



Certificate of Advanced Studies

Generative KI

Erwerben Sie fundierte Kompetenzen in den modernsten KI-Technologien. Entwickeln Sie damit innovative Lösungen und setzen Sie praxisorientierte Projekte um. Lernen Sie Deep-Learning-Architekturen wie GAN und Transformer kennen und wie diese neue KI-Anwendungen wie Bildgeneratoren, Large Language Models (LLMs) und neue Informationssysteme ermöglichen.

Inhaltsverzeichnis

1	Umfeld	3
2	Zielpublikum	3
3	Ausbildungsziele	3
4	Voraussetzungen	3
5	Unterrichtssprache	4
6	Durchführungsort	4
7	Kosten	4
8	Kompetenzprofil	5
9	Kursübersicht	6
10	Kursbeschreibungen	7
	10.1 Einführung in Maschinelles Lernen & Deep Learning	7
	10.2 ML-Engineering und Technologien	7
	10.3 Vorbereitung Lab-Infrastruktur (Optional)	8
	10.4 Einführung in Textverarbeitung und Textanalyse	8
	10.5 Einführung in Bildverarbeitung	9
	10.6 Large Language Models verstehen und nutzen: Architektur, Methoden und Praxis	9
	10.7 Generative KI für Bilder	11
	10.8 Semesterarbeit	12
11	Kompetenznachweis	12
12	Lehrmittel	12
13	Dozierende	13
14	Organisation	13

Stand: 18.02.2025

1 Umfeld

Generative künstliche Intelligenz (KI) revolutioniert die Datenwissenschaften. Sie treibt die Entwicklung neuartiger Informations- und Kommunikationssysteme voran, übernimmt Aufgaben, die bisher ausser Reichweite von IT-Systemen waren und unterstützt uns bei Alltagsaufgaben ebenso wie beim kreativen und innovativen Arbeiten. Daten und Informationen aus klassischen IT-/OT-Systemen in den Unternehmen können mit KI verknüpft und ausgewertet werden. In allen Branchen und Berufen entstehen unterstützende und agierende Systeme, die sich wesentlich auf generative KI abstützen.

2 Zielpublikum

Dieses CAS richtet sich an

- IT-Engineers, Entwickler*innen und alle, die in der Umsetzung von Lösungen mit generativer KI auf dem neusten Stand sein wollen.
- Mitarbeitende in Data-Science-Teams, die ihre Fähigkeiten im Bereich der künstlichen Intelligenz ausbauen und vertiefen wollen.

3 Ausbildungsziele

Dieses CAS vermittelt ein tiefes Verständnis generativer Modelle und Architekturen. Die Teilnehmenden lernen Theorie, Konzepte und praktische Anwendungen kennen, um KI-gestützte Innovationen in ihren Projekten zu nutzen.

Nach Abschluss dieses Moduls sind Sie in der Lage:

- die Grundprinzipien des maschinellen Lernens (ML) und Deep Learnings (DL) zu verstehen und anzuwenden.
- die verschiedenen Arten des maschinellen Lernens (überwacht, semi-überwacht, unüberwacht) zu unterscheiden und zu nutzen.
- die Funktionsweise komplexer neuronaler Netzwerke wie Feedforward Neural Networks (FNN), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) und Transformer-Architekturen zu verstehen und in eigenen Anwendungen zu nutzen.
- die Prinzipien und Eigenschaften generative Modelle wie Autoencoder, GANs, Stable Diffusion und deren Anwendung in verschiedenen Bereichen zu verstehen.
- die Arbeitsweise von Large Language Modellen zu verstehen und damit neue Anwendungen im Unternehmen zu entwickeln.
- mit Python-basierten Tools für die Implementierung und das Training von generativen Modellen zu arbeiten.
- die Grundlagen der Verarbeitung natürlicher Sprache (NLP) und deren Anwendungen in Textklassifikation, Frage-Antwort-Systemen und Textgenerierung zu verstehen.
- Bildverarbeitungsmethoden und deren Anwendungen in der Bildklassifikation und Segmentierung zu implementieren.
- ein selbstdefiniertes, praktisches Projekt im Bereich generative KI eigenständig zu planen und durchzuführen.

