AI 安全系列报告 第一期

AI正在制造新的安全代差

——从“防不防得住”到“来不来得及”

发布机构：360 AI安全研究院

发布日期：2026年5月

保密级别：公开发布

目录

执行摘要

2026 年，人工智能正在改变网络安全的基本逻辑。

过去，企业面对漏洞，通常还有一个相对完整的响应窗口：等待披露、接收通报、排查资产、评估影响、安排修复。这个流程虽然缓慢，但在很多场景下仍能勉强运转。

但 AI 正在改变这个前提。

一方面，AI 编程工具快速进入软件开发流程，代码生产效率被显著放大。代码产出变快，并不必然意味着代码更安全。Veracode 2025 年测试显示，在其设定的编码任务中，AI 生成代码安全测试失败率为 45%，部分语言场景最高达到 72%。这提示一个重要风险：AI 并不是主动“制造漏洞”，但它正在规模化放大代码缺陷与漏洞风险，使漏洞从偶发缺陷逐渐变成规模化副产物。

另一方面，AI 也正在推动漏洞发现能力进入机器化阶段。以大型语言模型为核心的安全智能体体系，正在把过去依赖少数顶尖专家的漏洞发现能力，转化为可复制、可扩展、可持续运行的系统能力。以 360 漏洞挖掘智能体为例，其已累计挖掘近千个漏洞，覆盖操作系统、办公软件、AI 工具、物联网设备等九大核心领域，其中经 CNNVD、CNVD 及厂商确认的高危/严重漏洞超过 50 项。

这两条趋势交汇在一起，正在带来一个更深层的变化：网络安全正在从“能力竞争”转向“时间竞争”。过去的问题是“防不防得住”；现在的问题正在变成“来不来得及”。

当攻击者可以借助 AI 更快分析漏洞、复现漏洞、构建 PoC 甚至串联攻击链，而大量企业的防御流程仍停留在人工排查、逐级审批、排期修复的节奏中，组织之间的安全差距就不再只是技术能力强弱，而是“机器速度”与“人类流程”之间的代际差。

本报告将这一差距定义为：AI 安全时间差（ASTG, AI Security Time Gap）。

ASTG = 企业高危漏洞平均修复时间 - 高关注漏洞可用 PoC 出现时间。

例如，当一个企业高危漏洞平均修复周期为 21 天，而高关注漏洞 PoC 出现窗口已压缩至 3 天以内时，该企业面对的不是简单的 18 天时间差，而是一整代安全体系的差距。

为解释这一变化，本报告提出三项原创分析框架：一是漏洞武器化时间压缩三阶段模型，解释为什么漏洞利用窗口正在缩短；二是 AI 可发现性分类法（AID-T），解释哪些漏洞最适合由 AI 优先发现；三是安全时间预算模型，解释企业应如何围绕“发现—判断—修复”重构安全能力。

从行业实践看，这一变化已经在政企安全运营中产生直接影响。党政与大型机构面临终端、文档系统和 OA 入口风险；央国企与大型企业面临供应链系统和核心业务平台的漏洞暴露风险；制造与工业企业需要关注漏洞向生产连续性外溢；能源、交通、运营商等关键信息基础设施行业，则需要在高可用前提下建立更快的漏洞缓解和应急响应能力。

进一步看，随着智能体能力向工业控制、自动驾驶、机器人和具身智能系统延伸，安全风险还可能从网络空间外溢到现实世界。未来安全问题不再只影响数据和系统，也可能影响生产设备、交通系统和现实空间中的执行行为。

基于上述判断，本报告认为，政企组织需要尽快完成安全运营体系转型：建立多源漏洞时序监测能力，将 AI 辅助漏洞发现前移至研发与上线阶段，围绕核心资产构建“发现—判断—修复—复盘”的闭环体系，并逐步建立以安全智能体为核心的主动发现和智能响应能力。

本报告的核心结论是：

AI 正在制造新的安全代差。

这个代差不只是技术差，更是时间差。

未来最大的安全差距，不是有没有漏洞，而是谁能更早发现漏洞、更快判断风险、更准完成修复。

需要说明的是，本报告并不试图将个别案例直接外推为全行业结论。报告中的结论分为四类：公开数据复算的事实、360 内部实践观察、基于原创框架的趋势判断，以及仍需后续验证的研究假设。

一、漏洞利用时间序列研究：来自 CISA KEV 的真实数据

本章使用 CISA 官方 KEV（Known Exploited Vulnerabilities）数据库的真实数据进行分析。截至 2026 年 5 月 7 日，KEV 目录累计收录 1589 条已知被利用漏洞。以下所有数据均可通过公开接口独立复算。

1.1 KEV 年度趋势分析

以下图表展示了 CISA KEV 目录自 2021 年建立以来的年度新增趋势。数据来源为 CISA 官方 KEV Catalog。

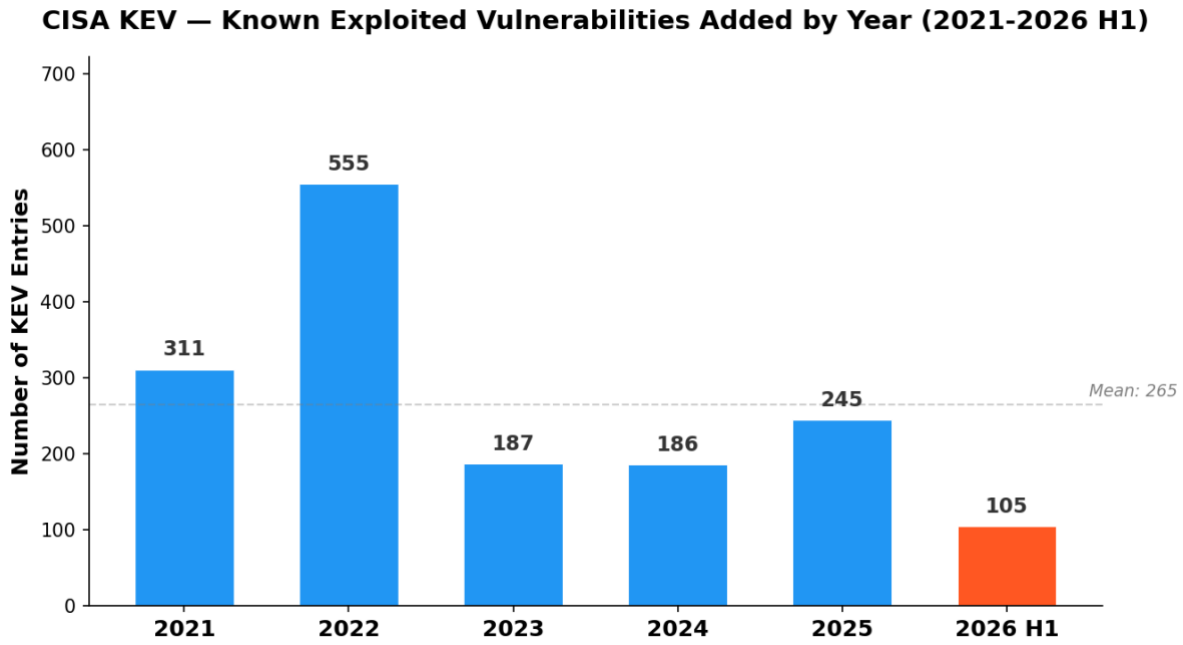

图 1：CISA KEV 已知被利用漏洞年度新增趋势（2021–2026 年 5 月）

数据来源：CISA 官方 KEV Catalog，截至 2026 年 5 月 7 日，共 1589 条

数据解读：

KEV 目录在 2022 年出现峰值（555 条），这与 CISA 在 BOD 22-01 指令生效后大幅扩充目录有关。2023-2024 年回落到年均约 186 条，2025 年回升至 245 条， 2026 年截至 5 月 7 日已达 105 条。如果保持当前节奏，2026 年全年可能超过 250 条。

需要注意的是，KEV 数量的增加不完全代表漏洞利用活动的增加——部分原因是 CISA 的收录策略在持续优化。但趋势本身值得关注：已知被利用漏洞的数量并未随时间推移而减少，反而维持在较高水平。

1.2 KEV 月度趋势

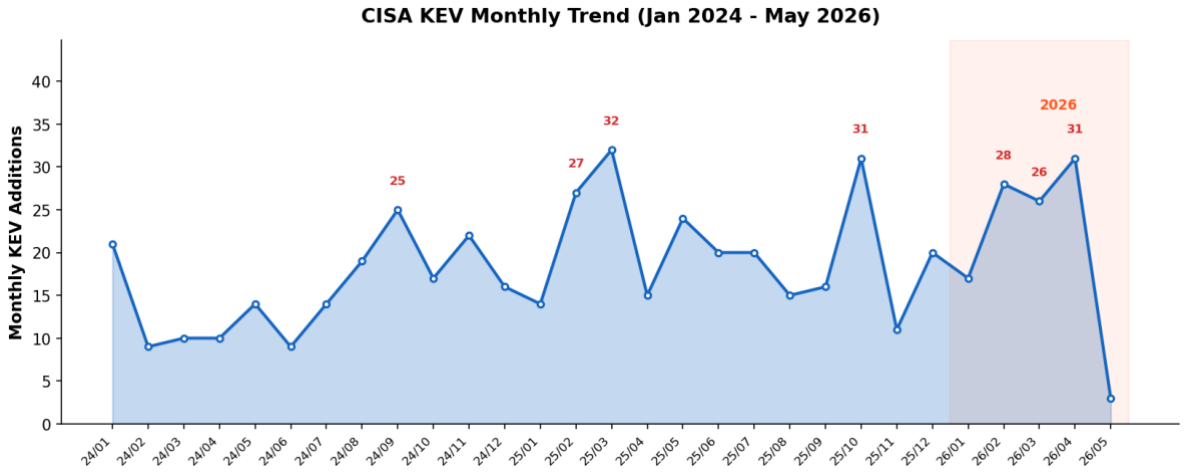

图 2：CISA KEV 月度新增趋势（2024 年 1 月–2026 年 5 月）

数据来源：CISA 官方 KEV Catalog

数据解读：

2026 年 1-4 月 KEV 新增量分别为 17、28、26、31 条，呈现稳定增长态势。2026 年 4 月达到 31 条，新增包括 Linux 内核 CopyFail（CVE-2026-31431）、Palo Alto PAN-OS 远程代码执行（CVE-2026-0300）等高影响漏洞。

1.3 漏洞生命周期：从 CVE 发布到被利用

以下对比展示了漏洞的 CVE 发布年份（漏洞被公开披露的时间）与 KEV 收录年份（被确认在野外利用的时间）的分布差异。

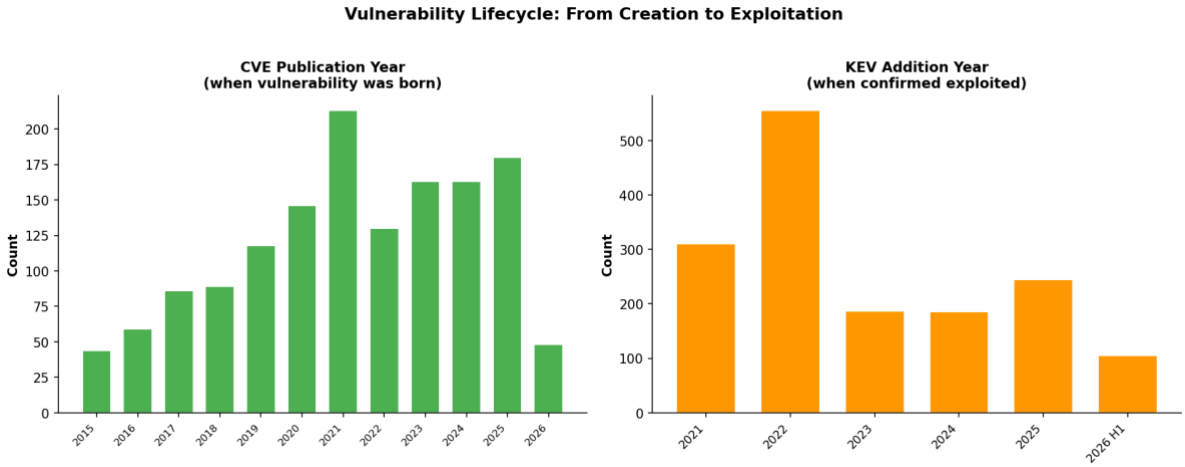

图 3：漏洞生命周期对比：CVE 发布年份 vs KEV 收录年份

左图：被利用漏洞的 CVE 发布年份 | 右图：CISA KEV 收录年份

数据来源：CISA KEV Catalog

CVE 发布年份的分布显示，当前被利用的漏洞中，相当一部分来自 2018 年之前——这意味着这些漏洞在被披露后多年仍然在野外被利用。2024 年和 2025 年的 CVE 各有 163 条和 180 条被收录入 KEV，表明新披露的漏洞从发现到被利用的"转化速度"在加快。

一个值得注意的信号：2026 年收录的 105 条 KEV 中，有约 30%的 CVE 来自 2020 年及更早。这说明漏洞的"半衰期"远超许多组织的预期——一个十年前披露的漏洞，在今天仍然可能构成实际威胁。

