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 KI-gestützter Anwendungen wie Chatbots, RAG-Systeme und Agentic-AI-Lösungen mit robusten Evaluations-, Monitoring- und Performance-Frameworks
- Entwicklung und Pflege skalierbarer Backend-Services, APIs und Datenpipelines mit Python, FastAPI, Apache Spark und verteilten Verarbeitungstechnologien
- Umsetzung von MLOps- und Observability-Praktiken für Modell-Deployment, Monitoring, Optimierung, Governance und Lifecycle-Management
- Etablierung von Engineering-Standards und CI/CD-Frameworks für Codequalität, Testing, Reproduzierbarkeit, automatisierte Releases und produktionsreife AI-Deployments
- Funktionsübergreifende Zusammenarbeit mit Business-, Produkt-, Daten-, Security- und Operations-Teams in agilen Umgebungen
- Technische Leitung, Production Support und Mentoring von Junior AI Engineers zur Förderung kontinuierlichen Lernens und Wissensaustauschs

Was Sie ausmacht:
- Fundierte Expertise in Python und modernen AI-Ökosystemen, inkl. LangChain, LangGraph, PyTorch, Ollama, vLLM und verwandten Frameworks
- Umfangreiche Erfahrung im Design, Deployment und Betrieb von Enterprise-Generative-AI-Lösungen, inkl. Chatbots, RAG-Architekturen, LLM-Anwendungen und lokal gehosteten LLM-Deployments
- Kenntnisse in Azure Cloud AI Plattformen, Oracle-Datenbanken und Enterprise-Integrationstechnologien für skalierbare und sichere AI-Lösungen
- Tiefes Wissen in Machine Learning und NLP, inkl. Sentiment-Analyse, NER, Dokumentenklassifikation, Topic Modelling und Sprachverständnissystemen
- Starke MLOps/AIOps-Expertise, inkl. reproduzierbarer Pipelines, automatisierter Evaluation, Modellversionierung, Feature Stores, Observability, CI/CD-Automatisierung und AI-Lifecycle-Management
- Praktische Erfahrung mit grossskaligem Data Engineering und verteiltem Computing mittels Apache Spark, Ray, Kafka und Qlik
- Ausgewiesene technische Führungsstärke, Stakeholder-Management, Eigenverantwortung und operative Exzellenz
- Erfahrung in Financial Services, Banking, Insurance 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
- Innovatives Arbeitsumfeld geprägt von Digitalisierung, neuen mobilen Zahlungsmethoden und Fokus auf Kundenerlebnis