人工智能在网络空间安全学科教学中的应用探索*

叶登攀 李珉

武汉大学国家网络安全学院，武汉 430072

摘 要 在数字化时代，网络空间安全变得日益重要，而网络威胁也日趋复杂和多样。本文探讨了人工智能在网络空间安全学科教学中的应用，并分析其带来的双重效应：一方面，人工智能技术的发展带来了新的安全挑战；另一方面，人工智能技术也为网络空间安全提供了新的防御手段，增强了网络攻击检测的速度和准确性。本文简述了网络空间安全的重要性与挑战，并分析了人工智能在此领域的应用。通过案例分析，展示了 AI 在教学中的应用，并探讨了其在实战教学中的作用。文中提出了应对技术发展、资源和伦理法律问题的策略，并强调了更新教学内容、校企合作及伦理法律教育的必要性，旨在培养应对未来挑战的专业人才。

关键字 人工智能，网络空间安全，教学应用

The Application of Artificial Intelligence in the Teaching of Cyberspace Security

Ye Dengpan Li Min

School of Cyber Science and Engineering of Wuhan University,
Wuhan 430072, China
limin_brilliant@163.com

Abstract—In the digital age, security in cyberspace is becoming increasingly important, and cyber threats are becoming more complex and diverse. This paper discusses the application of artificial intelligence in the teaching of cyberspace security, and analyzes its dual effects: On the one hand, the development of artificial intelligence technology has brought new security challenges, such as adversarial attack and deep fake technology, which pose a threat to network security; On the other hand, artificial intelligence technology also provides a new means of defense for cyberspace security, enhancing the speed and accuracy of cyber attack detection. This paper briefly describes the importance and challenges of cyberspace security, and analyzes the application of artificial intelligence in this field. Through case analysis, the application of artificial intelligence in teaching is demonstrated, and its role in practical teaching is discussed. The paper proposes strategies to deal with technology development, resources, and ethical and legal issues, and emphasizes the need to update teaching content, school-enterprise cooperation, and ethical and legal education in order to train professionals to meet future challenges.

Keywords—Artificial intelligence, cyberspace security, teaching applications

1 引 言

在数字化时代，网络空间的重要性日益凸显，它不仅是现代社会的中枢神经系统，支撑着全球经济的运作，更是国家关键基础设施的运行平台，成为国家安全和社会发展的基石。然而，随着网络技术的飞速发展，网络空间安全面临的挑战也日益严峻，数据泄露、网络攻击、身份盗窃和信息战等安全威胁层出不穷，对个人隐私、企业资产乃至国家安全构成了严重威胁[1]。

全球范围内，网络空间安全已经上升到国家安全战略的高度。美国将保护数字基础设施作为国家安全的优先事项，英国在《网络安全战略》中强调了 21 世纪国家安全对网络空间安全的依赖，日本则早在 2013 年通过“信息安全政策会议”制定了“网络安全战略”的最终草案，提出了一系列强化网络安全的措施。我国同样高度重视网络空间安全，2014 年成立的中央网络安全和信息化领导小组，不仅体现了国家层面的重视，也标志着我国从网络大国向网络强国迈进的战略决心。这一战略的实施，旨在通过加强网络空间安全教育和人才培养，提升国家在全球网络空间的安全地位和影响力。

2 网络空间安全与人工智能

人工智能（AI）在近年来迎来了前所未有的流行和繁荣，其应用领域广泛，从语音识别到自动驾驶，从医疗诊断到智能家居，AI 的影响力无处不在[2]。随着大数据、云计算、互联网、物联网等信息技术的快

速发展，AI 技术正以惊人的速度跨越科学与应用之间的“技术鸿沟”，实现了从“不能用、不好用”到“可以用”的技术突破，迎来爆发式增长的新高潮。AI 的发展潜力巨大，预计将继续推动各行各业的技术革新和产业升级，成为新一轮科技革命和产业变革的核心驱动力。尽管 AI 的发展仍面临诸多挑战，但其前景广阔，未来可期。

