Short Advanced Studies (SAS)[1]

Generative KI sicher entwickeln und betreiben

Generative KI ermöglicht eine enorme Wertschöpfung und neue Chancen, schafft aber auch neue Risiken und Verantwortungen. Lernen Sie, KI-Systeme sicher, regelkonform und verantwortungsvoll einzusetzen.

Sie erwerben in diesem Lehrgang praxisnahes Wissen und Handlungskompetenzen zu AI Security, Governance, Regulatorien, Compliance und Datensouveränität für den Einsatz in ihrem Unternehmen.

[1]Short Advanced Studies (SAS) sind praxisnahe Kurzweiterbildungen, die sich an ein Fachpublikum richten, das sich im direkten Dialog mit Expert*innen neuen Herausforderungen stellen möchte (1–9 ECTS).

Berner
Fachhochschule

Inhaltsverzeichnis

Stand: 02.07.2026

Berner
Fachhochschule

1 Portrait

Generative KI-Anwendungen sind in vielen Unternehmen bereits unverzichtbar im Einsatz und nehmen rasant zu. Damit steigen auch die Risiken und Anforderungen an Sicherheit, Governance, Compliance, und Daten Souveränität.

Das SAS Generative KI sicher entwickeln und betreiben vermittelt die wesentlichen Konzepte, Methoden und regulatorischen Rahmenbedingungen für einen sicheren und nachhaltigen KI-Einsatz. Die Teilnehmenden lernen, Risiken zu erkennen, geeignete Schutzmassnahmen zu definieren und Governance-Strukturen für den produktiven Einsatz von KI aufzubauen.

Der Lehrgang verbindet technische, organisatorische und regulatorische Perspektiven und richtet sich an Fach- und Führungskräfte, die Verantwortung für KI-Initiativen übernehmen.

2 Berufsperspektiven

Die erworbenen Kompetenzen sind insbesondere für Organisationen relevant, die KI-Anwendungen entwickeln, beschaffen oder produktiv einsetzen. Der Lehrgang unterstützt Fach- und Führungskräfte bei der Übernahme von Verantwortung in den Bereichen

- AI Governance und AI Compliance
- Data & AI Management
- Datenschutz und Regulatotorischer Rahmen
- Digitale Transformation und Innovations-Management
- Projekt- und Programmleitung von KI-Initiativen

3 Zielpublikum

Der Lehrgang richtet sich an

- Fach- und Führungskräfte, die KI-Initiativen verantworten
- Data Scientists, AI Engineers, AI Developers
- Projekt- und Programmleitende
- Compliance-, Risiko- und Datenschutzverantwortliche
- Cyber-Security-Verantwortliche

4 Ausbildungsziele

Die Teilnehmenden

- verstehen die wichtigsten Sicherheitsrisiken moderner KI-Systeme und können geeignete Schutzmassnahmen ableiten.
- kennen relevante Governance-, Compliance- und Risikomanagement-Frameworks für KI-Anwendungen.
- können Anforderungen des EU AI Act sowie wichtiger Standards wie NIST AI RMF und ISO/IEC 42001 auf konkrete Anwendungsszenarien übertragen.
- beurteilen Herausforderungen bezüglich Datensouveränität, Cloud-Abhängigkeiten und verantwortungsvollem KI-Einsatz.
- sind in der Lage, Sicherheits- und Governance-Aspekte bei der Planung, Einführung und dem Betrieb von KI-Systemen systematisch zu berücksichtigen.

5 Voraussetzungen

Für die Teilnahme werden keine vertieften technischen Spezialkenntnisse vorausgesetzt.
Von Vorteil sind:

- Grundverständnis moderner Informations- und Kommunikationstechnologien
- Erfahrung im Umgang mit digitalen Systemen oder Datenprojekten
- Interesse an künstlicher Intelligenz, Cyber Security, Compliance oder Governance-Fragestellungen
- Erste Erfahrungen mit KI-Anwendungen wie ChatGPT, Copilot oder vergleichbaren Systemen

6 Steckbrief

Short Advanced Studies (SAS)	Generative KI sicher entwickeln und betreiben
Titel/Abschluss	Short Advanced Studies (SAS) Generative KI sicher entwickeln und betreiben
Anzahl ECTS	1 ECTS-Credit
Kosten	CHF 1'200
Anmeldefrist	Bis 1 Monat vor Kursbeginn
Unterrichtssprache	Deutsch / Literatur und Unterlagen teilweise in Englisch
Studienort	Aarbergstrasse 46, Biel / hybrid

7 Inhalte und Lernziele

1.1 AI Security

Lernziele	Die Teilnehmenden - verstehen die spezifischen Sicherheitsrisiken, die beim Einsatz generativer KI-Systeme im Unternehmenskontext entstehen, und können diese von klassischen IT-Sicherheitsrisiken abgrenzen. - kennen grundlegende Prinzipien für den sicheren Aufbau von KI-Architekturen (Identity- und Secret-Management, Logging & Monitoring, Defense-in-Depth) und können diese auf reale Szenarien anwenden. - verstehen Angriffsvektoren wie Prompt Injection und Confused-Deputy-Probleme und kennen geeignete Gegenmassnahmen (Guard Rails, Gateways). - kennen Methoden, wie KI-Systeme selbst für Security-Aufgaben eingesetzt werden können – etwa zur Schwachstellenanalyse oder zum Testen von Softwaresystemen.