1.4 2026 年 KEV 收录漏洞的 CVE 年份分布

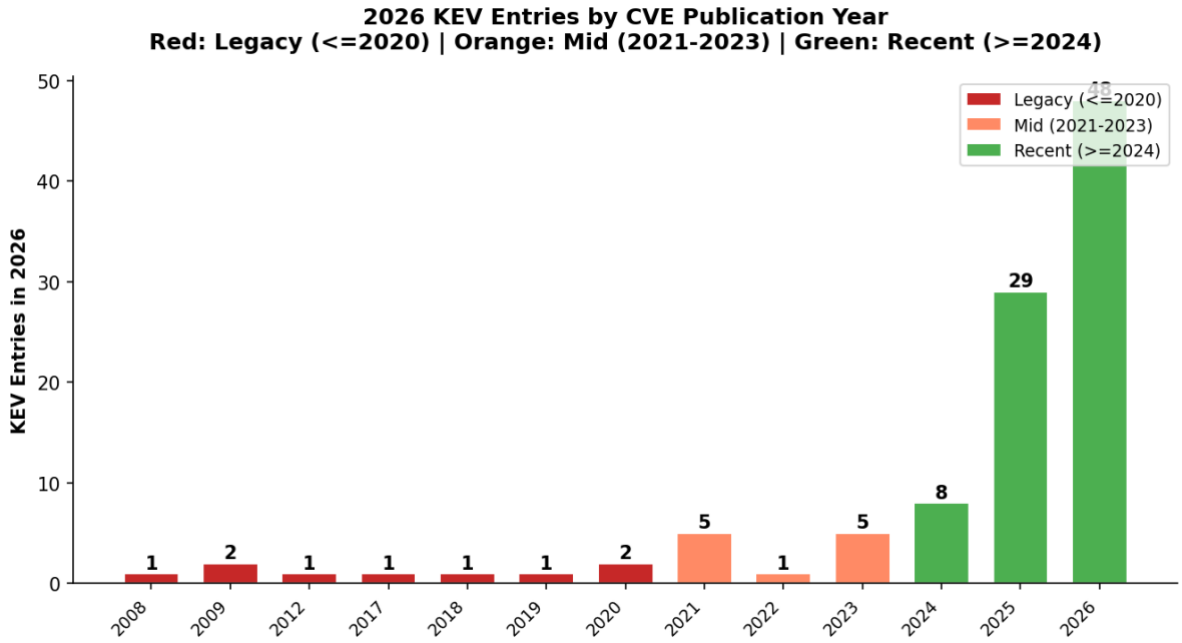

图 4：2026 年 KEV 收录漏洞的 CVE 年份分布

红色=陈旧漏洞(≤2020) | 橙色=中期(2021-2023) | 绿色=近期(≥2024)

数据来源：CISA KEV Catalog，2026 年 1-5 月

2026 年收录的 KEV 漏洞按 CVE 年份分布显示：约 30%来自 2020 年及更早的 CVE（红色部分），表明陈旧漏洞仍然是活跃威胁源。2021-2023 年的中期漏洞占比约 15%，而 2024 年及以后的近期漏洞占比约 55%。

这个分布说明了两件事：第一，组织对已知漏洞的修复存在系统性滞后——十年前的漏洞至今仍在被利用；第二，新披露的漏洞从发现到被利用的速度在加快。这正是本报告关注的核心问

题：当攻击者能够利用十年前的漏洞，同时也能够快速利用刚披露的漏洞时，防守方面临的是一个"两头受击"的态势。

二、AI 代码安全风险分析

2.1 Veracode 测试数据分析

以下数据来自 Veracode 2025 GenAI Code Security Report，该报告对 100+大语言模型进行了 80 种编码任务的安全测试。这是目前公开可获取的最系统的 AI 代码安全基准之一。

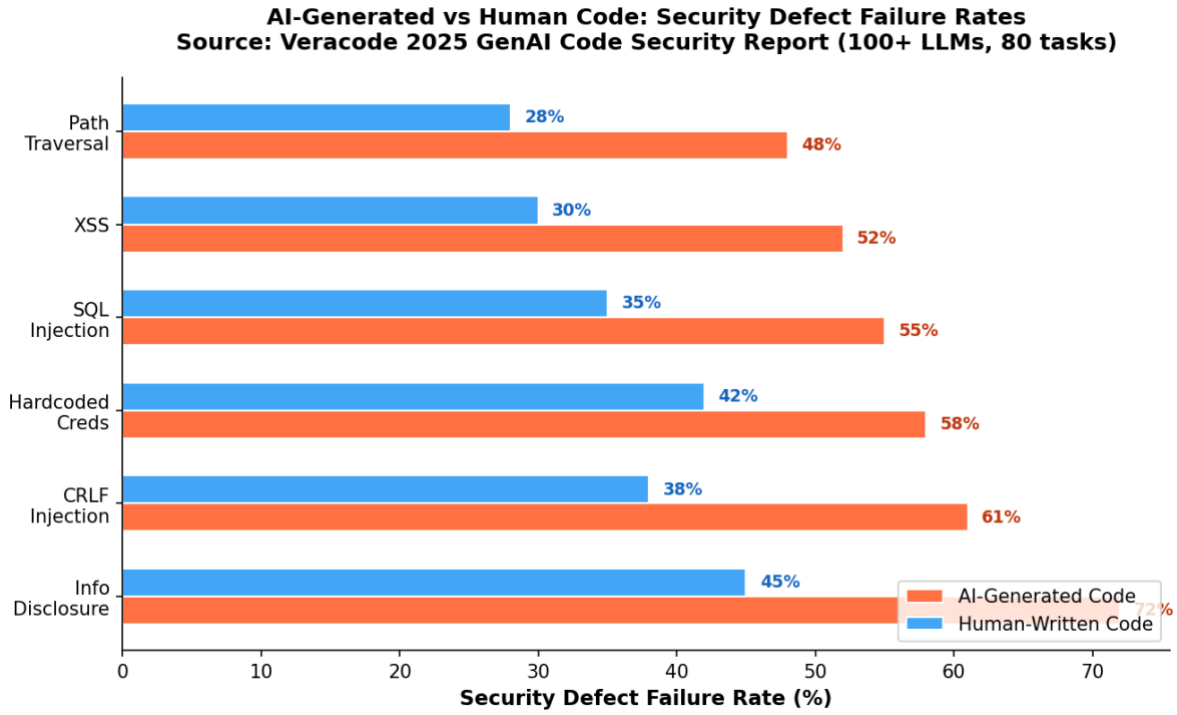

图 5：AI 生成代码 vs 人类代码：各类安全测试失败率对比

数据来源：Veracode 2025 GenAI Code Security Report（100+ LLM, 80 种编码任务）

证据边界说明：该测试衡量的是"在特定编码任务集上的安全测试失败率"，不是"AI 代码在生产环境中的实际安全表现"。测试任务的范围、难度、上下文复杂度都会影响结果。因此，本报告将此数据定位为"AI 代码安全存在显著风险"的信号，而非"AI 代码普遍不安全"的结论。

漏洞并不是 AI"主动制造"的结果，但在当前的技术路径下，AI 正在把漏洞从偶发缺陷变成规模化副产物。AI 代码工具的本质是"效率放大器"——它不会自动提升安全性，而是将开发者既有的能力与习惯等比例放大。如果安全意识不足，AI 只会以更快的速度生成更多潜在风险代码。

