生成式 AI 技术在网络安全防护中的应用

蒋帆
94750部队，福建 龙岩 366200
DOI: 10.61369/TACS.2026050019

摘　　要： 随着新型信息技术的发展与革新，网络攻击手段也在不断升级，并呈现出智能化、复杂化的发展趋势，对传统网络安全防御体系带来严峻挑战。生成式人工智能（Generative AI）依靠自身语义理解、内容生成与模式识别等功能优势，已经成为重塑网络安全防护的重要载体。本文通过分析生成式 AI 在网络安全防护中的核心优势，进而围绕威胁检测与狩猎、AI 钓鱼邮件防御、智能攻防演练与漏洞挖掘、代码安全与自动化修复等维度，提出基于生成式 AI 技术的新型网络安全防护范式构建策略，以此推动网络安全行业向智能理解、自主决策的新阶段发展。

关 键 词： 生成式 AI；网络安全；大语言模型；威胁检测；钓鱼防御；智能攻防

Application of Generative AI Technology in Network Security Protection

Jiang Fan
94750 Troops, Longyan, Fujian 366200

Abstract: With the development and innovation of new information technologies, cyber attack methods are also constantly upgrading, showing an intelligent and complicated development trend, which poses severe challenges to the traditional network security defense system. Relying on its advantages in semantic understanding, content generation, pattern recognition and other functions, Generative AI has become an important carrier for reshaping network security protection. By analyzing the core advantages of generative AI in network security protection, this paper proposes strategies for constructing a new network security protection paradigm based on generative AI technology around threat detection and hunting, AI phishing email defense, intelligent attack-defense exercises and vulnerability mining, code security and automatic repair, so as to promote the network security industry to a new stage of intelligent understanding and autonomous decision-making.

Keywords: generative AI; network security; large language model; threat detection; phishing defense; intelligent attack and defense

引言

在数字经济时代，网络安全已经成为国家战略与社会治理层面的重要议题。在人工智能技术辅助下，网络攻击从传统模式向智能漏洞挖掘、自动化侦察、恶意代码生成等形态转变升级，使得其攻击效率与隐蔽性显著提高。面对网络安全的威胁态势，生成式 AI 技术带来新的防护可能，其可以基于数据学习、大语言模型、扩散模型、生成对抗网络等不同技术路线，识别非结构化的安全文本，自动生成安全防护分析报告，模拟攻击对抗场景，甚至还可以自动编写安全补丁。现阶段，生成式 AI 技术在网络安全防护中的应用正在向规模化落地发展，尤其在威胁因子检测、网络安全运营、代码、身份与数据安全、合规审计等场景有着重要应用，进而成为网络安全防御侧的智能组件。

一、生成式 AI 技术在网络安全防护中的应用优势

（一）从被动防御到主动预测：防护范式的根本转变

在网络安全传统防御体系下，其防护功能采取“检测—响应”的基本逻辑，即只能在攻击出现后才能通过日志分析、警告信息、溯源研究等方式完成响应。该防御体系无法有效解决 APT 攻击、零日漏洞、多阶段复合攻击等威胁。生成式 AI 的应用重在推动防御体系从被动防御向主动预测升级，具体可以从三个层面分析其优势。第一，可以自动生成威胁样本，从而模拟不同类型的攻击行为，通过训练与测试提升防御系统的对抗水平。第二，可以展现出强大的推理能力，可以根据历史数据推算出潜在的攻击路径，进而挖掘出漏洞利用链[1]。第三，具备可持续性的安全评估与动态风险追踪能力，并生成针对性的防护建议。

（二）从规则匹配到语义理解：认知能力的质的飞跃

传统的安全检测与防护工具需要依赖提前设定的规则，并通过特征签名赋予权限，从而使得其防护功能具有较大局限，尤其

无法应对通过变种设计的攻击，或者全新类型的威胁。同时，在日常运行维护中还需要人力支持，需要花费较大的成本支出。生成式 AI 技术基于自然语言模型与语义理解能力，进而展现出多个维度的功能优势。其一可以智能解读安全告警，并对其关键信息进行关联分析，从而自动生成安全事件报告。其二可以自动挖掘与整合网络威胁情报，由此可以提取出威胁指标、攻击技术等相关信息，为防御体系优化提供参考。其三可以识别深层的攻击意图。针对钓鱼邮件、社交工程攻击等特殊攻击场景，其攻击意图通常隐藏在正常的文本或数据中[2]。而生成式 AI 可以理解其深层意图，进而精准识别伪装后的工程学攻击内容。

