人工智能驱动的计算机网络安全自动识别与防护策略

纪大伟

哈尔滨石油学院，黑龙江 哈尔滨 150028

摘　要：计算机网络安全是数字时代保障信息系统稳定运行的核心议题，随着网络攻击手段的智能化与复杂化，传统防护机制已难以应对新型安全威胁。本文聚焦人工智能驱动的计算机网络安全自动识别与防护策略，系统分析人工智能技术在网络安全领域的应用逻辑与实践路径。通过阐述智能识别技术对网络异常行为的精准捕捉、威胁特征的动态提取，以及自动化防护机制的实时响应原理，揭示人工智能为网络安全防护带来的范式革新。同时，探讨技术应用中面临的算法偏见、数据安全等挑战，并提出针对性优化策略，为构建自适应、智能化的网络安全防护体系提供理论参考。

关键词：人工智能；信息技术；计算机网络；安全防护

DOI:10.64649/yh.shygl.2025010012

0 引言

在信息技术高速发展的背景下，计算机网络已成为社会运转的关键基础设施，承载着海量数据传输与业务处理任务。然而，网络空间的开放性与复杂性使得安全威胁持续升级，从传统的病毒入侵、端口扫描，到新型的 APT 攻击、勒索软件、AI 生成式攻击等，攻击手段呈现出隐蔽化、协同化、智能化特征。传统网络安全防护依赖预设规则与特征库匹配，存在响应滞后、误报率高、难以应对未知威胁等局限，无法满足动态变化的安全需求。

人工智能技术的崛起为网络安全防护提供了全新思路。其通过机器学习、深度学习等算法对网络数据进行深度分析，可实现威胁的自动识别、实时预警与主动防御，打破传统“被动防御”模式。基于人工智能的网络安全系统能够从海量数据中学习攻击模式，自主更新防护策略，对未知威胁进行预判，为网络空间构建智能化安全屏障。

1 人工智能驱动的网络安全自动识别技术

1.1 异常行为检测机制

异常行为检测作为网络安全防护的第一道防线，核心在于建立正常网络行为基线以识别偏离基线的异常活动。人工智能技术从多方面提升其检测效能：借助深度学习算法对网络流量、用户操作日志、设备运行状态等多源数据展开特征挖掘，提取数据包大小、传输频率、协议类型、访问路径等深层特征，构建全面的行为特征模型 [1]；基于时序数据分析，通过强化学习算法实时调整正常行为基线，以适应业务高峰期流量波动、新设备接入等网络环境动态变化，减少因基线固化产生的误报；对网络行为进行实时监测，利用神经网络模型计算行为异常度评分，当评分超阈值时自动触发预警，实现从“事后分析”到“事中干预”的转变。

1.2 威胁特征智能学习

传统威胁检测依赖人工定义的特征库，难以应对变异迅速的恶意代码与攻击手段。人工智能凭借自主学习机制实现威胁特征动态更新：通过卷积神经网络（CNN）提取恶意代码的二进制序列、汇编指令等特征，识别变异代码中的核心攻击逻辑等不变特征，生成具备泛化能力的检测模型，有效对抗代码混淆、加壳等逃避技术；利用 K-means、DBSCAN 等无监督学习算法对历史攻击事件聚类，挖掘不同攻击手段的共性模式，如 APT 攻击的长期潜伏特征、分布式拒绝服务攻击的流量特征，形成攻击家族图谱，为新型攻击识别提供依据；基于知识图谱技术构建软件漏洞与攻击方法的关联模型，结合代码静态分析与动态执行数据，预测潜在漏洞的利用路径，实现对零日攻击的提前预警。

1.3 攻击溯源与归因分析

攻击溯源与归因是制定防护策略的关键，人工智能通过关联分析提升效率：整合防火墙日志、入侵检测告警、终端行为记录等分散数据，利用图神经网络构建攻击行为关联图，精准识别攻击路径中的关键节点，如攻击源 IP、跳板主机和目标资产；基于贝叶斯网络模型分析攻击步骤的时序关系，推断攻击者的目标（如数据窃取、系统破坏）与战术（如横向移动、

