浅谈人工智能时代计算机信息安全与防护

刘杰华
广西壮族自治区水利电力勘测设计研究院有限责任公司　广西　南宁　530023

摘　要：本文聚焦人工智能时代计算机信息安全面临的复杂威胁与防护策略。通过分析技术漏洞、应用场景及系统架构层面的安全风险，探讨同态加密、零信任架构、对抗性训练等核心技术防护手段，并提出跨学科融合、技术演进趋势及能力建设路径等关键提升方向，为构建智能化时代的信息安全体系提供理论支撑与实践指导。

关键词：人工智能；信息安全；防护策略；技术演进；能力建设

引言：随着人工智能技术的飞速发展，计算机信息已深度融入社会生活的各个角落，信息安全问题变得愈发关键。在人工智能时代，信息安全不仅关乎个人隐私保护，更与企业生存、社会稳定乃至国家安全紧密相连。新技术的应用带来了前所未有的安全挑战，从技术漏洞到应用场景风险，再到系统架构隐患，都对信息安全防护提出了更高要求。深入探讨人工智能时代计算机信息安全与防护策略具有重要的现实意义。

1 人工智能时代的信息安全威胁

1.1 技术层面的威胁

在人工智能时代，技术层面的安全威胁愈发复杂多样。算法漏洞中，模型偏见与对抗性攻击问题显著。模型偏见会致使决策结果不公平、不准确，影响用户体验，甚至造成重大损失。而对抗性攻击通过在输入数据中添加精心设计的微小扰动，使机器学习模型误判。比如在图像识别中，对抗样本能让图片被误分类，严重威胁依赖视觉识别的应用，如自动驾驶汽车。数据泄露风险同样严峻。大规模数据集的存储与传输安全防护难度大，无论是云端还是本地，一旦受攻击，敏感信息，如个人隐私、商业机密等可能被非法获取，这不仅损害用户信任，还可能引发法律纠纷与经济损失。自动化攻击也是隐患。借助AI技术，恶意软件能自我进化以突破传统安全防线。网络钓鱼攻击因自然语言处理能力变得更精准高效，能诱骗用户点击恶意链接或提供敏感信息[1]。分布式拒绝服务（DDoS）攻击借助AI实现智能化调度，可发动更大规模、更具针对性的流量冲击，瘫痪目标网站或服务。传统安全策略已难以应对这些新型威胁，亟待探索先进防护机制。

1.2 应用场景的威胁

人工智能技术在各个应用场景中的广泛应用，也带来了诸多安全风险。智能系统入侵成为新的威胁焦点。智能家居系统中，智能门锁、摄像头等设备可能因安全漏洞被攻击者控制，导致家庭隐私泄露、财产损失。自动驾驶汽车面临着复杂的网络攻击风险，攻击者可能入侵车辆控制系统，篡改行驶指令，引发交通事故。医疗AI系统一旦被攻击，可能导致医疗诊断错误、治疗延误，危及患者生命安全。深度伪造技术的兴起，对身份认证体系造成了巨大冲击。AI生成内容，如AI换脸、语音伪造等技术，能够以假乱真，使得传统的基于生物特征的身份认证方式面临挑战。攻击者可以利用深度伪造技术制作虚假视频、音频，冒充他人进行诈骗、窃取机密信息等违法活动，给个人、企业和社会带来严重损失。

1.3 系统架构的威胁

边缘计算与物联网设备的低安全性，是人工智能时代信息安全的重要挑战之一。边缘计算设备通常部署在资源受限的环境中，安全防护能力较弱，容易成为攻击者的突破口。物联网设备数量庞大、种类繁多，且安全标准不统一，大量设备存在默认密码、未及时更新补丁等安全问题，攻击者可以利用这些漏洞入侵设备，进而控制整个物联网网络，获取敏感数据或发起大规模攻击。云计算与混合云环境中的数据主权与隐私风险也不容小觑。在云计算环境下，用户数据存储在云服务提供商的服务器上，数据的所有权、控制权和使用权面临新的界定问题。不同国家和地区的法律法规存在差异，可能导致数据在跨境流动过程中面临合规风险。混合云环境中数据在不同云平台之间的迁移和共享，也增加了数据泄露和被滥用的风险。人工智能时代的信息安全威胁呈现出多样化、复杂化的特点，需要我们从技术、应用场景和系统架构等多个层面进行全面防范和应对。

