AhnLab

白皮书

携手 AhnLab AI PLUS 创新智能体 AI 安全

今天早上在上班路上，如果AI 已经自动帮你处理好了部分工作，你会怎么想?

有人会觉得轻松便利，但也难免有人会产生焦虑，担心AI 是否会取代自己的工作。但核心问题在于，这种顾虑早已不再是凭空想象。如今的AI 正在全速演进——它不仅能给出简单的回答，更能作为“协作者”深度参与实际业务处理。而这场变革的核心，正是“智能体AI（Agentic AI / AI Agent）”。

本文介绍智能体 AI 的概念、安全负责人的应对策略，以及 AhnLab 智能体 AI 安全平台 AhnLab AI PLUS 的价值与应用成效。

什么是智能体AI?

从宏观角度看，AI 的演进历程可分为以下三个阶段。

预测型AI (Predictive AI)	生成式AI (Generative AI)	智能体AI (Agentic AI)
基于历史数据进行分类或预测	根据提问创建答案， 能够创建文本、代码及图像	超越单纯生成答案， 能够接收目标并自主制定计划加以执行

简而言之，就是从“告知型AI”向“执行型AI”的转变。

而智能体AI 的核心价值体现在以下四个方面：

自主性 (Autonomy) 不需持续等待人类指令，能够自主判断	**目标导向性** (Goal-Driven) 为达成目标执行多阶段任务
工具使用 (Tool Use) 超越单纯的文本生成，主动与外部系统交互	**记忆与学习** (Memory) 记住先前结果，并将其用于后续行动

也就是说，智能体 AI 并非简单的自动化，而是一个以目标为中心的自主执行系统。

Agentic AI 有何不同?

现有生成式AI 与智能体AI 的最大区别在于其是反应型还是执行型。生成式AI 通常是在我们提问时给出答案。到目前为止，这仍是以人为中心的。相反，智能体AI 在人设定目标后，会自行安排顺序，并使用必要的工具来产出结果。

从安全角度来看，如果过去只是“请总结一下这些日志”的水平，那么现在正朝着“查找异常迹象、收集必要信息、进行隔离、创建工单，甚至撰写报告”这样的流程发展。安全领域过去也是以人为中心的控制环境，如今正逐渐转向以AI 为中心。



[图1] 生成式AI vs 智能体AI

近期，对智能体 AI 的关注度显著提升，这主要归因于以下三个方面的成熟度提高。

LLM 的推理能力	Agent 框架/协议	企业环境向API 中心化转型
已超越简单摘要， 能够处理复杂多阶段的问题	连接与执行工具的基础设施得到改善	安全解决方案、工单系统、 云服务、认证均可互联的架构

换言之，技术与连接环境已然就绪，实际需求也随之产生。

市场对智能体AI 的前景预测也极为乐观。市场研究机构Gartner 预测，60% 的品牌会通过智能体AI 提供一对一的个性化客户互动。关键在于，AI 会超越“辅助功能”的范畴，成为执行工作的“组织组成部分”。这也意味着，客户对组织的产品说明、运营方式等方面的期望水平本身很可能发生改变。

AI 时代下安全负责人的角色转变

从安全人员的角度来看，针对这些变化，有四大关键词值得关注。

智能体AI 监管 (Agentic AI Oversight) 随着隐形 AI Agent 的增加，新的攻击面随之产生	**AI 智能体身份与访问管理** (IAM for AI Agents) 不仅用户，AI Agent 也需作为行为主体进行管理
AI 驱动的SOC (AI-Driven SOC) 重复性分析工作预计将更多由AI 承担	**先发制人的安全** (Preemptive Security) 在事前征兆阶段进行防御，而非事发后应对

关键在于，运营、权限、检测和响应体系都可能发生改变。而且，这已不再是遥远未来的故事。
那么，安全人员的角色将如何转变?
过去，主要角色是作为操作员（operator），由人直接操作安全工具并逐一处理警报（alert）；而未来，作为编排者（orchestrator）的角色将变得更为重要，其职责包括设计AI Agent、制定策略、批准例外情况以及验证结果。



[图2] 智能体AI 安全模型

这并非AI 完全取代人类，而是人类的角色将得到提升。可以理解为，相较于处理重复性工作，“监督、判断、制定标准及验证”所占的比重将更大。
现在，让我们来看看在实际接触的安全领域中，预计会出现哪些变化。

