

# AI 技术每日分析：月之暗面推出 Kimi K3.1 Pro：百万 Token 上下文与强化学习扩展

中国高技术产业发展促进会新质生产力工作委员会

2026 年 10 月 25 日

## 摘要

10 月 24 日，月之暗面推出 Kimi K3.1 Pro，强化长上下文、代码开发与复杂智能体任务，并在开放权重之前主动调整网络攻击相关训练数据。开源工具方面，TanStack AI 1.0 进入稳定版本，思科推出 Agentic FAITH 多步骤安全智能体评测框架。与此同时，《华尔街日报》在 24 日的调查报道中关注 AI 辅助网络攻击扩散，显示模型能力提升正在同时推动开发效率与安全投入增长。模型、智能体执行框架和安全评测正在彼此牵动。

## Contents

|                                            |   |
|--------------------------------------------|---|
| 一、月之暗面推出 Kimi K3.1 Pro：百万 Token 上下文与强化学习扩展 | 2 |
| 二、TanStack AI 1.0 发布：76 个软件包统一进入稳定阶段       | 2 |
| 三、思科推出 Agentic FAITH：安全智能体评测开始记录全过程        | 3 |
| 四、《华尔街日报》调查 AI 辅助攻击：安全运营面对更快的攻防循环          | 4 |
| 趋势观察                                       | 4 |
| 参考资料                                       | 4 |

## 一、月之暗面推出 Kimi K3.1 Pro：百万 Token 上下文与强化学习扩展

月之暗面 10 月 24 日发布 Kimi K3.1 Pro 及 Highspeed 版本。新模型属于约 3 万亿参数量级的开放权重系列，支持 100 万 Token 上下文，面向代码开发、长周期智能体任务、三维内容生成与金融研究。官方公布的 Terminal-Bench 4.0 成绩为 65.7%，内部 Kimi Code Bench 3.0 Long-horizon 为 61.2%。这两组数字分别对应特定工具环境和任务集，比较模型时还需结合实际部署的推理预算、工具链与任务长度。

与模型同步推出的 Highspeed 版本，厂商称响应速度达到普通版的 2—3 倍。Kimi 网站、Kimi Work、Kimi Code 和 API 已经提供新模型服务，完整模型权重计划在 11 月 4 日前发布。由此，开发者可以先通过托管服务测试复杂任务，随后再根据硬件成本、数据管理需求选择自建部署。

本次发布的另一项技术细节是网络安全能力控制。月之暗面表示，在训练中清理部分网络攻击相关数据，并调整强化学习环境。其内部测试中，经过调整的模型在一组漏洞利用任务上的成绩为 81.0%，未经相同调整的对照模型为 90.5%。企业在追求更强编程能力时，也开始把开放权重后的滥用风险纳入模型研发流程。该结果来自厂商内部评测，后续开放权重后的第三方研究值得继续跟踪。

## 二、TanStack AI 1.0 发布：76 个软件包统一进入稳定阶段

TanStack 团队 10 月 23 日发布 TanStack AI 1.0，宣布 76 个软件包的核心 API 进入稳定阶段。项目目前提供 23 种模型服务商适配器、11 种沙箱服务商接入方式，并可与 React、Vue、Svelte、Angular 等前端技术栈协同工作。聊天、工具调用、多模态能力、MCP 和代码智能体可以采用相对统一的编程接口。

开源框架让应用团队更容易切换模型服务商。一个产品通常需要接入模型、检索、工具、代码运行环境及用户界面，长期维护成本更多来自接口变化和组件兼容。TanStack AI 试图用类型化接口和跨框架适配减少重复开发。

官方披露，1.0 分支已经具备覆盖 17 个服务商适配器的 943 项确定性端到端测试，并对文档中大量代码示例实施持续类型检查。新推出的 Agent Harness 仍处于实验阶段；采用 1.0 的现有项目也需要同步升级客户端和服务端相关软件包。项目成熟度的提升，对规模较小的 AI 应用团队尤为重要。

