# 1 AI-Enhanced Ethical Hacking: A Linux-Focused Experiment

《AI增强的道德黑客攻击：一项以Linux为重点的实验》论文探讨了将生成性人工智能（GenAI），特别是ChatGPT，整合到Linux系统渗透测试中的实验研究。研究通过模拟渗透测试的关键阶段，如信息收集、扫描、获取访问权限、维持访问权限和掩盖痕迹，验证了GenAI在提升效率、降低成本方面的潜力，同时强调了人机协作的重要性，并讨论了潜在风险。

# 2 Artificial Intelligence as the New Hacker: Developing Agents for Offensive Security

《人工智能作为新黑客：开发攻击性安全的智能体》论文探讨了将人工智能（AI）整合到攻击性网络安全中的应用，特别是通过开发一个名为ReaperAI的自主AI智能体来模拟和执行网络攻击。ReaperAI利用大型语言模型（LLM）如GPT-4的能力，能够自主识别、利用和分析安全漏洞。研究还探讨了AI在攻击性安全中的作用，并提出了未来研究方向，包括改进AI与网络安全工具的交互、增强学习机制以及讨论AI在攻击性角色中的伦理准则。

# 3、CyberMentor: AI Powered Learning Tool Platform to Address Diverse Student Needs in Cybersecurity Education

《CyberMentor：AI驱动的学习工具平台，以满足网络安全教育中的多样化学生需求》论文介绍了一个名为CyberMentor的学习工具平台，旨在解决网络安全教育中非传统学生面临的多样化需求。该平台利用生成式大型语言模型（LLM）和检索增强生成（RAG）技术，为学生提供知识、技能和职业准备建议。通过三个使用场景，展示了CyberMentor在促进知识获取、职业准备和提供基于技能的指导和支持方面的价值。

# 4、Generative AI in Cybersecurity: A Comprehensive Review of LLM Applications and Vulnerabilities

《生成式人工智能在网络安全中的应用：大型语言模型的全面综述》论文全面探讨了生成式人工智能（GenAI）和大型语言模型（LLMs）在网络安全中的应用，包括硬件设计安全、入侵检测、软件工程、威胁情报、恶意软件检测等领域。同时，论文分析了LLMs的漏洞，如提示注入、数据中毒、DDoS攻击等，并提出了缓解策略。此外，论文还评估了42种LLMs在网络安全知识和硬件安全方面的表现，探讨了LLMs在网络安全中的挑战和局限性，并提出了未来研究方向。

# 5、Generative AI in Cybersecurity

《生成式人工智能在网络安全中的应用》这篇论文由杜克大学等机构的研究人员撰写，探讨了生成式人工智能（GAI）在网络安全领域的双刃剑效应。一方面，GAI为网络安全防护提供了强大的工具，如威胁检测和响应；另一方面，它也为网络犯罪分子提供了更隐蔽、更具适应性的恶意软件开发手段。论文分析了GAI的发展趋势、对网络安全的影响，以及未来应对策略，强调了组织需主动识别并开发更复杂的防御策略以应对GAI驱动的网络威胁。

# 6、Jailbreaking Generative AI: Empowering Novices to Conduct Phishing Attacks

《破解生成式人工智能：使新手能够发动网络钓鱼攻击》研究了如何利用最新的生成式人工智能模型 ChatGPT-4o Mini 来发动网络钓鱼攻击。研究发现，即使是新手用户，也可以通过“越狱”（jailbreaking）技术绕过 ChatGPT 的道德限制，生成钓鱼内容并实施攻击。研究强调了生成式人工智能在网络安全领域带来的新风险，呼吁加强网络安全措施和用户教育。

# 7、PenTest++: Elevating Ethical Hacking with AI and Automation

《PenTest++：借助人工智能与自动化提升道德黑客水平》介绍了一个名为 PenTest++ 的系统。该系统将人工智能（AI）与自动化相结合，优化道德黑客的流程。它在虚拟环境中开发，可处理侦察、扫描、枚举、利用和文档化等关键渗透测试任务，同时保持模块化和适应性设计。该系统在提高效率、可扩展性和适应性方面具有显著优势，但也引发了隐私和 AI 生成错误（幻觉）等伦理问题。

# 8、Review of Generative AI Methods in Cybersecurity

《生成式人工智能在网络安全中的方法综述》全面探讨了生成式人工智能（GenAI）在网络安全中的应用现状。论文分析了 GenAI 在网络攻击（如社会工程学攻击、钓鱼邮件、恶意软件生成等）和网络防御（如漏洞检测、安全代码生成、威胁情报等）中的双重角色，并讨论了其带来的伦理、法律和社会问题。研究强调了开发稳健的伦理规范和创新防御机制的重要性，以应对 GenAI 带来的挑战，并促进其在网络安全中的公平应用。

# 9、Robustness and Cybersecurity in the EU Artificial Intelligence Act

《欧盟人工智能法案中的鲁棒性与网络安全》探讨了欧盟人工智能法案（AIA）中关于人工智能系统鲁棒性和网络安全的法律要求。研究分析了高风险人工智能系统（HRAIS）和通用人工智能模型（GPAIM）的相关条款，指出了法律术语与机器学习研究之间的差距，并提出了可能的解决方案。论文强调了技术标准和指南在确保人工智能系统合规性中的重要性，为未来的技术和法律研究提供了方向。

# 10、Survey Perspective: The Role of Explainable AI in Threat Intelligence

《可解释人工智能在威胁情报中的作用——调查视角》探讨了可解释人工智能（XAI）在安全运营中心（SOC）中的应用。研究通过行业调查分析了 SOC 分析师如何处理基于人工智能的安全警报，面临的挑战，以及对 XAI 的需求。研究发现，尽管大多数分析师尚未采用集成 XAI 的工具，但他们对攻击归因、置信度评分和特征贡献解释表现出浓厚兴趣。论文提出了针对 XAI 增强的安全警报系统的实用设计建议，以提高 AI 基于网络安全解决方案的透明度、可解释性和可操作性。

# 11、Towards Robust and Secure Embodied AI: A Survey on Vulnerabilities and Attacks

《迈向稳健和安全的具身智能：关于漏洞和攻击的综述》全面调研了具身智能系统（如机器人和自动驾驶车辆）面临的独特安全和稳健性挑战，包括传感器欺骗、对抗攻击、任务规划失败等漏洞。文章将漏洞分为外源性（如物理攻击、网络安全威胁）和内源性（如传感器故障、软件缺陷）两类，系统分析了针对具身智能的对抗攻击范式，探讨了攻击对感知、决策和交互的影响，还研究了针对大型视觉-语言模型（LVLM）和大型语言模型（LLM）的攻击向量，并提出了增强具身智能系统安全性和可靠性的策略。

# 12、Transparency, Security, and Workplace Training & Awareness in the Age of Generative AI

《生成式人工智能时代的透明度、安全性和工作场所培训与意识》探讨了生成式人工智能（AI）快速发展的背景下，组织如何应对AI使用政策的重新评估。研究重点关注了公开可访问的大型语言模型（LLMs），尤其是Gab AI，分析其在工作场所政策中的影响。研究强调了透明度、数据隐私和伦理考量的重要性，并提出了通过培训和政策更新来应对网络安全风险的必要性。