#1. 端点安全

在端点安全领域，AI Agent 的应用方案相对明确。

自主威胁检测与隔离	一旦发现可疑勒索软件行为，立即进行隔离，会提取相关IoC（指标），向其他端点部署阻断策略
漏洞优先级自动调整	不再仅依据简单的CVE 评分，而是基于实际暴露程度、资产重要性及攻击者利用可能性，进行实质性风险评估会调整补丁优先级
取证数据采集与报告自动化	自动收集内存转储、进程树、网络连接历史记录等取证信息，并自动创建报告草稿

从实务角度会看，在端点领域，AI 与“缩短平均修复时间（MTTR）和优化威胁分析资源”直接相关。

#2. 网络安全

其次，在网络安全领域的应用要点如下。

自主分段及策略优化	持续分析流量模式并优化策略
动态零信任策略执行	基于上下文动态执行零信任策略，而非静态规则
DDoS & C2 自主阻断	在 DDoS 或 C&C 通信等需快速阻断的领域执行自主阻断
配置错误的自主修正	自动发现并修正配置错误（例如：自动检测/修正云环境中的错误安全组、公开暴露的存储）

在网络安全领域，AI在“策略运营自动化与响应速度提升”方面发挥着核心作用。

#3. 安全运营中心（SOC）

在实际的 SOC 运营中，并非由单一 AI 处理所有事务，而是正朝着多个Agent 分工协作的结构发展。以下是Agent 间协作结构的示例。

检测 Agent	事件检测及异常迹象识别
威胁情报 Agent	通过IP/哈希信誉查询、攻击者画像分析等方式增强上下文信息
取证 Agent	自动收集日志、内存转储等相关证据
响应 Agent	执行隔离或阻断操作以最大限度降低内部影响
报告 Agent	生成工单和报告

最近，这种基于AI Agent 的安全运营模式正被定义为一个全新概念——“智能体SOC（Agentic SOC）”。该架构的核心优势在于高效响应与精细化分工。特别是它能有效解决以往由人工按部就班处理各项流程时，极易出现的流程瓶颈问题。



[图3] 智能体SOC 工作流

或许，安全人员在实际工作中最先能切身感受到的领域，正是基于AI Agent 的安全运营。过去，一旦触发警报，负责人必须亲自查询信誉评分、查找历史记录、核对类似案例，进而整理影响范围和应对方案。

然而，当智能体AI 融入安全运营后，它会分析并汇总决策所需的信息，并以可执行的形式呈现出来。因此，安全人员可以减少用于信息查询和整理的时间，从而专注于更重要的决策。

AI 带来的风险与应对方案

智能体AI 为企业及安全负责人提供了巨大的创新机遇。问题在于，攻击者同样拥有这些机遇。

AI Agent 对攻击者而言也是极具吸引力的工具。因为它能使攻击的速度、规模及复杂程度实现飞跃式提升。攻击者利用AI Agent 的方法如下。

基于AI 的自主攻击活动	利用AI Agent 自主执行从漏洞扫描、初始渗透、内部横向移动到数据泄露的整个攻击链（Kill Chain）
鱼叉式会络钓鱼的升级	通过学习攻击目标的公开信息，批量创建定制化信息，并针对受害者的回复进行互动的“交互式钓鱼Agent”行为
漏洞武器化	AI Agent 即时分析已公开的CVE 信息，自动创建利用代码——缩短补丁发布与攻击运行之间的“黄金时间”

不仅要考虑利用AI 来提升攻击水平，还需警惕针对组织内部AI Agent 发起的攻击所造成的损害。与此相关的具体威胁主要可归纳为以下五类。

提示注入 (Prompt Injection)	注入恶意提示，绕过LLM 的原始指令，诱导其执行错误操作
代理劫持 (Agent Hijacking)	窃取令牌或凭证，伪装成AI Agent 以夺取系统控制权
工具中毒 (Tool Poisoning)	污染 AI Agent 使用的外部工具或插件，从而操控其执行
内存中毒 (Memory Poisoning)	污染向量数据库等长期记忆，破坏判断逻辑
过度授权 (Excessive Permission)	向Agent 授予超出必要的权限，导致一次入侵即可扩大损害

