.amazon.com/sagemaker/latest/dg/sms-workforce-management-public.html>)

核心观点：本期新增的时间变量是2026-09-30已到，即 AWS 公布的 Mechanical Turk 永久关闭日期；此前通知也要求使用该人力选项的流程迁移。

为什么重要：数据标注与人工复核存在平台依赖，模型升级并不能自动替代这些环节。

影响：使用相关服务的团队应核对未完成任务及替代人力渠道，避免标注或评测流水线中断；不把关闭解释为人工审核需求消失。

验证状态：AWS 两处官方文档已交叉确认；未核验对应 X 原帖。文档未给出关闭时区与精确时刻，因此不声称上海截稿时已全部停服。

3. Cloudflare 日志可以区分 Worker 与人员设备的网络流量

信号标题：代理访问内网服务的来源开始可以直接定位

类型：官方渠道补充 / 平台日志能力变更

来源或链接：Cloudflare 2026-09-29变更记录 (<https://developers.cloudflare.com/changelog/post/2026-09-29-mesh-workers-vpc-network-logs/>)

核心观点：日志新增 Mesh 与 Workers VPC 流量识别，并可记录发起会话的 Worker 名称及稳定标识；已有 Worker 需要重新部署才能填充相应字段，历史会话不回填。

为什么重要：同一内网服务收到请求时，运维人员可以更直接地区分人员设备和自动化程序。

影响：事件排查可以定位具体 Worker 与隧道副本；上线日期以前的记录仍存在信息缺口，不能把新字段误认为自动获得了完整历史。

验证状态：官方变更记录的实时搜索全文已核验；网页正文直连重试仍失败，未进行日志实测，也未核验 X 原帖。

4. Copilot CLI 开始读取 Claude Code 的规则文件

信号标题：团队已有代码规则可以在另一个代理工具中复用

类型：官方渠道补充 / 开发工具兼容性变化

来源或链接：GitHub 官方 Copilot CLI 更新日志 (<https://github.com/github/copilot-cli/blob/main/changelog.md>)

核心观点：近期补充，2026-09-28的1.0.89版本新增对 .claude/rules 中 Claude Code 规则文件的支持，将其作为自定义指令读取。

为什么重要：团队试用不同代码代理时，不必为所有项目规范重新维护一份独立入口。

影响：切换工具的配置工作有望减少，但“能读取规则”不代表不同代理执行结果一致；迁移评估仍需使用相同代码任务。

验证状态：官方仓库更新记录已直接核验；未核验 X 原帖，未将文件兼容性表述为行为一致性保证。

5. Fabric IQ 在 Copilot Chat 与 Cowork 正式可用且默认开启

信号标题：Power BI 既有业务语义进入 Copilot 的默认可用范围

类型：官方渠道补充 / 访问范围与默认设置变化

来源或链接：Microsoft 官方发布 (<https://blogs.microsoft.com/blog/2026/09/28/new-microsoft-data-innovations-unlock-what-only-your-business-knows/>)

核心观点：近期补充，页面标注2026-09-28；Fabric IQ 在 Copilot Chat 与 Cowork 正式可用，并对 Fabric、Power BI 客户默认开启，沿用既有访问控制。Code 集成则仍通过 Frontier 项目推进。

为什么重要：企业问答能够调用既有 Power BI 语义模型，而不必为每次分析重新定义指标。

影响：管理员应检查哪些现有数据集已可被业务用户查询；Code 的开放阶段不同，不能据此向所有开发团队承诺相同能力。

验证状态：官方正文与默认设置表述已核验；未核验 X 原帖，未访问企业私有租户。采用正文日期，不以新闻聚合页的次日日期重标为新发布。

前沿研究速递

Quine：把生物模型、科研工具与实验反馈接到一起

官方研究发布：2026-09-29。Microsoft Research 原文

(<https://www.microsoft.com/en-us/research/blog/introducing-quine-an-ai-research-system-designed-for-the-complexity-of-biology/>)

做了什么：提出多模态生物世界模型及交互系统，连接模型、文献、科研工具与研究人员；团队与 Broad Institute 合作筛选可能改变肿瘤细胞状态的化合物，并报告对部分高排名候选开展了多种湿实验验证。

新在哪里：研究目标从单独预测某一类数据，延伸到跨生物尺度整合证据、选择候选并获得实验反馈；Quine Fellows 将向一批科学家提供访问机会。

潜在应用方向：药物发现中的候选优先级排序、实验设计与跨模态证据整理，可能缩短进入验证环节前的筛选工作。

一句话判断：出现湿实验反馈比单纯生成科研建议更有说服力，但目前仍是实验性研究系统，不能解读为临床有效性或商业药物成功。

BaRe-Mem：让代理学会判断“该听谁的意见”

论文提交：2026-09-29 00:25 (Asia/Shanghai)；预印本。论文与提交记录

(<https://arxiv.org/abs/2609.35551>)

做了什么：为多代理咨询建立在线贝叶斯可靠性记忆，根据中心模型的内部信念表示及历史交互，更新对不同顾问可靠性的估计。

新在哪里：这些估计既决定顾问回答的影响权重，也决定是否应咨询或独立推理。作者在九个基准、六个中心模型上测试，报告其面对误导建议时比辩论和多数投票更稳健。

潜在应用方向：多模型分析、专业顾问代理的选择，以及代理团队中的任务分配；重点是减少错误意见的连锁放大。

一句话判断：多代理的价值未必来自更多投票，可能来自更可靠的意见选择；现有结果属于论文内评测，生产环境长期效果尚需复现。