

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:

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

Jobbeschreibung:

AI Engineer_Senior (w/m/d)

Ihr Aufgabengebiet:

- Entwicklung von KI-Komponenten fuer Produkte und Enterprise-Plattformen, darunter Agenten, RAG-Pipelines, Inferenzdienste, LLM-Orchestrierung und Integration von vortrainierten oder feinabgestimmten Modellen
- Implementierung von Applikationslogik in Python, TypeScript oder vergleichbaren Sprachen unter Sicherstellung von Codequalitaet, Skalierbarkeit und Wartbarkeit
- Integration von AI/ML-APIs, KI-Clouddiensten und individuellen Modellen in Applikations- und Cloud-native-Architekturen
- Zusammenarbeit mit AI Architect und AI Data Scientist zur Ueberfuehrung von Modellen, Prototypen und technischen Spezifikationen in produktionsreife Komponenten
- Mitwirkung an der technischen Konzeption von KI-Modulen, inklusive Implementierungsmustern, Schnittstellen und Integrationslogik
- Durchfuehrung von Testing, Debugging, Profiling und Performance-Optimierung der entwickelten KI-Komponenten
- Anwendung von Software-Engineering-Praktiken wie Code-Reviews, Versionierung, technischer Dokumentation und Testautomatisierung
- Sicherstellung, dass entwickelte Loesungen den Standards bezueglich Sicherheit, Zuverlaessigkeit, Wiederverwendbarkeit und Compliance entsprechen
- Zusammenarbeit mit DevOps/MLOps Engineers zur Integration von KI-Komponenten in CI/CD-Prozesse, Deployment und Monitoring in Cloud-Umgebungen
- Mitwirkung am Lifecycle-Management von Modellen, inklusive Versionierung, Packaging, Konfiguration, Release und Betriebsueberwachung
- Unterstuetzung bei der Industrialisierung von KI-Prototypen und wiederverwendbaren Assets bis zu stabilen, dokumentierten Komponenten
- Mitwirkung am Aufbau der internen Bibliothek an Komponenten, Templates, Konnektoren und Acceleratoren der Factory
- Zusammenarbeit in multidisziplinaeren Teams mit Product Managern, FullStack Developern, UI/UX Engineers, Data Engineers, AI Data Scientists und AI Architects
- Uebersetzung von Produktanforderungen in umsetzbare technische Loesungen unter Beruecksichtigung von Einschränkungen, Abhaengigkeiten und Risiken
- Fachliche Unterstuetzung junionerer Teammitglieder durch Weitergabe von Best Practices
- Teilnahme an Planning-, Refinement-, Technical-Review- und Retrospektiven-Sitzungen zur kontinuierlichen Verbesserung der Delivery

Was Sie ausmacht:

- 3-5 Jahre Erfahrung in AI Engineering, Machine Learning Engineering, Software Engineering oder der Entwicklung von AI/ML-Lösungen in Enterprise- oder Technologieumgebungen
- Fundierte Kenntnisse in Python sowie den gängigen AI/ML-Frameworks wie PyTorch, TensorFlow, scikit-learn oder vergleichbaren
- Erfahrung mit LLMs, RAG-Pipelines, Embeddings, Prompt Engineering, Model Serving oder der Integration generativer Modelle
- Erfahrung in Entwicklung und Integration von REST-APIs, Applikationsdiensten, Microservices oder Cloud-native-Komponenten
- Vertrautheit mit Containerisierung, Docker, Kubernetes und Deployment-Praktiken in Cloud-Umgebungen
- Kenntnisse der MLOps-Prinzipien sowie gängiger Tools für Model-Lifecycle-Management, Experiment-Tracking oder Model-Monitoring
- Fähigkeit, sauberen, testbaren, dokumentierten und wiederverwendbaren Code zu schreiben
- Fähigkeit zur Zusammenarbeit in agilen, funktionsübergreifenden Teams mit technischen und produktbezogenen Ansprechpartnern
- Gute Englischkenntnisse
- Von Vorteil: Erfahrung mit GenAI- und Agentic-AI-Frameworks wie LangChain, LlamaIndex, Semantic Kernel oder vergleichbaren
- Von Vorteil: Erfahrung mit KI-Clouddiensten wie Azure AI, AWS Bedrock, AWS SageMaker, Google Vertex AI oder vergleichbaren Plattformen
- Von Vorteil: Vertrautheit mit Vector-Datenbanken und Retrieval-Technologien wie Pinecone, Weaviate, FAISS, pgvector oder vergleichbaren
- Von Vorteil: Kenntnisse verteilter Architekturen, Event-driven Architecture, API-Gateways und Microservice-Systemen
- Von Vorteil: Erfahrung mit MLOps-Tools wie MLflow, DVC, Kubeflow oder vergleichbaren Plattformen
- Von Vorteil: Kenntnisse in Responsible AI, Model Evaluation, Explainability, Applikationssicherheit und Datenschutz
- Von Vorteil: Erfahrung in Beratungs-, Enterprise- oder produktzentrierten Kontexten
- Von Vorteil: Zertifizierungen im Bereich Cloud, AI/ML oder Software Engineering

Was diese Stelle ausmacht:

- Investition in Zukunftskompetenzen ab dem ersten Tag durch Veranstaltungen, Tools und Kurse zu Künstlicher Intelligenz
- Über 50 Stunden technische und Management-Weiterbildung pro Jahr
- Klare Karriereentwicklungs- und Aufstiegsmöglichkeiten
- Internationale Mobilität, inklusive bis zu 20 Tage Smart Working aus dem Ausland ohne Dienstreise
- Hybrides Arbeiten ohne zentral vorgegebene Präsenztage
- Internationales Netzwerk und globale Zusammenarbeit
- Unterstützung durch persönliche Counselor ab dem ersten Tag
- Vielfältige Benefits- und Corporate-Welfare-Programme
- Teilnahme an Leadership-Programmen