Special thanks to "Learn Prompting" for their help in making this roadmap

learnprompting.org

Find the detailed version of this roadmap and other similar roadmaps

roadmap.sh

- AI and Data Scientist Roadmap
- AI Engineer Roadmap
- Data Analyst Roadmap
- MLOps Roadmap

AI Red Teaming

Introduction

- AI Security Fundamentals
- Why Red Team AI Systems?
- Ethical Considerations
- Role of Red Teams

Foundational Knowledge

AI / ML Fundamentals

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Neural Networks
- Generative Models
- Large Language Models
- Prompt Engineering

Cybsecurity Principles

- Confidentiality, Integrity, Availability
- Threat Modeling
- Risk Management
- Vulnerability Assessment

Prompt Hacking

- Jailbreak Techniques
- Safety Filter Bypasses
- Prompt Injection
  - Direct
  - Indirect
- Countermeasures

Model Vulnerabilities

- Model Extraction
  - Model Weight Stealing
  - Unauthorized Access
- Model Manipulation
  - Data Poisoning
  - Adversarial Examples
  - Model Inversion
- Defense Strategies
  - Adversarial Training
  - Robust Model Design
  - Continuous Monitoring

System Security

- Code Injection
  - Insecure Deserialization
  - Remote Code Execution
- Infrastructure Security
  - API Protection
  - Authentication
  - Authentication

Testing Methodologies

- Black Box Testing
- White Box Testing
- Grey Box Testing
- Automated vs Manual
- Continuous Testing

Professional Development

- Community Engagement
  - Conferences
  - Research Groups
  - Forums
- Practical Experience
  - Lab Environments
  - CTF Challenges
  - Red Team Simulations
- Certifications
  - Specialized Courses
  - Industry Credentials

Tools and Frameworks

- Testing Platforms
- Monitoring Solutions
- Benchmark Datasets
- Custom Testing Scripts
- Reporting Tools

Real-world Applications

- LLM Security Testing
- Agentic AI Security
- Responsible Disclosure

Future Directions

- Emerging Threats
- Advanced Techniques
- Research Opportunities
- Industry Standards

Visit the following relevant tracks to keep learning

- AI Engineer
- AI & Data Scientist
- Data Analyst