VITU ACADEMIC
PUBLISHING PTE.LTD

面向网络空间安全的开源大模型RAG框架构建与应用研究

郑明才　吴瑜儒　罗　磊
江西软件职业技术大学　江西南昌　330041

摘　要：生成式人工智能技术的快速发展，让大语言模型成为网络空间安全领域的重要技术支撑，检索增强生成（RAG）技术通过融合外部知识库与模型推理能力，为打造领域专属智能系统提供了可行路径。开源生态的成熟降低了技术落地门槛，安全社区得以基于开放架构自主搭建适配实际业务需求的RAG框架。当前该领域亟待解决的问题，在于如何高效融合多源异构安全数据，优化检索与生成机制的协同性，同时保障知识更新的实时性和推理过程的可追溯性。本文结合网络空间安全场景的实际需求，设计并构建开源大语言模型RAG框架，从核心构成、场景适配、知识库搭建、关键技术实现等方面开展研究，通过实验验证了框架在检索性能与推理效果上的优势，为网络安全领域智能系统建设提供了技术参考与实践方案。

关键词：网络空间安全；开源大语言模型；检索增强生成；知识图谱；向量数据库

引言

当前网络安全形势日趋复杂，病毒攻击手段层出不穷，对知识储备和快速反应能力提出更高的要求。传统的基于规则或特征库的安全工具存在知识迭代速度慢、泛化能力差等瓶颈，难以有效应对新型网络攻击。封闭的商用智能平台面临难以定制化、安全性风险高的问题，无法满足多样化的安全社区需求。

检索增强生成模式以外部权威知识库约束模型输出，可以在保留大数据模型一般性推理能力的前提下，提升其领域专业性与输出可信性，是解决大数据环境下大模型错觉问题、适应垂直领域应用的核心技术途径。随着Llama2、通义千问等开源模型的日趋成熟，以及向量数据库、知识图谱等基本构件的不断完善，本项目的研究成果将为构建透明、可扩展的RAG系统打下坚实基础。

针对上述问题，本文立足于网络空间安全领域的现实需求，基于开源开放源语言模型，构建面向开放源语言的开放源语言模型，围绕多源异构安全数据融合、检索与生成机制协同优化、全维度安全防护体系构建三个关键问题展开研究，实现网络安全知识的准确获取与专业推理，提升威胁检测的准确性、响应的时效性以及知识的智能化管理水平，为构建人机协同的智能安全防御系统提供技术支持。

资助项目：2025年江西省教育厅科学技术研究项目：面向网络空间安全的开源大模型增强检索生成（RAG）框架应用研究与实践（编号：GJJ2504605）。

一、开源大语言模型RAG框架核心构成与适配逻辑

此次研究以知识精确赋能、检索—推理协同为核心设计思想，结合网络安全领域的知识特征，进行定制化设计。

（一）构架核心构件的设计

该框架构建了知识存储层、检索引擎层、大模型推理层、知识更新层、交互适配器等五个核心模块，每个模块分工协作，形成一条从知识存储到人机交互的完整技术链条。知识存储层采用混合存储结构，选择Milvus、Chroma等向量数据库实现毫秒级语义匹配；基于PostgreSQL对结构化数据（如漏洞编号、CVSS分值等）进行关联存储，支持多维度联合检索。

在检索引擎层，融合稀疏稠密检索技术，基于BM25稀疏检索，实现核心关键字的准确匹配；基于Sentence-BERT的稠密检索，实现语义层次的深度关联检索；同时，针对查询类型，设计动态加权融合策略，自动调整检索权重，提高检索精度和召回率。在大模型推理层，选取Llama2、通义千问等开源模型，通过对网络安全场景的特定提示词的工程化优化，设计规范化的Prompt模板，包含检索结果摘要、场景指导和格式约束，引导模型自动生成结构化的、专业化的响应，减少无依据的生成。

在知识更新层面，构建“增量爬取—自动化处理—人工审核—合规入库”的循环迭代机制，利用网络爬虫对核心数据源进行定期的抓取，并将其自动处理后交由专家评审，以保证知识库的时效性和准确性。交互适配层提供了标准化的API接口和可视化的操作接口，该接口可以与已有的安全态势感知平台、IDS、防火墙等设备

进行无缝对接。

（二）适应网络安全情景

针对网络安全领域知识专业、结构多样、更新迅速等特点，从知识表示、检索策略和推理逻辑三个方面进行情景自适应优化研究。在知识表达层面，研究基于领域数据优化的词向量模型，从漏洞类型、攻击模式和防御策略三个方面提升专业词汇语义关联度，优化向量表达效果，提高检索精度。

