

AI 技术每日分析：Ant International 把 Agent 接入金融运营

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 19 日

摘要

今天 AI 技术方向最值得关注的变化，集中在“金融业务原生 Agent、预测型 AI、安全滥用治理、隐私保护和 Agent 攻击面”五个层面。Ant International 在 9 月 18 日公布面向全球企业的全栈 AI-native 金融方案，把支付、账户、外汇、资金管理与增长运营统一到 Agent workflows 中，并首次提出 Account for Agent；英国创业公司 Mantic 完成 2500 万美元种子融资，其预测系统在 Metaculus Cup 中击败人类参赛者，显示“面向决策结果优化”的垂直推理系统正在成为独立赛道；Anthropic 发布 9 月威胁情报报告，披露 AI 在网络攻击、欺诈、监控和非法模型蒸馏中的使用方式，并指出自主 Agent 正在明显压缩攻击者的时间和人力成本；AgentCloak 推出免费的桌面端 Sovereign AI 隐私工具，在提示发送给主流 AI 服务之前先对本地敏感数据做识别和脱敏；Hacktron 研究人员则披露，他们借助 Claude 完成漏洞研究并串联两处缺陷，最终接管 OpenAI 员工 ChatGPT/Codex 账户，凸显 Agent 与外部服务连接后身份和供应链边界的新风险。AI 产业正在从模型能力竞争转向“可控部署、端到端工作、工程交付和可验证任务完成率”的系统竞争。

Contents

一、Ant International 发布全栈 AI-native 金融方案：Agent 开始直接进入支付、账户与资金运营	2
二、Mantic 融资 2500 万美元：预测 AI 从“聊天回答”走向可验证的未来判断	2
三、Anthropic 发布 9 月威胁情报：自主 Agent 正在降低攻击活动的单位成本	3
四、AgentCloak 推出桌面版：隐私保护开始前移到“提示离开电脑之前”	4
五、研究人员借助 Claude 攻破 OpenAI 员工账户：Agent 连接器放大身份边界风险	5
趋势观察	5
参考文献	6

一、Ant International 发布全栈 AI-native 金融方案:Agent 开始直接进入支付、账户与资金运营

9 月 18 日, Ant International 在全球商户活动上发布面向支付、账户、外汇、资金和增长运营的全栈 AI-native 解决方案, 覆盖近 100 项产品能力。底层包括支付基础模型 Antom 3-in-1 Transformer 和面向外汇及流动性预测的时间序列模型 FalconTST, 并以双层信任与安全架构支撑 Agent 化操作。

其中最值得关注的是 Account for Agent (AFA)。过去企业 Agent 可以读取财务数据、推荐资金操作, 但真正执行付款、账户配置和资金调度仍需回到传统系统; AFA 试图给 Agent 建立可控、可追踪的账户与交易身份, 使其能够在授权范围内完成业务动作。公司还在支付、账户和资金管理中加入 Autopilot, 让用户通过自然语言发起和编排操作。

Ant International 披露, 其主要商户支付服务客户中已有 89.5% 使用 FinAI 能力, 81.4% 的支付任务得到 AI 辅助。上述数据属于企业口径, 仍需结合具体任务定义理解, 但它说明 Agent 商业化已经不只停留在客服和代码领域, 而是进入强监管、强资金约束的核心金融流程。

这一方向对企业 Agent 具有代表性: 未来“会不会调用 API”不是关键, 关键是 Agent 是否有明确账户身份、额度、权限、审计记录和异常处理机制。支付系统正在成为检验 Agent 治理成熟度最严格的场景之一。

二、Mantic 融资 2500 万美元: 预测 AI 从“聊天回答”走向可验证的未来判断

Reuters 9 月 18 日报道, 伦敦 AI 创业公司 Mantic 完成 2500 万美元种子轮融资, 由 Radical Ventures 领投, Microsoft 旗下 M12、Thinking Machines Lab 和 Balderton Capital 等参与。

