

Zum Klienten:

Weitere Informationen zum Klienten und laufend neue Jobchancen erhalten Sie nach Ihrer Anmeldung und dem Onboarding durch Ihren Fachberater per Mail oder online. Wir freuen uns, Sie persönlich kennenzulernen!

Arbeitsort:

Zug (X Kilometer und X Minuten von Ihrer Wohnadresse entfernt)

Jobbeschreibung:

QA Azure AI Engineer (w/m/d)

Ihr Aufgabengebiet:

- Aufbau und Implementierung von KI-Lösungen mit generativen Modellen und RAG-Pipelines
- Bereitstellung, Verwaltung und Wartung von KI-Modellen auf Cloud-Plattformen mit skalierbarer Architektur
- Schreiben komplexer Python-Skripte für Datenverarbeitung, Tests und Modellautomatisierung
- Evaluierung von Modellen anhand fortgeschrittener Metriken, inklusive LLM-as-a-Judge und deterministischer Assertions
- Entwicklung automatisierter Pipelines für KI-Tests und Deployment auf Cloud-Plattformen
- Fachliche Führung bei der Gestaltung von KI-Modellarchitekturen und Best Practices für das Deployment
- Zusammenarbeit mit Daten-, Backend- und Produktteams zur Integration von KI-Lösungen in die Produktion
- Pflege reproduzierbarer Workflows, technischer Dokumentation und Experimentresultate

Was Sie ausmacht:

- 1-2 Jahre praktische Erfahrung mit generativer KI und RAG-Architektur (Vektorsuche + LLM)
- 5+ Jahre Erfahrung im Deployment von KI-Modellen auf Cloud-AI-Services, nicht nur lokal
- 7+ Jahre Python-Erfahrung, inklusive komplexer Skripterstellung (Pandas, PyTest) über einfache Automatisierung hinaus
- 8+ Jahre QA-Erfahrung mit LLM-Evaluierung und Testautomatisierung
- 5+ Jahre Erfahrung im Aufbau von Pipelines für automatisierte KI-Modelltests und Deployment
- Technischer Abschluss in Informatik, Ingenieurwesen, Mathematik oder vergleichbare umfassende Erfahrung
- Praktische Erfahrung im Deployment und Management von KI-Modellen in Azure-AI-Cloud-Services
- Fähigkeit, fortgeschrittene KI-Evaluierungsmetriken und Regressionstests zu gestalten und zu interpretieren
- Erfahrung mit agentenbasierten KI-Frameworks oder Multi-Agenten-Systemen von Vorteil
- Erfahrung mit Cloud-Suchdiensten, inklusive Indexierung, Vektorprofilen und Hybrid-Suche von Vorteil
- Erfahrung mit LangChain oder Semantic Kernel für KI-Test-Harnesses von Vorteil
- Erfahrung mit Prompt Flow zur Erstellung von Modell-Evaluierungspipelines von Vorteil
- Kenntnisse in der Erstellung von Golden-Datasets für Regressionstests von Vorteil
- Master-Abschluss in Data Science oder verwandtem Bereich von Vorteil

- Erfahrung im Fine-Tuning kleiner Sprachmodelle wie Phi-3 oder LLaMA von Vorteil
- Erfahrung in der Verwaltung von ML-Workspaces, Compute-Instanzen und Experiment-Tracking von Vorteil
- Hintergrund in C#/.NET für die Integration von KI in Backend-Systeme von Vorteil
- Erfahrung mit Red-Teaming, adversarialen oder Jailbreak-Tests für KI-Modelle von Vorteil

Was diese Stelle ausmacht:

- Attraktives Gehalt mit leistungsbezogenen Boni
- Möglichkeiten zur fachlichen Weiterentwicklung und Aufstieg
- Dynamisches und kollaboratives Arbeitsumfeld