综上所述，虽然引入并善用AI Agent 很重要，但首要任务是构建“能够控制AI 所关联权限及执行路径的架构”。

面对智能体AI 时代，组织最需要做的准备是“AI 安全与治理（AI security & governance）”。这是近期在全球市场也备受关注的概念。

智能体清单管理 (Agent Inventory)	确保组织内所有运营中的AI Agent （包括SaaS）的可视性，并明确其所有者/管理员
最小权限 (Least Privilege)	设置只读默认值，限制Agent 访问的网络分段，按任务单元分离权限
建立“人机协同”标准 (HITL Policy)	设定自动化与必须经批准的操作范围标准，并配备可在Agent 发生故障时立即切断的“紧急关闭开关”
AI 专用认证体系 (Machine Identity)	将Agent 视为机器行为主体，并运营独立的凭证及审计日志体系

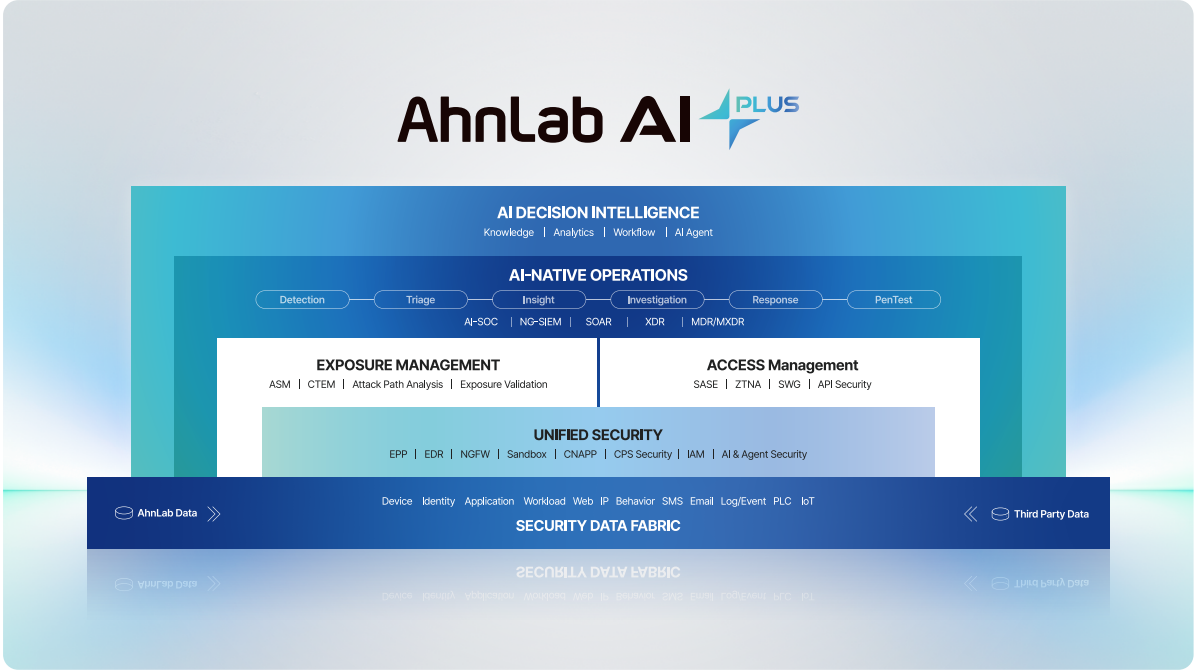
携手 AhnLab AI PLUS 创新智能体 AI 安全

随着攻击者越来越多地使用 AI，防御方应用 AI 已成为必然。AhnLab 以自主构建的智能体 AI 安全平台 AhnLab AI PLUS 为核心，携手客户推进 AI 安全转型。

AhnLab AI PLUS 是代表 AhnLab AI 原生安全战略与方向的统一品牌，也是有机连接技术、产品和服务的平台及架构。AhnLab 以此为核心，从将 AI 应用于现有产品和服务、增强检测、分析、判断与响应能力的“AI 赋能安全”（AI-powered Security），向以 AI 为中心设计产品、服务及安全运营的“AI 原生安全”（AI-native Security）演进。

AhnLab AI PLUS 将分散在不同产品和领域的数据关联到统一的安全上下文，构建基于智能体 AI 的架构，由多个 AI 智能体协作完成安全运营任务。AI 理解重要风险并确定优先级，帮助人员更快、更准确地做出判断，从而构建 AI 原生安全运营体系。

