

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:

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

Jobbeschreibung:

Senior AI Machine Learning Engineer (w/m/d)

Ihr Aufgabengebiet:

- Leitung von Design, Entwicklung und Deployment von End-to-End-AI/ML- und Generative-AI-Lösungen
- Aufbau von KI-gestützten Anwendungen wie Chatbots, RAG-Systemen und Agentic-AI-Lösungen inklusive Evaluations-, Monitoring- und Performance-Frameworks
- Entwicklung und Pflege skalierbarer Backend-Services, APIs und Datenpipelines mit Python, FastAPI, Apache Spark und verteilten Verarbeitungstechnologien
- Implementierung von MLOps- und Observability-Praktiken für Deployment, Monitoring, Optimierung, Governance und Lifecycle-Management von Modellen
- Etablierung von Engineering-Standards und CI/CD-Frameworks zur Sicherstellung von Codequalität, Testing, Reproduzierbarkeit, automatisierten Releases und produktionsreifen AI-Deployments
- Funktionsübergreifende Zusammenarbeit mit Business-, Produkt-, Daten-, Security- und Operations-Teams in agilen Umgebungen
- Technische Führung, Produktionssupport und Mentoring von Junior AI Engineers zur Förderung einer Kultur des kontinuierlichen Lernens

Was Sie ausmacht:

- Fundierte Expertise in Python und modernen AI-Ökosystemen wie LangChain, LangGraph, PyTorch, Ollama, vLLM und verwandten Frameworks
- Umfassende Erfahrung im Design, Deployment und Betrieb von Enterprise-Generative-AI-Lösungen, inkl. Chatbots, RAG-Architekturen, LLM-Anwendungen und lokal gehosteten LLM-Deployments
- Fundierte Kenntnisse in Azure Cloud AI-Plattformen, Oracle-Datenbanken und Enterprise-Integrationstechnologien
- Tiefes Wissen in Machine Learning und NLP, einschliesslich Sentiment-Analyse, NER, Dokumentenklassifikation, Topic Modelling und Sprachverständnis
- Starke MLOps/AIOps-Expertise, u. a. reproduzierbare Pipelines, automatisierte Evaluation, Modellversionierung, Feature Stores, Observability, CI/CD-Automatisierung und AI-Lifecycle-Management
- Praktische Erfahrung mit Data Engineering und verteiltem Computing im grossen Massstab, mit Apache Spark, Ray, Kafka und Qlik
- Ausgewiesene technische Führungsstärke, Stakeholder-Management, Eigenverantwortung und operative Exzellenz
- Erfahrung im Finanzdienstleistungssektor, Banking, Versicherung oder Wealth Management von Vorteil
- Erfahrung in regulierten und hochsicheren Umgebungen von Vorteil

- Cloud- oder AI-bezogene Zertifizierung von Vorteil

Was diese Stelle ausmacht:

- Flexible Arbeitsmodelle wie Teilzeit, Job-Sharing und hybrides Arbeiten
- Purpose-getriebene Kultur und globale Infrastruktur für agile Zusammenarbeit
- Innovatives Arbeitsumfeld geprägt von Digitalisierung, neuen mobilen Zahlungsmethoden und Kundenerlebnis
- Vielfältige Entwicklungsmöglichkeiten durch neue Rollen, Wissen und Fähigkeiten