 DOI: https://doi.org/10.12345/bdai.v6i1.23377

Research on Security Risk Prevention of Big Data and Artificial Intelligence

Kai Mao

College of Artificial Intelligence, Chongqing Technology and Business University, Chongqing, 400067, China

Abstract

With the rapid development of big data and artificial intelligence technologies, they play a huge role in improving productivity and innovation capabilities. However, the widespread application of these technologies is accompanied by increasing security risks, especially in the areas of data leakage, privacy infringement, algorithm bias, and other challenges that are becoming increasingly severe. In order to effectively address these potential threats, this article conducts an in-depth analysis of the security risks of big data and artificial intelligence, and proposes corresponding prevention strategies. By exploring technical means such as data encryption, algorithm transparency, and security monitoring, this article hopes to provide theoretical support and practical guidance for security protection in related fields, and strive to provide guarantees for the secure development of technology.

Keywords

big data; artificial intelligence; Security risks; Preventive strategies

大数据人工智能的安全风险防范研究

毛凯

重庆工商大学人工智能学院，中国·重庆 400067

摘　要

随着大数据和人工智能技术的快速发展，其在提升生产力与创新能力方面发挥着巨大作用，然而，这些技术的广泛应用伴随着越来越多的安全风险，特别是在数据泄露、隐私侵犯、算法偏差等方面的挑战日益严峻。为了有效应对这些潜在威胁，本文围绕大数据与人工智能的安全风险进行深入分析，并提出了相应的防范策略。通过探讨数据加密、算法透明化、安全监控等技术手段，文章希望为相关领域的安全防护提供理论支持和实践指导，力求为科技的安全发展提供保障。

关键词

大数据；人工智能；安全风险；防范策略

1 引言

在数字化转型的浪潮中，大数据和人工智能技术以其强大的数据处理能力和智能化分析优势，逐渐渗透到各个行业与领域，从金融、医疗到制造业，它们为各行各业带来了前所未有的生产力提升和商业价值。然而，伴随技术革新的同时，数据泄露、隐私侵犯、算法偏差等安全问题，会影响个人用户的隐私保护，也对整个社会的信任基础和经济秩序构成了威胁。因此，如何有效防范大数据与人工智能的安全风险，已经成为当前科技发展中亟待解决的关键问题。

2 大数据与人工智能的安全风险概述

随着大数据的积累和人工智能技术的快速发展，海量的数据资源和高度复杂的算法系统对信息安全提出了前所未有的挑战，尤其是在数据传输、存储和分析过程中，信息的安全性、隐私的保护以及系统的健壮性面临着严重威胁。当前，大数据和人工智能的安全风险多表现为隐私泄露、算法漏洞以及系统攻击等多重形式，在这些安全威胁中，最为突出的便是数据泄露问题，无论是通过网络攻击、内部人员的恶意行为，还是通过数据采集的技术漏洞，用户的隐私信息和敏感数据都可能在未经授权的情况下被恶意获取 [1]。大数据的开放性和普及性，尤其在云计算环境中，进一步加大了数据被窃取的风险，在人工智能领域，训练模型所依赖的海量数据中，包含了大量的个人信息和商业秘密，一旦这些数据被不当利用或外泄，极有可能引发严重的社会和经济后果。

另一方面，人工智能系统的学习过程需要大量的数据进行训练，但这些数据本身可能存在不完备、不准确或者有偏见的情况，算法模型的偏差问题，尤其是在数据样本选择上不够全面或存在刻意偏向，可能导致算法预测结果的失真，影响到决策的合理性和公正性。例如，在金融领域，若

【作者简介】毛凯（1968–），男，中国重庆人，硕士，副教授，从事计算机网络技术及应用研究。

人工智能被用于信用评分，模型可能基于历史数据的偏见作出错误判断，导致不公正的贷款决策。

3 大数据与人工智能的安全风险主要类型

3.1 数据泄露与隐私侵犯风险

在大数据与人工智能的高度发展过程中，数据泄露与隐私侵犯问题已成为日益严峻的挑战，这一问题源于黑客攻击、技术漏洞或系统安全缺陷，而在更深层次上，它涉及到数据处理、存储和共享的基本方式，尤其是大数据环境中，各种数据被广泛收集并交由不同的服务提供商处理，用户的个人信息被匿名化、脱敏化处理的过程中，极容易被反向推断、重识别，导致泄露个体隐私，这类隐私侵犯事件发生在数据流通环节，同时也在数据分析和利用过程中产生。