权限提升），为针对性防护提供决策支持；随着攻击持续展开，通过在线学习算法实时更新溯源模型，动态修正初始判断，不断提高归因准确性，从而为后续防护行动提供精准依据。

2 人工智能驱动的网络安全自动化防护策略

2.1 实时响应与动态拦截

自动化防护的核心在于将识别结果快速转化为防护行动，人工智能技术通过多重机制实现实时响应：构建基于威胁等级（低危、中危、高危）的分级响应策略，针对低危威胁（如端口扫描），系统自动执行 IP 封禁、流量限速等拦截规则；面对中高危威胁（如恶意代码注入），则触发联动响应，如隔离受感染终端、切断攻击链路，同时同步向安全人员发出告警 [2]。借助强化学习算法实现自适应防护规则生成，以“最小防护代价”（对正常业务影响最小）为目标自主生成规则，例如针对新型 DDoS 攻击，可动态调整流量清洗策略，在过滤攻击流量时保障合法请求通行。利用联邦学习技术促成不同安全设备（防火墙、入侵防御系统、终端安全软件等）的模型共享与协同决策，形成分布式防护体系，有效避免单点防御失效。

2.2 漏洞修复与补丁管理

漏洞作为网络安全的主要风险点，其修复效率直接影响防护体系的有效性，人工智能通过智能化管理显著提升这一过程的效能。在漏洞优先级排序方面，系统结合漏洞的危险等级、影响范围（如核心业务系统、普通终端等不同层级资产）以及被攻击利用的可能性等多维度因素，运用多目标决策算法进行综合评估与排序，确保修复资源优先投向高风险漏洞，避免资源错配导致的防护短板。

针对部分已知漏洞，借助生成式 AI 技术可自动生成修复补丁，并在虚拟仿真环境中完成兼容性测试（如与现有软件、系统的适配性）和有效性验证（如漏洞是否被彻底封堵），大幅缩短从漏洞发现到修复的时间差，降低被攻击利用的窗口风险。同时，基于历史补丁部署数据训练的预测模型，能够提前评估补丁部署可能引发的系统冲突，如软件兼容性问题、运行性能下降等，为技术人员提供针对性应对方案，保障修复过程中业务的连续性与稳定性。

2.3 安全态势感知与预警

构建全局安全态势感知能力是实现主动防御的核心基础，人工智能技术通过多维度应用显著提升感知与预警效能。在态势可视化建模方面，系统将网络拓扑结构、资产分布信息、实时威胁事件等多类数据转化为动态可视化模型，借助时序神经网络对海量历史与实时数据进行深度分析，精准预测安全态势的演化趋势，如攻击行为的潜在扩散路径、可能受影响的网络区域等，为防御决策提供直观参考。

多维度风险评估则从资产价值（如核心服务器、普通终端的差异化权重）、漏洞数量与严重程度、攻击发生频率与类型、现有防御体系的防护能力等维度，构建全面的风险评估指标体系，通过模糊综合评价算法将多维度数据转化为量化的整体风险值，为安全资源的精准调配提供科学依据，避免资源浪费或防御薄弱环节 [3]。此外，针对网络环境的动态变化（如重大会议期间的高安全需求、业务高峰期的流量波动），系统通过自适应算法实时调整预警阈值，在保障威胁捕捉灵敏度的同时减少误报，有效避免安全人员因频繁无效预警产生的“预警疲劳”，确保关键威胁信息得到及时响应。

3 人工智能在网络安全应用中的挑战

3.1 算法局限性与鲁棒性问题

人工智能模型的性能依赖于训练数据的质量与数量，若训练数据存在偏见（如缺乏特定类型攻击样本），会导致模型检测精度下降。攻击者可能利用对抗性样本技术欺骗 AI 模型（如通过细微修改恶意代码逃避检测），降低防护系统的可靠性。复杂网络环境中的噪声数据（如突发流量波动）可能干扰模型判断，增加误报率。

3.2 数据安全与隐私风险

人工智能训练需要大量网络数据（包括用户操作记录、敏感业务数据等），数据收集与使用过程中存在隐私泄露风险。若训练数据未经过脱敏处理，可能导致敏感信息被窃取；同时，模型本身可能成为攻击目标，攻击者通过模型反演技术还原训练数据中的隐私信息，引发数据安全问题。

3.3 技术协同与标准化不足

现有网络安全设备多采用封闭架构，不同厂商的 AI 模型难以实现数据共享与协同决策，形成“信息孤岛”。此外，人工智能在网络安全领域的应用缺乏统一标准（如特征提取规范、响应策略接口），导致不同系统间兼容性差，影响防护体系的整体性与扩展性。

3.4 人才与伦理挑战

人工智能与网络安全的交叉领域需要复合型人才，既掌握 AI 算法原理，又熟悉网络攻击技术，但目前此类人才供给不足。同时，AI 驱动的自动化防护可能引发伦理争议，如自主决策系统误判导致正常业务中断，或因算法黑箱