2 信息安全防护的核心策略

2.1 技术防护体系

在人工智能时代，保护数据的安全性至关重要。同态加密是一种允许直接对加密数据进行计算而不需解密的技术，这使得即使在处理过程中数据也能保持加密状

态，极大地增强了隐私保护能力。差分隐私则是通过向查询结果中添加噪声来确保个体信息不被泄露，仍然能够从整体数据集中提取有价值的信息[2]。这两种技术在AI数据保护中的应用，不仅提高了数据的安全性，还为合法的数据使用提供了保障。多因素认证（MFA）作为一种增强的安全措施，要求用户提供多种验证方式，如密码、生物特征或一次性验证码，从而增加了未经授权访问的难度。零信任架构（ZTA）则假设网络内部和外部都存在潜在威胁，并要求对每次访问请求进行严格验证，无论其来源如何。这种架构打破了传统的基于边界的防御模式，转而关注于用户身份和设备的持续验证，显著提升了系统的安全性。对抗性训练是提高机器学习模型鲁棒性的关键技术之一。通过对模型进行针对性攻击训练，使其学会识别并抵御对抗样本，从而增强其在实际应用中的可靠性。采用诸如正则化等方法来限制模型复杂度，可以有效减少过拟合现象，进一步提升模型的稳定性。这些技术共同作用，构建了一个更加坚固的防护屏障，以应对不断演变的安全威胁。

2.2 架构设计优化

为了有效应对复杂的网络安全挑战，建立一个从网络层到应用层的立体化防护体系显得尤为重要。在网络层，防火墙、入侵检测系统（IDS）等基础安全设施能够过滤恶意流量，阻止非法访问。而在应用层，则需要部署专门的安全软件，如Web应用防火墙（WAF）、反病毒程序等，以保护应用程序免受各种类型的攻击。这种多层次的防御策略形成了一个全方位的安全网，大大降低了被攻破的风险。区块链技术凭借其不可篡改性和透明性，在数据完整性和可信性方面展现出巨大潜力。通过将关键交易记录存储在分布式账本上，任何试图篡改数据的行为都会被立即发现并记录下来，确保了数据的真实性和一致性。利用智能合约可以自动化执行预设规则，简化业务流程的同时也减少了人为干预带来的风险。去中心化架构不仅是提升信息安全的有效手段，也为构建更高效、可靠的数字生态系统奠定了基础。面对日益复杂且难以预测的安全威胁，采用弹性安全设计变得尤为必要。这意味着不仅要具备强大的防御能力，还需拥有快速响应和恢复的能力。动态调整防护策略，根据最新的威胁情报及时更新防御措施，可以有效应对未知威胁。例如，当检测到异常行为时，系统能够自动采取隔离措施，并通知相关人员进行进一步分析和处理。这样既保证了系统的连续运行，又最大限度地减少了损失。

2.3 人工智能的双向利用

借助AI技术，可以开发出一系列先进的安全工具，用于威胁检测、异常行为分析及自动化响应。AI能够实时监控网络活动，识别潜在的安全威胁，并迅速做出反应。例如，基于机器学习的入侵检测系统能够学习正常行为模式，并据此判断是否存在异常活动。一旦发现可疑行为，系统会立即触发警报并启动相应的防护措施。这种方式不仅提高了威胁检测的速度和准确性，还能减轻安全人员的工作负担。尽管AI在许多领域表现出色，但在信息安全领域，人机协同仍然是不可或缺的。安全运营中心（SOC）通常由专业的安全分析师组成，他们负责监控和管理企业内部的安全状况[3]。通过引入AI技术，可以显著增强SOC的功能。例如，AI可以帮助分析师筛选海量的日志文件，找出值得关注的关键事件；而分析师则可以根据自己的经验和专业知识，对AI提出的建议进行深入分析和决策。这种人机协作模式充分发挥了双方的优势，提高了整体的安全防护水平。

3 防护能力提升的关键方向

3.1 跨学科融合

密码学、机器学习与系统工程的协同创新，为防护能力提升奠定坚实基础。密码学作为信息安全的基石，提供加密、解密等核心技术，保障数据传输与存储的保密性。机器学习则凭借强大的数据处理与模式识别能力，能对大量安全相关数据进行分析，发现潜在安全威胁模式。系统工程从整体架构层面，将密码学与机器学习有机整合到各类系统中。例如在网络安全防护系统中，利用密码学对数据进行加密传输，借助机器学习算法实时监测网络流量，分析是否存在异常行为。当机器学习检测到疑似攻击行为时，系统工程所构建的架构能迅速调配资源，运用密码学手段加强关键数据保护，阻止攻击进一步蔓延。这种协同创新，打破了学科间的壁垒，让不同领域技术相互赋能，显著提升系统整体防护能力。安全工程与AI伦理的交叉研究同样至关重要。安全工程致力于构建安全可靠的系统，防范各类安全风险。而AI伦理关注人工智能技术应用中的道德与伦理问题。在防护能力提升过程中，两者交叉融合。比如在设计基于人工智能的安全监控系统时，安全工程确保系统具备强大的检测与防御能力，而AI伦理则要求系统在数据收集、处理和决策过程中遵循道德准则。不能因追求检测效果而过度收集用户隐私数据，要保障用户数据的合理使用与安全。通过这种交叉研究，使得防护系统不仅技术过硬，而且在伦理层面也无懈可击，避免因伦理问题引发信任危机，确保防护能力的可持续提升。

