

“中国电子杯”第三届高校 ICT 产教融合创新大赛

赛题九

一、赛题名称

智能体驱动的 AI 安全知识情报系统设计与实现

二、命题单位

奇安信科技集团股份有限公司

三、所需知识背景

目标专业：网络空间安全、人工智能、计算机

能力层级：本科生、研究生、博士生

基础知识：网络空间安全、人工智能、软件工程、数据分析等

四、赛题背景

随着大语言模型、智能体为代表的人工智能技术迅猛发展，AI 应用呈爆发式增长，AI 安全漏洞日益成为产业界和学术界关注的焦点。当前，AI 安全漏洞信息分散在 CVE/NVD、安全社区、厂商公告、学术论文、技术博客等多个渠道，安全从业者面临严重的信息过载问题，难以快速获取、消化和关联关键情报。现有安全情报系统多依赖人工采集和规则匹配，缺乏对 AI 安全这一新兴领域的深度覆盖，更缺少基于自然语言的智能问答能力，无法有效支撑安全应急响应的决策需求。

五、赛题描述

本赛题为半开放性赛题，考察学生对真实安全问题的理解能力与自主创新设计能力，要求参赛团队基于智能体技术，设计并实现一套自动化的 AI 安全知识情报系统。系统应具备以下三大核心能力：

1. 情报监测模块

对 CVE/NVD、安全社区、厂商公告、安全博客等主流渠道的最新 AI 安全漏洞信息进行自动化监测与采集，实现多源数据的持续发现与汇聚。

2. 情报富化模块

对采集到的漏洞信息进行多维扩展与深度加工，包括但不限于：关联技术论文与分析文章、定位受影响互联网资产、搜集可用的 POC 验证代码、补充 CVSS 评分及影响评估等。

3. 情报问答模块

基于采集与富化后构建的知识库，提供自然语言交互式深度问答服务。系统应能理解复杂的安全查询意图，支持跨文档的语义检索与关联推理，精准回答如漏洞影响范围、修复建议、相关攻击链等问题。

此外，系统应同步采集最新的 AI 安全学术论文、技术标准与政策法规等信息，构建完整的 AI 安全知识底座。选手可根据自身理解，在上述模块基础上增加其他有价值的功能。智能体的角色与架构不做强制限制，鼓励选手充分发挥智能体技术在任务编排、自主决策、工具调用等方面的优势。

附：

AI 安全风险分类参考：① AI 基础设施/推理框架风险（如 Ollama、vLLM、Triton 等框架的漏洞）；② 模型与数据层风险（如提示词注入、系统提示词泄漏、数据投毒等）；③ AI 应用与智能体风险；④ 供应链风险等。上述分类仅供参考，不作为评分依据，参赛团队可根据实际情况自行分类，言之成理即可。

六、评价指标

本赛题评价指标围绕三大功能模块设置，兼顾功能完整性、性能表现与工程能力。具体如下：

1. 功能指标

- （1）情报监测功能完整性：数据源覆盖广度（CVE/NVD、安全社区、厂商公告、博客等）、采集自动化程度
- （2）情报富化功能完整性：富化维度的丰富度（论文关联、受影响资产、POC、CVSS 评分等）
- （3）情报问答功能完整性：问答覆盖的知识范围、对复杂安全查询的理解能力

2. 性能指标

- （1）监测时效性：漏洞信息从发布到系统采集的时间延迟
- （2）富化准确率：富化信息与原始漏洞的相关性与准确性
- （3）问答深度与准确率：回答的准确性与跨文档关联推理能力

3. 优化指标

- （1）智能体架构设计：Agent 的自主决策、任务编排与工具调用能力
- （2）系统自动化程度：从采集到问答的全流程自动化水平

4. 文档质量：技术方案的完整性、文档结构的规范性、表述清晰度

5. 演示效果：系统功能的现场呈现效果，含方案阐述与系统实操

比赛分为校内初赛和决赛，校内初赛由各校自行组织推荐，决赛由组委会统一线下组织。

七、阶段作品要求及评分标准

1. 校内初赛

1.1. 校内初赛作品提交要求

（1）技术报告：提交 Word 版本及 PDF 版本各 1 份，建议封面或目录后的第 1-2 页为重要内容预览页，包含作品重要核心亮点。内容需涵盖：系统总体架构设计、智能体方案设计、情报监测/富化/问答各模块详细设计、关键技术选型说明、系统工作流程图、系统实现细节、核心代码说明、各功能模块实现效果展示、性能测试结果与分析、创新点总结。

（2）操作演示视频。

（3）技术讲解 PPT：以 PDF 格式提交。

注：源代码由企业在参赛平台上审阅，不单独提交，避免泄露

1.2. 校内初赛评分参考标准（总分 100 分）

大项	内容	分值	评分要求
功能指标	1. 情报监测功能完整性 多源数据采集覆盖度（涵盖 CVE/NVD、安全社区、厂商公告、博客等）、采集自动化程度	15	至少支持 3 类以上数据源为合格（9 分）； 支持 5 类以上为良好（12 分）； 支持 7 类以上且含实时监测为优秀（15 分）
	2. 情报富化功能完整性 富化维度丰富度（论文关联、受影响资产、POC、CVSS 评分等）	15	至少支持 3 种富化维度为合格（9 分）； 支持 5 种以上为良好（12 分）； 富化信息准确且含自动关联推理为优秀（15 分）