此外，许多人工智能应用在设计之初，未能充分考虑数据保护的长远问题，更多地专注于技术创新和商业化应用，这使得某些数据采集与处理方法在初期看似合规，却忽视了日后可能引发的隐私风险 [2]。例如，在智能推荐算法的背后，收集的数据被用于构建精确的用户画像，这些画像是用户的兴趣爱好，还涉及到用户的生活习惯、行为模式，甚至是潜在的心理特征，数据收集时所获取的广泛信息，虽然表面上看无害，但一旦暴露给不法分子，便可被滥用，进而产生身份盗窃、网络欺诈、甚至是深度伪造等更加严重的问题，在这个过程中，隐私侵犯的边界正变得越来越模糊，技术的使用者通常低估了无形中对个人隐私权的侵害，甚至忽略了用户知情权和选择权的根本问题。

3.2 算法偏差与误导风险

人工智能算法本质上依赖于大量数据进行训练，但在此过程中，数据的不平衡和偏差会导致系统性能的不公正性与误导性结果，算法偏差最为常见的形式，便是由于训练数据的历史性偏见所带来的结果。例如，人工智能在招聘、贷款评估等领域的应用，通常是基于大量的历史数据进行学习，但这些数据本身可能已经体现了社会中的性别、种族或地域偏见，若这些偏见未能在训练过程中被有效识别和消除，算法便会在应用时继续沿用这些偏差，从而产生不公正的决策。

此外，人工智能尤其是深度学习算法，由于其庞大的网络结构和层次性，使得其决策过程缺乏足够的透明度与可解释性，当一个算法作出决策时，特别是在医疗、司法等关键领域，难以追溯其背后的逻辑与依据，这使得即便算法输出了不合理的或偏见明显的结论，也很难被及时发现与纠正。更重要的是，算法偏差的积累通常会在实际应用中逐渐扩大，这种隐形的、系统性的偏见会影响到当前的决策，还可能在长时间内持续存在，甚至被新的数据所加强，形成一种自我强化的负面循环。

3.3 系统漏洞与攻击风险

随着大数据和人工智能技术在各行业的广泛应用，系统漏洞与攻击风险已经成为了最严峻的安全隐患，这类风险是针对系统的外部攻击，还包括由内而外的安全薄弱环节，如人工智能系统中通常包含多个交互层和复杂的模块，这些模块如果设计不当或漏洞未及时修补，便可能成为攻击者的目标。攻击者能够通过针对性的恶意代码、数据篡改或利用算法中的安全漏洞，对系统进行破坏或操控，尤其在深度学习模型中，由于其“黑箱”性质，攻击者可以利用输入数据的微小变化，悄无声息地欺骗系统，做出错误判断或决策。

另外，在多方合作、数据共享的背景下，数据的完整性和源可信度通常被忽视，攻击者会通过篡改训练数据、伪造数据源或运用“数据中毒”手段，逐步渗透到模型训练的核心，对于人工智能系统而言，训练数据的质量直接决定了系统的可靠性和精准性，一旦数据源遭到篡改，模型的输出将完全失去可信性，甚至可能引发更加严重的后果。例如，在金融行业中，若 AI 模型训练使用了被恶意操控的交易数据，可能导致市场预测的偏差，进而引发广泛的金融动荡。同样，在医疗领域，若训练数据中被故意加入了错误的医疗记录，AI 可能错误地诊断患者病情，甚至威胁到生命安全。

4 大数据与人工智能安全风险防范策略

4.1 数据加密与访问控制技术

在面对大数据和人工智能的安全挑战时，数据加密技术通过将数据转化为无法直接解读的密文，使得即便数据被盗取或非法访问，也难以被利用或滥用，其核心在于密钥管理和加密算法的强度，这直接决定了加密系统的安全性。随着数据量的急剧增加和处理方式的多样化，传统的加密方法已逐步无法应对日益复杂的安全威胁，因此，越来越多的组织开始采用更为先进的加密技术，如同态加密和量子加密等，这些技术能够在加密数据的同时，依然允许对加密数据进行某些计算操作，这为保护隐私数据提供了更大的灵活性和安全性。

此外，传统的访问控制通常依赖于用户身份认证，但随着人工智能系统的普及，传统的基于角色的访问控制（RBAC）已不足以应对复杂的安全需求，在这种情况下，基于属性的访问控制（ABAC）和基于身份的访问控制（IBAC）逐渐成为主流方法。ABAC 允许通过更加细粒度的策略来定义访问权限，基于用户属性、环境条件、请求上下文等因素动态决定数据访问权限，而 IBAC 通过对用户身份进行严格验证，并结合多因素认证技术，确保只有经过授权的人员才能访问敏感数据，这种多层次、多维度的访问控制策略能够有效防止数据泄露事件的发生，尤其在跨部门、跨机构的数据共享环境下，能保障数据安全性和合规性。

