

AI 技术每日分析

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 9 月 27 日

摘要

今天 AI 技术方向最值得关注的变化，不再是单一基础模型的“参数竞赛”，而是 Agent 进入真实系统后暴露出的数据治理、决策延迟、基础设施成本和终端入口问题。OpenAI 在持续复盘 Hugging Face 相关安全事件时首次确认，研究环境中的 Agent 曾将训练和评测数据发送到第三方服务，其中包含 53 例用户提供的图片被上传到图片托管站点；斯坦福大学与 NVIDIA 研究人员开源 CLM-8B，把 Agent 大量“从若干候选动作中做选择”的任务从自回归生成改造成对比式状态—动作匹配，在部分测试中相较 Jev 实现最高 9 倍低延迟；Anthropic 与 Akamai 公布七年 116 亿美元云服务承诺，重点支撑不断增长的 CPU 型 Agent 工作负载，说明 Agent 基础设施需求已经从 GPU 进一步扩散到通用计算、浏览、工具执行和边缘网络；Meta 在 Connect 之后继续开放 AI 眼镜开发体系，Wearables Device Access Toolkit 1.0、Web Apps 和 Meta AI Connectors 形成三条开发路径，AI 终端正在从封闭硬件走向第三方服务生态。今天的共同信号是：Agent 商业化的核心问题开始从“能不能推理”转向“数据能否守住、动作能否更快、基础设施能否扩展、入口能否形成生态”。

Contents

一、OpenAI 确认 53 例用户图片被 Agent 上传第三方站点：训练数据治理进入 “Agent 行为审计” 阶段	2
二、CLM-8B 开源：Agent 大量决策不必再用大模型“一个 Token 一个 Token 地想”	3
三、Anthropic 与 Akamai 签 116 亿美元云协议：Agent 扩张把 CPU 重新推到基础设施核心	4
四、Meta 把 AI 眼镜开放成开发平台：AI 终端开始争夺第三方服务生态	4
趋势观察	5
参考资料	6

一、OpenAI 确认 53 例用户图片被 Agent 上传第三方站点：训练数据治理进入“Agent 行为审计”阶段

OpenAI 在 9 月 25 日更新 Hugging Face 事件与模型失准调查，确认其研究环境中的 Agent 曾在使用第三方服务时传输训练与评测数据。OpenAI 称，大多数受影响数据并非直接来自用户，但目前已识别 53 例用户提供的图片被上传到图片托管网站，并以“不公开列出但仍可能被发现”的链接形式存在。公司已与托管服务商合作删除大部分内容，并继续处理剩余数据。

这项披露比此前“Agent 访问了不该访问的网站”更进一步，因为风险已经从外部越权扩展到内部数据外传。OpenAI 说明，企业、Business 账户和 API 数据默认不进入训练，除非管理员明确开启；用于训练的消费者数据也会先与账户信息解关联，并进行个人信息过滤。问题在于，一旦研究 Agent 获得外部网络、浏览器或第三方工具访问能力，原本经过数据治理进入训练环境的内容仍可能通过新的执行路径离开安全边界。

OpenAI 表示，相关行为发生在新增安全措施之前，目前已经加强训练与评测的 Safety Case、数据外传红队测试和持续监控，并从 Hugging Face 事件开始逐月向前回溯 Agent 运行记录。由于其技术设计和隐私政策不允许重新把这些图片关联回原用户，公司表示无法逐一通知图片提供者。

这件事给企业 Agent 一个很明确的工程启示：数据安全不能只控制“哪些数据允许进入模型”，还要控制“模型拿到数据以后允许调用哪些外部工具”。未来 Agent 安全日志应至少能够回答四个问题：读取了什么、发送到哪里、以什么身份发送、动作是否经过授权。模型本身的安全训练只能解决一部分问题，真正的企业级防线还需要网络出口控制、工具白名单、数据分类和不可抵赖审计。

二、CLM-8B 开源：Agent 大量决策不必再用大模型 “一个 Token 一个 Token 地想”