2.2 AI 安全成熟度模型（AI-SMM）

基于对公开信息的综合分析，本报告提出 AI 安全成熟度模型（AI-SMM），将组织的 AI 安全能力分为五个等级。

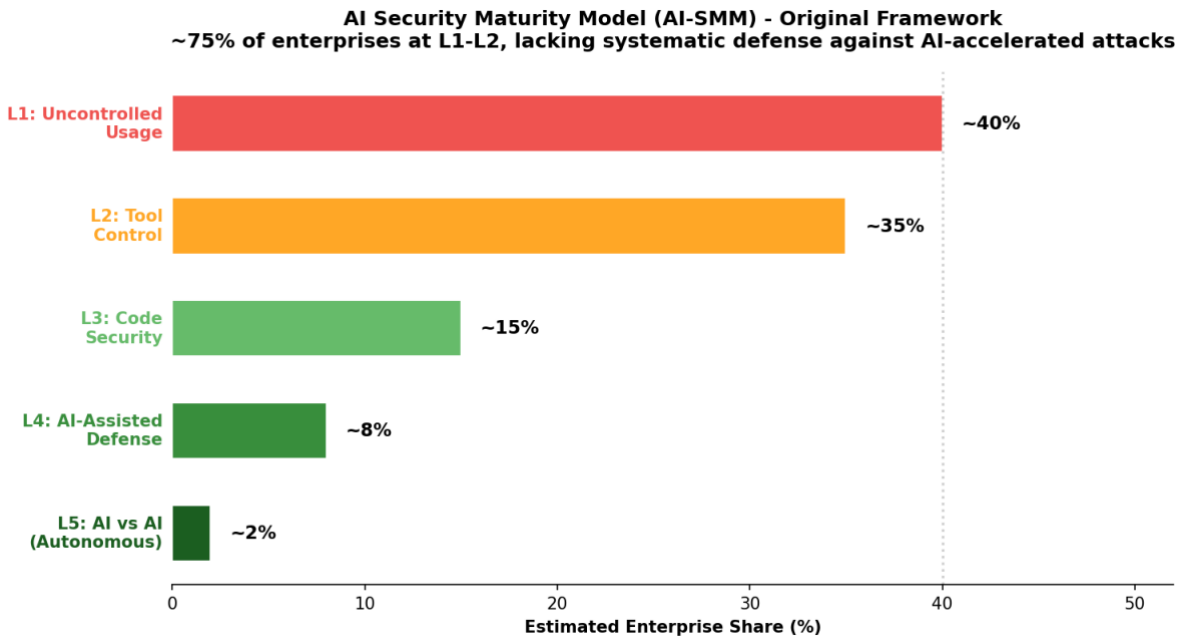

图 6：AI 安全成熟度阶梯（AI-SMM）— 原创框架

估算分布基于公开企业调研与行业报告综合判断，非精确测量

等级	名称	核心特征	关键指标	估算占比
L1	无感使用	员工自由使用 AI，无统一管控	AI 工具使用率>60%，安全管控=0	~40%
L2	工具管控	限制工具、账号、数据上传范围	有工具白名单，无代码审计	~35%
L3	代码安全	AI 生成代码纳入安全流程	AI 代码有专项 SAST/DAST	~15%
L4	智能防御	AI 辅助漏洞发现和响应	有 AI 安全工具，MTTR<7 天	~8%
L5	AI 对抗 AI	全自动攻防体系	防御 AI 响应≤攻击 AI	~2%

证据边界说明：各等级占比为基于公开企业调研与行业报告的综合估算，不是精确测量值。后续需要通过企业调研问卷进行实证验证。

2.3 AI 安全时间差（ASTG）指标定义

基于前述分析，本报告引入一项原创量化指标：AI 安全时间差（ASTG, AI Security Time Gap）。在本报告中，ASTG 默认以漏洞公开披露时间作为共同起点；若企业掌握更早的内部发现时间或厂商预警时间，应在组织内部模型中替换为实际起点。

ASTG = 企业高危漏洞平均修复时间 - 高关注漏洞可用 PoC 出现时间

该指标的含义是：当一个企业高危漏洞平均修复周期为 21 天，而高关注漏洞 PoC 出现窗口已压缩至 3 天以内时，该企业面临的是约 18 天的"AI 安全时间差"。这 18 天不是简单的数字差距，而是一整代安全体系的差距。

企业修复周期	PoC 出现窗口	ASTG 值	安全状态	建议行动
7 天	3 天以内	4 天+	红色警戒	立即启动应急修复通道
14 天	3 天以内	11 天+	高度危险	建立高危漏洞快速响应机制
21 天	3 天以内	18 天+	代差级别	全面重构漏洞管理体系
30 天+	3 天以内	27 天+	体系高风险	现有响应机制难以适应 AI 加速攻击节奏

三、【原创分析框架】本报告的三项底层模型

为解释 AI 时代安全时间差的形成机制，本报告提出三项原创分析框架，是基于公开重大漏洞案例、AI 辅助漏洞发现实践、企业漏洞响应流程观察，以及 360 在漏洞挖掘智能体方向的实战经验，提炼出的解释工具。

需要说明的是，以下框架属于"原创分析模型"，并不等同于已经完成的大规模统计结论。其价值在于为本报告后续章节提供统一的分析语言，并为未来量化研究提供可检验的假设基础。

这三项框架分别回答三个问题：

- 第一，为什么漏洞利用的时间窗口正在缩短？
- 第二，哪些类型的漏洞最适合由 AI 优先发现？
- 第三，企业应该把安全资源优先投入到哪个环节？

原创框架一：漏洞武器化时间压缩的三阶段模型

本报告认为，漏洞从"被发现"到"被利用"的时间压缩，并不是简单的线性变化，而是经历了三个主导模式的演进。其背后的核心机制，是漏洞发现、漏洞理解、利用代码编写和攻击链构建之间的技能壁垒被逐步拆除。

阶段	主导模式	核心特征	代表性变化
第一阶段（约 2015 年前后及以前）	手工业主导	漏洞发现高度依赖个人经验；利用代码编写是独立技术能力；复杂漏洞从发现到稳定利用往往需要较长周期	发现漏洞与编写利用代码之间存在明显技能门槛
第二阶段（约 2015—2023 年）	工具化主导	自动化扫描、Fuzzing、PoC 模板、漏洞复现工具逐渐成熟；漏洞披露后的复现速度明显提升	漏洞利用从专家手艺转向工具辅助
第三阶段（2024 年以来加速显性化）	机器速度主导	AI 参与代码理解、漏洞模式识别、利用链推理和攻击路径规划；在高价值场景中，分析与复现窗口进一步压缩	发现、分析、复现、武器化之间的流程边界被进一步打通

