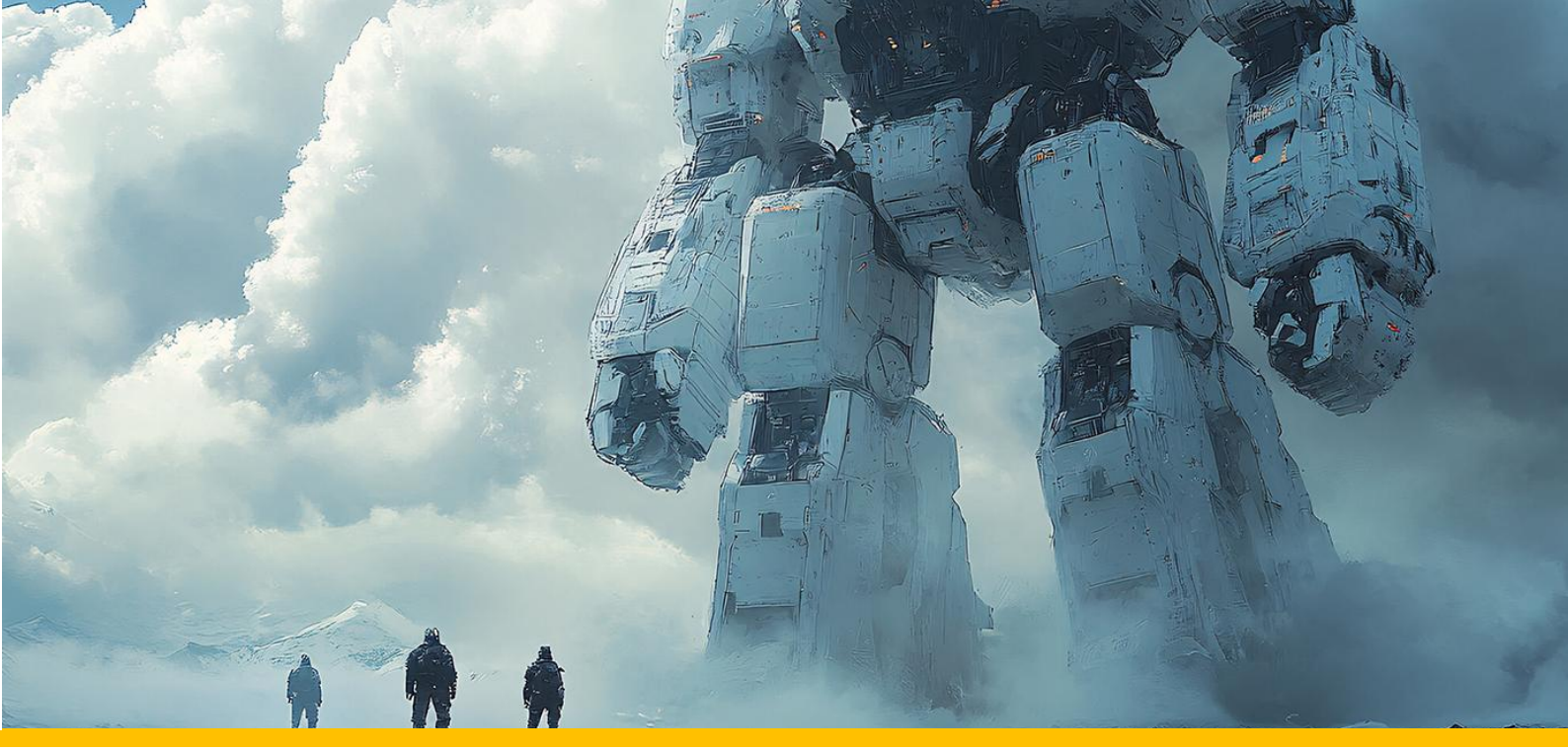

Short Advanced Studies (SAS)[1]

AI Security

Artificial Intelligence is being adopted and leveraged by businesses, and seamlessly integrated into various tools, products, and platforms for improving efficiency and providing additional capabilities. But the new risks and security of these systems and the information they process is different from the traditional IT/Cybersecurity domain.

This SAS is designed for security analysts, software architects, and system engineers with an interest in the security of Artificial Intelligence infrastructure.

Students will receive a practical introduction and overview to the topic of AI security and risk. They will explore the challenges around AI security, learn how vulnerabilities can be exploited, and discover effective strategies to safeguard these systems, ensuring reliable performance without unintended consequences.

[1]Short Advanced Studies (SAS) are short qualifying training courses designed for a specialist audience seeking to face new challenges in direct dialogue with experts (1-9 ECTS).

