

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 10 月 3 日

摘要

今天的几条消息放在一起看，Agent 系统的分工越来越具体。苹果计划收紧 macOS “完全磁盘访问”，处理 Agent 接触本机数据的风险；Strands Agents 开源 20 亿参数的 Strands Decider，把模型路由、工具选择、策略分类等高频判断交给小模型，在本地 RTX 3090 上的中位决策延迟约 115 毫秒；Photon 获得 450 万美元种子轮融资，想用同一套基础设施接入 iMessage、WhatsApp、Telegram、短信、邮件和语音；Armadin 完成 2.555 亿美元 B 轮融资，用 “Agent swarm” 持续模拟攻击；微软的 2026 Digital Defense Report 则把 Agent 身份、权限、数据和工具纳入安全管理的讨论。大模型仍负责难题，但入口、决策和权限控制，正由不同的系统分别承担。

Contents

一、苹果收紧 macOS 完全磁盘访问：Agent 安全开始进入操作系统权限层	2
二、Strands Decider 2B 开源：Agent 开始用 “小模型做决策、大模型解难题”	2

三、Photon 获 450 万美元融资：Agent 入口从“下载一个 App” 转向“进入已有消息渠道”	3
四、Armadin 融资 2.555 亿美元：红队测试从周期性人工项目走 向持续 Agent 攻击模拟	4
五、微软 2026 数字防御报告：Agent 身份、权限与工具必须进入 持续安全管理	5
趋势观察	5
参考资料	6

一、苹果收紧 macOS 完全磁盘访问：Agent 安全开始进入操作系统权限层

10 月 2 日，多家科技媒体披露，苹果计划加强 macOS 的“Full Disk Access（完全磁盘访问）”控制，回应 AI Agent 带来的风险。这项权限原本主要供备份、安全等少数应用使用。拿到权限的应用可能读取邮件、消息、浏览历史和大量用户文件，因此它一直是 macOS 隐私体系里的高权限入口。

Agent 让这个旧权限有了新问题。传统应用取得授权后，执行的功能通常相对固定；Agent 则会随任务改变读取的文件、调用的工具，以及结果的去向。若它同时能访问本地文件、浏览器、消息和网络，即使没有恶意代码，一次错误决策也可能泄露大量数据。

苹果的计划是让用户更明确、主动地授权完全磁盘访问，减少应用凭宽泛权限长期接触敏感数据的机会。完整的上线时间表尚未公布，眼下不能把它当作已经全面生效的新规则。

开发者需要留意的，是权限控制的位置变了。高权限 Agent 能否读取本机文件、读取范围有多大、授权持续多久，以及能否外发数据，今后可能同时受操作系统、企业 MDM 和 Agent 运行时约束。Coding Agent、个人助理和桌面自动化都会遇到这道权限关。

二、Strands Decider 2B 开源：Agent 开始用“小模型做决策、大模型解难题”

Strands Agents 于 10 月 2 日介绍并开源 Strands Decider 2B。这是一个约 20 亿参数的“Decision Model”，不是用来聊天或生成自由文本的。给它几个候选选项，它会直接打分并给出可靠度，适合分类、模型路由、工具选择和策略判断等需要反复执行的任务。

项目以 Qwen3.5-2B 为主体，去掉传统语言模型的输出头，换上比较候选答案的 pointer head，并用 LoRA 微调。项目方公布的测试结果是：在本地 RTX 3090 上，小任务的中位延迟约 115 毫秒；在 M3 MacBook 上约 153 毫秒。模型、训练数据和训练脚本也已公开。这些延迟来自项目方测试，实际部署仍要另行复核。

不少 Agent 现在连“要不要调用工具”“选哪个模型”“是否转交人工”这类反复出现的判断，也要调用成本更高的大模型。Decider 试着接手这些判断，把大模型留给需要推理或生成的任务；能省下多少延迟和 Token 成本，仍取决于具体部署。