随着人工智能技术的快速发展，其与网络空间安全的结合也日益紧密。人工智能（AI）与网络空间安全的结合产生了双重效应[3]，如图 1：一方面，AI 技术的兴起带来了伴生的安全挑战，包括内生的安全问题和衍生安全问题。这些问题源于 AI 自身的脆弱性，可预测性、可解释性不足，导致安全威胁的转移和新兴安全威胁的产生。攻击者能够利用对抗样本和数据投毒等技术对 AI 系统发起攻击，影响人脸识别、车牌识别等功能，甚至可能引发网络攻击和物理攻击。

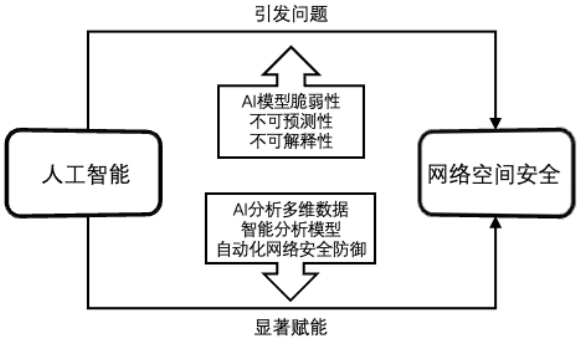

图 1　人工智能与网络空间安全结合产生双重效应

另一方面，AI 为网络空间安全领域提供了显著的赋能效应。机器学习和深度搜索等 AI 方法增强了网络攻击能力，同时提升了网络安全防御的自动化水平，实现了从被动防御到主动防御的转变。AI 技术能够通过分析多维数据，构建智能分析模型，提高对网络攻击的检测速度和准确性。AI 在网络空间安全领域的应用是一把双刃剑，它既带来了新的安全威胁，也提供了新的防御手段。

3　人工智能自身安全对于网络空间安全教学的示范案例

在人工智能（AI）技术与网络空间安全紧密融合的背景下，AI 展现出了其双刃剑的特性：一方面，它带来了伴生和衍生的新的安全威胁，如对抗样本攻击和深度伪造技术，这些技术能够误导人脸识别模型，甚至制造出足以欺骗人类视觉的假脸，对网络空间安全构成严重威胁。另一方面，AI 也提供了新的防御手段，通过分析多维数据，可以训练出高精度的模型，实现对网络攻击的快速响应和自动化防御。

当前网络空间安全教学的改革思路对提升网络安全人才培养质量具有重要意义[4]，但其实践手段往往生动性不足，未能充分体现网络空间安全教学的本质——网络攻击与网络防御相互制衡的过程。这种局限性影响了学习者对网络安全概念的深入理解和技能的掌握。为了解决这一问题，AI 的应用在网络空间安全教学中显得尤为重要。它不仅可以增强教学的生动性和互动性，而且通过模拟攻击和防御场景，能够有效提升学习者对网络安全威胁的认识和应对能力。

人工智能自身的安全问题对网络空间安全产生了广泛而深远的影响，特别是在对抗样本攻防、深度伪造技术的检测与取证、人脸识别技术中的隐私保护等方面尤为突出。本文将从上述几个方面进行详细的教学探索和分析。

3．1　对抗样本实践

对抗样本（Adversarial Examples）[5]是机器学习领域中一种特殊类型的输入，它们被设计为在模型的决策边界附近轻微扰动，目的是使模型以高置信度来做出错误的预测。AI 算法的脆弱性导致了对抗样本攻击的风险，这些攻击通过对样本进行微小的改动来欺骗 AI 系统，从而产生误判或漏判。在网络安全领域，对抗样本的存在对深度学习模型的可靠性和安全性构成了严重威胁。随着深度学习在图像识别、语音处理和自然语言理解等领域的广泛应用，对抗样本的攻击已成为一个不容忽视的问题。为了应对这一挑战，研究者们提出了多种安全机器学习算法，如使用多个分类器的技术、具有隐私保护能力的算法，以及博弈论在数据挖掘中的应用，以增强 AI 系统的鲁棒性。

