



Unser Mandant betreibt ein hochmodernes Dual-Rechenzentrum in der Technologieregion Karlsruhe und steht für höchste Verfügbarkeit, Datensicherheit und leistungsstarke IT-Services. Er unterstützt seine Kunden mit einem breiten Portfolio an Cloud-, IT-, Personalwirtschafts-, Banking- und Telekommunikationslösungen. Für diesen Kunden sind wir am Standort Eggenstein-Leopoldshafen zum nächstmöglichen Zeitpunkt im Rahmen einer Festeinstellung auf der Suche nach einem qualifizierten:

AI Platform Engineer mit Architektur-Know-how (m/w/d)

Job-ID: CF-00007301

Job-Ort: Eggenstein-Leopoldshafen

Remoteanteil / Home Office: 60 %

Gesucht ist kein reiner KI-Engineer, sondern ein Architekt für den produktiven Einsatz von KI. In dieser Rolle entwickeln Sie die zentrale Plattform, auf der aus leistungsfähigen Modellen sichere, skalierbare und werteorientierte Anwendungen entstehen. Gemeinsam mit Ihrem Team gestalten Sie eine souveräne KI-Plattform mit dem Anspruch, technologische Exzellenz, Datenschutz und gesellschaftliche Verantwortung in Einklang zu bringen. Sie bewegen sich an der Schnittstelle von Infrastruktur, Anwendung und Governance und sorgen dafür, dass KI nicht nur funktioniert, sondern im Alltag echten Mehrwert schafft.

Ihre Hauptaufgaben:

- Weiterentwicklung und Betrieb von KI-Plattform - von der Modellintegration bis zur produktiven Nutzung
- Entwicklung der zentralen Schicht zwischen KI-Infrastruktur und konkreten Anwendungen zur Sicherstellung skalierbarer Use Cases
- Integration, Test und Validierung neuer KI-Modelle für den produktiven Einsatz
- Konzeption und Umsetzung von RAG-Architekturen zur Nutzung organisationsspezifischer Wissensdaten
- Entwicklung und Integration von Schnittstellen und Tools zur Erweiterung von KI-Workflows (z.B. über MCP und Tool-Calling)
- Implementierung von Sicherheitsmechanismen und Guardrails zur Minimierung von Risiken im KI-Betrieb
- Sicherstellung von Governance-, Compliance- und Datenschutzerfordernungen im Kontext moderner KI-Systeme
- Monitoring und kontinuierliche Optimierung der Modellperformance und Systemqualität
- Enge Zusammenarbeit mit IT, Architektur, Betrieb sowie Fachbereichen zur Weiterentwicklung der Plattform
- Unterstützung von Nutzern und Mandanten im Umgang mit KI-Lösungen

Ihre fachlichen Voraussetzungen:

- Erfolgreich abgeschlossenes Studium in Informatik, Data Science oder eine vergleichbare Qualifikation
- Praktische Erfahrung im Aufbau und Betrieb von KI-Anwendungen, insbesondere im Umgang mit LLMs und Prompt Engineering
- Erfahrung mit RAG-Architekturen und Vektordatenbanken (z.B. ChromaDB oder vergleichbar)
- Gute Kenntnisse in Python
- Erfahrung im Umgang mit Linux und Container-Technologien (z.B. Docker, idealerweise Kubernetes)
- Erfahrung mit der Integration von Tools und Schnittstellen in KI-Systemen (z.B. MCP oder vergleichbare Frameworks)
- Verständnis von typischen Risiken im KI-Umfeld (z.B. Prompt Injection, Halluzinationen, Model Drift) und deren Absicherungen
- Erfahrung mit KI-Gateway-Lösungen oder vergleichbaren Systemen (z.B. OpenWebUI, LiteLLM)
- Kenntnisse in Identity- und Access-Management (z.B. OIDC, Keycloak) sowie im Umgang mit Zugriffs- und Berechtigungskonzepten
- Erfahrung in regulierten Umfeldern
- Grundkenntnisse in relevanten regulatorischen Anforderungen (z.B. EU AI Act, DSGVO, BSI-Grundschutz)
- Sehr gute Deutschkenntnisse in Wort und Schrift

Ihre persönlichen Voraussetzungen:

- Analytische, strukturierte und eigenständige Arbeitsweise
- Kommunikationsstärke und Teamfähigkeit
- Fähigkeit, komplexe technische Themen verständlich zu vermitteln

Unser Mandant bietet:

- Mobiles Arbeiten
- Sehr flexibles Arbeitszeitmodell
- 33,5 Tage Erholungsurlaub
- Bezahlte Freistellung am 24. u 31.12, am Geburtstag sowie ½ Tag am Gründonnerstag
- Sicherer unbefristeter Arbeitsplatz
- Vergütung nach TVöD Bund
- Moderne Arbeitsplatzausstattung
- Faire Bonuszahlung
- Weiterbildungs- und Entwicklungsmöglichkeiten
- Bike-Leasing
- Mitarbeiter-Events
- Kantine
- Betriebliche Altersvorsorge

- Gesundheitsförderung
- E-Ladesäulen



Ihr Ansprechpartner:

Adrian Weinast

auteega GmbH

Kaiserring 14-16

68161 Mannheim

Telefon: +49 170 2175 230

E-Mail: adrian.weinast@auteega.com

Jetzt bewerben