本报告将这一现象定义为：

漏洞武器化的技能门槛消解（Skill Barrier Dissolution）

过去，发现漏洞和编写利用代码往往是两类能力。前者依赖代码审计、系统理解和漏洞直觉，后者依赖利用技巧、环境调试和攻击链构建。两者之间存在明显的人才门槛和协作门槛。

AI 介入后，这两个环节开始被整合到同一个自动化分析流程中。攻击者不再必须拥有完整红队能力，只要具备基础操作能力并借助智能体，就可能完成过去需要多人协作的任务。

这才是"机器速度"对安全体系构成威胁的根本原因：不是单纯因为攻击更快，而是因为攻击能力的准入门槛被降低，漏洞武器化能力开始从少数专家手中扩散到更广泛的行动者手中。

换句话说，AI 带来的真正变化，不只是"快"，而是"会的人变多了，能做的人变多了，攻击链条被自动化压缩了"。

原创框架二：AI 可发现性分类法（AID-T）

AI 并不是对所有漏洞都同样有效。不同类型漏洞对上下文理解、语义推理、环境复现和触发条件搜索的要求不同，因此其"AI 可发现性"也存在明显差异。

本报告提出"AI 可发现性分类法"（AI Discoverability Taxonomy，简称 AID-T），将漏洞按照 AI 自动化发现的可接近程度分为四类。

层级	类型	漏洞特征	AI 作用方式	典型场景
L1	模式匹配型	漏洞表现为可识别代码模式、危险 API 调用或常见不安全写法；触发条件相对确定	AI 结合静态分析工具，显著提高覆盖率和解释效率	格式化字符串、缓冲区边界错误、硬编码密钥、不安全反序列化
L2	逻辑推理型	漏洞依赖跨函数、跨模块调用链，需要理解权限、状态和数据流变化	AI 可辅助路径理解、权限关系建模和异常逻辑发现	权限绕过、鉴权缺失、TOCTOU、业务逻辑漏洞
L3	语义理解型	漏洞隐藏在复杂格式、协议或规范实现差异中，传统工具难以理解语义约束	AI 可补足格式语义理解短板，但仍需要样本构造和验证工具	文档解析、协议处理、复杂文件格式解析
L4	时序耦合型	漏洞依赖多线程时序、硬件状态、竞争条件或极端边界条件	AI 可辅助生成假设、缩小搜索空间，但高度依赖符号执行、Fuzzing 和专家验证	内核竞争条件、驱动漏洞、硬件交互类漏洞

这一分类法的行业价值在于回答一个非常实际的问题：

企业应该优先让 AI 去扫描哪些代码和漏洞类型？

本报告的判断是：AI 安全能力不应被平均投放，而应按照漏洞类型和发现难度进行分层部署。

- 对于 L1 类漏洞，AI 的投入产出比最高，应尽可能实现自动化覆盖。
- 对于 L2 类漏洞，多智能体协同分析已经具备较高价值，适合重点部署。
- 对于 L3 类漏洞，AI 的语义理解能力可以弥补传统 Fuzzing 和规则扫描的短板，但仍需要结合样本构造、格式解析和人工验证。

- 对于 L4 类漏洞，不应期待 AI 单独完成发现，而应将其定位为假设生成器、路径分析助手和搜索空间压缩工具。

AI 漏洞挖掘的正确使用方式，不是"让 AI 替代所有安全专家"，而是让 AI 在最适合它的漏洞类型上形成规模化覆盖，同时把人类专家释放到更复杂、更高价值的判断环节。

原创框架三：安全时间预算模型

传统漏洞管理关注的是"修复速度"，但在 AI 加速攻防的背景下，仅关注修复已经不够。

本报告提出"安全时间预算模型"，用于刻画组织在漏洞攻防中的真实时间压力。

在防守侧，组织应对漏洞的总时间可表示为：

T_defense = T_detect + T_decide + T_repair

其中：T_detect（从漏洞出现、披露或进入攻击视野，到组织识别自身受影响的时间）；T_decide（从识别风险到确定处置方案、完成审批并启动修复流程的时间）；T_repair（从启动修复到补丁、缓解措施或配置调整完成部署的时间）。

在攻击侧，漏洞武器化时间可表示为：

T_attack = T_discover + T_weaponize + T_deploy

其中：T_discover（攻击方发现或获取漏洞信息的时间）；T_weaponize（攻击方完成漏洞复现、PoC 构建和利用链打磨的时间）；T_deploy（攻击方将利用能力部署到实际攻击中的时间）。

安全时间窗口成立的基本条件是：

T_defense < T_attack

当防守侧总时间大于攻击侧总时间时，组织就会进入"安全时间赤字"状态：

STD（Security Time Deficit）= T_defense - T_attack

如果 STD > 0，说明攻击方更快，组织已经失去时间优势；如果 STD < 0，说明防守方仍有机会在攻击完成前完成修复或缓解。

AI 带来的最大变化，是系统性压缩了攻击侧的 T_weaponize。过去，攻击方从漏洞信息到可用攻击链之间往往存在较高技术门槛；AI 辅助分析、代码生成和攻击路径规划正在压缩这一过程。

与此同时，许多企业的防守侧时间结构并没有同步变化：T_detect 仍依赖外部通报，T_decide 仍依赖层层审批，T_repair 虽然在技术上可以压缩，但往往被业务连续性、变更流程和组织协同拖慢。

AI 时代安全体系建设的首要任务，不只是"修得更快"，而是"发现得更早、决策得更快、修复得更准"。其中，主动发现能力是压缩安全时间预算的关键前置变量。

这也解释了漏洞挖掘智能体的核心价值：它并不仅仅是提高漏洞发现效率，而是把组织的安全行动前移到攻击者完成武器化之前，从而改变组织在攻防时间线中的位置。

换句话说，漏洞挖掘智能体的价值不是"让企业在被攻击之后更快补救"，而是"让企业在攻击者完成武器化之前就已经开始行动"。

三项框架的逻辑关系

框架	回答的问题	核心结论
漏洞武器化时间压缩三阶段模型	为什么漏洞利用窗口正在缩短？	发现与利用之间的技能壁垒正在被 AI 拆除
AI 可发现性分类法（AID-T）	哪些漏洞最适合 AI 优先发现？	AI 能力应按漏洞类型分层部署，而不是平均使用
安全时间预算模型	企业应该优先压缩哪个环节？	主动发现能力是压缩防守时间预算的关键前置变量

三者共同指向一个结论：未来安全竞争的核心，不再只是修复速度，而是围绕"发现—决策—修复"全链路形成时间优势。谁能更早发现、更快判断、更准修复，谁就能在机器速度攻防中掌握主动权。AI 时代的安全分水岭，不在于企业是否拥有更多安全工具，而在于企业能否把安全体系从人类响应节奏，切换到机器速度节奏。

四、360 漏洞挖掘智能体定性分析

本节对 360 漏洞挖掘智能体发现的漏洞样本进行定性分析。

数据说明：以下分析基于 360 已公开披露的案例和行业观察，仅做定性讨论，不提供具体数量或占比数字。精确的结构化分析需等待 360 内部数据字典和样本定义正式公开。