（1）　攻击方面的教学

首先，学生需要了解如何生成对抗样本。这包括但不限于快速梯度符号方法（FGSM）、基本迭代方法（BIM）和投影梯度下降（PGD）等。通过实践这些技术，学生可以直观地看到模型在面对精心设计的输入时的脆弱性。随后介绍不同类型的攻击策略，如白盒攻击（攻击者完全了解模型结构和参数）和黑盒攻击（攻击者对模型一无所知），以及它们在现实世界中的应用场景。通过分析历史上著名的对抗样本攻击案例，让学生了解这些攻击在现实世界中的影响和后果。

（2）　防御方面的教学

教授学生如何通过数据增强、模型正则化和集成学习等方法来增强模型的鲁棒性。接着介绍和实践各种防御机制，如对抗训练、输入预处理和模型蒸馏等，以提高模型对对抗样本的抵抗力。此外，教授

学生如何对模型进行安全评估，包括使用对抗样本进行测试和评估模型的防御效果。

3．2 深度伪造实践

深度伪造（Deepfake）技术通过深度学习，尤其是生成对抗网络（GANs），实现视频和图像中人脸的生成与操纵，使得通过 AI 生成的伪造视频和音频越来越逼真，这对个人隐私和信息安全构成了严重威胁，引发多方面的社会关注。检测 deepfake 内容的方法主要分为两类：一类是通过识别伪造过程中的底层伪影[6]，如不自然的面部表情；另一类是分析难以伪造的高层语义信息[7]，例如人物的行为模式。Deepfake 检测技术分为图像级和视频级。图像级的检测通常使用 CNN 来识别静态图像的异常，而视频级的检测则利用 RNN 等技术分析帧的连贯性。多模态检测方法[8]通过结合视觉与音频信息，提升检测的准确性和鲁棒性，如通过分析唇动与语音同步性来识别 deepfake。

在深度伪造的教学中，我们教授学生理解 GANs 的基本原理，探讨其对社会的负面影响，并教授使用 CNN 和 RNN 等技术进行图像和视频的伪造检测。同时，强调多模态检测的重要性和面对技术迭代挑战时的主动防御策略，以及 AI 在取证中的应用，以培养学生在网络安全领域的综合能力。

3．3 人脸隐私保护实践

人脸隐私保护是 AI 自身安全影响网络空间安全的另一个重要领域。随着人脸识别技术的普及，个人隐私泄露的风险也随之增加。人脸隐私保护技术的发展对于平衡人脸识别系统的便利性和个人隐私安全至关重要。AI 技术被用于开发更为安全的人脸识别系统，同时，也在研究如何保护个人面部数据不被滥用，包括匿名化技术和隐私保护算法的应用。当前，人脸识别技术的隐私保护策略多样，研究者们采取了多种创新策略来强化个人隐私的安全性。一种策略是构建“影子模型”来模拟潜在攻击者的行为[9]，这种方法能够在不降低识别精度的情况下，显著提升系统对未知重构攻击的防御能力。此外，通过修剪图像的低频分量，研究者们训练模型以利用局部频域信息来映射整体人脸特征[10]，这一技术有效减少了对高频分量的依赖，从而增强了隐私保护的性能。还有研究提出了基于流模型的隐私保护方法[11]，这种方法不仅保持了图像的视觉连贯性，还实现了人脸信息的匿名化和可逆性，为隐私保护提供了新的解决方案。

在人脸隐私保护的教学中，我们强调人脸识别技术与隐私安全之间的平衡，教授学生如何利用 AI 技术来增强人脸识别系统的安全性，同时探讨匿名化和隐私保护算法的应用。这些教学内容旨在让学生掌握关键技术，以应对人脸识别中的隐私挑战。

4 人工智能应用于实战教学

网络空间安全学科在数字化时代扮演着关键的角色。然而，目前教学方式和方法很多情况下仍然停留在传统的讲授模式，缺乏互动性和实践性，难以激发学生的学习兴趣和参与度。网络空间安全学科的教材和课程体系需要不断更新，以避免理论与实际应用的脱节。另外，网络空间安全学科教学中还存在缺乏与行业实际紧密结合的资源，缺乏安全攻防实际场景的实践以及网络空间安全学科的评估方法不够科学等问题。