4 Voraussetzungen

- Grundkenntnisse im Bereich des maschinellen Lernens, supervised & unsupervised Learning, Feature Engineering
- Grundkenntnisse in Python oder anderer Programmiersprache
- Für Studierende des MAS Data Science: [CAS Practical Machine Learning](#) besucht oder entsprechende Vorkenntnisse

Für die Durchführung von Übungen und Projektarbeiten wird mit [Google Colab](#), Jupyter Notebooks o.ä. und Open AI (kostenfreie Version) gearbeitet. Für die Semesterarbeit fallen ev. Kosten für Pro-Versionen oder Lizenzen mit mehr Optionen an, je nach eingesetzten Produkten.

5 Unterrichtssprache

Die Unterrichtssprache ist Deutsch, die Unterlagen sind teilweise in Englisch.

6 Durchführungsort

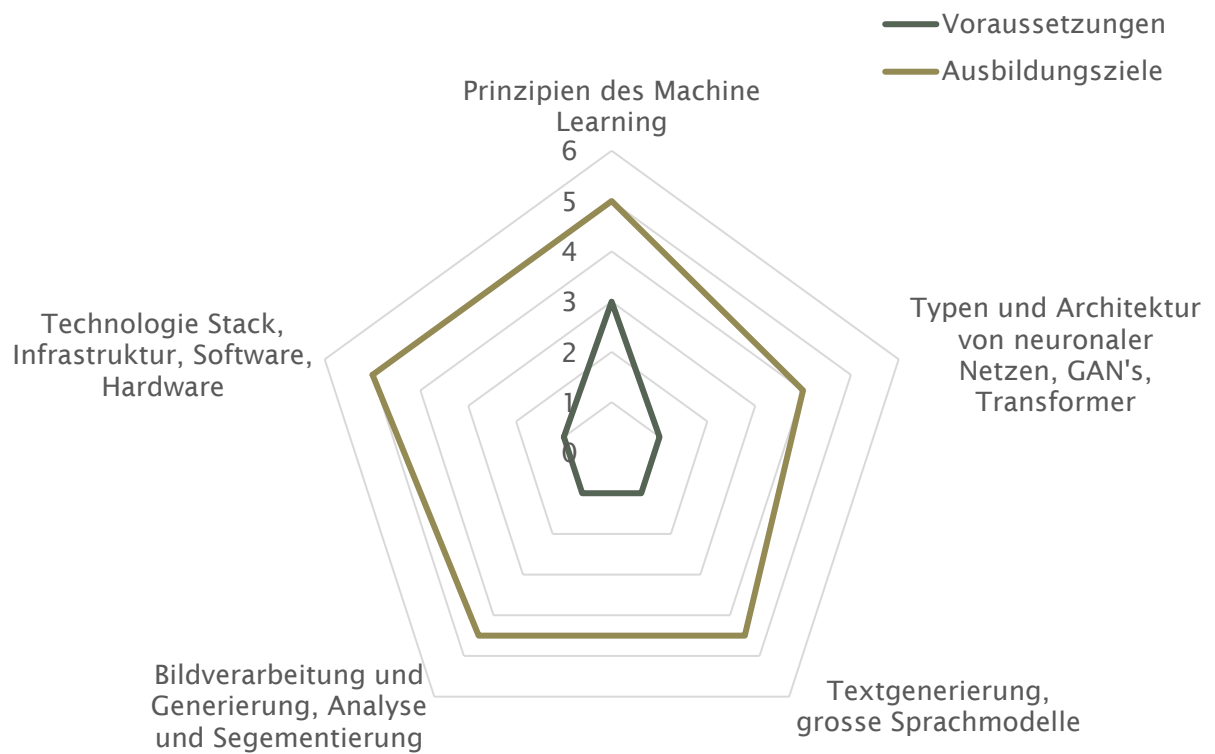
Berner Fachhochschule, Weiterbildung, Aarbergstrasse 46 (Switzerland Innovation Park Biel/Bienne), 2503 Biel/Bienne,
Telefon +41 31 848 31 11, E-Mail weiterbildung.ti@bfh.ch.

