人工智能赋能下的网络安全威胁检测与应对策略

纪　亮
中海油信息科技有限公司　广东深圳　518000

摘　要：本研究意在针对那人工智能（AI）给予能力支持背景下、网络安全威胁检测以及应对策略所发生的演变展开探讨。鉴于网络安全威胁朝着日益复杂的态势发展，致使传统方法暴露出一定程度的局限性之时，人工智能技术的运用给提升威胁检测的效率、准确性以及自动化程度提供了崭新路径。此研究先是对包含机器学习、深度学习还有自然语言处理之类技术在内的、人工智能于网络安全威胁检测方面的应用加以综述，其次针对诸如对抗性攻击和AI模型投毒这类由人工智能驱动产生的新型网络安全威胁予以分析，进而深入研究面对这些威胁所采用的像是自动化响应、威胁情报分析与预测性防御等有效应对策略。该研究还对人工智能在网络安全领域所产生的伦理和社会方面影响进行评估，着重关注隐私保护、算法偏见以及可解释性这些问题。研究得出人工智能虽可显著提高网络安全威胁检测的效率与准确性，然而也带来新的安全风险，必须采取对应防御措施的结论。除此之外，人工智能赋能之下的网络安全应对策略应注重自动化、智能化与预测性以达成主动防御。本研究构建出一个供企业和组织参考的、人工智能赋能的网络安全威胁检测与应对框架，并对人工智能在网络安全领域未来的发展趋势进行展望。

关键词：人工智能；网络安全；威胁检测；对抗性攻击；自动化响应；伦理影响

引言

处于数字化时代，网络已深度渗透到社会生活的各个方面，随之网络安全问题也愈发重要。当传统的网络安全威胁检测方法面临日益复杂且多样化的攻击手段之际，渐渐显露出局限性。凭借机器学习、深度学习以及自然语言处理等技术，人工智能技术的快速发展给网络安全领域带来新机遇，进而提升威胁检测的效率与准确性，实现更为智能化的安全防御。不过，人工智能的应用同时也引发一系列诸如对抗性攻击和AI模型投毒等新型威胁这类的新问题与挑战。因此针对人工智能赋能之下的网络安全威胁检测以及应对策略展开深入研究具有重要现实意义。

一、人工智能在网络安全威胁检测中的应用

机器学习作为人工智能的重要分支，在网络安全威胁检测领域作用关键。它运用决策树、支持向量机、随机森林等算法，对网络流量数据和系统日志进行分类与异常检测。机器学习模型通过学习正常网络流量模式，能精准识别DDoS攻击、恶意软件传播等异常流量行为，大幅提升检测效率，降低误报率。在用户行为分析上，机器学习构建用户正常行为模型，一旦检测到非法登录、未授权数据访问等异常操作，系统会迅速响应并处理。

深度学习是机器学习的进阶形式，具备强大的特征提取与模式识别能力，特别适合处理复杂网络数据。卷积神经网络（CNN）可检测网络流量中恶意软件特征，循环神经网络（RNN）擅长分析网络日志的时间序列数据，预测潜在安全威胁。深度学习还可用于图像和视频数据的分析，检测网络钓鱼图片、视频等恶意内容，增强网络安全防护。

自然语言处理（NLP）技术能让计算机理解和处理人类语言，在网络安全部署中意义重大。利用NLP技术分析网络文本数据，如社交媒体安全信息、安全论坛讨论等，可及时发现潜在安全威胁与漏洞信息。企业借助自动化工具，从安全公告、博客文章、社交媒体讨论等公开资源中提取关键信息。NLP技术自动解析文本，识别关键术语、攻击模式和技术细节，帮助企业快速掌握最新安全漏洞及修复办法，加快信息处理速度，提高准确性，提前做好威胁应对准备。例如，新的零日漏洞披露时，安全团队借助NLP技术能迅速评估风险，采取防护措施。

机器学习、深度学习和自然语言处理技术，共同为网络安全构建起多层次防御机制。它们提升了威胁检测的准确性与效率，增强了企业应对网络安全挑战的能力。企业持续优化这些技术应用，在保障信息安全的同时，