Berner
Fachhochschule

Table of Contents

1 Portrait

The adoption of artificial intelligence in business is growing rapidly, leading to new challenges and opportunities in cyber security. This five-day course provides participants with an introduction to the skills needed in the field of AI security, incorporating academic papers, technical reports, and other relevant resources throughout the course. The curriculum is organized into five parts:

1. Foundations: Participants will explore the AI lifecycle and understand the unique security challenges presented by its probabilistic nature.
2. Threat Landscape: Strategies for penetration testing and defending AI models, and the complexities involved in "patching" them.
3. Advanced AI Threats: Examination of sophisticated threats such as data poisoning and the implications of external connections via the Model Connection Protocol (MCP).
4. AI & Security incidents: Overview of investigating incidents involving AI systems, and how AI systems can help security investigations.
5. Governance and Human-AI Interaction: Discussion of the importance of policy, user dynamics, and the risks of over-reliance on AI, along with methods to enhance safety in human-AI interactions.

2 Career opportunities

- AI Security Analyst: Assessing and mitigating security risks in AI systems.
- AI Security Architect: Designing secure architectures for AI applications.
- Data Scientist with a focus on Security: Integrating security best practices into machine learning workflows.

3 Target audience

This course is designed for staff and managers in security, IT, and operations roles who are interested in the fundamentals of AI security. In addition, AI developers and project leaders can benefit. It is especially relevant for individuals with a technical background seeking to understand the current landscape of AI security practices.

4 Education goals

- Gain a solid understanding of the basic principles of AI security, including key terminology and frameworks.
- Recognize common vulnerabilities and attack vectors associated with AI systems.
- Evaluate existing security measures in AI and understand best practices for mitigating risks.
- Familiarize themselves with relevant academic literature and technical reports, improving their ability to stay informed about the latest developments in AI security.

5 Requirements

This course is designed for beginners in the field of AI security. Familiarity with basic AI concepts is assumed, AI security and risk knowledge is not required, but it will enhance understanding of the course content. Participants are expected to have:

- Basic IT security knowledge: understanding of common web vulnerabilities and IT terminology
- Basic AI knowledge: understanding of models and prompts
- A laptop or notebook with internet access for course activities

6 Factsheet

Short Advanced Studies (SAS)	AI Security
Degree/Certificate	Short Advanced Studies (SAS) in AI Security
Duration	5 days
Application deadline	Up to 1 month before the start of the course
ECTS credits	4 ECTS Credits
Costs	CHF 2'800
Teaching language	English
Location	Biel, Aarbergstrasse 46

7 Content + Learning objectives

7.1 AI foundations and attack surface

- AI fundamentals: Machine Learning, Deep Learning, and Generative AI. How LLMs predict the next to-ken, and why that matters for security.
- "Black box" problem: Understanding non-determinism and the lack of explainability.
- Attack surface: Where things go wrong: data poisoning, model theft, and supply chain vulnerabilities

7.2 AI threat Landscape

- Adversarial Machine Learning (AML): How some noise can fool a classifier, and real-world impact on systems like autonomous vehicles.
- LLM-Specific Threats: Prompt Injection and Jailbreaking, plus advanced techniques like the Crescendo Attack.
- Intro to MCP: the standard that allows to connect AI models to external data and tools.
- Examples of threats: prompt injection via data, the "Confused Deputy" problem.

7.3 Attacking and defending AI models

- Why "patching" a model is hard? Look at the current state of defensive and offensive approaches.
- **Lab**: workshop where students will try to "break" AI models with the techniques learned.

7.4 AI & Security incidents

- How to investigate and detect a compromised model?
- Analyzing prompt logs, model checksums, vector database states, and “thought chains”.
- Leveraging AI for automating forensic tasks and accelerating analysis.

7.5 Governance and Human-AI Interaction

- Security is also about policy, people, and how they interact with AI.
- Why is it a problem if humans trust AI too much? How to design it for safety?

8 Proof of proficiency

A multiple-choice test will be conducted to evaluate participants’ proficiency in the course material. Students who prefer not to provide proof of proficiency will still receive a course certificate upon completion.
Students who do not wish to provide proof of proficiency will receive a course certificate.

9 Lecturer

Name	Organisation	E-mail
Andrea Palmieri	SWITCH	

10 Organisation

SAS Supervisor:
Prof. Dr. Bruce Nikkel
E-mail: bruce.nikkel@bfh.ch
Threema: DC2JN4YK
Mobile: +41 79 255 63 16

SAS coordination:
Miriam Patwa
Phone: +41 31 848 58 68
E-mail: miriam.patwa@bfh.ch

Bern University of Applied Sciences
School of Engineering and Computer Science
Continuing Education
Aarbergstrasse 46 (Switzerland Innovation Park Biel/Bienne)
2503 Biel/Bienne

Phone +41 31 848 31 11
E-mail: weiterbildung.ti@bfh.ch

bfh.ch/ti/weiterbildung

Berner
Fachhochschule