7 Kosten

Siehe [Durchführungsdaten und Preise](#)

Als Entwicklungs- und Trainingsplattform werden u.a. kostenpflichtige Clouddienste von openAI, HuggingFace und weitere genutzt. Rechnen Sie Kosten von CHF 100 bis 500 ein, insbesondere auch je nach Thema der Semesterarbeit.

8 Kompetenzprofil



Kompetenzstufen

1. Kenntnisse/Wissen
2. Verstehen
3. Anwenden
4. Analyse
5. Synthese
6. Beurteilung

9 Kursübersicht

Nr	Kurs	Tage	Stunden	Dozierende
1	Einführung in ML & Deep Learning	1		Jürgen Vogel
2	Engineering und Technologien	1		Peter von Niederhäusern
3	Lab Vorbereitungstag (optional)	1		Peter von Niederhäusern
4	Einführung in Textverarbeitung und Textanalyse	2		Jürgen Vogel
5	Einführung in Bildverarbeitung	2		Marcus Hudritsch
6	Large Language Models verstehen und nutzen: Architektur, Methoden und Praxis	5		Yannis Schmutz, Jürgen Vogel
7	Generative KI für Bilder	5		Peter von Niederhäusern
8	Semesterarbeit	1	90	Expert*innen aus Lehre und Forschung, Wirtschaft und Praxis.
	Total	18		

Das CAS umfasst insgesamt 12 ECTS-Credits (300–360 Stunden Workload). Für die einzelnen Kurse ist entsprechend Zeit für Selbststudium, Prüfungsvorbereitung etc. einzurechnen.

10 Kursbeschreibungen

10.1 Einführung in Maschinelles Lernen & Deep Learning

Am ersten Kurstag starten wir mit einer Übersicht zu den wichtigsten Methoden, Algorithmen und Herausforderungen in der Entwicklung, der Evaluierung und dem Betrieb von Lösungen auf Basis von maschinellem Lernen und generativer KI.

Lernziele	Die Teilnehmenden frischen ihr Vorwissen zu maschinellem Lernen auf und erhalten einen ersten Überblick zu den Besonderheiten von Deep Learning und Generativer KI.
Themen und Inhalte	<ul style="list-style-type: none">- Grundlagen zu maschinellem Lernen<ul style="list-style-type: none">- ML-Algorithmen für überwachtes, semi-überwachtes und unüberwachtes Lernen- Vorbereitung von Daten- Feature Engineering- Trainieren und Evaluieren von ML-Modellen- Neuronale Netze: FNN, RNN- Deep Learning: CNN, LSTM, Transformer, GAN- Generative KI
Lehrmittel	<ul style="list-style-type: none">- Folien/Skript- Literaturempfehlung Nr 1- Praktische Übungen mit Python (Jupyter Notebooks, Scikit-Learn, PyTorch)

10.2 ML-Engineering und Technologien

Sie lernen die wichtigsten Tools und Frameworks für maschinelles Lernen und KI kennen. Sie erfahren, wie Sie die richtigen Tools auswählen, diese in bestehende Systeme einbinden und Ihre KI-Modelle optimieren können. Zusätzlich behandeln wir kurz Entwicklungsumgebungen, Cloud-Dienste und skalierbare Infrastrukturen für Ihre KI-Projekte.