4.1 按资产类型的定性观察

根据 360 公开披露的案例和行业信息，其漏洞挖掘智能体覆盖了以下九大核心领域：

资产类型	公开案例/信息	定性观察
Windows 操作系统	CVE-2026-24293 内核提权漏洞（潜伏近 5 年）	高复杂触发路径，传统扫描和简单 Fuzzing 因依赖多线程时序未能发现，智能体通过知识推理结合符号执行实现突破
Office 办公软件	Office 远程代码执行漏洞（潜伏 8 年）	传统 Fuzzing 工具因缺乏对文档格式语义的深度理解而无法生成有效触发样本，智能体凭借完整语义理解实现发现
AI 智能体/工具	OpenClaw MEDIA 协议 Prompt 注入绕过（影响 17 万+实例）	新型攻击面，传统安全工具无对应检测能力，智能体在 AI 系统自身安全缺陷发现上具有天然优势
AI 编程工具	AI 代码助手安全漏洞	AI 工具自身的安全盲点，反映出 AI 工具链安全的重要性
国产操作系统	若干系统级漏洞	国产生态安全能力建设的重点领域
安卓系统	若干移动端漏洞	移动安全持续受到关注
邮件服务器	若干邮件系统漏洞	企业通信安全的关键入口
物联网设备	若干 IoT 设备漏洞	IoT 安全面持续扩大，设备碎片化增加发现难度
OA 系统	若干 OA 系统漏洞	企业办公入口安全

4.2 按漏洞类型的定性观察

从公开案例来看，360 漏洞挖掘智能体发现的漏洞涵盖多种类型：

- 内存破坏类漏洞：传统 Fuzzing 覆盖不足，智能体通过语义理解增强 Fuzzing，提高了对复杂格式漏洞的发现能力。
- 权限提升类漏洞：触发路径依赖复杂时序和权限状态，智能体通过知识推理结合符号执行，实现了对传统工具难以覆盖的漏洞的发现。
- 远程代码执行（RCE）漏洞：文档/协议格式复杂，传统工具无法理解语义，智能体可补足格式语义理解短板。
- Prompt 注入/智能体安全漏洞：新型攻击面，传统安全工具对此几乎没有检测能力。
- TOCTOU/竞争条件类漏洞：高度依赖时序，智能体可作为假设生成器和搜索空间压缩工具，辅助人工分析。

4.3 按严重程度的定性观察

根据公开信息，360 漏洞挖掘智能体发现的漏洞中，经 CNNVD、CNVD 及厂商确认的高危/严重漏洞超过 50 项，累计发现近千个漏洞（含中低危）。这表明智能体不仅在发现高价值漏洞方面有突破，在中低危漏洞的规模化覆盖上也展现出能力。

4.4 行业洞察

基于上述定性观察，本报告提出以下行业判断：

- AI 智能体在语义理解型漏洞（L3 类）上表现出独特优势。传统 Fuzzing 工具因缺乏对文档格式、协议规范的深度语义理解而无法生成有效触发样本，而 AI 可以补足这一短板。
- AI 智能体在时序耦合型漏洞（L4 类）上需要结合符号执行和专家验证。AI 可以作为假设生成器和搜索空间压缩工具，显著提升人工分析效率。
- AI 智能体在新型攻击面（Prompt 注入、智能体安全）上具有天然优势。这类漏洞不涉及传统意义上的代码缺陷，而是 AI 系统自身的设计缺陷。
- 九大核心领域的覆盖表明，AI 漏洞挖掘正在从"单点突破"走向"体系化发现"。

五、政企行业影响：不同行业面临不同的安全时间差

AI 正在压缩漏洞发现、复现和武器化的时间窗口，但这种压力并不会平均落在所有行业身上。不同政企组织的资产结构、业务连续性要求、漏洞修复流程和安全成熟度不同，因此面临的“安全时间差”也不同。

对于政企客户而言，AI 时代的漏洞风险不再只是“有没有高危漏洞”，而是高危漏洞出现后，组织是否能在攻击者完成武器化之前完成发现、判断和处置。

5.1 党政与大型机构

党政与大型机构的典型高风险资产包括办公终端、文档系统、OA 系统、门户网站、邮件系统和内部协同平台。

这类组织面临的主要风险，是办公软件漏洞、文档解析漏洞、终端提权漏洞和 OA 入口漏洞被快速利用。一旦攻击者通过钓鱼文档、恶意附件或 OA 漏洞进入内部系统，后续可能造成数据泄露、横向移动和政务系统中断。

对这类机构而言，最大问题往往不是缺少安全设备，而是资产复杂、终端数量庞大、补丁流程谨慎，导致 T_detect 和 T_decide 过长。

因此，党政与大型机构需要重点建设三类能力：

1. 终端与文档系统的主动漏洞检测；
2. OA、邮件、门户等高暴露入口的持续监测；
3. 高危漏洞的快速评估和应急修复通道。

5.2 央国企与大型企业

央国企与大型企业的典型高风险资产包括 ERP、供应链系统、业务中台、内部 OA、VPN、云平台、数据库和核心业务系统。

这类组织的漏洞风险通常具有链式放大效应。一个供应链系统漏洞，可能影响多个业务部门；一个共享组件漏洞，可能同时影响多个下游应用；一个边界设备漏洞，可能成为攻击者进入内网的关键入口。

对央国企而言，安全时间差主要体现在两个方面：

第一，系统复杂，资产关联难。企业可能知道某个漏洞很危险，但无法快速判断哪些系统受影响。
第二，修复协调慢。核心业务系统涉及多个部门、供应商和运维窗口，导致 T_decide 和 T_repair 被显著拉长。

因此，央国企应重点建设：

1. 软件供应链审计能力；
2. 漏洞情报与资产清单自动关联能力；
3. 面向核心业务系统的漏洞优先级排序机制；
4. 对高危漏洞的跨部门快速决策机制。

5.3 制造与工业企业

制造企业的典型高风险资产包括工控系统、生产控制网络、MES 系统、工业网关、PLC、边缘设备和供应链管理平台。

与传统 IT 系统不同，工业企业面对的不是简单的数据泄露风险，而是生产连续性风险。一旦漏洞被利用，后果可能从系统异常扩展为产线停摆、设备异常、生产中断甚至安全事故。

制造企业的安全时间差更加特殊：很多工业系统不能频繁打补丁，不能随意重启，也不能简单套用互联网企业的快速迭代模式。这导致 T_repair 天然偏长。

因此，制造企业的关键不是简单追求“最快修复”，而是建立分层缓解能力：

1. IT/OT 一体化资产识别；
2. 工控系统漏洞主动发现；
3. 虚拟补丁和边界隔离；
4. 生产不中断前提下的分阶段修复方案；
5. 面向关键产线的应急回滚机制。

5.4 关键信息基础设施行业

能源、交通、运营商、金融等关键信息基础设施行业，对系统可用性和连续性要求极高。一旦高危漏洞被快速利用，可能影响大规模服务可用性，甚至产生社会影响。

这类行业的安全时间差具有更强的公共属性。对普通企业而言，漏洞利用可能造成业务损失；对关基行业而言，漏洞利用可能影响公共服务、通信网络、能源调度和社会运行稳定。

因此，关基行业需要把 ASTG 纳入安全运营指标体系，重点关注：

1. 高危漏洞从通报到确认影响范围的时间；
2. 从确认影响到启动应急机制的时间；
3. 从启动应急到完成缓解措施的时间；
4. 对无法立即修复系统的临时防护能力。