			分)
	3. 情报问答功能完整性 自然语言问答的知识覆盖范围、 对复杂安全查询的理解能力	15	支持基础关键词检索问答为合格（9分）； 支持语义理解与多轮对话为良好（12分）； 支持跨文档推理与深度分析为优秀（15分）
性能指标	1. 监测时效性 漏洞信息从发布到系统采集的 时间延迟	5	延迟≤24 小时为合格（3分）； 延迟≤12 小时为良好（4分）； 延迟≤6 小时为优秀（5分）
	2. 富化准确率 富化信息与原始漏洞的相关性 与准确性	5	准确率≥70%为合格（3分）； ≥85%为良好（4分）； ≥95%为优秀（5分）
	3. 问答深度与准确率 回答的准确性及跨文档关联推 理能力	5	基础问答准确率≥70%为合格（3分）； 深度问答含推理链为良好（4分）； 多跳推理准确率≥85%为优秀（5分）
优化指标	1. 智能体架构设计 Agent 自主决策、任务编排、工 具调用的合理性	15	基本使用 Agent 框架为合格（9分）； Agent 多角色协作与自主编排为良好（12分）； 架构设计精良且具可扩展性为优秀（15分）
	2. 系统自动化程度 从采集、富化到问答的全流程自 动化水平	10	核心流程实现自动化为合格（6分）； 全流程自动化且含异常处理为良好（8分）； 全流程自动化+自愈能力为优秀（10分）
文档质量	技术方案完整性、文档结构规范 性、表述清晰度	15	文档结构完整、表述清晰为合格（9-12分）； 含详细设计图与流程图且逻辑严谨为优秀 （12-15分）

2. 线下决赛

2.1 决赛作品提交要求

（1）答辩 PPT：凝练作品内容、方案、亮点、创新点，系统阐述相较于现有方案的优势。注意：为确保比赛公平公正，PPT 中不能出现队伍单位、人名等识别信息。答辩 PPT 模板由主办单位提供。

（2）技术报告：提交 Word 版本及 PDF 版本各 1 份，要求同初赛，完整介绍最终技术方案。

（3）操作演示视频。

2.2 决赛阶段评分参考标准（总分 100 分）

大项	内容	分值	评分要求
功能指标	1. 情报监测功能完整性 多源数据采集覆盖度与自动化 程度	10	实现多源自动采集为合格（6分）； 覆盖广泛且含增量监测为良好（8分）； 全自动+实时+去重为优秀（10分）
	2. 情报富化功能完整性 富化维度丰富度与准确性	10	基本富化维度实现为合格（6分）； 多维度自动富化为良好（8分）； 精准富化+智能关联为优秀（10分）
	3. 情报问答功能完整性 深度语义理解与交互问答能力	10	语义问答实现为合格（6分）； 多轮对话+推理为良好（8分）； 深度推理+个性化推荐为优秀（10分）
性能指标	1. 监测时效性与稳定性	5	延迟≤24 小时为合格（3分）；

	漏洞发现到采集的延迟及系统运行稳定性		延迟≤12 小时为良好（4 分）； 延迟≤6 小时为优秀（5 分）
	2. 富化准确率与召回率 富化结果的准确率与漏洞覆盖率	5	准确率≥80%且召回率≥70%为合格（3 分）； 均≥90%为良好（4 分）； 均≥95%为优秀（5 分）
	3. 问答质量 回答准确率、响应速度、推理深度	10	准确率≥75%且响应≤5s 为合格（6 分）； 准确率≥90%且支持多跳推理为良好（8 分）； 准确率≥95%且推理链可解释为优秀（10 分）
优化指标	1. 智能体运用水平 Agent 自主决策、多智能体协作、工具调用能力	15	有效运用 Agent 技术为合格（9 分）； 多 Agent 协作编排为良好（12 分）； Agent 架构创新且效果显著为优秀（15 分）
	2. 系统自动化与工程化 CI/CD、容器化、监控告警等工程实践	10	基本可部署运行为合格（6 分）； 容器化部署+日志监控为良好（8 分）； 完整 DevOps+自动运维为优秀（10 分）
文档质量	技术文档完整性与规范性	10	文档完整规范为合格（6-8 分）； 文档精良且可作为技术参考为优秀（8-10 分）
演示效果	答辩表现 陈述逻辑、问题回答质量、现场应变	10	陈述清晰、能回答基础问题为合格（6-8 分）； 逻辑严密、技术理解深入、回答精准为优秀（8-10 分）
其他	创新性、系统性、实用性、团队协作	5	综合考察作品创新程度、系统完整度、实际应用价值及团队分工协作情况。由评委综合主观评分

八、参赛平台

主办方不提供统一参赛平台，学生自行部署运行。

九、参考资料

无