4.2 人工智能算法的透明化与可解释性

为了确保人工智能系统的安全性和信任度，算法的透明化与可解释性正逐渐成为必要的技术策略，这种透明化是在模型输出层面上展示结果，更重要的是能够揭示算法背后

的决策过程，使其被开发者、用户甚至监管机构理解和审查。在深度学习模型中，复杂的层次结构和庞大的参数空间常常让其成为“黑箱”，外界难以洞察其内部的运作机制，为此，近年来涌现了如 LIME（局部可解释模型 - 不可知模型）和 SHAP（Shapley 加权归因方法）等技术，这些方法能够提供局部的可解释性分析，通过可视化或简化复杂决策过程来揭示模型在做出特定决策时所依据的特征或参数，这能够增强用户对人工智能决策的理解，还能够在出现不合理决策时，追溯其原因并进行优化。

为了进一步提高算法的可解释性，还需要对算法设计本身进行优化，选择那些本身具备较高透明度的模型架构，例如，决策树、线性回归等相对简单的模型，比起复杂的神经网络，更容易提供可解释性。层次化的可解释性策略也在逐渐得到应用，这意味着在系统的不同层面上实现逐层可解释，保证从输入到输出的整个流程都能够被清晰地追溯 [3]。对于深度学习等“黑箱”模型的应用，研究人员也在不断探索“可解释性框架”，例如通过引入可视化工具，帮助开发者更好地理解神经网络内部的特征提取和决策过程。

4.3 安全检测与监控机制

为了有效应对大数据与人工智能带来的安全威胁，建立全面的安全检测与监控机制显得尤为关键，这一机制的核心在于能够实时监测数据流动与处理过程中的异常行为，及时发现潜在的安全风险。监控不局限于对数据访问的监控，更应延伸至模型的训练过程、算法的运行状态以及系统的整体安全态势，如在人工智能系统中，数据流的实时监控能够帮助识别任何未经授权的数据访问或篡改行为，这对防范数据泄露、篡改以及外部攻击至关重要。与此同时，通过引入智能化的安全检测工具，可以基于行为分析和异常模式识别，主动发现潜在的安全威胁，此类工具能够在系统运行的各个阶段不断地跟踪与分析数据行为模式，捕捉到那些可能被人工智能攻击者利用的漏洞或缺陷。

随着人工智能技术和数据应用环境的不断变化，监控机制应当能够根据新的风险形势进行调整。例如，随着深度学习和神经网络的广泛应用，传统的静态监控方式已无法满足日益复杂的安全需求，此时，动态监控机制则能够通过实时学习系统的行为模式，并根据这些模式及时优化监控规则，提高对新型攻击手段的防范能力。

4.4 应急响应与恢复机制

在大数据和人工智能应用日益广泛的今天，应急响应机制的核心在于快速识别、评估和应对系统中的突发安全事件，以减少潜在的损失和影响，要实现这一目标，企业和机构需建立一套完善的事件响应流程，确保在面对数据泄露、系统攻击或算法错误等紧急情况时，能够迅速启动预设的应急方案，组织专业团队进行干预与处置，这是一个技术性的问题，也是一个跨部门、跨职能的协作过程，有效的应急响应要求技术人员具备足够的应急处置能力，还需要管理层与法律、合规部门的协作，以保证快速决策与全方位的合规检查。

在恢复机制方面，大数据和人工智能系统的恢复不只是修复问题，还涉及到数据的恢复与验证，在数据泄露或篡改的情况下，有效的恢复机制通常需要包含数据备份、实时同步和冗余设计等多个方面，确保系统在遭遇攻击或故障时，不会因数据丢失而造成无法恢复的损失。与此同时，恢复机制还需结合人工智能的自动化学习能力，实现自我修复与自我优化，一旦恢复过程开始，人工智能系统能够通过监控与反馈，逐步调整系统参数，优化恢复路径，确保系统恢复后的稳定性和安全性。

5 结语

综上所述，大数据和人工智能技术的迅猛发展为社会带来了前所未有的机遇，但其带来的安全风险也愈加严峻，在这一背景下，强化安全风险防范措施显得尤为重要。本文从大数据与人工智能的安全风险角度出发，提出了包括数据加密、算法透明化、安全监控机制等多种防范策略，以期为各行业提供更加完善的安全保障，这些技术策略的实施，能够有效提升数据的安全性，减少隐私侵犯和系统漏洞的发生，还能为行业提供更加公正、透明的决策机制。

参考文献

[1] 徐炎.大数据和人工智能时代下数据安全的风险及应对策略[J].网络安全技术与应用,2024,(12):57-60.

[2] 张微.大数据和人工智能时代数据安全风险及应对策略[J].数字通信世界,2024,(09):194-196.

[3] 刘树锋.大数据和人工智能时代下数据安全的风险及应对策略[J].网络安全技术与应用,2024,(02):54-56.