Generative-KI-Entwicklung in der Cloud

Generative KI & ChatGPT · Anbieter: NEXIO Online Akademie GmbH

FÖRDERUNG & KOSTEN

Kurspreis: 0 €

Förderfähig über: Bildungsgutschein

AZAV-zertifizierter Bildungsträger — Förderung durch die Agentur für Arbeit möglich

Kursdaten auf einen Blick

Durchführung	Online
Lernformat	Zeitgebunden
Dauer	12 Wochen in Vollzeit
Unterrichtssprache	Deutsch
Kursstart	Startet jeden Monat
Vorkenntnisse	Grundlegende Kenntnisse
Abschluss	Internationales Zertifikat AWS (nach bestandener Prüfung) und trägerinternes Zertifikat bzw. Teilnahmebescheinigung

Kurzbeschreibung

In diesem Kurs erwerben Sie End-to-End-Kompetenzen zur Entwicklung produktionsreifer Generative-KI-Anwendungen auf einer Managed-GenAI-Cloud-Plattform. Sie lernen, Foundation Models zu nutzen, Kernservices wie Guardrails und Streaming einzusetzen und Integrationsmuster wie API, …

Bildungsträger & Kontakt

NEXIO Online Akademie GmbH

Am Herrschaftsweiher 11, 67071 Ludwigshafen am Rhein

Telefon: 0800 100 87 34

E-Mail: beratung@nexio-akademie.de

Web: https://nexio-akademie.de

Kurs online ansehen

01 Kursbeschreibung

In diesem Kurs erwerben Sie End-to-End-Kompetenzen zur Entwicklung produktionsreifer Generative-KI-Anwendungen auf einer Managed-GenAI-Cloud-Plattform. Sie lernen, Foundation Models zu nutzen, Kernservices wie Guardrails und Streaming einzusetzen und Integrationsmuster wie API, SDK, Event-Driven und Serverless umzusetzen. Sie vertiefen Prompt Engineering, Prompt Management und Prompt-Tests. Retrieval-Augmented-Generation mit Vektordatenbanken, Embeddings, Chunking und Hybrid Search wird praxisnah behandelt. Sie erfahren, wie Sie Modelle durch Fine-Tuning und PEFT anpassen und evaluieren. Agentische Workflows mit ReAct und Tool-Invocation werden ebenso vermittelt wie Responsible AI und KI-Sicherheit, einschließlich Bias, Toxizität und Jailbreak-Prevention. Cloud-Security und Governance für GenAI, Betriebsdesign, Kostenoptimierung, Performance und Observability runden das Programm ab.

02 Lehrplan & Module

1. Foundation Models auswählen, vergleichen, bewerten

Du lernst, Foundation Models nach Aufgabe, Kontextlänge, Kosten und Qualität auszuwählen und miteinander zu vergleichen. Mit Benchmarks und eigenen Testfällen bewertest du ihre Eignung.

2. GenAI-Apps mit Managed-APIs, SDKs und Streaming entwickeln

Du entwickelst GenAI-Anwendungen über Managed-APIs und SDKs und setzt Streaming ein, um Antworten schrittweise auszugeben. Dabei behandelst du Authentifizierung, Fehlerbehandlung und Ratenbegrenzungen.

3. Prompt Engineering, Function Calling und Tool Use einsetzen

Du gestaltest Prompts systematisch, versionierst und testest sie. Außerdem bindest du über Function Calling und Tool Use externe Funktionen und Datenquellen an Sprachmodelle an.

4. RAG mit Embeddings, Vektordaten und Hybrid Search aufbauen

Du baust Retrieval-Augmented-Generation-Systeme mit Embeddings, Vektordatenbanken und hybrider Suche auf. Dabei lernst du, Dokumente sinnvoll zu zerlegen und relevante Inhalte zuverlässig abzurufen.

5. Fine-Tuning, PEFT und Continued Pre-Training planen

Du lernst, wann sich Fine-Tuning, parametereffiziente Verfahren wie LoRA oder Continued Pre-Training lohnen. Zudem planst du Datenaufbereitung, Training und Bewertung angepasster Modelle.

6. Agentische Workflows serverless und event-getrieben umsetzen

Du setzt agentische Workflows um, in denen Modelle mehrere Schritte planen und Werkzeuge aufrufen. Dazu nutzt du serverlose Funktionen und ereignisgesteuerte Architekturen.

7. Guardrails, PII-Schutz und Responsible AI verankern

Du richtest Guardrails ein, um unerwünschte Inhalte zu filtern und personenbezogene Daten zu schützen. Ebenso lernst du Prinzipien von Responsible AI wie Transparenz und Fairness kennen.

8. Cloud-Security: IAM, private Endpunkte, Verschlüsselung

Du sicherst GenAI-Anwendungen mit Identitäts- und Zugriffsmanagement, privaten Netzwerkendpunkten und Verschlüsselung von Daten ab.

9. Evaluation, Drift, Kosten, Latenz und Caching im Betrieb

Du bewertest Modellausgaben mit Evaluationsverfahren und überwachst Drift im laufenden Betrieb. Zudem optimierst du Kosten und Latenz, etwa durch Caching und Modellauswahl.

03 Enthaltene Zertifikate

Trägerzertifikat: Internationales Zertifikat AWS (nach bestandener Prüfung) und trägerinternes Zertifikat bzw. Teilnahmebescheinigung

04 Voraussetzungen

Vorkenntnisse

Niveau: Grundlegende Kenntnisse

Vorausgesetzt werden fundierte Grundlagen in IT, Cloud Computing und Softwareentwicklung. Sie sollten über Programmierkenntnisse verfügen, idealerweise in Python. Ein Grundverständnis von APIs, Datenformaten wie JSON und Cloud-Diensten ist erforderlich. Darüber hinaus benötigen Sie Englischkenntnisse zum Lesen technischer Dokumentation. Sie können Ihren eigenen PC oder Laptop für die praktischen Übungen nutzen. Empfohlen: bestandene AWS-Associate-Zertifizierung oder gleichwertige Cloud-Erfahrung im Umfang von etwa einem Jahr Praxis.

05 Zielgruppen & Berufsperspektiven

- Arbeitssuchende

Mögliche Berufsbilder: KI-Entwickler, Machine Learning Engineer, Cloud-Softwareentwickler

NÄCHSTER SCHRITT

Kursdetails online ansehen und Anbieter kontaktieren

NEXIO Online Akademie GmbH · Am Herrschaftsweiher 11, 67071 Ludwigshafen am Rhein
0800 100 87 34 · beratung@nexio-akademie.de · https://nexio-akademie.de

Online ansehen