这指向一种更细的模型分工：快速决策模型处理路由与 Guardrail，通用模型处理复杂推理，专业模型处理代码、法律或行业任务。对企业 Agent 而言，选对每一步该用的模型，比让单个大模型处理所有事更有机会控制成本和稳定性。

三、Photon 获 450 万美元融资：Agent 入口从“下载一个 App”转向“进入已有消息渠道”

消息基础设施初创公司 Photon 完成 450 万美元种子轮融资，由 Gradient 和 A* 共同领投，Vercel、HongShan 等参与。它押注的是一个很实际的使用习惯：人们未必愿意再装一个 AI 应用，却可能直接在每天使用的 iMessage、WhatsApp、Telegram、SMS/RCS、邮件或语音渠道里找 Agent 办事。

Photon 提供统一 API、可扩展 Channel Framework、CLI 和可观测工具，让开发者用一套基础设施把同一个 Agent 接到多个消息渠道。公司称，平台已有超过 4 万名开发者注册，过去四个月收入增长 10 倍，近一个月消息量增长 5 倍。这些都是公司披露的运营数据，还需要后续财务和客户数据验证。

这笔融资关注的不是又多了一个聊天应用，而是 Agent 从哪里接触用户。移动互联网时代，产品往往要靠 App Store 分发；在 Photon 设想的路径里，用户给号码发消息或向 WhatsApp 账号下指令，就可能使用机票预订、保险服务或银行业务。这个入口是否能成立，还要看实际产品如何落地。

Photon 还在看 Agent-to-Agent 通信。主 Agent 可以把机票、酒店、金融等任务交给专业 Agent，消息基础设施连接的对象便可能从“人—Agent”扩展到“Agent—Agent—人”。

四、Armadin 融资 2.555 亿美元：红队测试从周期性人工项目走向持续 Agent 攻击模拟

10 月 1 日，AI 安全公司 Armadin 宣布完成 2.555 亿美元 B 轮融资，估值超过 25 亿美元，总融资额达到 4.45 亿美元，由 Andreessen Horowitz 与 Accel 共同领投。公司由 Mandiant 创始人 Kevin Mandia 创办，打算用自主 Agent 持续模拟攻击者，而不是只在一年中做几次人工渗透测试。

漏洞扫描器会报出许多“存在漏洞，但未必能利用”的问题；人工红队质量高，却很难频繁、大范围地运行。Armadin 要让大量 Agent 尝试把漏洞、身份与配置缺陷连成可利用的攻击路径，帮助安全团队先处理那些能够造成突破的组合风险。

持续测试是 Agent 适合做的事：侦察、尝试、验证、回滚，再按结果重新规划，都可以反复运行，高频执行的成本结构也有别于人工团队。但运行攻击工具并能自主规划的平台，本身必须严格隔离，否则它也会成为新的风险面。

这轮融资显示，市场愿意为持续的 Agent 攻防能力投入资金。企业部署的自主 Agent 越多，防守方越可能需要持续检查暴露面、权限和攻击路径。至于网络安全会不会成为 Agent 最早形成闭环 ROI 的行业之一，

还得看产品效果。

五、微软 2026 数字防御报告：Agent 身份、权限与工具必须进入持续安全管理

微软于 10 月 1 日发布 2026 Digital Defense Report。报告指出，AI 正在缩短网络攻击的时间：2026 年上半年公开 CVE 接近 4 万个；攻击者用 AI 加快侦察、社会工程、恶意代码开发和漏洞利用，一些攻击链环节从过去的“天”缩短到“秒”。

报告特别提醒，不能只管 Agent 背后的模型。Agent 还会接入企业数据、API、工具与业务系统。安全团队需要弄清它以什么身份行动、有哪些权限、能调用哪些工具，何时撤销访问权限，以及事后能否追溯行为。

AI 没有让原有安全问题消失。有效账号、过度权限、暴露服务和供应链仍是攻击者常用的入口；AI 只是让人更快地发现并成规模地利用这些弱点。企业每上线一个 Agent，也就多了一个需要管理的数字身份。

