**Document Generated: 11/14/2026**

**Learning Style: Online Instructor-Led**

**Technology:**

**Difficulty: Advanced**

**Course Duration: 3 Days**

# AI Security Deep Dive (TTAI2800)



## About This Course:

AI and machine learning systems introduce unprecedented security challenges that traditional cybersecurity practices cannot adequately address. AI Security Deep Dive delivers the specialized knowledge and hands-on experience needed to secure AI/ML systems against sophisticated attacks, protect sensitive training data, and implement robust defenses for AI-integrated applications. This intensive course is designed for programmers building AI-enabled applications, security analysts

responsible for protecting AI systems, cybersecurity professionals expanding into AI security, and technical managers overseeing AI implementation projects.

Throughout three intensive days, you will master the fundamentals of machine learning from a security perspective, identify and exploit vulnerabilities in AI systems through hands-on exercises, and implement practical defenses against data poisoning, adversarial attacks, and privacy breaches. You will gain critical experience securing traditional applications that integrate AI models, including LLM-powered features, and learn to validate inputs and outputs to prevent prompt injection and other AI-specific attacks. The course combines essential AI/ML concepts with real-world security scenarios, ensuring you understand both the technical foundations and practical implementation challenges.

## Course Objectives:

- Master AI/ML security fundamentals from the ground up. Understand how machine learning works, identify attack vectors unique to AI systems, and assess security
- implications of different ML model types and deployment patterns.
- Identify and exploit AI-specific vulnerabilities through hands-on exercises. Conduct data poisoning attacks, implement adversarial examples, perform model inversion and membership inference attacks, and understand the mechanics of AI system compromise.
- Implement comprehensive defenses against AI security threats. Design and deploy robust input validation, output filtering, differential privacy mechanisms, and secure training pipelines to protect against known attack vectors.
- Secure traditional applications integrating AI models and APIs. Build secure interfaces to LLM APIs, implement prompt injection defenses, validate AI-generated content, and establish secure authentication and authorization patterns.
- Protect sensitive information in AI training and inference. Apply privacy-preserving techniques, detect and prevent data leakage through model behavior, and implement secure data handling practices for AI systems.
- Establish enterprise-grade AI security governance and incident response. Develop AI security policies, create monitoring and detection capabilities, design incident response procedures for AI breaches, and build security-first AI development workflows.

## Audience:

- This intermediate-level course is designed for programmers and developers building AI-enabled applications, security analysts and cybersecurity

professionals expanding into AI security, and technical leads responsible for securing AI implementations. Software engineers integrating machine learning models, security architects designing AI system defenses, and incident response teams preparing for AI-related threats will gain essential skills to identify vulnerabilities, implement robust security measures, and respond to sophisticated AI attacks.

- Technical managers, DevSecOps professionals, and compliance officers overseeing AI security initiatives will also benefit from this course by gaining insights into AI-specific risk management, security governance frameworks, and regulatory compliance considerations. Whether you are directly developing AI systems, securing existing AI implementations, or establishing organizational AI security practices, this course provides the technical depth and practical experience needed to protect against emerging AI threats and build resilient AI-powered solutions

## Prerequisites:

- Read code and understand basic programming concepts. The course provides hands-on opportunities using interactive Python and optionally other platforms.
- Successful students will need to setup a basic development environment, read and follow program logic and make minor modifications to code.
- Awareness of traditional cybersecurity issues.  The successful student will have some prior knowledge of security issues in an IT environment.
- Basic understanding of web applications.  Students should have some experience and exposure to basic HTTP based web technology.
- Familiarity with data handling and basic statistical concepts. Understanding of data formats, databases, and basic data analysis principles.
- Experience with software development lifecycle and security practices. Knowledge of testing, deployment, and security integration in development processes.

## Course Outline:

Day 1: AI/ML Foundations and Attack Fundamentals

AI/ML Security Foundations

Understanding artificial intelligence and machine learning from a security perspective - establishing the essential knowledge base for identifying and defending against AI-specific threats.