能维持业务连续性和客户信任。

二、人工智能驱动的新型网络安全威胁

在网络安全这片领域之中，对抗性攻击以及AI模型投毒这两者，乃是极具挑战性的两种威胁形式；其中的对抗性攻击，其含义为攻击者通过故意地制造那些对抗性样本，借此去欺骗人工智能模型，进而使该模型产生错误判断的情况，比如在恶意软件检测领域内，攻击者能对恶意软件代码做出细微的修改动作，致使基于机器学习的检测系统无法识别其威胁的本质，从而成功绕过相关防护措施这般；同样的，在图像识别系统方面，攻击者向输入数据添加几乎难以察觉的扰动，这样的操作会导致深度学习模型将恶意图像误判成正常图像，这种技术于实际应用当中极有可能引发严重的安全漏洞。而另一方面，AI模型投毒则是涉及到攻击者在模型训练阶段注入恶意数据，以此对模型的学习过程以及最终性能造成影响，例如在网络流量分析过程里，要是攻击者能够在用于训练的网络日志之中插入伪造的正常流量记录，便可能致使模型学习到错误行为模式，即会认为这些异常流量属于正常，那么当真正的攻击发生之际，模型或许会因为先前被投毒的数据而未能及时发出警报，使得系统防御能力被极大削弱。此两种攻击方式，不仅展现出人工智能技术在网络安全领域所存在的脆弱性，而且也提醒着我们必须采取更为严密的安全措施，从而保护AI系统避免受到此类威胁的影响，所以为了应对这些复杂挑战，研究者们正在努力开发新的算法与技术，其目的在于增强模型的鲁棒性和抗干扰能力，确保即便是面对精心设计出的对抗性样本亦或是被污染的训练数据，AI系统依旧可以保持高效且准确的威胁检测能力。

三、针对新型威胁的应对策略

（一）自动化响应

鉴于当下网络安全威胁呈现出愈发复杂的态势，故而自动化响应机制所具备的重要性便不言而喻了。借助人工智能技术，能够达成针对安全事件进行自动应对的目的，例如实现像自动隔离遭遇感染的设备以及切断恶意网络连接之类的操作。举例来讲，在一家规模较大公司的网络环境当中，若发现存在一台电脑不幸被恶意软件所感染，那么已部署的自动化系统便能够立刻识别并且对该电脑进行隔离，从而有效防止病毒的进一步扩散。不仅如此，该系统还能够自动对攻击究竟源自何处展开分析，迅速阻断恶意的IP地址抑或是域名，以此减少遭受攻击的可能性。自动化响应系统依据提前精心设定好的安全策略以及规则，仅需几秒钟就可以针对安全事件做出相应的反应，大大缩短从发现威胁直至采取实际行动所需耗费的时间。

在大公司复杂的网络环境里，当面对错综复杂的网络安全威胁时，自动化响应机制所起的作用极大。曾有一日，公司的安全监控系统察觉到一名员工所使用的电脑被一种全新的恶意软件所感染，且此恶意软件企图通过网络传染给其他设备。幸而提前妥善部署了自动化响应系统，一旦检测到异常情况，系统立刻识别并且隔离受感染的电脑，成功有效阻止病毒的继续扩散。与此同时，系统自动针对攻击来源展开分析，迅速切断与这个恶意软件相关联的所有外部IP地址以及域名的连接，显著缩小了被攻击的范围。所有这些操作皆是按照提前确定好的安全策略以及规则，在短短几秒钟之内便完成了，明显缩短从发现威胁到采取行动的时间跨度。这种快速做出反应的方式，不仅提高应对威胁的效率，还降低对人工的依赖程度，进一步降低运营成本。更为关键的是，自动化响应系统能够持续不断监控网络，即便是处于非工作时间，也能够及时处理潜在安全威胁。这使得公司在面对新出现的安全问题之时，可以迅速调整防御方法，保证业务正常以及数据安全。由这个具体例子不难看出，自动化响应机制与人工智能技术相结合，为公司构建起一个既强大又具备灵活性的保护屏障，让公司在面对网络威胁时能够表现得更加从容淡定。

（二）威胁情报分析

威胁情报分析是应对网络安全威胁的重要手段，通过收集、分析多渠道信息，助力企业提前做好防御准备。信息源包括安全厂商报告、黑客论坛、公开漏洞数据库等。比如，新零日漏洞披露时，威胁情报分析可帮组织快速评估风险，采取防护措施。

人工智能的自然语言处理（NLP）技术，能高效处理大量文本数据，提取有价值的威胁情报。它可自动解析安全公告、博客文章、社交媒体讨论，识别关键术语、攻击模式与技术细节。

全球运营的金融机构中，威胁情报分析是应对网络安全威胁的关键。新零日漏洞公开披露后，安全团队迅速启动分析流程，从多信息源收集数据，评估漏洞对企业系统的潜在风险，了解攻击者利用漏洞的方式与时机。借助NLP技术，安全团队高效处理海量文本数据，解析各类信息。分析知名黑客论坛帖子时，团队发现新型攻击工具使用趋势，预测到针对金融行业的大型网络攻击

