R1kkoSec

AI 驱动的网络安全实训与自动化研判平台

技术白皮书 v1.0

2026 上海职业院校学生技能大赛 · 人工智能赛道
文档版本: 2026-04-05 | 内部技术文档

目录

1. 项目概述

R1kkoSec 是一个 AI 驱动的网络安全实训与自动化研判平台，旨在解决网络安全教育中的三大核心痛点：师资不足、学生依赖 WriteUp、专业工具门槛过高。平台通过将大语言模型的推理能力与 Kali Linux 终端沙箱深度结合，实现了从自然语言输入到真实命令执行、从证据采集到攻击链还原的完整自动化闭环。

项目由 Web 平台（R1kkoSec）和全平台移动端（R1kkoSecGo）两部分组成，总计超过 50,000 行生产代码，268 个自动化测试全部通过，覆盖 6 个终端平台。

指标	数值	说明
总代码量	50,000+	Web 20,078 + Go 29,767
自动化测试	268 passed	66 测试文件，9,176 行测试代码
知识库规模	120,000+	题库 3,200 + Writeup 8,500 + CVE 15,000
单次检索延迟	<85ms	缓存命中 40ms，冷查询 <200ms
分析时间	42s	从样本输入到完整报告输出
最终置信度	94%	四阶段推理收敛结果
LLM Provider	8 种	OpenAI/Claude/Gemini/DeepSeek/Azure 等
目标平台	6 端	Android/iOS/macOS/Linux/Windows/Web
CVE 发现	8 个	覆盖 6 个开源项目，最高 CVSS 8.2

2. 系统架构

R1kkoSec 采用五层分层架构，通过依赖注入容器（40+ 注册组件）实现模块间的松耦合和可热替换。

层级	组件	职责
前端层	Web / TUI / CLI	三入口统一，SSE 实时推送，Live2D 角色集成
后端层	Flask + DI 容器	40+ 依赖注入组件，26 个功能模块，RESTful API
记忆层	Hybrid 检索引擎	Keyword + Vector + Graph 三路融合检索
智能层	Agent + 8 Provider	多模型统一路由，多轮工具调用闭环
执行层	Kali 隔离沙箱	真实命令执行，命令证据可回放，安全隔离

2.1 依赖注入与模块化

系统通过 createRuntime() 总入口初始化，container.register() 注册 40+ 核心依赖。构建分两层：registerCoreDependencies（memory、retrieval、store、embedder）和 registerRuntimeDependencies（tool executor、agent、proactive timer）。模块替换成本极低，符合开闭原则。

2.2 持久化策略

系统使用 SQLite 作为主持久化层，包含四个独立数据库：app_state（全局状态）、auth_sessions（认证会话）、sessions（聊天历史）、teaching（教学系统 6 张表）。采用 WAL 日志模式和 busy_timeout 保证并发安全，同时保留 JSON 文件备份兼容。

2.3 多模型路由

Agent 层通过 Provider Factory 统一接入 8 种 LLM：OpenAI、Claude (Anthropic)、Gemini (Google)、DeepSeek Reasoner、Azure OpenAI、OpenRouter 以及 OpenAI-Compatible 接口。工厂在配置不足时直接抛出可读错误，避免静默降级。

3. 核心引擎：四阶段递进推理

R1kkoSec
的推理引擎采用四阶段递进架构，每一步结论都建立在前一步的可验证证据之上，从根本上杜绝 AI 幻觉。

阶段	名称	核心操作	输出
Stage 1	Evidence	文件元数据、字符串提取、导入表分析、哈希计算	结构化证据集
Stage 2	Hypothesis	从 12 万条知识库三路融合检索，自动生成并排序假设	假设列表 + 置信度
Stage 3	Verification	逐条假设执行验证命令，MITRE ATT&CK; 映射	验证结果 + CVSS
Stage 4	Conclusion	攻击链完整还原，结果持久化存档	研判报告 + 处置建议

3.1 主动唤醒机制

系统内置三类主动触发器，让 AI 从被动应答升级为主动追问：Relation Trigger（关联知识主动注入）、Topic-Shift Trigger（话题偏移追问）、Low-Entropy Trigger（信息不足预警）。当检索结果显示上下文不充分时，AI 会主动向用户提问，而不是生成低质量的猜测性回答。

4. Hybrid 三阶段检索管线

检索管线是推理引擎的核心驱动力，负责在 85 毫秒内从 12 万条安全知识中召回最相关的证据。

4.1 检索流程

阶段	操作	性能参数
A: Hash 粗筛	全量 hashVecs 计算 hash 相似度	prescreenRatio=0.05, min=20, max=100
B: Local Rerank	ONNX LocalEmbedder 批量重排	rerankMultiplier=3, hardCap=16, timeout=350ms
C: TopK 输出	local cosine 优先，不可用时回退 hash	localCacheMaxEntries=2000, TTL=300s

4.2 写入策略

写入分三档：普通 block 走 hash-only（保速度），important/conflict 标签走 hybrid（保精度），查询向量强制 hybrid（保 query/block 子空间兼容）。

4.3 鲁棒性保障

查询向量维度异常记录 trace 不静默吞掉。local 维度不匹配回退 hash。local batch 失败或超时全链路回退 hash。空文本候选跳过 local 批处理。block 更新/删除失效 localCache。LocalEmbedder
具备队列背压能力：queueMaxPending=1024，auto-batch，overflow 抛错。