- Overview of the OWASP Top 10 Application Security Vulnerabilities. Since AI models are frequently embedded within traditional web or enterprise applications, they inherit many of the same security risks identified by the OWASP Top 10. Understanding these common vulnerabilities is essential for developers and security professionals to protect both traditional and AI-powered applications from cyber threats.
- Essential AI/ML concepts for security professionals: supervised vs unsupervised learning, neural networks, deep learning fundamentals
- AI system architecture and deployment patterns: training vs inference, model serving, API endpoints
- The AI threat landscape: why traditional security approaches fail with AI systems
- Understanding the AI attack surface: training data, models, inference endpoints, and integration points
- Hands-on Lab (Jupyter Notebook): Setting up an AI security testing environment and exploring vulnerable ML models

Data Poisoning and Training Attacks

Deep dive into attacks targeting the AI training process, including practical implementation of data poisoning techniques and defense strategies.

- Data poisoning fundamentals: targeted vs untargeted attacks, clean-label attacks
- Training data vulnerabilities: data sources, collection pipelines, and validation gaps
- Backdoor attacks in machine learning models: trigger insertion and activation
- Supply chain security for AI: malicious datasets, compromised pre-trained models
- Hands-on Lab (Jupyter Notebook): Implementing data poisoning attacks against image classifiers and text models
- Hands-on Lab (Jupyter Notebook): Building data validation pipelines and poisoning detection systems

Day 2: Adversarial Attacks and Model Security

Adversarial Examples and Model Manipulation

Comprehensive exploration of adversarial attacks against deployed AI models, including hands-on generation of adversarial examples and evasion techniques.

- Adversarial examples: perturbation-based attacks, gradient-based methods (FGSM, PGD)
- Model evasion techniques: black-box vs white-box attacks, query-based optimization
- Physical world adversarial attacks: adversarial patches, real-world evasion
- Transferability of adversarial examples across different models and architectures

- Hands-on Lab (Jupyter Notebook): Generating adversarial examples using popular attack frameworks
- Hands-on Lab (Jupyter Notebook): Testing adversarial robustness of production AI systems

Privacy Attacks and Information Extraction

Understanding how attackers can extract sensitive information from AI models, including membership inference and model inversion attacks.

- Membership inference attacks: determining if specific data was used in training
- Model inversion attacks: reconstructing training data from model parameters
- Property inference: extracting global properties about training datasets
- Model extraction and stealing: replicating proprietary models through queries
- Hands-on Lab (Jupyter Notebook): Conducting membership inference attacks against machine learning models
- Hands-on Lab (Jupyter Notebook): Implementing model inversion techniques to extract sensitive information
- Differential privacy fundamentals and implementation strategies for AI systems

Day 3: Secure AI Integration and Enterprise Defense

Securing AI-Integrated Applications

Practical security implementation for traditional applications that leverage AI models and services, including LLM integration patterns.

- Secure API integration patterns for AI services: authentication, rate limiting, input validation
- LLM integration security: prompt injection attacks, output validation, context isolation
- Building secure AI microservices: containerization, network isolation, monitoring
- Input sanitization for AI systems: handling untrusted data, format validation
- Hands-on Lab: Implementing secure LLM integration using the Hugging Face Inference API (Python/Flask, Java/Spring, ASP.Net, Node.js options)
- Hands-on Lab: Building input validation pipelines for AI-powered web applications in your chosen language

Enterprise AI Security Strategy

Comprehensive approach to building organizational AI security capabilities, including governance, monitoring, and incident response.

- AI security governance frameworks: risk assessment, policy development,

compliance
- Continuous monitoring for AI systems: model drift detection, anomaly identification
- AI security testing and red teaming: automated testing, adversarial validation
- Incident response for AI breaches: containment strategies, forensic analysis
- Hands-on Lab: Setting up AI security monitoring dashboards and alerting systems
- Hands-on Lab: Conducting AI security assessments and building remediation plans

Advanced Topics and Emerging Threats

Exploration of cutting-edge AI security challenges and future threat vectors.

- Large Language Model (LLM) specific attacks: jailbreaking, instruction following exploits
- Multi-modal AI security challenges: vision-language models, cross-modal attacks
- AI supply chain security: model provenance, dependency management
- Regulatory compliance for AI systems: GDPR, algorithmic auditing requirements

Course Wrap-up and Resources

- Next steps in your AI security journey
- Essential tools and frameworks for ongoing AI security work
- Building and maintaining AI security expertise within your organization
- Community resources and continued learning opportunities