

ICA Specialist Certificate in AI Threats and Risk Mitigation



Syllabus

Unit 1

The evolution of AI threats

- Adversarial AI 101
- Identifying the threats: Who are the attackers?
- AI safety, security and AI ethics
- The uniqueness of AI systems
- Traditional AI security vs AI cybersecurity

Unit 2

Adversarial attacks

- Introduction to AI threats
- Data layer attacks
 - Model evasion
 - Data poisoning
- Model level attacks
 - Model poisoning
 - Model inversion
 - Model extraction
 - Membership inference

Unit 3

Language-level attacks

- Prompt injection
 - Direct prompt injection
 - Indirect prompt injection
- Jailbreaking
- Tool manipulation
- Semantic obfuscation attack

Unit 4

AI-enabled attacks

- Deepfake and synthetic identity risks
- The automation and scaling of financial crime with AI tech
 - AI chatbots
 - Personalisation at scale
 - Automation of message creation
 - Rapid creation of synthetic identities
 - Fraud infrastructure and 'fraud-as-a-service'
- AI-enabled fraud
- Deepfake fraud
- Deepfake extortion
- AI-enabled disinformation
- AI-powered compliance evasion

**Unit
5****AI-misalignment risk**

- What is AI misalignment
- AI misalignment and the financial industry
- Outer misalignment
- Inner misalignment
- Deceptive misalignment
- Agentic misalignment

**Unit
6****Navigating the evolving regulatory landscape**

- EU AI Act's response to AI threats
 - Cybersecurity and robustness requirements
 - Data governance and integrity
 - Risk management systems
 - Human oversight and governance
 - Post-market monitoring and incident reporting
 - Implications for AI-enabled financial crime
- The EU NIS2 Directive on Cybersecurity and Adversarial Threats
- Singapore's response to AI security threats
- MAS Technology Risk Management (TRM) Guidelines
- US regulatory perspectives on AI security
- Prompt injection attack scenario

**Unit
7****AI security, governance and defence strategies
– Part 1**

- NIST AI Risk Management Framework (RMF)
- NIST IR 8596 Cyber AI Profile
- ISO/IEC 42001
- Defensive techniques
- Red teaming
- Anomaly detection
- Robustness testing
- Combatting AI-enabled attacks

**Unit
8****AI security, governance and defence strategies
– Part 2**

- Shadow AI risk
- Stakeholder cooperation to combat AI threats
- The AI security and governance strategy
 - Establish an AI risk governance baseline
 - Defending against adversarial attacks (machine learning systems)
 - Securing against language-level attacks (LLMs)
 - Combatting AI-enabled fraud
 - Managing AI misalignment risk and agentic risk
 - Managing regulatory risk
 - Managing shadow AI risk
 - Full training in AI threats and security risks
 - Continuous monitoring and incident response
 - Stakeholder cooperation