检索策略层支持关键词+属性组合的检索方式，用户可以通过CVSS值、发表时间、受影响设备等指标，快速筛选出风险较大的知识；然后，设计多级递进的检索机制，根据第一轮的检索结果产生中间检索指令，层层递进，最终形成一个完备的知识库。在推演适配器层次，引入领域规则约束机制，构建专业规则库，保证模型输出满足行业规范要求，并对核心场景参数进行调整，使检索召回率大于85%，满足实际业务需求。

二、开源大语言模型RAG框架安全知识库构建

本项目拟从数据源选择、结构化处理和存储优化三个层面，构建高质量的网络安全知识库。

（一）知识库中的数据源和筛选准则

拟采用权威、多元和可溯源的方法，对已有的官方漏洞库（CVE）、CNNVD（CNVD）、奇安信（CNAS）、安恒信息（CNVD）等安全公司的威胁智能平台（如Metasploit、Nmap），开源工具（如Metasploit、Nmap）等，以及网络安全国家标准、国家应急响应案例库和核心安全技术论坛等。

数据筛选严格遵循“专业”“及时”“准确”和“实用”四个方面的要求：“专业”指的是由权威机构或资深专家发布的，其中包括了具体的技术细节；时效性：对动态数据如漏洞、威胁智能等设定不超过6个月的更新周期，逾期的将作为历史参考；真实性方面，通过多源信息的交叉验证，剔除虚假和无效的内容；以实用性为导向，筛选出可直接应用于实践的知识，以提高知识库的实用性。

（二）知识的结构化加工和存储优化

采用“文本块—实体提取—关系构建—向量转换”的标准化处理流程：在单一知识主题分解文本基础上，采用BERT模型抽取核心实体编号和CVSS分值，利用规则引擎+机器学习方法建立多维度实体关联，并采用领域微调词向量模型向量存储到矢量数据库中。

此次研究针对现有漏洞修复数据格式不统一的问题，研究标准化转化规范，实现不同版本CVSS得分、攻击类型MITREATT&CK框架的统一，并对漏洞修复状态进行规范化标注。其中，矢量数据库采用乘积量化+矢量压缩的方法减少了存储开销，而关系型数据库则为核心属性建立了专门的索引，提高了检索的效率。为保证数据的安全性和抗灾性，采用异地多副本备份+细粒度访问权限控制策略。

表1　知识库核心数据类型与存储配置

数据类型	存储方式	更新周期	核心字段
漏洞描述	向量数据库+关系型数据库	≤6个月	漏洞编号、漏洞原理、利用条件、修复方案、CVSS3.1评分、影响设备
威胁情报	混合存储（向量+关系）	实时增量	攻击类型、攻击特征、传播路径、防御策略、威胁等级、攻击组织
安全规范	关系型数据库	年度更新	标准编号、合规要求、适用场景、落地细则、更新时间
应急案例	向量数据库	季度更新	事件类型、攻击路径、处置流程、责任分工、恢复措施、案例总结

三、开源大语言模型RAG框架关键技术实现

在增强检索、安全自适应推理、安全性保护等方面取得突破性进展，为框架的性能和运行安全性提供保障。

（一）优化检索技术

针对关键字精确查询与语义模糊查询两类典型需求，设计差异化权重策略，设计差异化加权策略，实现精确查询条件下的稀疏检索权重提高到60% ~ 70%。该算法以语义模糊查询为基础，使稠密检索的权重提高到60% ~ 70%，同时兼顾准确性和全面性。在此基础上，提出了一种基于梯度提升的排序方法，将知识优先级和查询相关性相结合，对检索结果进行二次排序，减少冗余信息对检索结果的影响。设计层次化的粗检索与细检索机制，对大类进行准确定位，并对子主题进行细化，降低检索复杂性。本项目拟将检索式缓存技术引入到高频查询场景中，将失效时间控制在30分钟以内，以减少系统的响应时间，提高系统的运行效率。

（二）大模型安全适配与推理优化

以高质量网络安全语料为基础，采用领域增量调优+场景优化相结合的方法，对模型进行增量式预训练，优化模型对专业术语和技术逻辑的理解，实现小样本调优，提高模型输出精度。

制定一套标准化程序模板，包括知识局限、形式要求、专家指导等；明晰模型根据检索结果产生响应，并根据不同场景设置固定的输出结构。在此基础上，引入

安全规则校验机制，构建包含漏洞修复有效性和防御策略合规的规则库，形成“生成—测试—优化”的闭环推理过程，保证结果的专业性和可靠性。

（三）架构安全保护机制

针对数据安全问题，采用基于角色的权限控制机制，对不同角色进行不同的权限分配，对敏感知识进行限制，采用TLS1.3对数据进行加密，并对加密后的日志进行至少6个月的保存，支持安全审计。