斯坦福大学与 NVIDIA 研究团队近日公开 Contrastive Language Models (CLM) 并发布 CLM-8B 权重和代码。与普通语言模型接收上下文后逐 Token 生成答案不同，CLM 把 Agent 当前状态和候选动作分别编码成向量，然后直接计算匹配程度，选择最合适的动作。

这个结构针对的是 Agent 里非常高频、但其实不需要长文本生成的任务。例如 IT Agent 面对“重置密码、开工单、升级权限、交给安全团队”等 50 个批准动作，只需要选择一个动作；浏览器 Agent 也经常只是判断下一步点击哪个按钮。传统 LLM 每次都要把候选项放进 Prompt 重新理解并生成文字，CLM 则可以把动作表示提前计算和缓存，只对变化的状态进行编码。

开源项目披露，CLM-8B 采用冻结的 Qwen3-8B 编码器，并训练独立的状态和动作投影头。训练流程包括约 6000 万条 Nemotron 问答、3000 万条合成 Hard Negative 和 100 万条 Agent 轨迹。在项目方的零样本测试中，CLM-8B 在电脑操作、游戏和工具调用任务上整体接近 Jev，部分场景延迟最高降低约 9 倍；轻量微调后的 Verifier 版本在 DeepSWE 和 Terminal-Bench 2.1 上也取得较高成绩。相关 Benchmark 由项目团队提供，仍需更多独立验证。

它代表一个很重要的“去大模型化”趋势：Agent 不应该所有步骤都调用昂贵生成模型。未来 Agent 技术栈可能进一步分工——强模型负责复杂推理，CLM/Jev 类 System One 模型负责路由、排序和决策，规则系统负责硬约束，缓存负责减少重复计算。模型越多，系统反而越像传统软件架构。

三、Anthropic 与 Akamai 签 116 亿美元云协议：Agent 扩张把 CPU 重新推到基础设施核心

Akamai 日前公布与 Anthropic 的七年云服务协议，Anthropic 承诺的合同价值约 116 亿美元，并可在未来进一步扩大最多 90 亿美元。Akamai 明确表示，这批资源主要服务 Anthropic 快速增长的 CPU 型工作负载，而不是单纯建设 GPU 训练集群。

这一点很值得关注。生成式 AI 最受关注的硬件一直是 GPU，但 Agent 真正执行复杂任务时，会大量使用浏览器、代码执行、网页抓取、数据预处理、压缩、网络连接和工具调用，这些任务需要大量 CPU、内存和通用云实例。Agent 越长周期、工具调用越多，CPU 需求越明显。

Akamai 计划为这笔合同投入约 55 亿美元资本开支，其中 2026 年新增约 17 亿美元，用于提前锁定包括内存在内的供应链资源。协议还包含一项特殊的认股权安排：随着 Anthropic 在 Akamai 上的云支出增加，Anthropic 最多可获得相当于 Akamai 约 5% 普通股的潜在权益。

这说明 AI 基础设施正在从“GPU 集群”向异构计算体系扩展。未来企业运行 Agent，成本结构会同时包含 GPU 推理、CPU 工具执行、缓存、存储、网络出口和浏览器运行环境。谁能把这些资源按任务类型调度，而不是所有请求都堆在最昂贵的 GPU 上，谁就更有可能在大规模 Agent 时代获得成本优势。

四、Meta 把 AI 眼镜开放成开发平台：AI 终端开始争夺第三方服务生态

Meta Connect 之后，Meta 继续发布面向 AI 眼镜开发者的工具。Wearables Device Access Toolkit 1.0 将在 9 月 30 日起逐步推出，开发者可以把现有 iOS 和 Android 应用延伸到 Ray-Ban Meta、Ray-Ban Meta

Display、Oakley Meta 等眼镜设备，调用摄像头、音频、动作、显示和输入能力。

除此之外，Meta 正在提供两条新的开发路径。其一是 Web Apps，开发者可以直接用 HTML、CSS 和 JavaScript 为带显示的 AI 眼镜开发独立体验，不必配套原生手机应用；其二是 Meta AI Connectors 开发者预览，可以把企业自己的 API 变成 Meta AI 可调用的工具，让用户通过眼镜里的 AI 助手直接访问第三方服务。

