人工智能时代下网络信息安全问题和防护对策

华　泽
临沧市人民政府办公室　云南　临沧　677000

摘　要：随着人工智能（AI）技术的迅猛发展，网络信息安全面临着前所未有的挑战。本文聚焦于人工智能时代下网络信息安全问题及其防护对策。在问题方面，涵盖算法安全、数据安全以及新型网络攻击形式等。针对这些问题，防护对策包括技术层面、法律法规与监管以及人员与意识培养等多方面措施。综合来看，保障人工智能时代的网络信息安全需要多管齐下，协同应对不断演变的安全威胁。

关键词：人工智能；网络信息安全；算法安全；数据安全；防护对策

引言：在当今时代，人工智能（AI）正以前所未有的速度迅猛发展并深度融入社会的各个层面。在医疗领域，AI技术被广泛应用于疾病诊断，例如通过深度学习算法对海量医疗影像进行分析，辅助医生精准判断病情；金融行业中，智能算法用于风险评估、欺诈检测等，能够快速处理大量交易数据并做出决策；交通方面，自动驾驶技术借助AI实现车辆的智能导航与交通流量优化[1]。这些应用场景的背后，网络信息处于核心地位[2-3]。无论是医疗数据的传输与存储、金融交易信息的处理，还是交通系统中的车辆和路况数据交互，都依赖网络信息的高效、准确传递。然而，随着人工智能与网络的深度融合，网络信息安全面临着全新的严峻挑战。对于个人而言，个人隐私信息可能因安全漏洞而泄露，遭受诈骗等风险；企业方面，商业机密、客户数据等一旦被侵犯，会导致巨大的经济损失和声誉损害；从国家层面看，关键基础设施相关的网络信息安全受到威胁，可能影响国家安全、社会稳定和经济发展[4]。因此，在人工智能时代下，网络信息安全问题的重要性愈发凸显，亟待深入研究与应对。

1 人工智能时代下网络信息安全问题

1.1 算法安全问题

1.1.1 算法偏见

算法偏见的产生根源在于多种因素。一方面，训练数据的不均衡是一个关键因素。当训练数据不能全面、均衡地代表所有相关群体时，人工智能算法就容易产生偏见。例如，在招聘算法中，如果训练数据中男性的比例过高，这种数据的不均衡会使算法学习到的模式偏向男性群体。在筛选简历时，可能会对女性候选人存在不公平对待，如不合理地降低女性候选人的评分，尽管她们可能具备与男性候选人相当甚至更优秀的能力。此外，算法设计过程中的先验假设也可能导致偏见。如果算法开发者在设计算法时存在某些潜意识的偏见，并将其融入算法结构中，也会使算法在决策过程中对某些群体产生不公平的结果[5]。

1.1.2 算法可解释性低

低可解释性给算法的安全性带来诸多风险。在许多人工智能算法中，尤其是深度神经网络等复杂模型，其内部决策过程如同一个“黑箱”。由于难以确定算法决策过程中的错误来源，这给安全保障带来了巨大挑战。以医疗诊断中的人工智能算法为例，在医疗场景中，准确性和可靠性至关重要。若算法无法解释其诊断依据，当出现错误诊断时，医疗人员很难追溯到问题所在。这可能导致错误的治疗方案被采用，对患者的健康造成严重威胁[6]。

1.2 数据安全问题

1.2.1 数据泄露风险

在人工智能系统中，数据存储和传输过程充满了泄露风险。从数据存储角度看，数据库是黑客攻击的重要目标。黑客可能利用数据库系统的漏洞，如SQL注入漏洞等，非法获取存储在数据库中的数据。对于人工智能系统而言，这些数据往往包含大量的敏感信息，如用户的个人信息、企业的商业机密等。在数据传输过程中，数据可能在网络传输过程中被窃取。例如，在未采用足够加密措施的网络环境下，数据以明文形式传输时，攻击者可以通过网络嗅探等手段截获数据。

1.2.2 数据滥用

数据被恶意使用的情况日益严重。其中，未经授权的数据挖掘和分析是常见的数据滥用形式。在一些情况下，数据收集者可能在用户未明确同意的情况下，对收集到的数据进行深度挖掘和分析，以获取更多有价值的信息。此外，数据滥用还可能涉及对用户数据的重新组合和利用，以达到一些不良目的，如政治操纵等。

