人工智能驱动的信息安全防护技术与应用研究

刘　洋　张昕航*
深信服科技股份有限公司杭州分公司　浙江杭州　310000

摘　要：随着信息技术的发展，网络安全威胁日益复杂，传统防护手段难以全面应对。人工智能凭借卓越的数据处理能力和自适应学习能力，能够精准识别威胁并动态防护，为信息安全提供了全新的解决方案。本文围绕人工智能在信息安全防护中的技术支撑与实际应用展开研究，首先介绍人工智能的基本概念和核心技术，探讨其在入侵检测、恶意代码识别、用户行为分析等安全防护场景中的应用，进一步分析当前面临的技术挑战与发展趋势。通过案例分析与技术比较，揭示人工智能驱动的信息安全系统在准确性、实时性和智能性方面的优势，并提出未来可行的发展路径。研究表明，人工智能技术的引入将有效提升信息安全防护的整体水平，具有重要的理论价值与现实意义。

关键词：人工智能；信息安全；防护技术；恶意代码识别

序言

信息化时代背景下，数据资源已成为国家战略资源，信息安全的重要性日益凸显。然而，传统的信息安全防护手段如防火墙、病毒库等方式多依赖静态规则匹配，面对新型攻击手段和不断变化的威胁模式时往往显得力不从心。人工智能技术的快速发展为信息安全领域注入了新的活力。通过机器学习、深度学习、自然语言处理等手段，人工智能不仅能够实现对大量数据的实时分析与智能识别，还能在未知威胁出现时进行预测与响应，从而显著提升防护系统的动态适应能力与智能决策能力。近年来，AI在金融、医疗、制造等多个行业取得了突破性进展，其在网络安全中的探索也愈加深入。

一、人工智能技术概述

（一）人工智能的基本概念

人工智能是指通过计算机程序模拟人类思维过程，使机器能够具备感知、识别、推理、学习和决策等智能行为的技术系统。其本质在于利用算法与数据实现“机器智能”，以支持复杂环境下的自动处理和反馈。AI的发展经历了符号主义、连接主义到当今的深度学习阶段，逐渐从依赖规则驱动转向以数据驱动为核心[1]。当前主流的人工智能研究方向包括机器学习、深度学习、知识图谱、自然语言处理等，均以提高机器的自主学习与推理能力为目标。人工智能与传统程序化技术最大的不同在于其具备“学习能力”，能从历史数据中提取规律，不断优化决策模型，从而在面对未知数据或变化场景时展现出强大的泛化能力[2]。

（二）人工智能的关键技术

人工智能之所以能够在多个领域取得实质性突破，核心在于一系列关键技术的成熟与融合，首先是机器学习技术，尤其是监督学习与无监督学习在分类、聚类和预测任务中的广泛应用[3]。其次是深度学习技术，它通过多层神经网络结构对图像、语音、文本等非结构化数据进行特征提取与建模，大幅提高了AI的感知能力[4]。再者是自然语言处理技术，使机器能够理解和生成自然语言，从而在人机交互与语义分析中发挥作用；此外，还有强化学习、联邦学习等新兴方向，推动AI向更高水平的智能化演进。这些关键技术构成了人工智能系统的技术基石，为其在信息安全防护中的建模分析、威胁识别与自适应防御提供了强大支持[5]。

二、信息安全防护的现状与挑战

（一）当前信息安全的主要威胁类型

当前，信息安全面临的威胁种类日益复杂和多样，主要包括恶意软件攻击、网络钓鱼、数据泄露、勒索软件、拒绝服务攻击（DDoS）等。其中，恶意软件通过植入病毒、木马或间谍软件，窃取用户信息或破坏系统运行；网络钓鱼则通过伪装为合法网站或电子邮件诱骗用户泄露敏感信息；数据泄露则可能源于内部人员泄密、

作者简介：
刘洋，男，汉，湖南省新化县，1991年12月9日，硕士，信息安全；
张昕航，男，汉，陕西省汉中市，1993年12月16日，本科，信息安全。

系统漏洞或第三方平台管理不善，给用户隐私和企业声誉带来巨大风险。近年来，勒索软件攻击愈演愈烈，攻击者通过加密用户重要文件并索要赎金牟利，不仅造成财产损失，还会导致业务中断。DDoS攻击则通过大量请求挤占目标服务器资源，使其无法正常提供服务。此外，随着物联网设备、云计算平台和移动终端的大规模普及，攻击面也显著扩大，信息安全威胁已渗透到社会生活的方方面面。面对这些不断演化的威胁，传统安全手段已难以独立应对，亟需引入更加智能、高效的防护机制。图1为网络病毒和网络攻击多发性和多样性的特征。