（三）从人工依赖到智能决策：安全运营的效率革命

网络安全领域面临着高端人才短缺的问题，但传统安全防护运营需要专业人力支持，由此使得安全防护领域发展受限。生成式 AI 的应用一定程度上释放了大量人力，并且提供了智能决策辅助，显著提升了安全防护运营工作的效率。第一，针对安全事件可以提供自动化响应与编排，从而形成标准化的处置流程，有效降低损失。第二，有序传承和调用安全相关知识与数据。生成式 AI 可以通过大规模学习相关报告、文档与处置记录，在安全分析师解决问题时提供辅助[3]。第三，持续学习与自主进化。生成式 AI 可以通过持续学习不断优化自身的检测模型与响应策略，从而具备不断更新和应对新攻击、新环境的能力。

二、生成式 AI 技术在网络安全防护中的应用策略

（一）智能化威胁检测与威胁狩猎：从被动响应到主动出击

第一，多源数据融合与行为异常检测。在信息大爆炸背景下，现代企业的网络安全数据种类繁多且数量庞大，生成式 AI 可以将多源异构数据融合，从而建立正常的行为基线，以此实时监督基线运行状态，提取异常数据。以 HyperGLLM 框架为例，其可以精准提取基础行为特征，并智能识别跨事件、长时间尺度下的复杂行为模式与隐藏依赖，从而检测出高级威胁[4]。

第二，智能化威胁狩猎。生成式 AI 具备完整的工作流，可以独立完成“生成假设、验证假设、关联证据”等工作，进而发现潜在的威胁痕迹。在大语言模型支持下，生成式 AI 可以运用语义推理引擎，分析告警、工作日志、威胁情报、钓鱼内容和代码工件等数据信息，从而快速完成检测、分类与解释，并为检索工具提供辅助。

第三，多模态威胁检测。生成式 AI 可以通过多模态分析应对多样化的攻击手段。例如 SAGE 框架中，可以通过相互解耦的模态专家网络，确保不同模态保持独立的语义表达，进而在全局专家协商与“仲裁庭”机制下，可以实现更高质量的检查效果[5]。

（二）AI 驱动的钓鱼邮件检测与社会工程学防御：对抗语义伪装

第一，基于语义分析的钓鱼邮件检测。AI 生成的钓鱼内容依靠语义伪装与社会工程学，可以欺骗传统以规则匹配为基础的防御机制，因此需要建立以语义理解与上下文验证为基础的检测机制。生成式 AI 可以依托大语言模型对邮件文本进行检测分析，进而理解其意图、语气特征和隐含的情感倾向，完成威胁识别。

第二，多模态钓鱼检测系统。生成式 AI 技术可以辅助处理通信语义，也可以基于计算机视觉识别界面是否为伪造，甚至可以利用行为生物特征检测通话模式是否异常，从而达到从短信、电话、网页等多维度的检测效果。

第三，新型载荷载体的检测与防御。新型钓鱼攻击还会利用 SVG 图像、ICS 日历邀请等方式绕过安全防护系统。生成式 AI 技术可以结合计算机视觉技术，深入分析 SVG 文件是否存在隐藏指令；也可以识别 ICS 文件中的邀请链接是否异常[6]。

（三）智能攻防演练与自动化漏洞挖掘：以攻促防的 AI 实践

第一，自主化 AI 红队测试。当前大量企业在部署 AI 大模型，以此提供客服机器人、自动工作流、智能工作助手等服务，带来便捷服务的同时也产生了越狱攻击、提示词注入、数据窃取、智能体操控等新型攻击方式。生成式 AI 可以辅助模拟渗透测试，创建不同的复杂攻击场景，从而拓宽 AI 应用的安全边界。

第二，基于 AI 攻防的漏洞挖掘。生成式 AI 有着突出的代码漏洞挖掘能力。例如 Mythos Preview 模型可以通过 AI 自动设计攻击与防御方案，进而在攻防过程中不断发现其代码漏洞，有效减少危险的攻击入口[7]。

第三，常态化与自动化的攻防演练。传统安全防护体系下，红蓝对抗演练采用特殊“活动”的形式开展，无法高频率实施。但生成式 AI 可以建立攻击模拟系统，从而在非工作时间不断进行模拟演练，测试防御系统的各个环节，发现脆弱点即可生成完整的演练报告和修复建议[8]。

（四）代码安全审计与自动化漏洞修复：从发现到修复的智能闭环

第一，智能代码审查。生成式 AI 可以分析和理解代码的语义逻辑，因此可以用于直接审查计算机系统及其软件设备的代码。例如 Aardvark 智能体可以监控代码的提交与变更活动，进而自动识别潜在的漏洞，并推断可能的攻击路径，并生成完善的修复建议[9]。

