Faysal Aziz

Essentials of AI Security

Learn to defend, detect, and deter security risks in AI environments in a day

FA-0195

Agentic & Generative AI · AI for Leaders & Business · Data & Analytics

---

Introduction

In the burgeoning landscape of technological advancement, artificial intelligence (AI) emerges as a revolutionary force, reshaping industries, enhancing human capabilities, and redefining the bounds of what machines can achieve. However, with great power comes great responsibility, and the rapid deployment of AI technologies has ushered in a host of unique security challenges that demand immediate and comprehensive attention.

The relevance of this training is underscored by a slew of recent incidents where AI systems were compromised, leading to significant financial, reputational, and ethical consequences. From data poisoning attacks that skew machine learning models, to adversarial attacks that subtly manipulate AI perceptions, the spectrum of AI vulnerabilities is wide and the potential damage, profound. These incidents highlight the pressing need for robust security measures and risk management strategies in AI development and deployment.

The objective for this course is straightforward yet ambitious: to not only understand these risks but also to master the strategies to address them effectively. By the end of this training, you will be equipped with the tools and knowledge to not only react to AI security challenges but to proactively integrate security into the very fabric of AI development processes. You will learn about the latest advancements in AI security, engage with real-world case studies, and participate in discussions that will sharpen your ability to think critically about AI security.

Learning outcomes

By the end of this training, participants will be able to:

- Understand the basic principles of AI and ML, including how these technologies work.
- Identify key security risks associated with AI systems.
- Apply best practices to mitigate security risks in AI applications.

- Understand the specific challenges posed by generative AI technologies.
- Develop strategies for safe and secure deployment of AI technologies.

Prerequisites

- Basic knowledge of Artificial Intelligence and Machine Learning concepts.
- Familiarity with computer programming (preferably Python).
- Understanding of cybersecurity fundamentals.
- Prior exposure to data handling and processing.

Training outline

1. Introduction to AI and Machine Learning

1. Definition of AI and ML
2. Brief history and evolution of AI
3. Types of AI: Narrow AI, General AI, and Super AI
4. Key concepts in ML: Supervised Learning, Unsupervised Learning, Reinforcement Learning

2. Fundamentals of AI Security

1. Concept of AI Security
2. Importance of Security in AI

   Common vulnerabilities in AI systems

   1. Data poisoning
   2. Model stealing
   3. Adversarial attacks
4. Ethical considerations in AI

3. Deep Dive into Generative AI

   Understanding Generative Models

   1. Generative Adversarial Networks (GANs)
   2. Variational Autoencoders (VAEs)
   3. Transformer models used in generative tasks
2. Use cases of generative AI

   Security risks specific to generative AI

1. Deepfakes and misinformation
2. Intellectual property issues
3. Biases in generated content

4. AI Security Risk Management

1. Risk identification and assessment in AI projects

   Frameworks and tools for AI risk management

   1. AI Risk Framework
   2. Security assessment tools specific to AI

   Developing an AI security strategy

   1. Security by design
   2. Continuous monitoring
   3. Incident response planning

5. Best Practices and Mitigation Strategies

Data security practices for AI

1. Secure data storage and transmission
2. Data anonymization techniques

Model security enhancements

1. Robust training techniques
2. Model hardening practices

Policy and governance for AI security

1. Regulatory compliance
2. Ethical guidelines for AI use

6. Case Studies and Group Discussions

1. Analysis of real-world AI security breaches
2. Group discussion on hypothetical AI risk scenarios
3. Strategies for mitigating discussed risks

7. Wrap-Up and Future Directions

1. Summary of key takeaways
2. Emerging trends in AI security

3. Q&A session

This outline is designed to provide a comprehensive understanding of AI security risks, applicable mitigation strategies, and the importance of integrating security practices into the AI development process.

© 2026 Faysal Aziz · Kuala Lumpur