关基行业的目标，不是简单做到“有漏洞就修”，而是建立高可用前提下的快速缓解和持续防御能力。

5.5 行业影响对比表

行业	高风险资产	主要时间差风险	应对重点
党政与大型机构	终端、文档系统、OA、邮件	通报晚、排查慢、横向移动快	文档安全、终端检测、OA 入口防护、主动漏洞发现
央国企与大型企业	ERP、供应链系统、业务平台、VPN	资产关联慢、跨部门修复协调慢	软件供应链审计、资产自动关联、优先级响应
制造与工业企业	工控系统、生产网络、工业网关	系统不可随意停机，T_repair 天然偏长	IT/OT 一体化安全、虚拟补丁、分阶段修复
关键信息基础设施	调度系统、通信网络、核心业务平台	高危漏洞影响社会运行，容错空间小	高可用安全体系、快速缓解、应急通道

总体来看，AI 时代的政企安全挑战，本质上不是某一种漏洞类型的挑战，而是不同组织能否根据自身资产结构，建立适配自身业务的安全时间预算管理能力。

六、企业应对路径与行动建议

基于前述研究框架和实证分析，本报告认为，企业应对 AI 时代安全风险，不能只是在原有漏洞管理流程上“加快一点”，而是要围绕“发现—决策—修复”全链路重构安全时间预算。核心目标是：缩短 T_detect，压缩 T_decide，优化 T_repair，最终降低企业自身的 AI 安全时间差（ASTG）。

6.1 立即行动：建立多源漏洞时序监测能力

第一，接入多源漏洞情报，建立自动化监控机制。

企业可优先接入 CISA KEV 官方 CSV/JSON 数据源，将其作为“全球已知被利用漏洞”的重要输入。CISA KEV 的价值在于，它聚焦已经被确认存在利用活动的漏洞，适合作为高优先级漏洞治理参考。

但对于中国企业而言，CISA KEV 不应作为唯一依据。企业还应同步接入或持续跟踪 CNVD、CNNVD、国家级漏洞通报、行业监管通报、厂商安全公告、开源组件漏洞库，以及 360 自有漏洞情报和威胁情报，形成“全球已知利用漏洞 + 本土漏洞通报 + 企业自身资产暴露面”的多源监测体系。

第二，将漏洞情报与自身资产清单自动关联。

漏洞管理的关键，不是知道世界上出现了多少漏洞，而是知道这些漏洞是否影响自己。企业应将新增漏洞条目与自身资产清单进行自动匹配，覆盖操作系统、数据库、中间件、网络设备、边界设备、VPN、OA、邮件服务器、Web 应用、开源组件、云服务、工业控制系统以及 AI 工具链。

对于“已被利用 + 有公开 PoC + 影响自身暴露资产”的漏洞，应直接进入最高优先级处置队列。

第三，建立 ASTG 内部基线。

企业应统计自身从漏洞公开披露、厂商公告、CNVD/CNNVD 通报、监管通报或内部首次发现，到完成修复或缓解的平均时间，计算自身的 AI 安全时间差。

ASTG = 企业高危漏洞平均修复时间 - 高关注漏洞可用 PoC 出现时间

在实际落地中，ASTG 的起点可以根据企业情报能力灵活设置：可以采用 CVE 披露时间，也可以采用 CNVD/CNNVD 发布时间、厂商公告时间、CISA KEV 入表时间、内部首次发现时间或威胁情报首次命中时间。对成熟度较高的企业，更建议使用“内部首次感知时间”作为起点，以衡量真实响应能力。

第四，将 ASTG 与行业基准和内部历史基线对比。

企业应按漏洞类型、资产类型、业务系统和暴露程度分别计算 ASTG，识别时间差最大的风险区域。对中国企业而言，应重点关注 OA 系统、VPN、邮件服务器、边界设备、Web 中间件、国产操作系统、工业控制系统、开源组件和 AI 工具链等高暴露资产。

6.2 中期建设：将 AI 代码安全嵌入研发流程

第一，在 CI/CD 流程中嵌入 AI 代码专项扫描。

企业应将 AI 生成代码纳入统一安全检测流程，在代码提交、合并、构建和上线环节引入 CodeQL、Semgrep、SCA、Secrets 扫描、依赖漏洞检测等工具，建立 AI 代码安全基线。

需要强调的是，AI 生成代码不应被默认为“更规范”或“更安全”。AI 代码工具本质上是效率放大器，如果安全规范不足，它会以更快速度放大已有风险。

第二，建立 AI 生成代码使用规范。

企业应明确哪些场景允许直接使用 AI 生成代码，哪些场景必须经过人工审计，哪些核心模块禁止直接引入 AI 生成代码。尤其是认证鉴权、支付交易、数据加密、权限控制、日志审计、模型调用、Agent 执行链路等高风险模块，应建立更严格的审查要求。

第三，建立 AI 代码安全责任矩阵。

AI 生成代码不能成为责任真空地带。企业需要明确开发者、安全团队、架构负责人、业务负责人和管理层在 AI 代码使用中的责任边界。原则上，AI 可以生成代码，但最终责任仍应归属于使用者和审批流程，而不能推给模型本身。

第四，将 AI 相关漏洞类型纳入下一次攻防演练的重点测试范围。

企业应在下一次攻防演练中，专门加入 AI 生成代码缺陷、Prompt 注入、Agent 越权执行、工具调用滥用、模型输出污染、AI 插件供应链风险等攻击场景，验证现有检测与响应体系是否有效。

6.3 企业自检表：快速评估 AI 安全成熟度

企业可对照以下自检表，快速判断自身 AI 安全成熟度，并明确下一步建设方向。

企业问题	自检指标	对应成熟度等级
AI 代码是否可见	是否知道 AI 生成代码占比	L1→L2
AI 工具是否可管	是否建立 AI 工具白名单、账号权限和数据上传边界	L1→L2
AI 代码是否可审	是否进入 SAST/DAST/SCA/人工审计流程	L2→L3
高危漏洞是否可抢修	是否有 72 小时应急修复机制	L3→L4
漏洞是否可提前发现	是否使用 AI 辅助漏洞挖掘或主动检测能力	L4→L5
响应是否机器化	是否具备自动分诊、自动验证、自动处置能力	L5