Lernziele	Die Teilnehmenden verstehen die grundlegenden Werkzeuge, Tools und die gängigen Frameworks für die Entwicklung und das Training neuronaler Netze. Sie können entscheiden, ob Modelle in der Cloud oder lokal gehostet werden sollen und kennen die relevanten Kriterien für diese Auswahl.
Themen und Inhalte	<ul style="list-style-type: none">- Technology Stack, Tools,- Google Colab- Einblick in andere Entwicklungs-Frameworks- PyTorch, TensorFlow, Transformers- High-Performance-Computing:<ul style="list-style-type: none">o On-Premiseo Cloudo Experimentell (bspw. Colab)- Huggingface als ModelHub- Evaluierung, Deployment (TensorBoard, TFX, Weights & Biases, MLFlow)
Lehrmittel	<ul style="list-style-type: none">- Folien/Skript

10.3 Vorbereitung Lab-Infrastruktur (Optional)

Dieser Kurstag ist dem Einrichten der persönlichen Arbeitsgeräte (Laptops) resp. der Arbeitsumgebung gewidmet. Sind Sie schon vertraut mit diesen Themen oder verfügen Sie bereits über eine lauffähige Umgebung, so ist dieser Kursteil optional.

Lernziele	Die Teilnehmenden haben eine funktionierende Entwicklungsumgebung, mit welcher im Anschluss die Übungen durchgeführt werden. Sie kennen zudem die Vor- und Nachteile der Anbieter von KI-Plattformen.
Themen und Inhalte	<ul style="list-style-type: none">– Editoren, IDEs– Zugang zu KI-Plattformen– Tipps und Tricks
Lehrmittel	<ul style="list-style-type: none">– Folien/Skript

10.4 Einführung in Textverarbeitung und Textanalyse

Die Verarbeitung natürlicher Sprache (Natural Language Processing NLP) hat trotz komplexer Herausforderungen wie der Handhabung von Mehrdeutigkeiten («Die **Maus** muss aufgeladen werden») in den letzten Jahren enorme Fortschritte gemacht. In dieser Einführung wird zunächst ein Überblick zu den grundlegenden Methoden zur Analyse und Klassifikation von Textdaten für verschiedene Anwendungsfelder gegeben. Zudem werden Herausforderungen wie Mehrdeutigkeiten, Mehrsprachigkeit und Datenqualität thematisiert.

Lernziele	Die Teilnehmenden <ul style="list-style-type: none">– verstehen die grundlegenden Methoden der Textverarbeitung, einschliesslich Preprocessing, Tokenisierung und Vektorisierung.– verstehen statische Sprachmodelle und können Textklassifikation und Embeddings nutzen.– Kennen die Herausforderungen der Textanalyse und geeignete Verfahren zur Verbesserung von Datenqualität und Modellgenauigkeit.– sind vorbereitet für den vertiefenden Kursteil zu Large Language Modellen
Themen und Inhalte	<ul style="list-style-type: none">– Methoden<ul style="list-style-type: none">– Preprocessing: Reinigung und Vorbereitung von Textdaten.– Tokenisierung: Zerlegung von Texten in kleinere Einheiten wie Wörter oder Sätze.– Vektorisierung und Embeddings: Repräsentation von Texten für maschinelles Lernen mit gängigen Verfahren wie TF-IDF und word2vec.– Statistische Sprachmodelle: Modellierung von syntaktischen Strukturen und semantischen Zusammenhängen in Texten mit verschiedenen Methoden wie Naive Bayes, N-Gram, repräsentativen und generativen LLMs.– Herausforderungen von natürlicher Sprache: Mehrdeutigkeiten, Mehrsprachigkeit, Datenqualität und Bias– Anwendungen<ul style="list-style-type: none">– Textklassifikation, z.B. zur Kategorisierung von Inhalten oder Analyse von Kundenfeedback.– Frage-Antwort-Systeme zur Beantwortung spezifischer Fragen.

