AI Security

26HS

Programme	MSc in Business Information Systems
Degree	Master
ECTS	6
Module type	elective module
Module coordinator	Dr. Pascal Moriggl
Compulsory attendance	Attendance is recommended
Leading principle / Short description	This module follows a flipped classroom approach with a strong practical focus, enabling students to develop and apply AI security skills through hands-on exercises, while foundational concepts are acquired through guided self-study. The amount of powerpoint and text is kept low, and the amount of hands-on experience is maximized. Leading principles in cybersecurity are not adressed in this course, but are a fundamental part of the underlying theory that enables the concepts shown in this course.
Module content	This module will include the following topics 1. LLM and Agents Infrastructure Setup 2. LLM and Agents Basic Interactions 3. Traditional Cybersecurity and AI 4. AI-dedicated Security [... on Infrastructure, Model, Interaction] 5. Attacker-side AI [... on regular systems, on AI setups] 6. Defender-side AI [... on regular systems, on AI setups]
Competencies to be achieved	**Knowledge and Understanding** • Students **explain** the fundamental concepts of AI security across infrastructure, models, and interactions. • Students **describe** common vulnerabilities in LLMs, agents, and AI-integrated systems. • Students **identify** relationships between traditional cybersecurity principles and AI-specific security challenges. **Applying Knowledge and Understanding** • Students **apply** AI security techniques to configure, test, and secure LLM and agent-based systems. • Students **analyze** attack vectors on both conventional and AI-driven systems using practical scenarios. • Students **implement** defensive mechanisms leveraging AI for threat detection and mitigation. **Making Judgements** • Students **evaluate** the effectiveness of security measures in AI and non-AI environments. • Students **assess** risks, trade-offs, and limitations of AI-based attack and defense strategies. • Students **critically justify** security design decisions in AI systems based on practical evidence and testing outcomes.
Prerequisites	none
Teaching and learning methods	**Contact Hours:** On-site hands-on activities focused on guided exercises, system setup, attack and defense simulations, and iterative problem-solving. **Guided Self-Study:** Preparation through structured readings, tutorials, and short exercises prior to class sessions Lab activities as graded homework, requiring independent implementation, testing, and documentation of AI security tasks
Literature	Required Reading: Ken Thompson. 1984. Reflections on trusting trust. Commun. ACM 27, 8 (Aug 1984), 761–763. https://doi.org/10.1145/358198.358210
Remarks	This module follows the structure of preparing ahead in time for the upcoming class to then maximizing in-class practical experience with AI from a security perspective. The lab component serves as the central assessment element,requiring continuous application of concepts. Prior completion of modules the "GRC", "Risk and Compliance Lab" or "Cybersecurity in Health" is beneficial but not mandatory

Grading

Grade 1-6 (half grades)

Assessment

Assignments 100%	
Oral / Written	Written
Grading Scale	Grade 1-6
Remarks	The course includes several assignments that are part of the lab activities (homework).

Module details

26HS.AI Security - Tue - Olten	
Time	8:15 AM - 12:00 PM
Language	Englisch
Periodicity	Weekly
Lecturers	Dr. Pascal Moriggl
Number	3-26HS.W-M-BIS-E-AIS.EN/a