Mantic 并不主攻通用对话，而是针对政治、经济、文化和商业事件进行概率预测。公司系统在 2026 年 Metaculus Cup 中整体表现超过人类参赛者，随后开始吸引企业和政府客户。其方法是基于前沿模型叠加专门的预测训练、历史数据回测和校准机制，让模型不仅给出结论，还给出可以后续验证的概率判断。

这类系统的重要性在于，AI 输出终于可以直接用现实结果进行评价。普通企业问答很难定义“答案到底有多好”，而预测问题可以在未来事件发生后测量 Brier Score、校准度和长期胜率。随着 Agent 进入供应链、金融、项目管理和风险决策，企业会越来越需要这种可量化的预测能力。

当然，“在预测竞赛中超过人类”并不等于能够稳定预测所有重大事件。真正商业化仍需要解决数据泄漏、分布变化、黑天鹅以及模型对市场共识的依赖问题。但它提供了一个新方向：AI 产品可以围绕可验证结果优化，而不是围绕语言流畅度优化。

三、Anthropic 发布 9 月威胁情报：自主 Agent 正在降低攻击活动的单位成本

Anthropic 最新发布《Detecting and countering misuse of AI: September 2026》。报告总结了 2025 年 12 月至 2026 年 8 月期间发现并处置的恶意使用案例，涵盖网络攻击、监控、影响行动、欺诈、生物滥用、武器相关活动和非法模型蒸馏。

从技术趋势看，最值得关注的是“从助手到编排器”的变化。Anthropic 表示，部分攻击者已经不只是让模型写脚本，而是让 Agent 执行命令、收集凭据、处理多个目标，甚至让多 Agent 框架连续运行数小时或数天。报告称，一些原本需要团队完成的攻击活动被压缩到两三小时，并可以同时处理多个目标。

报告也承认，安全机制并非总能阻止恶意操作。攻击者会把完整目标

拆成多个看似无害的小任务，跨不同会话组合结果。这与传统内容安全的逻辑不同：单次请求安全，并不意味着长周期工具链最终行为安全。

因此，Agent 安全下一步必须更重视跨会话行为、工具调用链和最终结果，而不能只做逐条 Prompt 审核。模型厂商、安全平台和企业部署方都需要建立更强的身份验证、速率控制、异常检测与事件追踪。

四、AgentCloak 推出桌面版：隐私保护开始前移到“提示离开电脑之前”

9 月 18 日，AgentCloak 发布 AgentCloak Desktop，定位为免费的 Sovereign AI 隐私工具。它运行在用户本地，在用户把内容发送给 ChatGPT、Claude 等服务之前，先识别和替换敏感数据，尽量避免姓名、账号、密钥、公司数据等内容直接进入第三方 AI 平台。

这个思路的价值在于，它不要求企业自建完整大模型。很多中小企业希望继续使用成熟云模型，但又担心员工把合同、客户信息和内部数据粘贴到聊天工具。AgentCloak 希望通过本地“隐私代理层”解决两者之间的矛盾。

这类产品是否真正安全，最终要看识别准确率、日志处理、更新机制以及自身是否引入新的权限风险。但它代表一个越来越重要的产品类别：AI 安全不只是模型侧的内容过滤，也可以在数据离开终端之前建立一层独立控制。

未来企业 AI 入口可能像网络安全一样分层：终端侧做数据识别与脱敏，身份层做权限，Agent 层做动作治理，云端再做模型安全和审计。

五、研究人员借助 Claude 攻破 OpenAI 员工账户：Agent 连接器放大身份边界风险

9 月 18 日，The Register 报道，Hacktron 研究人员在漏洞研究过程中借助 Claude 发现并利用图像处理链中的漏洞，随后串联账号认证问题，接管了 OpenAI 员工的 ChatGPT 与 Codex 账户，并最终证明可以让被接管账户的 Codex 在 OpenAI 内部 GitHub 组织中发起一个无害 PR。研究人员随后停止测试并上报问题。