AhnLab AI PLUS 的具体架构及各层的作用如下。

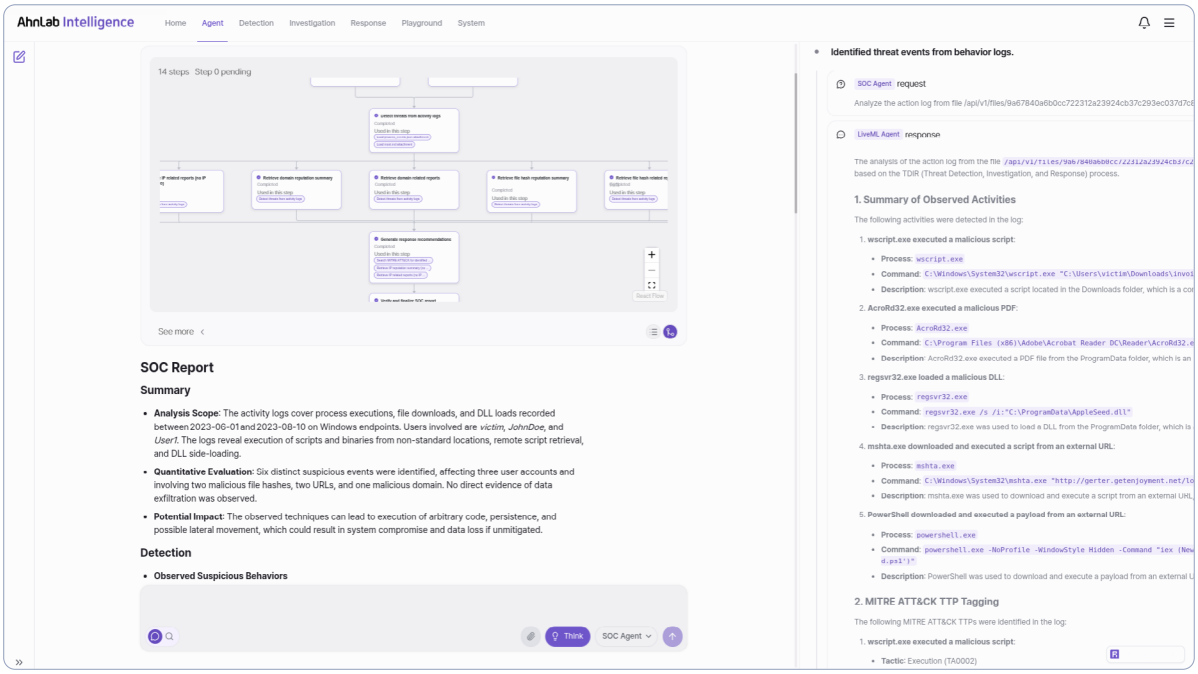


[图 4] AhnLab AI PLUS 架构图

层级	作用
AI Decision Intelligence	• 超越单纯的威胁分析，为整个安全运营设置优先级并支持决策 • 理解并判断海量告警和事件中哪些最重要
AI-Native Operations	• 通过 AI 原生运营自动化、智能化处理检测、分类、分析、响应、攻击模拟及验证等全过程 • 提高安全运营速度并扩大处理规模
Access/Exposure Management	• 通过持续的暴露面管理，减少可能遭受攻击的资产和路径 • 基于信任持续验证用户、设备和应用程序的访问，降低潜在风险
Unified Security	• 将整合现有防护功能的安全平台与 AI 有机结合，增强协同效果
Security Data Fabric	• 从 AhnLab 自有传感器及第三方等多种来源收集和关联安全数据

AhnLab 不仅在单个安全产品中增加 AI 功能，还从一开始就以 AI 为中心连接并重新设计产品、数据和整个安全运营，打造“AI 原生安全”。为此，AhnLab 基于 AhnLab AI PLUS，将端点、网络、云、OT、威胁情报和安全监控等安全领域的数据与功能关联到统一的安全上下文。

各产品和服务产生的告警、事件和遥测数据在统一数据结构中加以利用。AI 分析各信号之间的关联、攻击流程及影响范围，将多个产品产生的信号重构为一个安全事件，并确定响应优先级。安全运营人员无需逐一查看各产品的告警，即可根据整合后的事件和上下文了解重要风险以及所需的调查和响应措施。



