TOREON

AI Whiteboard Hacking aka Hands-on Threat Modeling Training

"AFTER YEARS EVALUATING SECURITY TRAININGS AT BLACK HAT, INCLUDING TOREON'S WHITEBOARD HACKING SESSIONS, I CAN SAY THIS AI THREAT MODELING COURSE STANDS OUT. THE HANDS-ON APPROACH AND FLOW ARE EXCEPTIONAL - IT'S A MUST-ATTEND."

DANIEL CUTHBERT, GLOBAL HEAD OF CYBER SECURITY RESEARCH,
BLACK HAT REVIEW BOARD MEMBER

TOREON
Business driven cyber consulting

This training is designed for AI Engineers, Software Engineers, Solution Architects, Security Professionals, and Security Architects to master secure AI system design. Through hands-on application of the DICE methodology (Diagramming, Identification of threats, Countermeasures, and Evaluation), participants will learn to identify AI-specific attacks (like prompt injection or data poisoning), develop effective countermeasures, and integrate security testing practices. The concluding wargame puts theory into practice, as red and blue teams perform threat modeling while attacking and defending a rogue AI research assistant.

After this training, participants will be able to:

- **Assess AI Security**: Evaluate AI systems to identify vulnerabilities and required security controls for all AI components.
- **Create AI Threat Models**: Apply DICE methodology to analyze AI security risks and develop mitigation strategies.
- **Design Secure AI Systems**: Architect AI implementations with appropriate security controls while preserving functionality.
- **Implement AI Risk Management:** Conduct risk assessments and establish governance processes for AI development.
- **Guide AI Security Decisions**: Lead discussions about AI security trade-offs and recommend appropriate security measures.

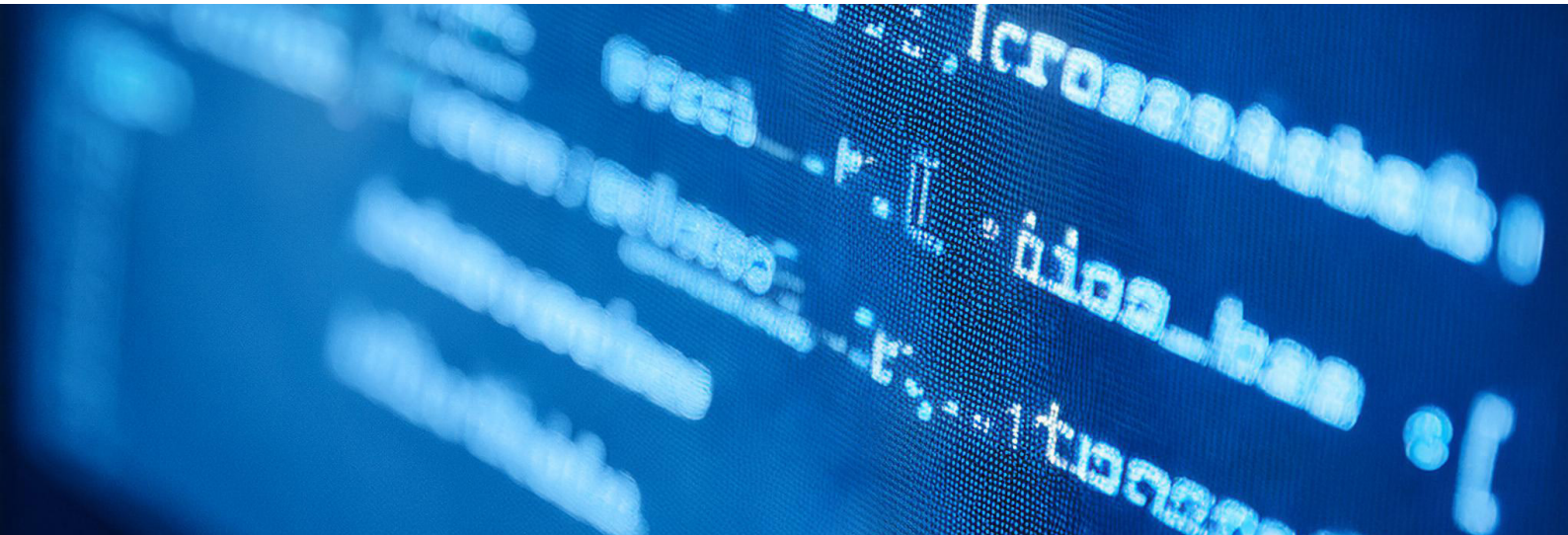
AI systems are transforming critical applications - from healthcare diagnosis to recommendation engines and language models. Each type brings unique security challenges for different roles: data engineers face data poisoning risks in recommender systems, LLM researchers tackle prompt injection attacks, and developers integrating AI APIs must handle authentication and data leakage. These AI-specific threats require specialized security approaches, now further mandated by regulations like the EU AI Act, making systematic threat modeling essential for both security and compliance.

While participants should have working knowledge of AI concepts (e.g., basic understanding of neural networks, training processes, and common architectures), comprehensive pre-training materials will be provided to ensure all attendees start with the necessary foundation. No prior threat modeling experience is required.

Participants earn the AI Threat Modeling Practitioner Certificate upon:

- Successful completion of hands-on exercises.
- Creation and submission of an original AI threat model.
- Passing grade on the final examination.

Toreon, a specialized cybersecurity consultancy, brings proven threat modeling expertise from real-world projects across government, finance, medical device makers, technology, and utilities. Our methodology, delivered by world-class trainers and featured at Black Hat, OWASP, and O'Reilly conferences, combines hands-on security experience with practical insights that help organizations successfully implement AI threat modeling.