计划。

基于这些情报，企业迅速行动，更新防火墙规则，强化入侵检测系统，开展关键服务器补丁管理。安全团队还模拟攻击场景，培训员工识别与响应类似威胁。这种前瞻性分析让企业在攻击前充分准备，安全措施更精准有效。威胁情报分析不仅是被动防御部分，更是主动出击关键，使安全团队在威胁萌芽时就能反应，大幅提升整体网络安全水平，保障客户数据安全与业务连续性。

（三）预测性防御

在网络安全防护这个广大领域之中，存在着一种被称之为预测性防御的防御方式，它依靠先进人工智能技术得以实现，其中机器学习以及深度学习算法起着尤为关键的作用，此方式乃是借助对历史数据与实时数据展开分析，以达对未来或许出现的网络安全威胁作出预测之目的，进而提前将防御措施精准落实到位。以网络流量数据为例，通过对其实施时间序列分析，即可对即将来临的分布式拒绝服务（DDoS）攻击进行预测，在此情形下，系统会对流量模式异常与否展开监测，诸如数据包数量突然增多或者特定IP地址频繁发出请求之类的情况皆会被监测到，而后依据利用历史数据训练得出的模型去判定是否潜藏威胁。同样，针对内部用户行为数据展开分析，也有助于对员工如未经授权的数据访问或文件下载等异常操作进行预测，通过构建用户行为基线模型，一旦出现任何偏离正常行为的状况，警报便会被触发，提示或许存在内部威胁。

于一个大型电子商务平台而言，预测性防御已然成为提升网络安全防护能力的关键策略，企业借助先进人工智能技术，特别是机器学习和深度学习算法，能够分析历史数据与实时数据，预测未来可能出现的网络安全威胁并提前落实防御措施。例如，系统针对网络流量进行时间序列分析时，监测到某一特定IP地址在短时间内发出大量请求，这种异常流量模式与以往DDoS攻击的数据特征相吻合，依据之前训练出的模型，系统迅速判定这极有可能是一次即将发生的DDoS攻击，并且自动开启防御机制，采取诸如增加带宽、过滤恶意流量等措施，从而有效避免服务中断。与此同时，企业还深入分析内部用户行为数据并构建用户行为基线模型，当一名员工试图未经授权访问敏感数据，此行为偏离其正常操作模式时，警报立刻被触发，安全团队迅速介入展开调查，发现是一名新入职员工因不熟悉权限设置而出现的误操作，即便如此，这也充分显示了预测性防御在识别潜在内部威胁方面的有效性。这种防御方式不仅使企业从传统被动防御转变为积极主动防御模式，而且在威胁实际发生之前就已采取防范措施，极大提升整体安全防护能力并降低事后补救的成本与复杂度，通过持续优化更新预测模型，企业能在快速变化的威胁环境中维持领先地位，确保关键业务资产安全，维护客户信任以及业务的连续性。

结论

本研究深入探讨了人工智能赋能下的网络安全威胁检测与应对策略。通过综述人工智能在网络安全威胁检测中的应用，分析了新型网络安全威胁，并提出了相应的应对策略。研究发现，人工智能能够显著提升网络安全威胁检测的效率和准确性，但同时也带来了新的安全风险。构建的人工智能赋能的网络安全威胁检测与应对框架为企业和组织提供了有效的参考。未来，随着技术的不断发展，人工智能在网络安全领域将发挥更加重要的作用，同时也需要不断应对新的挑战和问题。企业和组织应积极采用人工智能技术，加强网络安全防护，实现更加安全可靠的数字化发展。

参考文献

[1]方滨兴，时金桥，王忠儒，等.人工智能赋能网络攻击的安全威胁及应对策略[J].中国工程科学，2021，23（3）：7.

[2]杨红梅，吴红红，戴爽.人工智能赋能网络安全新范式的研究[J].通信世界，2024（13）：30-32.

[3]赛博研究院.人工智能技术与网络空间安全[J].信息安全与通信保密，2019，000（006）：21-26.

[4]王玉立，张一鸣.人工智能技术加强网络安全建设的研究[J].网络空间安全，2021（Z5）：73-78.

[5]曹洪军，孔晶晶.人工智能赋能网络意识形态安全的逻辑探赜[J].理论月刊，2023（9）：46-53.