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 5 月 13 日

摘要

前沿模型正在直接进入网络安全防御产品，OpenAI 推出 Daybreak，Anthropic 的 Mythos 也被美国国防部用于漏洞修复；模型公司开始把能力推向更垂直的专业行业，Anthropic 扩大 Claude 法律工具生态；企业 Agent 落地的瓶颈从“会不会执行”转向“能否被身份治理、权限控制和审计系统约束”。

Contents

1 1. OpenAI 推出 Daybreak，网络安全成为前沿模型主战场

OpenAI 发布 Daybreak，将其定义为“让软件从设计阶段就具备韧性”的网络安全计划。官方页面强调，下一代网络防御不应只是在漏洞出现后修补，而应在软件建设流程中更早发现风险、验证漏洞、生成补丁并推动“resilient by design”。The Verge 和 CIO Dive 进一步披露，Daybreak 将 Codex Security 代理与 OpenAI 最新模型能力结合，用于威胁建模、攻击路径分析、漏洞验证和修复自动化。这一动作说明，AI 安全竞争正在从聊天模型安全，转向谁能进入代码仓库、安全运营中心和企业漏洞修复流程。

2 2. 美国国防部部署 Anthropic Mythos，但同时准备摆脱供应链依赖

Reuters 报道称，美国国防部正在通过 Project Glasswing 部署 Anthropic 的高级网络安全模型 Mythos，用于发现并修补政府系统中长期存在的软件漏洞。但报道同时称，五角大楼正计划逐步转向其他 AI 供应商，原因涉及使用条款争议和供应链风险评估。这个案例很值得关注：它一方面证明前沿模型已经能进入国家级网络安全任务，另一方面也说明在关键基础设施中使用单一模型供应商，会迅速上升为采购、安全和主权能力问题。

3 3. Anthropic 扩大 Claude 法律工具，专业行业 Agent 开始加速

Reuters 报道，Anthropic 为法律专业人士扩展 Claude 功能，新增法律专题工具和外部服务连接，可在 Claude 中接入 Thomson Reuters Westlaw Primary Law、Practical Law、Everlaw、DocuSign、Box 等服务。Thomson Reuters 也宣布 CoCounsel Legal 与 Claude 连接。与通用问答不同，法律场景要求事实来源、权限、流程和责任链可控，因此这类更新代表 AI 正在进入“带工具、带来源、带行业工作流”的专业生产系统。

4 4. SAP 与 NVIDIA 强调企业级 Agent 运行安全

SAP 官方发布与 NVIDIA 围绕企业级 Agent 执行安全的合作内容，强调 OpenShell 作为 SAP Business AI Platform 中所有 SAP AI Agent 的运行时安全层，覆盖 Joule Studio 构建的自定义 Agent。NVIDIA 博客也指出，企业 Agent 需要运行在有治理、审计、身份和合规边界的环境里。这个方向与企业落地痛点高度一致：Agent 不是一个更聪明的聊天框，而是新的“非人类操作者”，必须被纳入企业 IAM、日志和权限体系。

5 5. Agent 身份安全报告揭示企业试点的隐性风险

Akeyless 发布的 2026 年 AI Agent 身份安全报告称，67

6 结语

今天 AI 行业的核心不是“模型又强了一点”，而是强模型如何进入安全、法律和企业流程。网络安全正在成为前沿模型公司正面竞争的主战场；行业工具链正在成为模型落地的新入口；企业 Agent 的下一道门槛则是身份治理、权限分层和审计闭环。

7 参考资料

1. OpenAI | Daybreak | OpenAI for cybersecurity | 2026 年 5 月。用于支撑 OpenAI Daybreak 官方定义、目标和安全设计理念。链接：https://openai.com/daybreak/
2. The Verge | OpenAI just released its answer to Claude Mythos | 2026 年 5 月 12 日。用于支撑 Daybreak 与 Codex Security、GPT-5.5-Cyber 相关功能介绍。链接：https://www.theverge.com/ai-artificial-intelligence/928342/openai-daybreak-security-ai
3. CIO Dive | OpenAI launches Daybreak to combat cyber threats | 2026 年 5 月 12 日。用于补充 Daybreak 合作伙伴和企业安全场景。链接：https://www.ciodive.com/news/OpenAI-Daybreak-cyber-threats/820036/
4. Reuters | Pentagon deploys Anthropic's Mythos to patch cyber gaps while planning to ditch firm | 2026 年 5 月 12 日。用于支撑美国国防部部署 Mythos 及供应链风险争议。链接：https://www.reuters.com/technology/pentadeploys-anthropics-mythos-patch-cyber-gaps-while-planning-ditch-firm-2026-05-12/
5. Reuters | Anthropic expands Claude's AI tools for law firms, lawyers | 2026 年 5 月 12 日。用于支撑 Claude 法律工具和法律平台集成。链接：https://www.reuters.com/legal/litigation/anthropic-expands-claudes-ai-tools-law-firms-lawyers-2026-05-12/
6. Thomson Reuters | Thomson Reuters and Anthropic Expand Partnership to Connect Claude with CoCounsel Legal | 2026 年 5 月 12 日。用于补充 Claude 与 CoCounsel、Westlaw 等法律工具连接。链接：https://www.morningstar.com/news/pr-newswire/20260512ny56719/thomson-reuters-and-anthropic-expand-partnership-to-connect-claude-with-cocounsel-

legal

7. SAP News | SAP and NVIDIA: Enterprise-Grade Agent Execution | 2026 年 5 月 12 日。用于支撑 SAP 与 NVIDIA 围绕企业级 Agent 治理和安全执行的合作。链接：https://news.sap.com/2026/05/secure-ai-agents-how-sap-and-nvidia-co-define-enterprise-grade-agent-execution/
8. NVIDIA Blog | NVIDIA and SAP Bring Trust to Specialized Agents | 2026 年 5 月 12 日。用于补充 OpenShell、Joule Studio 和企业 Agent 安全层信息。链接：https://blogs.nvidia.com/blog/sap-specialized-agents/
9. Akeyless | 2026 State of AI Agent Identity Security report | 2026 年 5 月 12 日。用于支撑企业 Agent 身份和凭据风险调研数据。链接：https://www.prnewswire.com/news-releases/two-thirds-of-enterprises-suspect-ai-agents-have-already-accessed-unauthorized-data-akeyless-finds-302769768.html
10. Reuters | OpenAI gives European companies access to its latest models to bolster resilience | 2026 年 5 月 12 日。用于补充 OpenAI 可信网络安全访问在欧洲落地情况。链接：https://www.reuters.com/sustainability/boards-policy-regulation/openai-gives-european-companies-access-its-latest-models-bolster-resilience-2026-05-12/

关注我们

扫码关注高促会新质生产力工委会

扫码关注工业智能算网平台

获取更多 AI 技术、产业趋势与研究报告