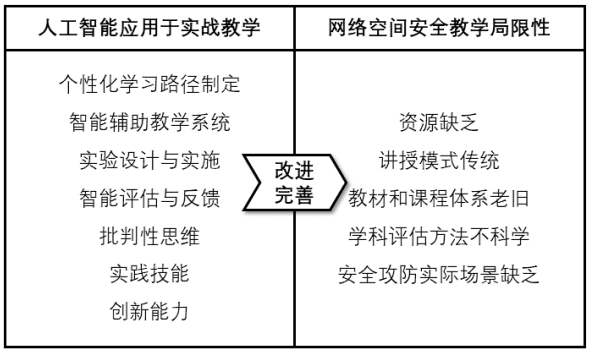

图 2 人工智能应用于实战教学

人工智能技术在教育教学领域得到广泛的应用，能够辅助教师和学生在网络空间安全领域进行知识生产，如图 2。

4．1 个性化学习路径制定

人工智能系统能够根据学生的学业成绩、认知能力、职业规划以及个人偏好等多个维度，提供个性化的学习体验。通过数据分析和机器学习算法，AI 能够为每位学生量身定制学习计划，从而激发学生的学习热情，提升学习效率，并促进知识的深入理解。

4．2 智能辅助教学系统

网络空间安全学科涉及的概念繁多且技术迭代迅速，实践操作中遇到的环境配置、代码调试等问题复杂多变。人工智能辅助教学系统能够提供实时的交互式问答帮助，辅助学生解决编程中的错误，解答复杂概念问题，从而提高学习效率。此外，AI 还能够筛选和整理互联网上的海量信息，为学生提供精准的学习资源，节省宝贵的学习时间。

4．3 实验设计与实施

网络空间安全领域的前沿技术，如对抗样本攻防、深度伪造检测与取证、人脸识别技术中的隐私保

护等，都是当前研究的热点。在教学中，将这些领域的最新研究成果和发展趋势介绍给学生至关重要。通过引入业界公认的基准模型（baseline）和开源项目，结合真实世界的网络对抗案例，学生不仅能够获得实际操作经验，还能够在此基础上进行创新和优化，这不仅有助于弥补教材的不足，还能增强教学的实践性和针对性。

4．4　智能评估与反馈

人工智能系统能够提供自动化的评估工具，如基于机器学习算法的代码评估、智能评分等，这些工具能够客观地评价学生的实验成果和理解程度。智能化的评估不仅提高了评价的准确性，还能够帮助学生更好地认识自己的学习状态，激发学生的学习动力，促进自我学习和自我提升。

此外，在网络空间安全教学中，培养学生的关键能力至关重要。通过人工智能自身安全对于网络空间安全的示范案例和 AI 运用到网络空间安全实战教学中的引入，有利于提升学生的综合素质。

4.5　批判性思维

通过分析真实的对抗样本案例，鼓励学生提出质疑并进行深入探讨。例如，在课程中引入对抗样本攻击案例，如通过对图像进行微小的扰动使 AI 系统误判。学生可以分析对抗样本生成的过程，理解其背后的原理，并讨论如何改进现有的防御机制。

4.6　实践技能

为了提升学生在深度伪造检测和取证方面的技能，可以设计丰富的实验和项目。例如，在实验课中设置深度伪造视频的检测任务，让学生使用深度学习模型对深度伪造视频进行识别，并进行取证分析，锻炼他们的实际操作能力。此外，还可以通过模拟真实取证场景，培养学生的取证思维和技能。

4.7　创新能力

鼓励学生探索新技术，培养他们在人工智能和人脸隐私保护方面的创新能力。例如，设立创新实验室，让学生自由探索人脸识别技术在隐私保护中的应用，如通过对人脸数据进行加密或模糊处理，确保个人隐私的安全。通过项目制学习，学生不仅能掌握前沿技术，还能锻炼他们的创新思维。