Meta 还在 Web Apps 中测试 WebMCP，让开发者明确声明“AI 可以在网页里调用哪些动作”。这实际上把 MCP 式工具调用进一步延伸到可穿戴终端：用户看到一个真实世界对象、说一句话，AI 可能就直接调用外部服务完成查询、记录、导航或交易。

AI 眼镜的竞争由此开始从硬件参数转向生态。真正决定它是否成为新计算入口的，不只是重量、显示和续航，而是第三方服务能否快速接入，权限能否足够细，AI 能否在“眼前场景—一个人上下文—外部服务”之间形成可靠闭环。

趋势观察

今天几条技术新闻分别补上了 Agent 系统的四个关键层：OpenAI 事件提醒企业必须治理数据外传；CLM-8B 说明大量 Agent 决策可以从生成式大模型剥离；Anthropic 与 Akamai 显示 Agent 执行正在创造新的 CPU 和通用云需求；Meta 则把 AI 入口推向眼镜与第三方服务生态。AI 应用正在从“一个模型回答一个问题”演变为一个复杂分布式系统，真正的竞争壁垒将逐渐落到数据边界、模型路由、执行基础设施和终端生态。

参考资料

1. **OpenAI** | 《The Hugging Face incident and other third-party impacts from misaligned models》更新 | 2026-09-25 | 核验 53 例用户图片、第三方数据传输及新增安全措施。

<https://openai.com/index/hugging-face-incident-and-misalignment/>

2. **TechCrunch** | 《Unsecured OpenAI agents posted 53 user images on the internet without the lab's knowledge》 | 2026-09-25 | 交叉核验图片外传事件及用户通知限制。

<https://techcrunch.com/2026/09/25/unsecured-openai-agents-posted-53-user-images-on-the-internet-without-the-labs-knowledge/>

3. **Contrastive-LM / GitHub** | CLM 开源项目 | 2026-09 | 核验训练配方、缓存机制和 Benchmark。

<https://github.com/Contrastive-LM/CLM>

4. **Hugging Face** | Contrastive-LM/CLM-v0.1-8B | 2026-09 | 核验 Apache 2.0、模型结构与使用限制。

<https://huggingface.co/Contrastive-LM/CLM-v0.1-8B>

5. **VentureBeat** | 《Stanford and Nvidia's open CLM-8B caches reusable agent actions and runs up to 9x faster than Jev in tests》 | 2026-09-25 | 补充 CLM 架构与测试结果。

<https://venturebeat.com/technology/stanford-and-nvidias-open-clm-8b-caches-reusable-agent-actions-and-runs-up-to-9>

x-faster-than-jev-in-tests

6. **Akamai** | 《Akamai Announces \$11.6 Billion Multi-year Agreement with Anthropic》 | 2026-09-24 | 核验合同规模、CPU 工作负载、资本开支和认股权安排。

<https://www.akamai.com/newsroom/press-release/akamai-announces-11-6-billion-multi-year-agreement-with-anthropic>

7. **TechCrunch** | 《Anthropic to pay Akamai \$11.6 billion over seven years in cloud deal》 | 2026-09-25 | 补充 Agent CPU 计算需求与交易结构。

<https://techcrunch.com/2026/09/25/anthropic-to-pay-akamai-11-6-billion-over-seven-years-in-cloud-deal/>

8. **Meta for Developers** | 《Building for AI glasses: a path for every builder》 | 2026-09-25 | 核验 AI 眼镜开发路径和工具。

<https://developers.meta.com/resources/videos/building-for-ai-glasses-a-path-for-every-builder/>

9. **Meta for Developers** | 《Meta Connect Recap: How To Build For AI Glasses And Reach an Audience》 | 2026-09-24 | 核验 Web Apps、WebMCP 和 Meta AI Connectors。

<https://developers.meta.com/blog/meta-connect-recap-ai-glasses/>

10. **Meta for Developers** | Wearables Developer Portal | 2026-09 访问 | 补充 Device Access Toolkit 1.0 和支持设备范围。

<https://developers.meta.com/wearables/>

联系我们，请扫描二维码



新质生产力工作委员会
官方公众号



工业智能算网
gyznsw.cn

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。

网站地址：<https://gyznsw.cn>