模型安全性层次构建恶意查询特征库，对恶意查询请求进行拦截，建立敏感信息识别模型，实现漏洞脚本和敏感IP的脱敏；部署模型沙盒将大型模型与业务核心网隔离开来，限制了执行权限。在访问安全方面，我们使用了多因素认证机制来实现认证。设置访问频率阈值，防止恶意获取和拒绝服务攻击；每月进行渗透测试和漏洞扫描，确保体系结构在可控状态下运行。

四、实验验证与结果分析

为验证框架性能优势，搭建专用实验环境，设计对比实验与消融实验，从检索层与推理层多维度开展量化验证。

（一）实验环境

实验采用云服务器集群搭建环境，硬件方面主服务器为1台IntelXeonGold6330，搭配128GB内存、NVIDIAA100显卡，从服务器2台IntelXeonE5-2680v4，64GB内存；软件方面采用Ubuntu20.04LTS系统，Milvus2.3.0向量数据库、PostgreSQL15.4关系型数据库，Llama2-7B-Chat与Qwen-7B-Chat模型，基于Elasticsearch8.8.0搭建检索引擎。

（二）实验数据集

构建网络安全领域多源异构融合数据集，总数据量12.6万条，按8：2划分为训练集与测试集，涵盖漏洞描述4.2万条、威胁情报5.8万条、安全规范0.6万条、应急案例2.0万条，同时选取MSMARCO与CSAWCTF作为对比数据集，验证框架领域适配性。

（三）实验指标与方法

检索层选取召回率、精确率、F1值、平均响应时间为指标，推理层选取推理准确率、结构化程度、专业合规性为指标。采用对比实验将本文框架与通用RAG框架、传统规则检索系统对比，消融实验移除动态加权检索、场景Prompt模板、安全规则校验三大核心模块，验证各模块有效性，所有测试重复5次取平均值，消除随机误差。

（四）实验结果与分析

对比实验结果显示，本文框架检索层召回率92.3%、精确率88.7%、F1值90.5%、平均响应时间86ms，相比通用RAG框架各项指标均显著提升，相比传统规则系统F1值提升28.7%；推理层准确率87.6%、结构化程度0.89、专业合规性0.91，大幅优于通用RAG框架。

消减实验显示，去掉动态加权搜索策略后，搜索F1下降8.2%，删除场景模板后，推理精度下降7.5%，结构化程度下降0.23，删除安全规则检查后，专业遵从度下降0.18，三大核心模块协同提升框架性能，验证了技术设计的有效性。

结束语

面向网络空间安全领域重大需求和难点问题，以5个核心组件协同设计和场景适配为切入点，以开源大语言模型为基础，开展多源异构安全数据高效融合与精确检索研究。为此，从增强检索能力、大模型自适应能力和推理优化三方面入手，提高框架的检索性能和推理性能。然后，从数据、模型和访问三个层面，建立一套完整的安全保障系统。通过研究，可有效提高智能安防系统中威胁检测、事件响应和知识管理等方面的智能化水平，为实现人-机协同智能安防系统提供核心技术支持。

当前框架仍存在优化空间，未来将重点解决三大问题：一是实现攻击流量、恶意代码样本等多模态安全数据的统一表示与高效索引，丰富知识库维度；二是推动检索结果与生成内容的可信对齐，构建知识溯源机制，提升输出可解释性；三是开展框架轻型化部署研究，通过模型量化、剪枝等技术，实现边缘端、嵌入式设备的快速落地。未来将持续优化技术与场景适配，推动RAG技术从安全分析辅助工具向自主决策智能体演化，为网络空间安全防御范式智能化转型提供更强支撑。

参考文献

[1]万敏.面向广播电视行业的生成式人工智能大模型数据安全治理框架构建研究[J].广播与电视技术，2025，52（10）：16-19.

[2]刘怡宁.面向监管合规的金融大模型可解释性框架构建——基于多模态融合与因果推理的路径研究[J].上海商业，2025（08）：113-115.

[3]罗妍，刘宇炀，李晓瑛，等.面向医学大模型的体系化人工智能框架构建与应用[J].北京邮电大学学报，2024，47（04）：98-104.

[4]钱峰，成建国，夏润亮，等.水利大模型的建设思路、构建框架与应用场景初探[J].中国水利，2024（09）：9-19.

[5]王明程，王高开，李勇男.基于大模型智能体的安全风险态势感知框架构建[J].情报理论与实践，2024，47（07）：190-198.