	<ul style="list-style-type: none"> - Chatbots zur automatisierten Kommunikation mit Nutzer*innen.
Lehrmittel	<ul style="list-style-type: none"> - Folien/Skript - Literatur (Kapitel 12, Lehrmittel): 2,3

10.5 Einführung in Bildverarbeitung

Bildgenerierende AI basiert auf neuronalen Netzen (NN), die ursprünglich für die Bildanalyse eingesetzt wurden. Dabei werden Bildmerkmale anhand vieler Beispielbilder gelernt, damit ein NN eine Klassifizierung durchführen kann. Um zu verstehen, was solche Bildmerkmale sind und wie sie für die Bildanalyse eingesetzt werden können, müssen wir einige Grundlagen der digitalen Bildverarbeitung verstehen.

Lernziele	<p>Die Teilnehmenden</p> <ul style="list-style-type: none"> - kennen die wichtigsten Prinzipien und Methoden der Bildverarbeitung, der klassischen Bildanalyse sowie der neuronalen Netze. - können grob abschätzen, welche Methoden für welche Aufgaben in Frage kommen. - Sind vorbereitet für die weitere Vertiefung «Generative KI für Bilder».
Themen und Inhalte	<ul style="list-style-type: none"> - Bildverarbeitung: <ul style="list-style-type: none"> - Punkt Operatoren: Histogramm, Helligkeits- und Kontrastkorrektur - Lokale Operatoren: Filter, Convolution, Morphologie - Globale Operatoren: Fourier Transform, Principal Component Analysis - Bildanalyse mit Feature Engineering: <ul style="list-style-type: none"> - Segmentierung - Feature Extraktion - Klassifikation und Evaluation - Bildanalyse mit Deep Learning: <ul style="list-style-type: none"> - Flache neuronale Netze - Convolutional Neural Nets
Lehrmittel	<ul style="list-style-type: none"> - Folien/Skript

10.6 Large Language Models verstehen und nutzen: Architektur, Methoden und Praxis

Grosse Sprachmodelle (Large Language Models LLMs) zielen darauf ab, komplexe textuelle Aufgaben mithilfe leistungsstarker neuronaler Netze zu bewältigen. Moderne LLMs wie GPT oder LLaMA haben den Bereich revolutioniert und bieten vielseitige Anwendungen. Durch Techniken wie Prompt Engineering, Feinabstimmung und semantische Suche können spezifische Fragestellungen rund um Textdaten effizient beantwortet werden.

Lernziele	<p>Die Teilnehmenden</p> <ul style="list-style-type: none"> - lernen die wichtigsten Konzepte moderner Sprachmodelle und deren praktische Anwendungen kennen. - Sie sind in der Lage, die Funktionsweise und Architektur von LLMs zu verstehen und deren Nutzen für verschiedene Aufgaben wie Textklassifikation, semantische Suche oder die Feinabstimmung auf spezifische Bedürfnisse einzuschätzen. - werden befähigt, Anwendungsfälle praktisch umzusetzen.
-----------	--

Themen und Inhalte	<ul style="list-style-type: none"> - Methoden <ul style="list-style-type: none"> - Foundation LLMs: Vortrainierte Sprachmodelle, die ein breites Spektrum an textbasierten Aufgaben lösen können. Sie können sowohl manuell durch menschliche Interaktion als auch programmatisch angesprochen werden (Beispiele: GPT, LLaMA). - Prompt Engineering: Die Kunst, gezielte Eingaben für Sprachmodelle zu entwickeln, um möglichst präzise und relevante Ausgaben zu erzielen (Techniken: Zero-shot, One-shot, Few-shot Learning, In-Context Learning). - Transformer-Architektur: Die Basis moderner LLMs, die durch ein besseres Verständnis ihrer Funktionsweise ermöglicht, gezielt aufgabenspezifische Modelle zu bewerten und anzupassen (Typen: Encoder-Decoder, Encoder-only, Decoder-only; Konzepte: Embeddings). - Topic Modelling: Verfahren zur Gruppierung ähnlicher Dokumente oder Klassifikation von Texten (Anwendungen: Textklassifikation, Clustering). - Semantische Suche und RAG (Retrieval-Augmented Generation): Ergänzt ein LLM um externes Wissen, um Aufgaben besser zu lösen. So können beispielsweise unternehmensspezifische Fragen beantwortet werden, ohne dass das Modell explizit auf diesen Daten trainiert wurde (Werkzeuge: LangChain). - Feinabstimmung (Finetuning): Weiterführendes Training eines Foundation-Modells auf spezifische Daten oder Aufgaben, um dessen Leistung zu optimieren (Methoden: Effizientes Lernen, Supervised Learning, Instruction Tuning, Reinforcement Learning mit menschlichem Feedback). - Herausforderungen und Grenzen: Probleme wie die Qualität der Trainingsdaten, Bias, Halluzinationen und der Schutz der Privatsphäre stellen zentrale Herausforderungen bei der Nutzung von LLMs dar. - Anwendungen <ul style="list-style-type: none"> - Programmatische Anbindung von Sprachmodellen und deren taskbezogene Konfiguration - Optimierung von Aufgabenstellungen für LLMs - Textklassifikation - Anwendungen von RAG für spezifische Wissensabfragen - Feinabstimmung von Sprachmodellen für spezifische Aufgaben
Lehrmittel	<ul style="list-style-type: none"> - Folien/Skript - Literatur (Kapitel 12, Lehrmittel): 2,3,4