第二，大规模漏洞检测与修复实证。LLM-GUARD 项目组对 ChatGPT-4、Claude 3和 LLaMA 4三种模型进行了系统测评，验证了其在C++和Python等不同类型代码中的缺陷检测能力[10]。实验表明，所有模型在基础语法与语义错误识别方面均有较好表现，但在高级安全漏洞检测与复杂生产环境代码分析方面还需要进一步提升[11]。

第三，供应链安全的 AI 化保障。开源组件和第三方库是现代软件开发的重要基础，这也使得供应链风险在日益提升。生成式 AI 可以分析开源仓库相关数据以及漏洞数据库，从而自动识别供应链是否存在风险点[12]。

（五）AI 内容安全治理与合规建设：构建可信 AI 应用环境

第一，生成式 AI 自身的风险面管理。生成式 AI 是现代网络安全攻击的重要工具，尤其在诱导模型注入、敏感信息挖掘、恶意指令输入、数据投毒等方式，会严重影响和干扰大模型。因此生成式 AI 安全工具的部署需要建立子啊完善的的防护体系之上，包括安全过滤模型输入输出数据、建立最小权限机制、定期开展红队测试等[13]。

第二，数据安全与隐私合规。生成式 AI 在数据学习过程中不可避免地会接触到敏感信息，比如网络日志、行为记录、漏洞信息等[14]。因此在 AI 防护工具使用时必须明确数据采集、存储与使用符合法律要求，并建立完善的数据防泄露措施，所有数据都要脱敏处理后输入模型，以此建立完善的数据流转审计机制。

第三，面向未来的治理框架构建。生成式 AI 防护工具的大规模应用还需解决“安全信任壁垒”问题。我国发布了首份《生成式人工智能网络安全等级保护与检测技术白皮书》，系统梳理了生成式 AI 可能面对的安全风险，进而确保监管部门明确 AI 安全的衡量方式、监督依据、具体影响与责任边界，为相关产业规范有序发展提供了技术支撑[15]。

三、结语

综上所述，生成式 AI 技术在网络安全防护领域的应用，标志着安全防御从“人防”走向“智防”的历史性转折。生成式 AI 一方面为网络防护体系带来了根本性的改变，另一方面也在被攻击者系统性地滥用，带来数据隐私、模型安全、结果可信等方面的全新危险。展望未来，生成式 AI 与网络安全的深度融合还需从智能体构建、AI 攻防对抗新常态、安全治理框架完善等方面切入，以此充分发挥其优势，规避其风险。

参考文献

[1] 施欣晨 . 基于人工智能的计算机网络安全防护系统设计研究 [J]. 中国信息化 ,2025,(12):38+11.

[2] 余龙 . 融合区块链与人工智能的智能家居网络安全防护体系 [J]. 中国宽带 ,2026,22(01):63-65.

[3] 高菲 . 基于 AI 告警降噪技术在网络安全防护中的应用 [J]. 网络安全技术与应用 ,2025,(11):57-59.

[4] 席旭 . 基于人工智能的计算机网络信息安全防护策略研究 [J]. 科技资讯 ,2025,23(18):52-54.

[5] 姚仕聪 . 基于人工智能的移动通信网络安全防护策略研究 [J]. 电脑编程技巧与维护 ,2025,(08):159-161.

[6] 杨信 . 基于人工智能技术的计算机网络安全防护系统设计研究 [J]. 中国自动识别技术 ,2025,(04):55-58.

[7] 何龙 . 人工智能在网络安全防护中的应用与挑战 [J]. 中国安防 ,2025,(08):91-94.

[8] 欧阳金福 . 人工智能视角下数据中心的网络运维与安全防护策略 [J]. 中国宽带 ,2025,21(06):46-48.

[9] 徐阳 , 徐良 , 张麾军 . 人工智能在网络安全防护中的实践应用及其风险防范 [J]. 保密科学技术 ,2025,(02):21-26.

[10] 黄哲 .AI 赋能千行百业网络安全防护任重道远 [N]. 中国计算机报 ,2024-08-12(010).

[11] 高洁 . 基于人工智能的网络安全防护策略研究 [J]. 信息与电脑 ,2024,36(23):98-100.

[12] 王志翔 . 基于人工智能的网络安全研究 [J]. 中国宽带 ,2024,20(12):43-45.

[13] 李芃达 . 筑牢网络安全防火墙 [N]. 经济日报 ,2024-10-10(006).

[14] 王俊岭 . 网络安全防护迎来新需求 [N]. 人民日报海外版 ,2024-08-07(011).

[15] 张彬 , 张亚勇 . 人工智能时代网络安全防护探赜 [J]. 网络安全和信息化 ,2024,(03):6-8.