3.2 技术演进趋势

在隐私计算领域，联邦学习与多方安全计算

（MPC）的落地，为数据安全与隐私保护开辟新路径。联邦学习使多个参与方在不共享原始数据的情况下联合训练模型。以金融领域为例，多家银行可借助联邦学习共同训练信用风险评估模型，各方数据保留本地，既规避数据泄露风险，又能提升模型准确性与泛化能力，强化金融风险防护。多方安全计算（MPC）在分布式计算环境中保障计算时的数据安全性。在供应链金融场景里，涉及众多企业的计算，MPC确保各企业数据保密性，安全完成复杂运算，提升供应链金融安全防护水平，推动业务稳健发展。量子安全方面，后量子密码学（PQC）的预研与部署迫在眉睫。随着量子计算技术发展，传统密码学面临被破解危机。许多国家和企业已投入大量资源预研后量子密码学，探寻将后量子密码算法融入现有网络安全体系的方法。在关键信息基础设施中，提前部署后量子密码技术，替换易受量子攻击的传统密码算法，可保障基础设施网络安全，防止因量子计算威胁致使关键信息泄露，维护国家和企业信息安全[4]。AI安全评估中，模型可解释性、透明度与责任归属成为重要趋势。模型可解释性助力人们理解人工智能模型的决策过程，当模型做出安全相关决策，如检测到网络攻击并实施防御时，可解释性便于判断决策合理性。透明度要求模型的数据来源、训练过程公开透明，便于监管与审计。责任归属则明确模型出现安全问题时的责任主体。在自动驾驶汽车安全防护中，若因人工智能模型误判引发交通事故，清晰的责任归属机制能确保责任被追究，促使企业持续完善模型，提升安全防护能力。

3.3 能力建设路径

安全人才培养是提升防护能力的核心。构建复合型AI安全专家的培养体系迫在眉睫。这类专家不仅要掌握传统安全知识，如网络安全、数据安全等，还要精通人工智能技术原理。在高校教育中，设置跨学科课程，融合计算机科学、信息安全与人工智能等专业知识。企业也应加强内部培训，通过实际项目锻炼，让员工在实践中积累经验。例如开展网络安全攻防演练，模拟真实攻击场景，让安全人才在实战中提升应对能力，培养出既懂安全技术又能运用人工智能优化防护策略的复合型人才。技术标准化对于防护能力提升不可或缺。制定安全协议、接口规范与互操作性标准，能确保不同安全系统之间的协同工作。在物联网安全防护中，统一的安全协议可保障不同厂家的物联网设备在数据传输、认证等方面的安全性与兼容性。接口规范让安全设备与系统之间的对接更加顺畅，提高整体防护效率。互操作性标准使不同安全产品能相互配合，形成强大的防护合力，避免因标准不统一导致的安全漏洞与防护短板。持续监控与响应能力建设同样关键。建立实时威胁情报系统，通过多种渠道收集安全威胁信息，及时发现新的攻击手段与安全漏洞。在网络安全领域，实时监测网络流量，一旦发现异常，迅速启动自动化修复机制。例如利用自动化脚本，及时更新系统补丁，封堵安全漏洞，防止攻击进一步扩大。通过持续监控与快速响应，能够在安全威胁发生的第一时间采取有效措施，将损失降到最低，切实提升防护能力，保障系统安全稳定运行。

结束语

人工智能时代计算机信息安全与防护面临诸多挑战，但也蕴含着巨大的发展机遇。通过深入研究信息安全威胁，采取有效的防护策略，并积极探索防护能力提升的关键方向，能够更好地应对复杂多变的安全形势，保障人工智能技术的健康、可持续发展，为社会的进步与繁荣提供坚实的安全保障。

参考文献

[1]魏敏.基于人工智能的计算机网络信息安全防护模式研究[J].信息记录材料,2024,25(11):130-132.

[2]王鲲.人工智能技术下计算机信息安全与防护探讨[J].信息与电脑(理论版),2024,36(16):183-185.

[3]安仲源.基于人工智能的计算机信息安全与防护研究[J].信息记录材料,2024,25(06):161-163.

[4]张慧珍.人工智能时代计算机信息安全与防护策略分析[J].中国新通信,2023,25(15):107-109.