Delinea 近期的调查也提到一道落差：几乎所有组织都有 AI 数据访问政策，但能在 Agent 行动时实时执行授权并追溯的比例低得多。企业下一步要做的，不是只把政策写出来，还要让最小权限、实时授权、自动撤权和会话审计在运行时真正起作用。

趋势观察

这几条消息分别落在不同的层上：苹果管本机权限，Strands Decider 处理高频决策，Photon 做消息入口，Armadin 持续测试攻击路径，微软提醒企业管好 Agent 身份和工具访问。把它们放在一起，Agent 的落地问题就不只是“选哪个模型”了。成本能否压住、权限能否收住、行为能否查清，都要在系统运行时回答。

参考资料

1. **The Verge** | 《Apple will limit Mac disk access as AI agents 'substantially' increase risk》 | 2026-10-02 | 用途：核验苹果收紧 macOS 完全磁盘访问及 Agent 风险背景。

<https://www.theverge.com/tech/1004295/apple-limit-mac-disk-access-ai-agents>

2. **WIRED** | 《A Flaw in ChatGPT's Mac App Could Have Let Hackers Grab Sensitive Data》 | 2026-10-02 | 用途：背景资料，说明桌面 AI 应用高权限带来的攻击面。

<https://www.wired.com/story/a-flaw-in-chatgpts-mac-app-could-have-let-hackers-grab-sensitive-data>

3. **Strands Agents** | 《Introducing Strands Decider 2B: a small, open source, decision model》 | 2026-10-02 | 用途：核验模型架构、开源方式、延迟及适用任务。

<https://strandsagents.com/blog/introducing-strands-decider/>

4. **TechCrunch** | 《Photon held a funeral for mobile apps. Now it has \$4.5M to help replace them with agents》 | 2026-10-01 | 用途：核验 Photon 融资、开发者规模与消息 Agent 定位。

<https://techcrunch.com/2026/10/01/photon-held-a-funeral-for-mobile-apps-now-it-has-4-5m-to-help-replace-them-with-agents/>

5. **Seedtable** | 《Photon Raises 4.5M USD in Seed Funding》 | 2026-

10-01 | 用途：交叉核验投资方和融资规模。

<https://seedtable.com/companies/photon/funding-rounds/seed-2026-10>

6. **Armadin** | 《Investors Back Armadin Platform Strategy with New Round of Funding》 | 2026-10-01 | 用途：核验 2.555 亿美元 B 轮与资金用途。

<https://www.armadin.com/blog-posts/investors-back-armadin-platform-strategy-with-new-round-of-funding>

7. **Reuters** | 《AI cybersecurity startup Armadin valued at over \$2.5 billion after new funding round》 | 2026-10-01 | 用途：交叉核验估值、融资与 Agent 安全市场。

<https://www.reuters.com/legal/transactional/ai-cybersecurity-startup-armadin-valued-over-25-billion-after-new-funding-round-2026-10-01/>

8. **Microsoft** | 《2026 Digital Defense Report》 | 2026-10-01 | 用途：核验 CVE、AI 攻击链和 Agent 身份权限治理趋势。

<https://www.microsoft.com/en-us/security/security-insider/threat-landscape/2026-digital-defense-report>

9. **Microsoft Security Blog** | 《Insights from the 2026 Microsoft Digital Defense Report》 | 2026-10-01 | 用途：补充 Agent 身份、数据、工具和持续安全治理。

<https://www.microsoft.com/en-us/security/blog/2026/10/01/insights-from-the-2026-microsoft-digital-defense-report/>

10. **Delinea** | 《2026 Identity Security Report: The AI Enforcement Gap》
| 2026-09-29 | 用途：背景资料，核验 AI 政策与实时执行之间的差距。

<https://delinea.com/news/2026-identity-security-report-the-ai-enforcement-gap>

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>