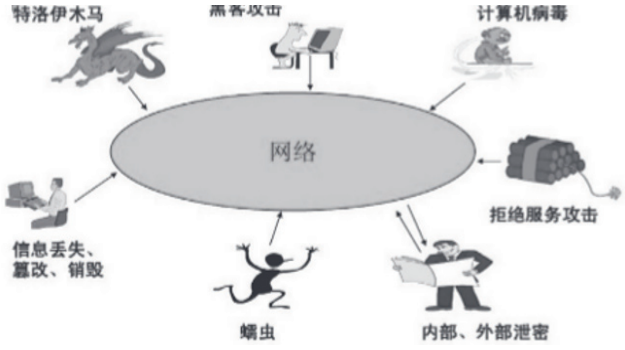

图1　网络病毒和网络攻击多发性和多样性的特征

（二）传统信息安全防护手段及其不足

传统的信息安全防护手段主要包括防火墙、杀毒软件、入侵检测系统（IDS）和访问控制策略等。这些技术在早期互联网环境中确实发挥了重要作用，但随着攻击技术的持续演进，其防护能力逐渐显现不足。例如，防火墙主要基于静态规则进行访问控制，难以应对隐蔽性强的新型攻击手段；杀毒软件依赖特征库识别病毒，对于未知或变种病毒的检测能力有限；入侵检测系统虽然可以监控网络异常，但误报率高，响应速度慢；而传统的访问控制策略主要依赖人工配置，难以适应动态变化的业务需求和用户行为。在当前大数据、高速互联、攻击智能化的背景下，传统手段普遍存在响应滞后、检测能力弱、依赖人工维护等问题，已难以全面保障信息系统的安全。此外，面对零日漏洞、APT（高级持续性威胁）等复杂攻击，传统技术手段缺乏自主学习与动态响应能力。因此，必须探索更为先进的技术路径来构建主动、智能、适应性强的信息安全防护体系。

（三）AI介入信息安全防护的必要性分析

随着网络攻击日益智能化与复杂化，传统以规则驱动、被动式响应为主的信息安全防护手段已难以应对新型威胁。人工智能技术凭借强大的数据处理、模式识别、自主学习与预测能力，能够实现主动防御与智能响应，为信息安全注入全新动力，提供更加高效精准的解决方案。首先，AI能够对海量的网络数据进行实时分析，快速识别异常行为与潜在威胁，显著提升攻击检测的准确率和响应速度。其次，AI技术具备自适应学习能力，可以持续优化安全策略，发现传统规则难以识别的高级持续性威胁（APT）和零日攻击。此外，在面临复杂多变的网络环境和多源异构的数据结构时，AI模型能够自动提取特征、构建预测模型，有效减轻人工负担并提高系统的整体防御水平。再者，AI还可用于构建智能身份验证、风险评估、日志分析等多个维度的安全机制，实现从“被动防御”向“主动预测与快速响应”的转变。因此，将人工智能技术引入信息安全防护体系，不仅是技术发展的趋势，更是应对未来安全挑战的必然选择。

三、人工智能在信息安全防护中的应用

人工智能在信息安全防护中的应用可从图中多个层面体现。在基础层，依托GPU、NPU等算力平台，结合深度学习、数据挖掘等算法，有效支撑大规模数据分析与威胁检测。技术层中，感知技术如自然语言处理与语

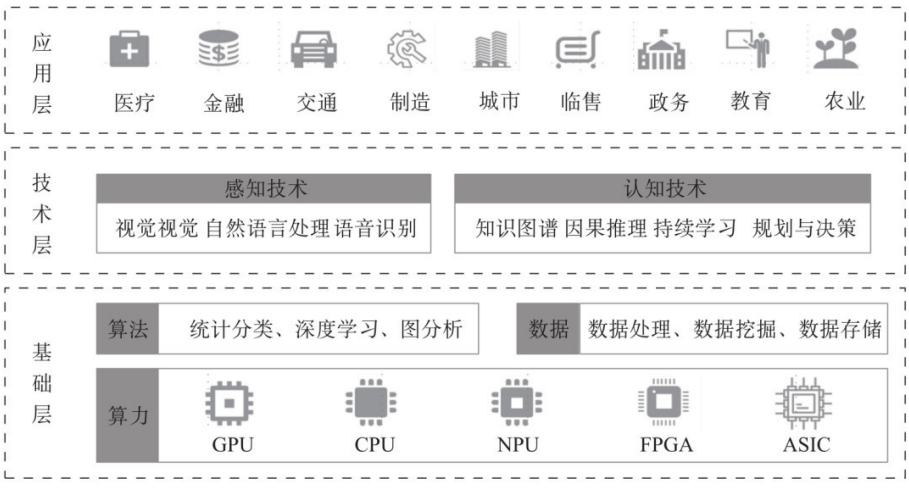

图2　人工智能在信息安全防护中主要应用

音识别可用于识别钓鱼邮件与语音诈骗，认知技术则通过因果推理和持续学习提升入侵检测与应急响应能力。在应用层，AI广泛用于政务、金融、医疗等领域的信息安全管理，提升防护效率和智能化水平。图2为人工智能在信息安全防护中主要应用。