据报道，OpenAI 收到报告后约 14 小时修复相关问题，并支付 6500 美元赏金；Discourse 也发布了图像处理沙箱相关修复。整个链路从发现到验证影响不到 72 小时。

这次事件最值得关注的并不是“Claude 攻击 OpenAI”的戏剧性，而是 AI 工具本身成为了安全研究的加速器。复杂漏洞发现、Exploit 生成和跨系统验证过去可能需要数周，现在可以在 Agent 辅助下快速推进。

与此同时，企业把 GitHub、Slack、邮箱和云系统连接给 Coding Agent 之后，一个身份系统漏洞可能产生更大的后果。因此，Agent 时代的安全边界需要从“保护账号”进一步升级到“限制已登录 Agent 能够代表用户做什么”。

趋势观察

今天五条新闻共同指向一个变化：AI 开始真正进入有资金、有身份、有数据、有责任的业务系统。Ant International 把 Agent 接入金融运营，Mantic 把 AI 输出变成可验证预测，Anthropic 展示攻击者如何使用 Agent 放大效率，AgentCloak 把隐私保护前移到终端，Hacktron 案例则暴露连接器与身份边界风险。下一阶段企业 AI 的竞争重点，将越来越接近交易权限、预测准确率、数据边界、攻击面和可追溯责任，而不只是模

型排行榜。

参考文献

- Ant International / PR Newswire | 《Ant International Launches Industry's First Full-Stack AI-Native Solutions for Payment, Account, FX, Treasury and Growth Operations for Global Businesses》 | 2026-09-18 | 核验全栈 FinAI、AFA、基础模型与企业使用数据。<https://www.prnewswire.com/apac/news-releases/ant-international-launches-industrys-first-full-stack-ai-native-solutions-for-payment-account-fx-treasury-and-growth-operations-for-global-businesses-302883018.html>
- Reuters | 《AI startup Mantic raises \$25 million for superhuman forecasting》 | 2026-09-18 | 核验融资、Metaculus Cup 表现与预测 AI 定位。<https://www.reuters.com/technology/ai-startup-mantic-raises-25-million-superhuman-forecasting-2026-09-18/>
- Anthropic | 《Detecting and countering misuse of AI: September 2026》 | 2026-09 | 核验自主网络攻击、跨会话规避与非法蒸馏等案例。<https://www.anthropic.com/threat-intelligence-report-september-2026>
- Business Wire / AgentCloak | 《AgentCloak Launches AgentCloak Desktop》 | 2026-09-18 | 核验桌面端 Sovereign AI 隐私工具。<https://www.businesswire.com/news/home/20260918884045/en/>
- The Register | 《Researchers used Claude to hack OpenAI employees' ChatGPT accounts》 | 2026-09-18 | 核验 Hacktron 漏洞链、修复时间与影响范围。<https://www.theregister.com/security/2026/09/18/researchers-used-claude-to-hack-openai-employees-chatgpt-accounts/5297517>

- Reuters | 《Anthropic quietly sets up biology lab as it ramps AI drug program》 | 2026-09-18 | 背景资料, 观察模型公司向实体科研基础设施延伸。<https://www.reuters.com/world/anthropic-quietly-sets-up-biology-lab-it-ramps-ai-drug-program-2026-09-18/>
- Reuters | 《Europe's AI firms, playing catch-up, challenge US calls for slowdown》 | 2026-09-18 | 背景资料, 观察欧洲 AI 竞争与安全政策争论。<https://www.reuters.com/business/europes-ai-firms-playing-catch-up-challenge-us-calls-slowdown-2026-09-18/>
- Info-Tech Research Group / PR Newswire | AI Agents 与产品决策治理报告 | 2026-09-18 | 长尾参考, 观察 Agent 人类授权边界。<https://www.prnewswire.com/news-releases/ai-agents-are-influencing-product-decisions-without-explicit-human-authorization-warns-info-tech-research-group-302883560.html>

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>