### 三、思科推出 Agentic FAITH：安全智能体评测开始记录全过程

思科 Foundation AI 团队 10 月 23 日发布 Agentic FAITH，将原有网络安全模型评测框架扩展到需要连续调用工具的智能体。首个接入的基准 VulnLoc 要求智能体阅读代码仓库、检索相关逻辑、定位存在漏洞的文件，并在修复后的仓库中识别问题已经消失。

评测被分为两个阶段：第一阶段以文件集合 F1 分数考察漏洞定位，第二阶段以真阴性率衡量智能体对修复结果的判断。框架同时保存模型对话、工具调用、执行错误、提交答案和结束原因，使研究人员能够追溯错误究竟出现在检索、分析还是任务停止环节。

安全智能体可能连续执行几十次文件检索和命令操作。即使最终答案正确，也要检查中间路径。Agentic FAITH 支持 Docker 或 Kubernetes 隔离环境，并限制评测任务对代码仓库的写入权限。其架构也为企业测试代码审查、运维诊断和自动化合规检查提供了可借鉴的方法。

## 四、《华尔街日报》调查 AI 辅助攻击：安全运营面对更快的攻防循环

《华尔街日报》10 月 24 日刊文讨论 AI 工具对网络攻击门槛和攻击频率的影响。随着自动化代码生成、漏洞分析和多步骤任务执行逐渐成熟，攻击者可以用更少的人力完成过去需要专业脚本和大量重复工作的活动。企业已经开始面对更密集的正常登录、API 滥用与自动化试探。

这类变化推动安全运营从定期规则检查走向持续监测。身份授权、沙箱隔离、令牌管理和工具调用日志比单纯依靠大模型输出审核更重要。下一阶段 AI 安全产品可能更多围绕智能体运行行为展开，例如工具调用溯源、执行策略校验及自动化事件复盘。面向外部开放的智能体应用尤其需要将“任务执行权限”与“用户查看权限”分别管理。

### 趋势观察

这些技术更新各有分工：大模型继续扩大可处理任务的复杂度，开源框架提供跨模型和跨环境的开发支撑，安全团队则把评测对象推进到智能体每一步实际操作。对企业研发而言，下一阶段值得投资的基础能力包括稳定的工具接口、可回放的执行记录、标准化评测任务，以及可独立审核的权限控制。

### 参考资料

1. Moonshot AI | Introducing Kimi K3.1 Pro | 2026-10-24 | 用途：核验模型发布、百万 Token、编程评测、Highspeed 与权重开放时间  
<https://www.kimi.ai/blog/kimi-k3-1-pro>
2. Reuters | Moonshot curbs hacking ability in new Kimi model as fears grow over AI-powered cyberattacks | 2026-10-24 | 用途：交叉核验训

### 练数据调整及网络安全能力评测

<https://www.reuters.com/technology/moonshot-curbs-hacking-ability-new-kimi-model-fears-grow-over-ai-powered-2026-10-24/>

3. TanStack 官方博客 | TanStack AI 1.0: Still Your AI, Still Your Way | 2026-10-23 | 用途: 核验 76 个稳定软件包、23 个适配器、11 种沙箱及实验性 Harness

<https://tanstack.com/blog/tanstack-ai-1-0>

4. Cisco 官方技术博客 | Introducing Agentic Evaluation in FAITH | 2026-10-23 | 用途: 核验 VulnLoc 评测、F1、真阴性率和执行轨迹记录

<https://blogs.cisco.com/ai/introducing-agentic-evaluation-in-faith>

5. Cisco FAITH 文档 | FAITH Evaluation Framework | 2026-10-23 访问 | 用途: 了解基准任务部署与扩展机制

<https://faith-evals.com/>

6. The Wall Street Journal | AI Hacks Are Coming Thick and Fast—and Overwhelming the Internet | 2026-10-24 | 用途: 观察 AI 辅助攻击的发展及企业防护压力

<https://www.wsj.com/tech/ai/ai-hacks-are-coming-thick-and-fast-and-overwhelming-the-internet-4529c26c>

7. TanStack 官方文档 | TanStack AI Docs | 2026-10-24 访问 | 用途: 核验开源框架接口及迁移资料

<https://tanstack.com/ai/latest/docs/blog>