5. 安全知识底座

R1kkoSec 与通用 AI 的本质区别在于：分析开始之前，12 万条专业安全知识已经装进了检索引擎。AI 的每一个推理步骤都有据可查。

数据类型	数量	覆盖范围	特点
CTF 题库	3,200+	Crypto/Pwn/Web/Reversing/Forensics/Stego/Misc	难度标定 + 解题路径标注
Writeup 解析	8,500+	多路径解法 + 工具链记录	关键命令逐行注释，按相似度可召回
CVE 漏洞	15,000+	NVD 同步，含 CVSS + MITRE 映射	PoC 代码片段 + 影响版本 + 修复方案

6. R1kkoSecGo 全平台生态

R1kkoSecGo 是主平台的全平台生态连接层。29,767 行 Dart 代码，一套代码覆盖 Android、iOS、macOS、Linux、Windows、Web 六个端。

模块	功能	技术实现
首页	项目动态、安全文章、论坛、全局搜索	StatefulWidget + BootstrapPreloadStore 预热
题库	筛选、进度、资源联动	QuestionBankController (ChangeNotifier)
解题工作区	PTY 终端 + AI Agent 协作 + 命令审批	xterm + WebSocket + SSE + ExecGuard
AI Chat	流式对话 + 多模型 + 会话管理	ChatHttpService SSE POST + CancelToken
个人中心	认证、签到、收藏、训练历史	AuthStore (flutter_secure_storage)

定位关系：主平台做深度能力中枢，R1kkoSecGo 做全平台生态连接层。两者共享同一后端 API 和认证体系，层级互补而非竞争。

7. 工程化实践

R1kkoSec 不是原型演示，而是通过完整质量门禁的可持续迭代产品。

维度	现状	证据
代码规模	50,000+ 行	Web 129 源文件 + Go 38 源文件
测试覆盖	268 Tests Pass	66 测试文件，含 Hybrid 向量存储、LocalEmbedder 批处理等关键边界测试
构建质量	Typecheck + Build 100%	三重门禁：类型检查、测试、构建全部通过
配置管理	99 参数全贯通	ENV -> TOML -> CLI -> Runtime Override 四级合并
存储后端	4 种可切换	Memory / SQLite / Lance / Chroma
可观测性	debugTraceRecorder	内存记录 + API 拉取/清空 + Debug Panel 可视化
持久化	SQLite WAL	auth_sessions + chat_sessions + app_state + teaching 四库独立
CI/CD	Gitea + 自动部署	push 到 master 自动 rsync + pip install + systemctl restart

8. 安全研究成果

在研发过程中，团队将 R1kkoSec 的自动化安全分析能力应用于真实开源项目审计。R1kkoSec 的 Hybrid 检索引擎自动匹配已知漏洞模式和代码特征，在审计过程中触发高置信度告警，经验证后按负责任披露流程提交。

#	项目	Stars	漏洞	CWE	CVSS	状态
1	Gradio	35K	CORS 全放行（非 localhost）	CWE-942	8.1 HIGH	GHSA 分诊中
2	Streamlit	35K	trustedUserHeaders 认证绕过	CWE-290	8.2 HIGH	已报告
3	Gradio	35K	Audio 组件 SSRF	CWE-918	7.5 HIGH	GHSA 分诊中
4	sqladmin	2.7K	全局 CSRF	CWE-352	7.1 HIGH	已报告
5	sqladmin	2.7K	ajax_lookup 权限绕过	CWE-862	4.3 MED	已报告
6	OpenClaw	热门	CDP 端口暴露 0.0.0.0	CWE-284	8.1 HIGH	已报告
7	FastAPI	72K	Swagger UI XSS	CWE-79	7.1 HIGH	待提交

#	项目	Stars	漏洞	CWE	CVSS	状态
8	Sanic	18K	响应头 CRLF 注入	CWE-113	7.5 HIGH	待提交

影响范围：覆盖 6 个主流开源项目，总计超过 170,000 GitHub star。最高 CVSS 8.2 (HIGH)。所有漏洞均按 90 天负责任披露时间线处理。

9. 应用场景与市场价值

场景	目标用户	核心价值	市场规模
高校安全实训	3,000+ 网安专业院校	替代传统人工教学，实训效率提升 10 倍	50 亿+
企业安全培训	企业安全团队	新人上手周期从 3 个月缩短到 2 周，成本降低 80%	大型
CTF 竞赛训练	CTF 战队和选手	教方法论而非答案，选手培养周期缩短 5 倍	增长中

核心竞争优势：全自动端到端解题（传统 30 分钟 -> R1kkoSec 42 秒）；Codex 单引擎中继（减少故障面）；RAG 知识库 + Kali 工具链（100+ 安全工具全栈集成）；6 端全平台覆盖。

10. 技术路线图

阶段	时间	目标
当前	2026 Q2	完善教学系统 6 张表，Mission Control 实时协同，CVE 持续发现
近期	2026 Q3	PartitionMemoryManager 拆分，Bench 预构建提速，Hybrid Trace 指标化
中期	2026 Q4	Go App 社交模块完善，企业 SaaS 版本发布，渗透测试自动化流水线
远期	2027	私有化部署方案成熟，行业认证（等保/ISO27001），海外市场拓展

R1kkoSec — 真始于执行。