这张表的意义，不是给企业贴标签，而是帮助企业判断：自身安全体系是否还停留在“人工响应节奏”，还是已经开始向“机器速度节奏”切换。

6.4 长期目标：向 AI-SMM 高阶能力演进

企业应定期使用 AI-SMM 模型进行自测，明确自身所处等级，并制定向下一等级演进的具体路径。

L1→L2：完成 AI 工具盘点，建立基础管控能力。

建议周期：0—3 个月。

重点任务包括：盘点企业内部正在使用的 AI 工具，建立 AI 工具白名单，明确账号权限、数据上传边界和敏感信息使用规则。

L2→L3：将 AI 生成代码纳入安全扫描流程。

建议周期：3—6 个月。

重点任务包括：将 AI 生成代码纳入 SAST、DAST、SCA、Secrets 扫描和人工 Review 流程，建立 AI 代码安全基线，并开始统计 AI 生成代码占比和安全缺陷密度。

L3→L4：部署 AI 辅助漏洞发现和响应工具，建立 ASTG 监测体系。

建议周期：6—12 个月。

重点任务包括：接入多源漏洞情报，建立资产自动关联机制，计算企业 ASTG 基线，部署 AI 辅助分诊、漏洞验证和补丁优先级排序能力。

L4→L5：构建 AI 对抗 AI 的自动化攻防体系。

建议周期：12—24 个月。

重点任务包括：建立安全智能体体系，实现漏洞发现、影响评估、攻击面分析、修复建议、响应编排的半自动化或自动化闭环。高危操作仍需保留人工审批，但常规分诊、验证和处置应逐步进入机器速度。

最终，企业安全体系的目标不是简单增加工具，而是完成一次节奏切换：从“发现后响应”转向“提前发现”，从“人工排队处理”转向“机器辅助决策”，从“被动修复漏洞”转向“主动压缩安全时间差”。

七、前瞻判断：三个可验证趋势

基于前述原创分析框架和实证数据，本报告提出以下三个可在未来 12-24 个月内验证的前瞻判断。

趋势一：企业漏洞管理 SLA 将被迫从"月级"压缩到"天级"

CISA KEV 数据显示，已知被利用漏洞的数量维持在高位，且新披露漏洞进入 KEV 的速度在加快。与此同时，AI 正在压缩攻击侧的 T_weaponize（漏洞武器化时间）。

当高关注漏洞的 PoC 出现窗口从"数周"压缩到"24-72 小时"（部分案例），而企业平均修复周期仍以"周"甚至"月"计算时，防守方的 ASTG（安全时间差）将持续扩大。

可验证指标：到 2027 年中，观察头部企业是否将高危漏洞修复 SLA 从"周级"正式调整为"天级"（如 72 小时、48 小时）。如果这一趋势出现，将验证本报告的前瞻判断。

趋势二：AI 生成代码安全检测将从 SAST 附属能力变成独立赛道

Veracode 测试数据显示，在其设定的编码任务中，AI 生成代码未通过安全测试的比例为 45%，部分语言场景最高达到 72%；相比之下，人类代码相关测试结果约为 28%—45%。随着 AI 生成代码在企业新代码中的占比持续提升，传统 SAST 工具对 AI 代码特有缺陷的检测覆盖率将越来越不足。

可验证指标：到 2027 年中，观察是否出现专门针对 AI 生成代码安全检测的独立产品或独立安全赛道。如果主流安全厂商开始推出"AI 代码安全专项检测"产品线，将验证本报告判断。

趋势三：安全厂商竞争将从"漏洞响应能力"转向"漏洞提前发现能力"

安全时间预算模型显示，T_detect（发现时间）是压缩防守总时间的关键前置变量。当攻击侧的 T_weaponize 被 AI 压缩后，防守方唯一的破局点就是将安全行动前移到攻击者完成武器化之前。

可验证指标：到 2027 年中，观察安全厂商的市场定位是否从"快速响应"转向"提前发现"。如果主要安全厂商开始在产品路线图中强调"主动漏洞发现"而非"快速响应修复"，将验证本报告判断。

八、局限性与研究方法说明

8.1 本报告的局限

- CISA KEV 数据反映的是"已知被利用"的漏洞，不代表真实的最早利用时间。灰色市场和私有利用链会让 Δt_poc 被系统性高估。
- Veracode 测试数据限于特定任务集，不能直接代表 AI 代码在生产环境中的整体安全表现。
- AI-SMM 成熟度模型的等级占比为估算值，需通过企业调研进行实证验证。
- 360 漏洞挖掘智能体相关内容来自内部实践观察，但完整数据字典、时间窗口、去重规则和逐项明细尚未公开，因此本报告将其作为定性观察，而非可复算统计结论。
- 报告中的三项原创分析框架属于解释工具，尚待大规模实证检验。

8.2 研究方法与结论分类

本报告将主要结论分为四类，并在关键章节中明确其证据边界。

结论类型	说明	本报告中的例子
公开数据复算	读者可通过公开数据源独立验证	第一章 CISA KEV 数据分析
内部案例观察	来自 360 内部实践，可引用但方法待公开	第四章 360 漏洞挖掘智能体定性分析
趋势判断	基于框架和分析的行业预判	第七章三个可验证趋势
待验证假设	提出研究方向，尚未完成大规模实证检验	AI 安全时间差（ASTG）及相关后续指标体系

本报告的定位：AI 安全系列报告第一期：趋势判断与研究框架报告。本报告的核心价值不在于给出终极答案，而在于提出了一套可验证、可复算、可检验的研究框架，使后续报告能够从"二手观点汇编"转向"原创证据驱动"。

附录：数据与方法说明

A. 数据来源

- CISA KEV Catalog: https://www.cisa.gov/known-exploited-vulnerabilities-catalog（CSV/JSON 公开接口，截至 2026 年 5 月 7 日，共 1,589 条）
- Veracode 2025 GenAI Code Security Report
- CopyFail 漏洞: https://copy.fail/

B. 证据等级说明

A 档：报告自身可直接复算（有数据、有方法、有附录）

B 档：可定位到公开一手来源（官方数据库、官方报告原文）

C 档：机构自报或商业摘要，方法不透明

D 档：缺乏可追溯原始证据或定义不清

本报告中，第一章（CISA KEV 数据分析）的所有图表和统计数字均达到 B 档标准。第二章的 Veracode 数据为 B 档，AI-SMM 成熟度模型和 360 智能体相关内容为 C 档（需后续方法公开验证）。原创分析框架为方法论贡献。

C. 可复现性附录清单

- 数据：CVE 清单、披露日期、KEV 入表日期、修复日期、任务分层标签
- 代码：数据抓取脚本、清洗脚本、统计脚本、作图脚本
- 环境：操作系统、CPU/GPU、模型版本、分析器版本
- 命令：全部可直接运行的 CLI 命令，含输入输出路径
- 误差报告：置信区间、bootstrap 区间、稳健性分析、失败案例列表

注：本报告部分内容由 360 AI 安全研究智能体辅助完成，并经 360 AI 安全研究院研究团队审核确认。

— 报告完 —