	– können die Sicherheitsanforderungen an MCP-Server-Architekturen einschätzen und kennen relevante Härtungsmassnahmen (u. a. für Vector Databases). – sind in der Lage, Grundzüge eines Incident-Response-Prozesses für KI-Systeme zu beschreiben.
Themen und Inhalte	Bedrohungsszenarien – Generative KI-Systeme bringen neuartige Angriffsflächen mit sich. Zentrale Vektoren sind Prompt Injection (direkt via Nutzereingabe oder indirekt über verarbeitete Daten) sowie das Confused-Deputy-Problem, bei dem ein KI-Agent mit weitreichenden Berechtigungen durch manipulierte Eingaben unerwünschte Aktionen ausführt. Sichere KI-Architektur – Aufbau einer gesicherten, unternehmensinternen KI-Architekturschicht mit KI Gateways, Identity- und Secret-Management (Least Privilege), Guardrails sowie Defense-in-Depth als Grundprinzip. MCP-Security und Vector-DB-Hardening – MCP-Server sind privilegierte Komponenten und benötigen Authentifizierung, Autorisierung und Sandboxing. Vector Databases sind anfällig für Data Poisoning. Zugriffskontrollen und Dokumenten-Integritätsprüfungen sind zentrale Gegenmassnahmen. Monitoring, Logging und Incident Response – Protokollierung von Prompts und Modellantworten, Anomalieerkennung sowie definierte Incident-Response-Prozesse (Isolation, Forensik, Eskalation) für KI-spezifische Vorfälle. KI als Security-Werkzeug – Einsatz generativer Modelle für automatisierte Code-Reviews, Penetrationstest-Szenarien und Log-Analyse inklusive Auseinandersetzung mit den Grenzen und Risiken dieses Ansatzes.
Lehrmittel	– Folien/Skript – Praktische Übung, Laptop erforderlich

2.1 Ethics, Regulations & Sovereignity

Lernziele	Die Teilnehmenden – Kennen verbindliches Regulierungswerk wie bspw. den EU AI Act und verstehen dessen konkrete Auswirkungen auf Engineering- und Architekturentscheidungen.

	- kennen NIST AI RMF und ISO/IEC 42001 als praxisorientierte Managementrahmen und Normen für den verantwortungsvollen KI-Einsatz. - können KI-Systeme nach Risikoklassen einordnen und wissen, welche Compliance-Anforderungen sich daraus ergeben. - verstehen die ethischen Herausforderungen beim Einsatz generativer KI, insbesondere Automation Bias, Trust Calibration und Deepfakes. - kennen die Konzepte Datensouveränität, Modellherkunft und Auftragsverarbeitung und können diese auf den Unternehmenskontext anwenden. - sind sensibilisiert für Abhängigkeitsrisiken gegenüber US-amerikanischen Cloud-Anbietern und kennen Strategien für einen souveränen KI-Betrieb.
Themen und Inhalte	EU AI Act - Weltweit erstes verbindliches Regulierungswerk für KI, relevant auch für Schweizer Unternehmen mit EU-Marktbezug - Klassifizierung von KI-Systemen nach Risikoklassen (minimal, begrenzt, hoch, inakzeptabel) mit entsprechenden Pflichten für Entwicklung, Dokumentation und Betrieb Managementrahmen und Normen - NIST AI RMF: praxisorientierter Rahmen für das Risikomanagement von KI-Systemen (Govern, Map, Measure, Manage) - ISO/IEC 42001: internationale Norm für KI-Managementsysteme, vergleichbar mit ISO 27001 für Informationssicherheit Ethik und Mensch-KI-Interaktion - Automation Bias: Tendenz, KI-Empfehlungen unkritisch zu übernehmen – mit potenziell schwerwiegenden Folgen in risikoreichen Anwendungsfeldern - Trust Calibration: Vertrauen von Nutzenden an die tatsächliche Zuverlässigkeit des Systems anpassen Deepfakes und synthetische Medien - Generative KI ermöglicht täuschend echte Bild-, Audio- und Videofälschungen mit wachsendem Missbrauchspotenzial - Erkennungsmethoden (Detection) und organisatorische Massnahmen als Teil des verantwortungsvollen KI-Einsatzes Datensouveränität und Auftragsverarbeitung - Provider-Rechte: Nutzungsbedingungen regeln, ob Prompts für Modelltraining verwendet werden dürfen - Auftragsverarbeitungsverträge (AVV) als datenschutzrechtliche Anforderung, inkl. Modellherkunft und Datenstandorte Technologische Souveränität

	- Strategische Abhängigkeit von US-amerikanischen Grossanbietern: Risiken bei Verfügbarkeit, Preisgestaltung und Datenschutz - Alternativen: Open-Weight-Modelle (z. B. Llama, Mistral), europäische Cloud-Anbieter, Eigenbetrieb auf unternehmenseigener Infrastruktur
Lehrmittel	- Folien/Skript

8 Kompetenznachweise

Bewertete Übungsaufgaben während des Kurses.

9 Dozierende

Vorname Name	Firma	E-Mail
Marius Bleif	eraneos / FPV	marius.bleif@bfh.ch

Organisation

SAS-Leitung:
Arno Schmidhauser
Tel: +41 31 848 32 75
E-Mail: arno.schmidhauser@bfh.ch

SAS-Administration:
Andrea Moser
Tel: +41 31 848 32 11
E-Mail: andrea.moser@bfh.ch

Berner Fachhochschule
Technik und Informatik
Weiterbildung
Aarbergstrasse 46 (Switzerland Innovation Park Biel/Bienne)
2503 Biel/Bienne

Telefon +41 31 848 31 11
E-Mail: weiterbildung.ti@bfh.ch

bfh.ch/ti/weiterbildung

Berner
Fachhochschule