AI Security Deep Dive (TTAI2800) Training

Duration: **3 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

Learn how to identify, exploit, and defend against modern AI and machine learning security threats while building secure AI-powered applications and enterprise defense strategies.

The **AI Security Deep Dive** course is a comprehensive three-day intensive training program designed for professionals who need to understand and defend against the rapidly evolving security challenges associated with artificial intelligence and machine learning systems.

As organizations increasingly integrate **AI**, **Machine Learning**, **LLMs**, and AI-powered applications into enterprise workflows, the attack surface expands dramatically, introducing vulnerabilities that traditional security models cannot fully address.

This expert-led course provides practical, hands-on experience in identifying and mitigating AI-specific threats including:

- **Data poisoning**
- **Adversarial attacks**
- **Model inversion**
- **Membership inference attacks**
- **Prompt injection**
- **LLM security vulnerabilities**
- **AI API security**
- **Privacy attacks**
- Enterprise AI governance and defense

Days 1 and 2 focus heavily on interactive **Jupyter Notebook labs**, while Day 3 emphasizes real-world AI integration, secure API development, and enterprise AI defense strategies.

Participants also gain practical integration experience using:

- **Python/Flask**
- **Java/Spring**
- **ASP.Net**
- **Node.js**

This course is ideal for professionals seeking advanced expertise in **AI security**, **secure AI development**, **offensive AI testing**, and enterprise AI risk management.

Technologies & Tools Covered

Participants will work with:

- **Python**
- **Jupyter Notebook**
- **Flask**
- **Java/Spring**
- **ASP.Net**
- **Node.js**

BILGINÇ IT Academy

info@bilginc.com

bilginc.com

- **Hugging Face Inference API**
- **OWASP AI Security Frameworks**
- **Differential Privacy Tools**
- **LLM Security Frameworks**

Prerequisites

To ensure a successful learning experience, participants should have:

- Basic programming knowledge
- Ability to read and understand code
- General cybersecurity awareness
- Basic understanding of HTTP and web applications
- Familiarity with data handling and statistical concepts
- Awareness of software development lifecycle (SDLC) practices

Who Should Attend

This course is ideal for:

- **Programmers and software developers**
- **Security analysts**
- **Cybersecurity professionals**
- **Technical leads**
- **DevSecOps engineers**
- **Security architects**
- **Incident response teams**
- **AI security engineers**
- Development teams building **AI-powered applications**

This training is especially valuable for professionals working in:

- **AI security**
- **LLM security**
- **AI risk management**
- **Secure AI development**
- **AI governance and compliance**

AI Security Deep Dive (TTAI2800) Training

Duration: **3 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

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

- Master **AI/ML security fundamentals**
- Identify attack vectors unique to AI systems
- Conduct and defend against **data poisoning attacks**
- Implement and mitigate **adversarial attacks**
- Perform **membership inference** and **model inversion attacks**
- Secure AI APIs and AI-integrated applications
- Prevent **prompt injection attacks**
- Apply **differential privacy** techniques
- Build enterprise-grade **AI security governance frameworks**
- Develop AI-focused incident response strategies

Training Outline

Day 1: AI/ML Foundations and Attack Fundamentals

1. AI/ML Security Foundations

- Understanding AI and machine learning from a security perspective
- Overview of the **OWASP Top 10** vulnerabilities
- AI system architectures and deployment models
- Training vs inference environments
- AI attack surface analysis
- Why traditional security approaches fail with AI systems

Hands-On Lab

- Setting up an AI security testing environment
- Exploring vulnerable machine learning models

2. Data Poisoning and Training Attacks

- **Targeted vs untargeted attacks**
- Clean-label attacks
- Training data vulnerabilities
- AI supply chain security risks
- Backdoor attacks in ML models
- Compromised pre-trained models

Hands-On Lab

- Performing poisoning attacks against image classifiers
- Poisoning attacks on text models

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- Building validation pipelines
- Detecting poisoned datasets

Day 2: Adversarial Attacks and Model Security

3. Adversarial Examples and Model Manipulation

- **FGSM and PGD attacks**
- Black-box vs white-box attacks
- Model evasion techniques
- Adversarial patches
- Transferability attacks across models

Hands-On Lab

- Generating adversarial examples
- Testing production AI robustness

4. Privacy Attacks and Information Extraction

- **Membership inference attacks**
- **Model inversion attacks**
- Property inference techniques
- Model extraction and model stealing
- Sensitive data leakage

Hands-On Lab

- Conducting membership inference attacks
- Implementing model inversion techniques
- Applying differential privacy defenses

Day 3: Secure AI Integration and Enterprise Defense

5. Securing AI-Integrated Applications

- Secure AI API integration
- Authentication and rate limiting
- **Prompt injection mitigation**
- Input and output validation
- Secure AI microservices
- Containerization and monitoring

Hands-On Lab

- Secure LLM integration using Hugging Face APIs
- Python/Flask, Java/Spring, ASP.Net, and Node.js integration options

BILGINC IT Academy

info@bilginc.com

bilginc.com

- Building secure AI-powered web applications

6. Enterprise AI Security Strategy

- AI governance frameworks
- AI risk assessment methodologies
- Continuous AI monitoring
- AI red teaming practices
- Incident response for AI breaches
- Building AI security dashboards

Hands-On Lab

- Configuring AI monitoring systems
- Conducting AI security assessments
- Creating remediation plans

7. Advanced Topics and Emerging Threats

- **LLM jailbreaking**
- Instruction-following exploits
- Multi-modal AI attacks
- AI supply chain security
- GDPR and AI regulatory compliance
- Algorithmic auditing