1.3 网络攻击新形式

1.3.1 对抗性攻击

对抗性攻击是一种针对人工智能模型的新型攻击方式。攻击者通过对输入数据进行微小修改来误导人工智能模型。这种微小的修改对于人类视觉或其他感知系统来说可能几乎不可察觉，但却能使人工智能模型产生错误的判断。对抗性攻击的存在表明人工智能模型在面对恶意攻击时的脆弱性，即使是在正常工作状态下准确率很高的模型，也可能在遭受对抗性攻击时产生严重的错误。

1.3.2 人工智能赋能的网络攻击

随着人工智能技术的发展，恶意行为者开始利用人工智能技术进行网络攻击。其中，自动化的网络钓鱼攻击是一种典型的方式。黑客可以利用人工智能生成高度逼真的钓鱼邮件。这些邮件在内容、格式、语言风格等方面都与正常邮件非常相似，甚至能够根据目标用户的特点进行个性化定制。通过这种方式，黑客可以提高攻击成功率，欺骗用户点击邮件中的恶意链接或提供敏感信息。此外，人工智能还可以被用于自动化地扫描网络漏洞，然后根据漏洞信息制定更有效的攻击策略，这使得网络攻击更加智能化、高效化，给网络安全防御带来了巨大的挑战。

2 人工智能时代下网络信息安全防护对策

2.1 技术层面的防护

2.1.1 加密技术

加密技术在保护数据安全方面起着至关重要的作用。对称加密算法是一种加密和解密使用相同密钥的加密方式。其优点在于加密和解密速度快，适用于大量数据的加密。在数据存储和传输过程中，只有拥有正确密钥的接收方才能将密文还原为明文。非对称加密算法则使用一对密钥，即公钥和私钥。公钥用于加密，私钥用于解密。这种方式解决了对称加密中密钥分发的难题。例如，在数字签名场景中，发送者使用自己的私钥对消息进行签名，接收者使用发送者的公钥来验证签名的真实性。

2.1.2 人工智能安全技术

专门用于保障人工智能算法安全的技术不断发展，其中对抗样本检测技术尤为重要。对抗样本是指那些经过精心设计、对输入数据进行微小修改后，能够使人工智能模型产生错误输出的样本。对抗样本检测技术旨在识别这些恶意的对抗样本。通过检测对抗样本来防止人工智能模型被误导的原理在于，检测技术可以分析输入数据的特征分布。正常的数据样本在特征空间中具有一定的分布规律，而对抗样本往往会偏离这种正常分布。例如，在图像识别的人工智能模型中，正常的图像数据在颜色、纹理等特征维度上有特定的分布模式。当对抗样本出现时，它可能会在这些特征维度上出现异常值或者异常的分布变化。检测技术可以通过构建特征模型或者使用统计方法来识别这些异常，一旦检测到对抗样本，就可以拒绝该输入或者对其进行修正，从而防止人工智能模型基于错误的输入做出错误的决策。

2.1.3 入侵检测与防御系统

入侵检测和防御系统（IDPS）的工作原理主要包括基于特征的检测和基于行为的检测。基于特征的检测是指系统预先定义了一系列已知攻击的特征模式，例如特定的恶意代码片段、网络攻击数据包的格式等。当网络流量或者系统活动中出现与这些特征模式匹配的情况时，系统就判定为可能存在入侵行为。基于行为的检测则是通过建立正常系统行为的模型，然后监测系统的实际行为与正常行为模型的偏差。例如，在企业网络环境中，入侵检测系统可以实时监测网络流量。当有黑客试图通过SQL注入攻击企业的数据库时，SQL注入攻击的数据包具有特定的格式和内容特征，如包含恶意的SQL语句构造。基于特征的入侵检测系统能够识别这些特征，及时发现这种攻击行为，并通知防火墙或者其他防御机制阻止该攻击流量进入企业内部网络，从而保护企业的数据库安全。

2.2 法律法规与监管

2.2.1 相关法律法规

在中国，《中华人民共和国网络安全法》等法律法规也对网络信息安全进行了全面规范。《中华人民共和国网络安全法》明确了网络运营者的安全义务，包括保障网络安全、保护用户信息等。在人工智能领域，这意味着开发和运营人工智能系统的企业需要确保其算法不会侵犯用户的权益，数据的使用符合法律法规要求[7]。这些法律法规为人工智能时代的网络信息安全提供了基本的法律框架，促使企业和组织在开发和应用人工智能技术时重视信息安全问题。

