Leveraging Large Language Models for Penetration Testing

Exploring AI's Role in Cybersecurity

AL MOCHKIN - HACK.LU 2024 - CTI SUMMIT

About myself

- Entrepreneur and Philanthropist
- Pioneering global computer literacy and privacy advocacy
- Founder of several successful startups
- 20+ years of experience in the tech industry
- Executive board member of European House of Artificial Intelligence

Artificial Intelligence 101

Intelligence is the ability to adapt to change.

– Steven Hawking

Machine intelligence is the last invention that humanity will ever need to make.

– Nick Bostrom

Offensive AI

01	Automated Social Engineering	**Deepfakes – Phishing**
02	Evading Detection	**Polymorphic Malware** – Reinforcement learning to bypass defenses
03	Intelligent Reconnaissance	Automated code analysis – **OSINT – Automating discovery of targets**
04	Credential Attacks	Biometric spoofing – **Credential stuffing**
05	Automated Exploit Development	**Application fuzz – CVE variability – Malware copilots**

How LLMs Can Enhance Pentesting

MITRE ATT&CK	LLM capabilities/tactics:
Reconnaissance	8/10
Resource Development	5/8
Initial Access	4/10
Execution	11/14
Persistence	8/20
Privilege Escalation	6/14
Defense Evasion	18/43
Credential Access	8/17
Discovery	27/32
Lateral Movement	4/9
Collection	9/10
Command and Control	7/17
Exfiltration	7/9
Impact	3/14

LLMs are capable of **125 out of 227** ATT&CK tactics

And also it is:

1. Faster
2. Cheaper

Real-World Applications

1. PentestGPT
   a. FOSS
   b. Open-AI backend
   c. https://github.com/GreyDGL/PentestGPT
2. Auto-Pentest-GPT-AI
   a. FOSS
   b. Mistral-based
   c. https://github.com/Armur-Ai/Auto-Pentest-GPT-AI
3. N5S
   a. FOSS
   b. Nemotron-based plus proprietary GAN
   c. https://n5s.ai

Challenges and Considerations

While promising, using LLMs for pentesting isn't without challenges:

1. Ethical Concerns
2. False Positives
3. Security of the LLMs

Best Practices for LLM-Enhanced Pentesting

To effectively leverage LLMs in pentesting:

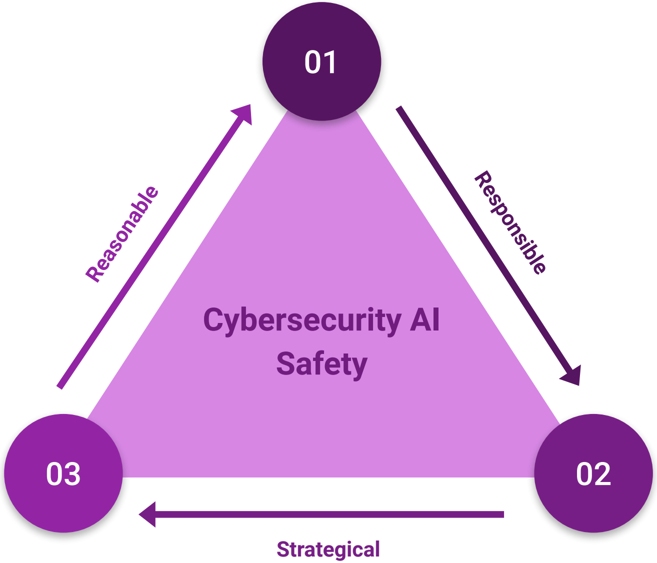
1. Combine AI with Human Expertise (Context enrichment)
2. Input Validation (Reduce false positives)
3. Monitoring (Avoid building SkyNet)

The Future of LLM-Powered Pentesting

As LLMs continue to evolve, we can expect:

- More sophisticated and targeted exploit generation
- Enhanced ability to identify complex, multi-step attack vectors
- Improved natural language processing for social engineering simulations
- AI-powered products require defenses themselves:
  - New attack vectors:
    - Prompt Injections
    - Training Data Poisoning

Conclusion


01
Responsible
02
Strategical
03
Reasonable
Cybersecurity AI Safety

Q&A, Contacts

Thank you for your attention. Are there any questions?

LinkedIn: