

50% remote – AI Engineer – AI Agents & Deployment auf Databricks und Azure (m/w/d)

Für unseren Kunden suchen wir einen AI Engineer - AI Agents & Deployment auf Databricks und Azure (m/w/d).

Rahmendaten:
Laufzeit: asap bis 31.12.2027 (längerfristige Zusammenarbeit gewünscht)
Leistungsort: Köln, nach Absprache remote -Anteile bis zu 50% möglich
Auslastung: Vollzeit

Profilanforderungen
Leistungsgegenstand ist die Konzeption, Entwicklung und produktive Bereitstellung intelligenter AI Agents auf Databricks und Azure – sic nachvollziehbar und skalierbar. Dazu gehören AI Engineering, LLMOps und Cloud Architecture sowie die Realisierung agentenbasierter Anwendungen zur Verbindung von Unternehmensdaten, Large Language Models und internen Systemen in Zusammenarbeit mit einem crossfunktionalen Team.

Leistungsbeschreibung
- Konzeption, Entwicklung und Betrieb produktiver AI -Agent -Lösungen auf Databricks und Azure – vom ersten Prototyp bis zum skalier Deployment.
- Entwicklung von KI -Agenten und Multi -Agent -Workflows zur zielgerichteten Verbindung von Large Language Models, Unternehmens und internen Systemen.
- Implementierung von RAG, Tool Calling und Agent -Orchestrierung sowie Entwicklung wiederverwendbarer Tools, Prompts sowie Zusto Kontextkomponenten.
- Integration von Azure OpenAI und weiteren Foundation Models in Databricks sowie in bestehende Applikations - und Datenlandschaf
- Entwicklung Python -basierter Agent -Services und APIs sowie deren Integration in bestehende Backend - und Cloud -Architekturen.
- Einsatz von Databricks Mosaic AI, Model Serving und MLflow zur Versionierung, Evaluation, Bereitstellung und Überwachung von Mod Agenten.
- Implementierung von Vector Search, Embeddings und Retrieval -Pipelines für den kontrollierten Zugriff auf relevante Unternehmensd
- Orchestrierung von Agenten -, Daten - und Automatisierungsprozessen mit Databricks Workflows und definierten Tool -Zugriffen.
- Automatisierung der Bereitstellung mit GitLab CI/CD und Terraform , einschließlich reproduzierbarer Umgebungen, Tests und nachvoll Releases.
- Entwicklung von Evaluationen, Tracing und Monitoring für Antwortqualität, Groundedness, Sicherheit, Latenz, Kosten und Zuverlässig
- Etablierung von Guardrails, Berechtigungskonzepten und Human -in-the -Loop -Prozessen für den verantwortungsvollen Einsatz von
- Sicherstellung von Security und Governance mit Unity Catalog, Managed Identities, Key Vault, RBAC, Private Endpoints und Logging.
- Enge Zusammenarbeit mit Product Owner, Fachbereichen, Data -, Backend - und Cloud - Teams sowie Security und Privacy bei prod Agent -Use -Cases.
- Dokumentation von Agentenlogik, Tool -Schnittstellen, Datenflüssen, Architektur, Evaluationsverfahren und Runbooks für den nachhal Betrieb

Erforderliche Qualifikationen und Erfahrungen

- Mehrjährige Erfahrung als AI Engineer, Machine Learning Engineer, Software Engineer oder in einer vergleichbaren technischen Rolle
- Praktische Erfahrung mit LLM -Anwendungen und AI Agents , idealerweise mit RAG, Agent -Orchestrierung, Tool Calling und produktiv Deployment.
- Sehr gute Kenntnisse in Python sowie in API -Integration, automatisierten Tests, Code Reviews und modularer Softwareentwicklung.
- Fundierte Erfahrung mit Databricks , beispielsweise mit MLflow, Mosaic AI, Model Serving, Vector Search und Unity Catalog.
- Gute Kenntnisse der Microsoft -Azure -Plattform , insbesondere Azure OpenAI, Managed Identities, Key Vault, Storage und Networkin
- Erfahrung mit Agenten - und LLM -Frameworks , beispielsweise LangGraph, LangChain oder Semantic Kernel, und zustandsbehaftet Workflows.
- Praktische Erfahrung mit GitLab CI/CD für Testing, Versionierung und automatisierte Bereitstellung von AI -Anwendungen.
- Praktische Erfa hrung mit Terraform (Infrastructure as Code) für die Bereitstellung von Azure - und Databricks -Ressourcen.
- Solides Verständnis von LLMOps, Agent Evaluation und Observability , einschließlich systematischer Qualitätskontrollen.
- Kenntnisse in AI Security, Datenschutz und Governance , sicherem Modellzugriff, Guardrails und dem Schutz sensibler Informationen.
- Fähigkeit, AI Agents produktiv, zuverlässig und skalierbar bereitzustellen und fachliche Anforderungen in robuste Lösungen zu übersetz

Weitere wünschenswerte Qualifikationen

- Kenntnisse des Model Context Protocol (MCP) und von Mustern zur Multi -Agent - oder Agent -to-Agent -Kommunikation.
- Erfahrung mit FastAPI, Docker und Kubernetes für den Betrieb containerisierter Agent -Services und APIs.
- Kenntnisse in Spark, SQL, Delta Lake und Databricks Workflows für Datenaufbereitung und KI -gestützte Prozesse.
- Erfahrung mit Human -in-the -Loop, strukturierten Ausgaben und kontrollierten Rollouts für AI -Anwendungen.
- Erfahrung in regulierten oder datenschutzsensiblen Unternehmensumfeldern