Harnessing Artificial Intelligence for Penetration Testing Syllabus

December 2023

Instructor Information

Kiera Conway Kiera.Conway@trojans.dsu.edu

General Information

Course Description:

This seminar-style course, "Harnessing Artificial Intelligence for Penetration Testing," provides an in-depth exploration of the integration of Artificial Intelligence (AI) techniques, including Machine Learning (ML), Reinforcement Learning (RL), and Deep Reinforcement Learning (DRL), into the field of Penetration Testing (PT). Drawing from recent advancements in AI technology, this course addresses the growing need for innovative applications and techniques in the rapidly evolving cybersecurity landscape. The primary focus is on leveraging AI to enhance PT methodologies, covering phases from information gathering to post-exploitation. Participants will engage in theoretical discussions, hands-on exercises, and real-world applications to equip themselves with practical skills to confidently navigate complex cybersecurity challenges, anticipate emerging threats, and effectively contribute to the dynamic and complex domain of offensive security.

Expectations and Goals:

By the end of the course, participants are expected to:

- Understand the role of AI in PT, from fundamentals to advanced techniques
- Gain practical proficiency in deploying popular PT tools
- Navigate the PT process, from information gathering to post-exploitation, using manual and AI techniques
- Develop critical thinking and problem-solving skills in the context of cybersecurity
- Reflect on ethical considerations associated with AI in cybersecurity

Course Materials

Required Materials:

- None (materials will be provided)
- An account on TryHackMe is required for hands-on exercises

Optional Materials:

- The Basics of Hacking and Penetration Testing, 2nd Edition by Patrick Engebretson
  (Recommended for additional insights, but not required)

Course Schedule

Lecture 1: Introduction to AI in PT

- Course Introduction
- Overview of the Penetration Testing (PT) phases
- The significance of AI in Penetration Testing
- Introduction to Machine Learning (ML) and Reinforcement Learning (RL)

Lecture 2: Gathering Information

- Definition, objectives, and importance of Gathering Information in PT
- Traditional information gathering methods
- Introduction to AI-driven data acquisition
- **PRACTICE**: WebOSINT, SoMeSINT
- **RESEARCH**: What are some creative challenges that real-world applications of these tools could help solve?

Lecture 3-1: Manual Scanning Methods

- Definition, objectives, and importance of Scanning in PT
- Types of Scanning and their significance
- Examples of manual tools associated with each scanning type
- Limitations of current scanning methods
- **PRACTICE**: Nmap Live Host Discovery, OpenVAS Basics, OWASP Juice Shop

Lecture 3-2: AI in Scanning

- Definition and analysis of AI in the context of Scanning
- Role of AI in enhancing scanning techniques
- Advantages AI offers across different scanning types
- AI tools: GyoiThon, Dark Trace, Shodan
- Ethical considerations in AI-powered Scanning
- **RESEARCH**: Imagine a scenario where a 'Strong AI' tool, designed to scan systems for security issues within legal and ethical boundaries, autonomously decides to exploit a known vulnerability in a critical system to gather more information. This action, while achieving its legal and ethical goal, raises concerns as it violates the penetration testing principle of not causing harm. The ethical breach occurs because the AI tool autonomously chose an action that goes beyond agreed-upon legal and ethical boundaries. In such a case, who should be held accountable for these actions? What mechanisms could be put in place to ensure clear lines of responsibility?

Lecture 4-1: Traditional Exploitation Methods

- Definition, objectives, and importance of Exploitation in PT
- Traditional exploitation methods: SQL Injection, XSS, and Buffer Overflows
- Introduction to Metasploit and its dual nature
- Limitations of current exploitation methods
- **PRACTICE**: Metasploit: Introduction, Metasploit: Meterpreter, SQL Injection

Lecture 4-2: AI in Exploitation

- Definition and analysis of AI and ML in Exploitation
- Role of AI in enhancing exploitation techniques
- Advantages and challenges of using AI for exploit development
- Real-world examples: Social Engineering Toolkit (SET), DeepExploit, and Python Libraries
- Emerging trends and technologies in AI for Exploitation
- **RESEARCH**: Considering the unpredictable nature of evolving cyber threats, how can AI models be trained to handle novel and unexpected attack scenarios effectively? What challenges arise when the threat landscape diverges from the training data? What kind of patterns should AI focus on to limit these challenges?

Lecture 5-1: Manual Post-Exploitation

- Definition, objectives, and importance of Post-Exploitation in PT
- Traditional post-exploitation methods: Netcat, JtR, AuditPol, Dradis, and more
- Objectives and importance of these practical tools
- **PRACTICE**: Password Attacks, Evading Logging and Monitoring

Lecture 5-2: AI in Post-Exploitation

- Definition and analysis of AI and ML in Post-Exploitation
- Emphasis on RL, DRL, and the Actor-Critic method, specifically A2C
- Role of AI in enhancing post-exploitation techniques
- Advantages and challenges of using AI for Post-Exploitation development
- Reference to novel studies and theoretical implantation
- **RESEARCH**: How can the RL agents introduced in the studies be effectively employed by future researchers to enhance and automate specific aspects of post-exploitation phases? In what ways did the authors of the study lay the groundwork for future researchers, and what specific insights or methodologies can be built upon in subsequent studies? What are some limitations that future researchers might encounter in the improvement of these methods?