[图 5] AhnLab AI 原生传感器：基于上下文识别风险优先级

通过 [图 5] 可以更直观地了解这一过程。AhnLab 的 20 个 AI 原生传感器在最近一个月共产生 1,969,908 条告警。按此估算，人工逐条确认约需 65,664 工时，工作量巨大。AhnLab AI 自动分析约 200 万条告警，筛选出 16 条需要实际调查的告警，并识别出 13 个真实威胁。发现这些威胁仅需 4 分 23 秒。用户点击 AI 识别出的威胁，即可查看详细分析和建议的响应措施，轻松采取行动。

AhnLab 记录并检查 AI 的判断、行动和数据访问过程，重要措施须经用户批准后执行。这体现了“受控自主”（Controlled Autonomy）原则：在赋予 AI 自主性的同时，通过权限、策略、审计和审批进行管理。该体系在提高安全运营速度和效率的同时，确保安全性与可靠性。

AhnLab 不仅提供智能体 AI 安全能力，还提供 AI 安全与治理能力。AhnLab 子公司 AhnLab Cloudmate 提供的 SecureBridge 可选择性地检测和阻止提示词中的机密及敏感信息，防止其用于 LLM 训练；安全的提示词则正常发送，让用户无需担忧数据泄露即可使用 AI。

AhnLab AI PLUS 带来的实际成效

用户可以切实感受到以下三方面的成效。

See What Matters	从海量告警中优先识别重要威胁和风险
Understand Faster	快速了解风险原因以及受影响的资产和用户
Act with Confidence	提供所需的调查和响应方法，迅速落实响应措施

AhnLab 利用 AhnLab AI PLUS，在事件处理、分析和安全监控等多个安全运营领域取得以下成效。

3x
年度事件处理量

60% ↓
平均分析时间

2.5x
安全监控效率

AhnLab AI PLUS 旨在缩短客户发现重要事项、理解其重要性并决定应采取何种措施所需的时间。借助智能 AI，客户可以实现安全运营创新。

FAQ

#1. 什么是 AhnLab AI PLUS?

它是 AhnLab 的 AI 安全平台及战略，以 AI 为中心连接 AhnLab 的技术、数据、产品和服务，实现智能化安全运营。

#2. 为什么选择 AhnLab AI PLUS?

应对 AI 时代的威胁，需要实现安全运营各环节的联动与协同。AhnLab AI PLUS 学习 AhnLab 积累 30 余年的安全数据和情报，提供差异化的 AI 安全能力。

#3. 用户将体验到哪些变化?

AI 快速理解并判断重要风险，进而开展调查与响应，使用户能够更快、更高效地进行安全运营。同时，用户可以最大限度减少手动操作中的人为错误，提高工作效率，专注于重要决策和工作。

#4. 如何应用于安全解决方案?

AhnLab 在 EDR、XDR 等解决方案中提供“AI Insight”等 AI 事件分析功能，并计划近期提供 [图 5] 所示的 AI 初步分析、调查和响应功能。该功能对每天产生的数十至数万条告警进行自动筛选，基于 AI 调查和风险评分自动识别真实威胁告警与误报告警，并优化处理优先级，让分析人员专注于真实威胁，从而大幅减少手动操作和工作疲劳。

结论

如果用一句话来界定智能体AI，那就是“能够自主规划、调用工具并执行行动以达成目标的‘自驱执行型AI’”。对安全负责人而言，它既是“强有力的自动化契机”，同时也带来了“全新的攻击面”。因此，熟练应用AI Agent固然重要，但如何对其进行合规管控、建立怎样的验证标准，同样是必须提上日程的关键课题。

智能体AI绝非遥不可及的未来科技，而是已经拉开落地序幕的“现在进行时”。读完本文后，您最需要立足自身组织思考的问题是：

“我们组织究竟需要什么样的AI Agent？又该如何去落地？此外，我们真的做好驾驭它的准备了吗？”

只有深刻理解这两个问题背后所代表的技术演进，才能在瞬息万变的AI时代中，破局成长为一名成功的安全负责人。

可在AhnLab官网查看有关**AhnLab AI PLUS**的详细信息。



AhnLab

北京市朝阳区望京阜通东大街1号望京SOHO塔2 B座 220502室 | 上海市闵行区新镇路1699弄E栋303室
https://www.ahnlab.com/cn | cn.sales@ahnlab.com
电话： +86 10 8260 0935（北京）| +86 21 6095 6780（上海）
© 2026 AhnLab, Inc. All rights reserved.