10.7 Generative KI für Bilder

Generative neuronale Netze für Bilder nutzen Verfahren, um komprimierte Bilddaten aus einem Merkmalsraum wiederherzustellen. Die Idee von generativen neuronalen Netzen für Bilder entstand, als man ein Verfahren gefunden hatte, um Bilder, die man in einen kondensierten Raum komprimiert hatte, wieder zu einem Bild hochzukalieren. Damit können wir Bilder segmentieren, Stiländerungen vornehmen, hochskalieren oder entrauschen. Um komplett neue Bilder aus einer Texteingabe zu generieren, wird zusätzlich ein sogenanntes Word-Embedding verwendet. Das ist ein Verfahren, bei welchem Worte mit Bildern verknüpft werden. Dies wird danach verwendet, um Bilder aus komplettem Rauschen zu einem Bild mit einem ähnlichen Word-Embedding zu entwickeln

Lernziele	<p>Die Teilnehmenden</p> <ul style="list-style-type: none"> - lernen die Prinzipien und typischen Arten generativer Netze kennen und können deren Einsatz grob einschätzen. - kennen die wichtigsten Prinzipien und einige typische Arten von generativen neuronalen Netzen für Bilder. Sie können grob abschätzen, welche Netze für welche Anwendungen in Frage kommen.
Themen und Inhalte	<ul style="list-style-type: none"> - Methoden: <ul style="list-style-type: none"> - Encoder-Decoder-Modelle sind die Basis fast aller nachfolgenden Technologien. In einem Decoder-Teil kann der komprimierte Endblock des Encoder-Teils wieder zu einem Bild hochgeneriert werden. - Semantische Segmentierung ist eine Klassifizierung auf Pixelebene und erlaubt die Detektion von Objekten. Sie wird in vielen Bereichen angewendet, u.a. in der Medizin und bei autonomen Fahrsystemen. - Auto-Encoder (Variational Autoencoder) erlauben es, Bilder zu komprimieren und wieder zu einem ähnlichen Bild zu dekomprimieren. - Generative Adversarial Networks (GANs) sind generative Netzwerke, worin sich ein Generator und ein Diskriminator bekämpfen und damit ein Netzwerk trainieren können. - Stable Diffusion: Stable Diffusion generiert Bilder, indem es iterativ Rauschen entschärft. Dabei wird es von dem Textencoder, der vorab auf Konzepte trainiert wurde, sowie dem Aufmerksamkeitsmechanismus angeleitet. - CLIP ist der Textencoder in Stable Diffusion, der Text und Bilder verbindet und so Bilder erkennen und kategorisieren kann, ohne dass für jede Kategorie ein spezielles Training erforderlich ist. - Anwendungen: <ul style="list-style-type: none"> - Denoising for Ray Tracing - Super Resolution (Nvidia DLSS) - Inpainting & Outpainting - Style Transfer - OpenAI: DALL-E - SORA: Video-Generierungsmodell von OpenAI, das in der Lage ist, realistische und detaillierte Videos aus einfachen Texteingaben zu erstellen.
Lehrmittel	<ul style="list-style-type: none"> - Folien/Skript