（一）入侵检测与行为分析

人工智能技术在入侵检测与行为分析中的应用大大提升了信息安全系统的主动防护能力。传统的入侵检测系统（IDS）多依赖于已知特征库，难以应对新型或变异攻击，而人工智能，特别是机器学习和深度学习算法，能够基于大量网络流量数据进行学习，从而自动识别异常行为。例如，利用神经网络、支持向量机（SVM）和聚类算法，可实现对用户行为模式的建模，识别潜在的异常访问、网络扫描、暴力破解等攻击行为。AI还可以结合上下文数据和行为路径进行动态分析，提升对“低慢小”攻击的识别能力。此外，通过持续学习和模型优化，AI入侵检测系统能适应复杂多变的网络环境，实现实时、精准的威胁识别与响应。

（二）恶意软件识别与分析

恶意软件持续演变，对信息系统构成严重威胁。人工智能技术，尤其是深度学习和自然语言处理（NLP），为恶意软件的自动识别与分析提供了强有力的工具。传统的静态分析方法依赖于签名库，无法识别变种病毒，而AI能够通过对大量已知恶意样本进行训练，提取其在代码结构、系统调用、行为日志等方面的特征，实现未知或变种恶意软件的识别。例如，卷积神经网络（CNN）可以将可执行文件转换为图像进行分析，而循环神经网络（RNN）可处理API调用序列，从中发现隐藏的恶意模式。此外，AI还可用于恶意软件的行为分析，如检测加密勒索、远程控制、数据窃取等行为路径，实现从检测到响应的全过程自动化。随着AI模型的不断优化，其对抗恶意软件的能力也日益增强，成为提升信息系统安全性的关键手段。

（三）钓鱼邮件与欺诈信息检测

钓鱼邮件与欺诈信息是网络攻击的重要手段，常以伪造身份、诱导点击链接或下载附件的形式窃取用户敏感信息。人工智能技术在这一领域的应用有效提高了识别准确率和响应速度。通过自然语言处理（NLP）和文本分类算法，AI能够分析邮件内容的语义特征、语言风格、情感倾向等，识别可疑信息。同时，结合发件人地址、邮件头信息、超链接跳转路径等结构化数据，AI可构建多维度模型，对钓鱼行为进行全面识别。深度学习模型如BERT或Transformer架构在邮件内容分析中展现出强大能力，尤其擅长识别伪装性强、逻辑欺骗性高的钓鱼信息。此外，AI还能结合用户历史行为数据，对接收者是否可能成为攻击目标进行预测，提前发出风险警告。通过持续学习和攻击样本积累，AI在打击钓鱼与欺诈方面具有显著优势。

（四）风险预测与态势感知

人工智能在信息安全风险预测与态势感知中的应用，极大地提升了网络防御的前瞻性和决策智能化水平。通过整合来自网络流量、日志信息、终端行为、社交媒体等多源数据，AI能够实现对网络空间的实时监测与动态分析。机器学习算法，如随机森林、集成学习、图神经网络等，能对历史攻击事件建模，预测可能出现的安全威胁及其演化趋势。与此同时，AI驱动的态势感知系统可通过可视化手段展示当前安全状态、关键资产风险等级与潜在攻击路径，帮助安全运维人员实现精准决策。深度强化学习的引入，还能模拟攻击者行为路径，从而优化防御部署策略。总体而言，人工智能使风险管理由“事后响应”向“事前预测”转变，为构建主动安全体系奠定了基础。

结语

综上所述，随着信息技术的飞速发展，网络空间安全面临着前所未有的挑战。人工智能作为新一代信息技术的核心力量，正在不断推动信息安全防护技术的革新与应用优化。通过本论文的研究可以看出，人工智能在入侵检测、恶意行为识别、数据加密与隐私保护等方面展现出强大的能力，极大提升了防护效率与智能化水平。然而，人工智能本身亦存在算法安全、数据依赖和可解释性等问题，仍然需要在实践中不断完善。

参考文献

[1] 贾昉.探究将网络信息安全教育融入中职学校计算机教学活动路径[C]//2022电脑校园网络论坛论文集，2022：97-99.

[2] 贾立伟，石晓明，李琨.安全风险时代计算机人工智能与防护策略探讨[J].信息记录材料，2021，22（07）：53-54.

[3] 杨振祺.大数据时代背景下人工智能在计算机网络技术中的应用[J].信息记录材料，2021，22（03）：161-162.

[4] 闫卫刚.人工智能时代计算机信息安全与防护策略探讨[J].电子测试，2020（19）：135-136.

[5] 刚轶金，李涛.大数据时代计算机网络信息安全及防护策略[C]//2020万知科学发展论坛论文集（智慧工程三），2020：121-129.