特性难以追溯决策责任。

4 优化策略与未来展望

4.1 技术层面的优化路径

增强模型鲁棒性是首要方向，通过采用联邦学习、迁移学习等技术，可减少对集中式训练数据的依赖，降低数据孤岛与隐私泄露风险；同时引入对抗训练方法，让模型在与模拟攻击样本的对抗中提升对异常样本的识别与抵抗能力。此外，推行多模型融合策略，如将深度学习的自主学习能力与传统规则引擎的确定性逻辑相结合，能有效弥补单一模型在复杂场景下的局限性，提升检测与防护的稳定性。

构建安全数据治理体系需双管齐下，一方面建立数据分类分级机制，按敏感度对网络日志、业务数据等进行层级划分，对高敏感数据实施脱敏处理（如数据 anonymization、Tokenization 等）；另一方面引入隐私计算技术，如差分隐私通过添加噪声保护数据隐私，安全多方计算实现数据“可用不可见”，在保障隐私的前提下支撑跨主体的模型训练与协同决策[4]。推动技术标准化与开放协作同样关键，需制定统一的行业标准，规范数据接口格式、模型评估指标（如检测准确率、响应时延）及安全合规要求，解决不同系统间的兼容性问题。同时鼓励开源社区发展，促进安全模型、算法工具的共享复用，打破厂商技术壁垒，形成协同创新的生态，提升整体防护体系的扩展性与互操作性。

4.2 管理与伦理层面的完善措施

加强人才培养与跨学科合作是基础保障，需推动高校与企业深度协同，联合开设人工智能与网络安全交叉学科课程，课程体系涵盖机器学习算法、网络攻击原理、数据安全治理等内容，定向培养既懂 AI 技术又精通网络安全的复合型人才。同时建立产学研合作机制，鼓励企业参与高校科研项目，推动实验室技术向产业应用转化，通过联合攻关解决 AI 在网络安全领域的实际难题，实现理论研究与实践应用的无缝衔接。

建立算法伦理与问责机制是规范应用的关键，需明确 AI 安全系统的决策边界，对于涉及核心业务中断、数据销毁等重大操作，必须保留人工干预通道，避免系统自主决策引发不可逆损失。定期对自动化防护系统开展伦理审查，重点排查算法设计中可能存在的偏见（如对特定类型网络行为的误判倾向），确保防护规则的公平性与合理性。此外，构建决策追溯机制，通过日志记录 AI 模型的训练数据、决策依据及执行过程，明确安全事件中技术系统、操作人员的责任划分标准，为事故追责提供可追溯的证据链，平衡技术创新与风险管控的关系。

4.3 未来发展趋势

未来，人工智能驱动的网络安全防护将呈现多维度演进趋势。其一，向“认知安全”深度拓展，借助类脑计算技术模拟人类认知逻辑，不仅能识别攻击行为的表面特征，更能深度解析攻击者的战略意图与战术路径，实现从“被动防御”到“主动预测”的质变，提前构筑针对性防御屏障[5]。其二，边缘 AI 应用全面铺开，在终端设备（如物联网终端、工业控制设备）部署轻量化 AI 模型，实现本地化威胁检测与实时响应，大幅减少数据上传云端的传输延迟，尤其适用于对实时性要求极高的工业互联网、车联网等场景，提升边缘节点的自主防护能力。其三，与区块链技术深度融合，利用区块链的不可篡改性与分布式记账特性，构建安全数据共享与模型协同的信任机制。

5 结论

人工智能技术为计算机网络安全防护带来了革命性突破，通过自动识别技术实现对网络威胁的精准捕捉，借助自动化防护策略构建实时响应机制，显著提升了网络安全防护的效率与智能化水平。然而，算法局限性、数据安全、技术协同等挑战仍需在实践中不断攻克。未来，需通过技术优化、标准制定、人才培养等多维度措施，推动人工智能与网络安全的深度融合，构建兼具防御效能与伦理合规的智能化安全体系，为数字经济的健康发展提供坚实保障。

参考文献：

[1] 王维 . 计算机网络安全中的风险与防范策略分析 [J]. 电子技术 ,2024,53(01):160-161.

[2] 张茜 . 基于人工智能的计算机网络通信安全风险评估与防护 [J]. 软件 ,2024,45(01):152-154.

[3] 刘智骁 , 李浩 . 计算机网络空间安全风险因素与防控策略分析 [J]. 网络空间安全 ,2023,14(06):76-80.

[4] 冯影 , 乔瑶瑶 . 人工智能及其在计算机网络技术中的应用分析 [J]. 中国高新科技 ,2023,(19):42-43+64.

[5] 从慧杰 , 王贺庆 , 陶思源 , 等 . 计算机网络安全与风险控制策略分析 [J]. 集成电路应用 ,2023,40(10):284-285.

作者简介：纪大伟（1986.10—），男，汉族，黑龙江省哈尔滨市，大学本科，副教授，计算机科学与技术。