人工智能在大数据环境中的安全性挑战与对策

公帅　李海涛　张久岩　段阳帆　郭明月
中电信数智科技有限公司　100035

【摘　要】随着大数据和人工智能技术的广泛应用，数据的价值不断被挖掘，人工智能在推动社会进步的同时，也面临严峻的安全性挑战。本文从数据隐私保护、算法安全、对抗性攻击和数据管理等方面分析了人工智能在大数据环境中所面临的安全性挑战，并提出相应的对策，以期为实现人工智能安全、可靠的发展提供参考。

【关键词】人工智能；大数据；安全性挑战；隐私保护；对抗性攻击

Security challenges and countermeasures of artificial intelligence in the big data environment

Gongshuai　Li Haitao　Zhang Jiuyan　Duan Yangfan　Guo Mingyue
China Telecom Digital intelligence Technology Co.，LTD　100035

【Abstract】With the wide application of big data and artificial intelligence technology，the value of data is constantly mined. While artificial intelligence promotes social progress，it also faces severe security challenges. This paper analyzes the security challenges faced by artificial intelligence in the big data environment from the aspects of data privacy protection，algorithm security，adversarial attacks and data management，and puts forward corresponding countermeasures in order to provide reference for the safe and reliable development of artificial intelligence.

【Key words】artificial intelligence，big data，security challenges，privacy protection，adversarial attack

一、引言

在大数据时代，人工智能技术得到了迅猛发展，已经广泛应用于医疗、金融、教育、交通等各个领域。然而，人工智能的广泛应用也带来了新的安全性挑战，特别是在数据隐私保护、算法透明性和对抗性攻击等方面，人工智能面临严峻的考验。这些安全性挑战不仅影响了人工智能的实际应用效果，还可能引发严重的社会伦理和法律问题。为此，分析人工智能在大数据环境中的安全性挑战，并探索切实可行的应对策略，已成为当前研究的热点与难点。

二、人工智能在大数据环境中的安全性挑战

（一）数据隐私保护的挑战

在大数据和人工智能的驱动下，海量的个人数据被广泛收集并用于算法训练和预测分析，但这些数据中包含大量的个人隐私信息。数据隐私保护成为了人工智能应用中的核心问题。首先，人工智能模型需要大量数据进行训练，在数据共享和传输过程中，隐私泄露风险加大。即便在加密传输的情况下，数据在存储和处理过程中也可能遭遇黑客攻击，从而导致隐私信息的泄露。其次，数据共享过程中往往难以完

全匿名化，逆向推理技术的存在使得即使去标识化的数据仍可能被重识别，进一步加剧了隐私泄露的风险。因此，数据隐私保护是人工智能在大数据环境中面临的首要安全挑战。

（二）算法安全问题

算法是人工智能的核心，算法的安全性直接关系到人工智能系统的可靠性。然而，当前的人工智能算法中普遍存在算法脆弱性问题，容易受到外部干扰和操控，产生错误判断或偏差。首先，算法可能会受到数据污染的影响，黑客通过植入恶意数据，干扰算法的正常训练，从而导致模型输出错误结果。其次，算法的黑箱特性使得很多人工智能模型的决策过程难以被解释和追踪，增加了模型被恶意利用的风险。此外，算法偏见问题也是一种重要的算法安全问题，数据偏差可能导致算法输出具有偏见的结果，这不仅会影响人工智能系统的可靠性，还可能引发社会伦理问题。

（三）对抗性攻击

对抗性攻击是人工智能面临的一种特殊安全威胁。对抗性攻击指的是通过向模型输入精心设计的对抗样本，以此干扰模型的判断结果，导致模型输出错误决策。例如，在图像识别系统中，攻击者可以对图像进行微小的像素修改，从而欺骗算法做出错误分类。对抗性攻击对人工智能系统的可靠性和安全性构成了严重威胁，特别是在自动驾驶、金融风控等高风险应用场景中，对抗性攻击可能带来灾难性后果。对抗性攻击的难以察觉性和不可预测性，使得它成为人工智能在大数据环境中的一大挑战。

（四）数据管理与数据安全风险

在大数据环境中，数据量的迅速增长和数据流动的频繁性增加了数据管理的难度。数据在采集、存储、传输和处理的过程中，都可能面临数据泄露、篡改等安全风险。一方面，数据存储设施可能面临黑客攻击，数据泄露事件时有发生；另一方面，数据的流动性也带来了合规风险，尤其在跨境数据流动中，数据管理面临更加复杂的法规约束。此外，数据质量问题也是一个重要的安全挑战，低质量或不可靠的数据可能影响算法训练的精度，导致人工智能系统输出不准确的结果。因此，如何在保障数据安全的同时提高数据质量，是人工智能应用的一个重要安全性挑战。