2.2.2 监管机制

监管机构对人工智能相关的网络信息安全监管主要体现在多个方面，尤其是对数据使用的监管。监管机构会制定数据使用的规范和标准，要求企业在使用数据时遵循合法、合规、安全的原则。例如，在数据收集方面，监管机构会要求企业明确告知用户数据收集的目的、范围和使用方式，并获得用户的同意。以某行业监管机构对人工智能企业的数据合规性检查为例，监管机构会检查企业的数据来源是否合法，是否存在未经授权

收集数据的情况。在数据存储方面，会检查企业是否采取了足够的安全措施，如数据加密、访问控制等。在数据使用过程中，是否按照预先声明的目的使用数据，是否存在数据滥用的情况。通过这种监管，可以促使人工智能企业建立健全的数据安全管理体系，保障用户的数据安全和权益，避免因数据安全问题引发的各种风险，如用户隐私泄露、数据被恶意利用等。

2.2.3 人员培养

（1）专业人才培养

培养网络信息安全专业人才，尤其是掌握人工智能和安全技术的复合型人才是非常必要的。随着人工智能技术在网络信息安全领域的广泛应用，单纯掌握传统安全技术或者人工智能技术的人才已经难以满足需求。例如，在对抗人工智能赋能的网络攻击时，需要既了解人工智能算法漏洞又能运用安全技术进行防御的人才[8]。高校在人才培养方面可以开设相关的交叉学科课程。例如，将计算机科学、人工智能、密码学、网络安全等课程有机结合。可以设置专门的实验室，让学生在实践中掌握人工智能安全技术的应用。企业也可以通过内部培训和与高校合作的方式培养人才。企业可以根据自身的业务需求，为员工提供定制化的培训课程，如针对企业特定的人工智能安全需求，培训员工掌握对抗样本检测技术等。同时，企业可以与高校建立实习基地，为学生提供实践机会，也为企业选拔优秀的复合型人才。

（2）安全意识教育

提高普通用户和企业员工的安全意识具有重要意义。在普通用户层面，例如防范网络钓鱼攻击的意识非常关键。网络钓鱼攻击往往利用用户的疏忽大意，通过伪装成合法的网站或者邮件来骗取用户的敏感信息。如果用户缺乏安全意识，很容易上当受骗。在企业员工层面，员工的安全意识薄弱可能会导致企业遭受严重的安全威胁。为了提升安全意识，可以通过多种宣传和培训方式。在宣传方面，可以制作安全知识宣传海报、视频等，通过社交媒体、企业内部平台等渠道进行传播。例如，可以制作关于如何识别网络钓鱼邮件的短视频，向普通用户和企业员工进行宣传。在培训方面，可以定期组织安全培训课程，针对不同的用户群体和企业部门设置不同的培训内容。例如，对企业的财务部门员工重点培训防范财务诈骗相关的安全知识，对普通用户可以重点培训如何设置安全的密码、如何识别安全的网站等知识。

3 结束语

人工智能时代网络信息安全问题有新特点，算法安全方面对抗样本攻击和模型误导问题常见，数据安全随大规模数据处理变得复杂，新攻击形式挑战传统防御手段。防护对策综合作用显著，技术提供基础保障，法律法规规范使用，人员意识培养提升防范能力，三者相互补充构建防护体系。在未来发展中，人工智能发展使网络安全面临更多挑战，如入侵检测、对抗样本检测技术需更智能先进，法律法规监管要更严格，且量子计算等会威胁传统加密技术，防护对策需不断优化完善。

参考文献

[1]黄健.人工智能时代下网络信息安全问题和防护策略[J].网络安全技术与应用,2024(7):124-126.

[2]赵薇.大数据时代人工智能在大学校园网络安全技术中的运用[J].软件,2024,45(1):158-160.

[3]董理君,刘超,张锋,等.大数据与人工智能时代背景下的网络安全课程体系研究[J].软件导刊,2024,23(8):37-42.

[4]钱建波,于正永,朱重龙.新时代背景下高校网络和信息安全分析研究[J].网络安全技术与应用,2022(11):64-65.

[5]陆芸.人工智能时代计算机信息技术安全与防护策略探讨[J].长江信息通信,2023,36(3):175-177.

[6]侯艳玲,谢兴昶,洪之坤,等.数字经济时代网络信息安全基本现状与发展趋势[J].山东工业技术,2023(3):60-64.

[7]代燕妮.大数据时代背景下人工智能在信息安全技术中的应用研究[J].长江信息通信,2023,36(7):145-147.

[8]张崇芳,卢斐.人工智能时代高校计算机网络信息安全问题研究[J].电脑爱好者(普及版),2021(2):50-51.