10.8 Semesterarbeit

Die Studierenden führen ein eigenständiges Projekt durch, bei dem sie eine Fragestellung aus den behandelten Themenbereichen bearbeiten. Sie erstellen eine Projektarbeit, die die Problemstellung, den Lösungsansatz, die Implementierung sowie die Ergebnisse dokumentiert.

11 Kompetenznachweis

Für die Anrechnung der 12 ECTS-Credits ist das erfolgreiche Bestehen der Qualifikationsnachweise (Prüfungen, Projektarbeiten) erforderlich, gemäss folgender Aufstellung:

Kompetenznachweis	Gewicht	Art der Qualifikation	Erfolgsquote Studierende
Semesterarbeit	10	Projektarbeit mit Dokumentation und Präsentation	0 - 100 %
Total	10		0 - 100 %

Der gewichtete Mittelwert der Erfolgsquoten der einzelnen Kompetenznachweise wird in eine Note zwischen 3 und 6 umgerechnet. Die Note 3 (gemittelte Erfolgsquote weniger als 50%) ist ungenügend. Die Noten 4, 4.5, 5, 5.5 und 6 (gemittelte Erfolgsquote zwischen 50% und 100%) sind genügend.

12 Lehrmittel

Die Kursinhalte stützen sich u.a. auf folgende Lehrmittel

Nr.	Titel	Autoren	Verlag	Jahr	ISBN/URL
1.	Machine Learning with PyTorch and Scikit-Learn	S. Raschka, Y. Liu, and V. Mirjalili	Packt	2022	978-1-80181-931-2
2.	Speech and Language Processing	D. Jurafsky and J. Martin		2025	Website
3.	Building LLMs for Production	L.-F. Bouchard and L. Peters	O'Reilly	2024	979-8324731472
4.	Hands-On Large Language Models	J. Alammam and M. Grootendorst	O'Reilly	2024	978-1098150969
5.	Generative Deep Learning	David Foster	O'Reilly	2024	978-1-09-813418-1
6.	Neuronal Networks from Scratch	Harrison Kinsley and Daniel Kukiela			Github
7.	Make Your Own Neural Network	Tariq Rashid	O'Reilly	2024	978-1492064046

13 Dozierende

Vorname Name	Firma	E-Mail
Jürgen Vogel	Berner Fachhochschule	juergen.vogel@bfh.ch
Marcus Hudritsch	Berner Fachhochschule	yannis.schmutz@bfh.ch
Peter von Niederhäusern	Berner Fachhochschule	peter.vonniederhausern@bfh.ch
Yannis Schmutz	Berner Fachhochschule	marcus.hudritsch@bfh.ch

14 Organisation

CAS-Leitung:

Arno Schmidhauser

Tel: [+41 31 848 32 75](tel:+41318483275)

E-Mail: arno.schmidhauser@bfh.ch

CAS-Administration:

Andrea Moser

Tel: [+41 31 848 32 11](tel:+41318483211)

E-Mail: andrea.moser@bfh.ch

Während der Durchführung des CAS können sich Anpassungen bezüglich Inhalten, Lernzielen, Dozierenden und Kompetenznachweisen ergeben. Es liegt in der Kompetenz der Dozierenden und der Studienleitung, aufgrund der aktuellen Entwicklungen in einem Fachgebiet, der konkreten Vorkenntnisse und Interessenslage der Teilnehmenden, sowie aus didaktischen und organisatorischen Gründen Anpassungen im Ablauf eines CAS vorzunehmen.

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

bfh.ch/cas-gen-ki