三、人工智能在大数据环境中的安全性对策

（一）加强数据隐私保护技术的应用

为应对数据隐私保护挑战，应在人工智能系统中加强数据保护技术的应用。一方面，企业和研究机构可以采用数据加密技术，如同态加密、联邦学习和差分隐私等方法，确保数据在处理过程中不泄露隐私。同态加密可以在数据加密状态下直接进行计算，联邦学习则通过多方联合训练模型，而不直接分享数据，从而有效保护数据隐私。另一方面，差分隐私技术在数据分析过程中能够加入随机噪声，使得个体信息难以被识别，提高数据保护的隐私性。通过多种隐私保护技术的结合，人工智能系统可以在确保数据隐私的前提下进行模型训练与推理。

（二）提升算法透明性与鲁棒性

算法的透明性与鲁棒性对于提升人工智能系统的安全性至关重要。首先，应积极研究可解释性人工智能技术，通过引入可解释性算法，使得人工智能决策过程透明化，便于追踪和监管。透明化的算法可以帮助用户理解人工智能的决策逻辑，增加系统的可信度。其次，为应对数据污染和对抗性攻击，企业应采取鲁棒性增强措施，如采用对抗性训练、数据清洗和模型加固等技术。对抗性训练可以通过在模型训练中加入对抗样本，提高模型对恶意攻击的抵抗能力。数据

清洗技术则有助于清除恶意数据干扰，提高模型的训练质量。

（三）建立多层次的对抗性攻击防御机制

为应对对抗性攻击，人工智能系统需要建立多层次的防御机制。一方面，企业可以在模型训练阶段采用对抗样本生成和对抗性训练方法，以增强模型的鲁棒性，使其能够识别和防御对抗样本。另一方面，在模型部署阶段，可以引入对抗性检测模块，对输入数据进行实时检测，判断是否包含对抗样本。此外，通过组合多模型、多层次的防御架构，可以在不同模型间共享信息、互相验证，进一步增强防御效果。对于高风险场景，应进行系统化的攻击模拟和应急响应测试，确保在攻击发生时能够快速应对。

（四）优化数据管理与数据安全体系

数据管理是保障人工智能安全性的重要基础，应通过建立规范的数据管理体系提升数据安全。一方面，企业应加强数据管理的合规性，严格遵守相关法律法规，特别是对跨境数据流动的合规要求。建立数据访问权限控制和日志监控机制，可以有效防止未经授权的访问和数据泄露。另一方面，企业可以通过区块链技术实现数据溯源和数据验证，提高数据的可信度和防篡改性。此外，加强数据质量管理，通过数据清洗和数据标注等措施，确保用于模型训练的数据可靠准确，从而提升人工智能系统的整体安全性。

（五）加强安全意识与人才培养

安全性挑战不仅是技术问题，还需要全员的安全意识和专业人才支持。企业应加强人工智能安全知识培训，提高员工的安全意识，使其在数据处理和系统使用中遵循安全规范。同时，加强人工智能安全专业人才的培养，建立专业的安全团队进行持续的技术研究与安全监测。企业还应积极参与行业安全标准的制定和实践，推动形成统一的人工智能安全标准与规范，通过全行业的协同合作，共同应对安全性挑战。

四、结论

在大数据环境中，人工智能的广泛应用不仅带来了技术进步和生产力提升，同时也带来了数据隐私保护、算法安全和对抗性攻击等方面的安全性挑战。为实现人工智能的安全、稳定发展，企业应采取多层次的技术对策，包括加强数据隐私保护、提升算法鲁棒性、构建对抗性攻击防御机制、优化数据管理体系等。此外，培养全员安全意识和专业技术人才，推动行业标准的完善，也有助于提升人工智能的安全保障水平。未来，随着人工智能与大数据技术的不断发展，只有加强技术创新与制度建设，才能构建安全可信的人工智能应用环境，确保技术造福于社会发展。

参考文献

[1]廖霄，李卓晖.人工智能与大数据分析在 IT 计算安全性与隐私保护中的应用探索[J].电子元器件与信息技术，2023，7（11）：137-140.DOI：10.19772/j.cnki.2096-4455.2023.11.035.

[2]左李景.大数据背景下信息通信网络安全管理策略研究[J].中国新通信，2022，24（14）：107-109.

[3]任致远，李江岱.大数据时代人工智能在计算机网络技术中的应用探讨[J].软件，2022，43（05）：110-112.