4．8　教学实践应用效果

这种教学应用的效果显著，学生在专业竞赛中屡获佳绩，如“华为杯”中国研究生网络安全创新大赛决赛一等奖，充分证明了学生的实操能力和专业技能。毕业生在网络安全公司、研究机构、高等教育机构和政府部门等领域就业，展现出扎实的专业技能和出色的工作能力。教学模式通过平衡理论知识与实践应用，增强了学生在专业领域的竞争力。

5　面临的挑战与解决方案

在网络空间安全教学中，存在一些挑战需要克服，以确保教学的有效性和时效性。

5.1　技术更新迅速

网络安全技术发展迅速，教学内容需要不断更新。为了解决这一问题，建议建立动态的教学资源库，定期更新教材和课程内容，邀请行业专家进行讲座和培训。此外，可以通过在线平台提供最新的技术资料和案例分析，让学生随时获取最新信息。

5.2　实践资源限制

由于设备、资金和场地的限制，实践资源可能不足。为解决这一问题，建议高校与企业合作，建立共享实验室和实践基地，提供学生更多的实践机会。例如，利用云计算技术搭建对抗样本实验平台，学生可以在虚拟环境中进行对抗样本生成和检测的实战操作。

5.3　伦理与法律问题

在教学中，如何处理与 AI 相关的伦理和法律问题至关重要。建议在课程中增加伦理与法律模块，介绍相关法律法规和道德规范，培养学生的法律意识和道德素养。例如，讨论深度伪造技术在隐私保护中的挑战，让学生了解如何在技术开发中遵守法律和伦理要求，避免技术滥用。

6　结束语

本文从网络空间安全的重要性出发，探讨了人工智能在该领域的教学应用，包括对抗样本、深度伪造技术和人脸隐私保护等方面的实践案例。通过这些案例，本文不仅展示了 AI 技术在网络空间安全教学中的双刃剑特性，也强调了其在提升学生实践技能和创新能力方面的潜力。本文进一步讨论了人工智能在个性化学习路径制定、智能辅助教学系统、实验设计与实施以及智能评估与反馈等方面的应用，这些应用为网络空间安全的教学提供了新的视角和方法。同时，本文指出了当前教学中面临的挑战，如技术更新迅速、实践资源限制以及伦理与法律问题，并提出了相应的解决方案。通过教育工作者、技术开发者和政策制定者的共同努力，可以培养出更多具备批判性思维、实践技能和创新能力的网络空间安全专业人才，为维护网络空间的安全和稳定做出贡献。

参 考 文 献

[1] Zhang H, Han W, Lai X, et al. Survey on cyberspace security[J]. Science China Information Sciences, 2015, 58: 1-43.

[2] 李晓理,张博,王康,等.人工智能的发展及应用[J].北京工业大学学报,2020,46(06):583-590.

[3] 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(03):60-66.

[4] 李震宇,刘琰,谭磊,朱玛,罗向阳.基于 5C 模型的网络安全类课程实践教学改革与探索[J].计算机技术与教育,2021,9(2):111-114.

[5] Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv:1412.6572, 2014.

[6] Cozzolino D, Pianese A, Nießner M, et al. Audio-visual person-of-interest deepfake detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 943-952.

[7] Agarwal S, Farid H, Gu Y, et al. Protecting World Leaders Against Deep Fakes[C]//CVPR workshops. 2019, 1: 38.

[8] Yin Z, Wang J, Xiao Y, et al. Improving Deepfake Detection Generalization by Invariant Risk Minimization[J]. IEEE Transactions on Multimedia, 2024.

[9] Wang Z, Wang H, Jin S, et al. Privacy-preserving Adversarial Facial Features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 8212-8221.

[10] Mi Y, Huang Y, Ji J, et al. Privacy-Preserving Face Recognition Using Random Frequency Components[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 19673-19684.

[11] Yuan L, Liang K, Pu X, et al. Invertible Image Obfuscation for Facial Privacy Protection via Secure Flow[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023.