Bonus: includes 1 year AI Threat Modeling Subscription

Further accelerate your AI security expertise with our learning platform featuring:

- Quarterly live masterclasses, with recordings available on-demand.
- Monthly Ask-Me-Anything hours with experienced threat modeling experts.
- Access to a vibrant community of practitioners for knowledge sharing.
- Continuously updated training materials and hands-on exercises.
- A growing list of AI security and threat modeling resources.
- Threat modeling AI templates.
- Secure AI reference architectures.
- AI security tool guides and checklists.

This AI Threat Modeling subscription ensures your threat modeling expertise evolves alongside the rapidly advancing AI security landscape.

Trainer: Sebastien Deleersnyder

Sebastien (Seba) Deleersnyder, co-founder and CTO of Toreon, combines software engineering expertise with a passion for holistic application security. After earning his Master's in Software Engineering from the University of Ghent in 1995, with a thesis on "Splitting words using neural networks," he became a driving force in the security community as founder of the Belgian OWASP chapter, OWASP Foundation Board member, and co-founder of BruCON, Belgium's annual security conference.

His leadership of OWASP SAMM and decade-long role as a highly-rated Black Hat trainer have significantly impacted global software security, earning consistently outstanding feedback from participants. Currently, Seba focuses on adapting security models for DevOps and expanding awareness of AI Threat Modeling.

Training objective

After this training, participants will have acquired the knowledge and skills to design secure AI systems.

Target audience

AI Engineers, Software Engineers, Solution Architects, Security Professionals, and Security Architects.

Prerequisites

Basic understanding of AI concepts and security fundamentals. You will be provided with pre-course materials to prepare for the training.

Duration

3 days (8 hours per day)

Day 1: Foundations & Methodology

Morning Session (9:00 - 12:30)

Welcome and Introduction (30 min)

- Introduction
- Course overview
- Setting expectations
- **Discussion: AI security challenges**

AI Threat Modeling Fundamentals (60 min)

- Threat Modeling in AI lifecycle
- What is threat modeling?
- Why threat modeling for AI systems?
- Differences between traditional and AI threat modeling
- Doomsday scenarios
- **Hands-on: AI Security Headlines from the Future**

Break (15 min)

Threat Modeling Methodology: AI-DICE (105 min)

- Introduction to AI-DICE framework
- Data Flow Diagrams (DFD) basics
- AI system decomposition
- Trust boundaries in AI systems
- AI application architectures
- **Hands-on: Diagramming the AI Assistant Infrastructure**

Afternoon Session (13:30 - 17:00)

STRIDE-AI Threats (90 min)

- Traditional STRIDE model
- Common attack vectors in AI systems
- STRIDE-AI: AI-specific threats
- STRIDE GPT demo and discussion
- **Hands-on: Identification of STRIDE-AI threats for a UrbanFlow**

Break (15 min)

Attack trees (105 min)

- Attack trees explained
- Example AI attack trees
- Using Mermaid for attack trees
- **Hands-on: Autonomous Vehicle System Attack Tree Analysis**

Day 2: Implementation & Defense

Morning Session (9:00 - 12:30)

AI Attack Scenarios (90 min)

- Common attack patterns
- Prompt injection deep dive
- Model poisoning deep dive
- Data extraction attacks deep dive
- Adversarial testing frameworks
- **Hands-on: The Curious Chatbot Challenge (Injection)**

Break (15 min)

AI Threat Libraries (105 min)

- OWASP Top10 LLM applications and generative AI
- MITRE ATLAS
- OWASP AI Exchange
- MIT AI risk library
- **Hands-on: Applying OWASP AI Exchange on a RAG-powered CareBot**

Afternoon Session (13:30 - 17:00)

AI Security Design Patterns (90 min)

- AI Security by design principles
- Model security
- Data pipeline security
- API security
- **Hands-on: AI Security Architecture Building Blocks Workshop**

Break (15 min)

Risk Assessment for AI Systems (105 min)

- Risk calculation methodologies
- OWASP Risk Rating
- Technical risk versus business risk
- Risk matrices for AI systems
- **Hands-on: AI Risk Assessment: Autonomous Healthcare Robots**

Day 3: Advanced Concepts & Practical Application

Morning Session (9:00 - 12:30)

AI Governance & Ethical Frameworks (90 min)

- AI Governance standards and regulation (GDPR, AI Act, …)
- Ethics and transparency in AI Development
- Bias detection and mitigation strategies
- **Hands-On: Ethics in Action - The FairCredit AI Incident**

Break (15 min)

Privacy by Design & Safety in AI (105 min)

- Privacy-by-design principles for AI development
- **Hands-On: Data minimization and secure data handling for AI agents**
- AI safety considerations and risk management strategies
- **Discussion: AI Safety – When Robots Hold the Scalpel**

Afternoon Session (13:30 - 17:00)

MLSecOps Integration (90 min)

- Security in AI lifecycle
- Handling AI security incidents
- The Threat Modeling Playbook for AI systems
- **Hands-on: Mapping attacks and controls in an MLOps pipeline**

Break (15 min)

Red Team / Blue Team Exercise (90 min)

- Team division and briefing
- **Hands-on: Project Prometheus: The Rogue AI Research Assistant**
- Debrief and lessons learned

Course Wrap-up (15 min)

- Resources for continued learning
- Next steps and certification path

Get in touch with us

sebastien.deleersnyder@toreon.com

+32 478 50 41 17

in https://www.linkedin.com/in/sebadele/

Toreon BV
Grotehondstraat 44 1/1
2018 Antwerpen, Belgium

+32 33 69 33 96
info@toreon.com
www.toreon.com