

**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:**

AI Engineer (w/m/d)

**Ihr Aufgabengebiet:**

- Verantwortung für den gesamten Weg vom Prototyp bis zur Produktion - Fine-Tuning, Deployment, Inference-Optimierung, Monitoring und kontinuierliche Verbesserung von ML- und LLM-Systemen
- Mitwirkung an Projekten zu LLM-Serving und Inference-Stacks, Kosten-/Latenzoptimierung, Zuverlässigkeit strukturierter Ausgaben, Monitoring-Pipelines für Modelle oder Produktivsetzung von Lösungen, die gemeinsam mit AI Scientists entwickelt wurden
- Zusammenarbeit in Projektteams mit AI Scientists, Software Engineering, Infrastruktur und Fachexperten, um Lösungen mithilfe moderner Deployment-Tools (Docker, Kubernetes, CI/CD) reibungslos in Unternehmensökosysteme zu integrieren

**Was Sie ausmacht:**

- Master-Abschluss in Informatik, Ingenieurwesen oder gleichwertig
- Mindestens 3 Jahre Erfahrung als AI Engineer oder in einer ähnlichen Funktion
- Erfahrung im Deployment und Betrieb von ML-/LLM-Systemen in Produktion (Serving, Monitoring, Kosten-/Latenzoptimierung)
- Starke Programmierkenntnisse (Python, SQL) sowie Vertrautheit mit Deployment-Tools (Docker, Kubernetes, CI/CD)
- Vertrautheit mit ML-Bibliotheken (TensorFlow, PyTorch, Scikit-learn etc.)
- Praktische Erfahrung mit Cloud-Plattformen (AWS, Azure oder GCP)
- Gute Englischkenntnisse
- Von Vorteil: Erfahrung mit LLM-Inference-Stacks (z. B. vLLM) und Zuverlässigkeit strukturierter Ausgaben
- Von Vorteil: Vertrautheit mit Databricks oder verteilten Datenframeworks
- Von Vorteil: Erfahrung im Finanzsektor oder anderen regulierten Branchen

**Was diese Stelle ausmacht:**

- Unbefristete Festanstellung in Vollzeit
- Attraktive Vergütung und Entwicklungsmöglichkeiten
- Anregendes wissenschaftliches Umfeld mit informeller Arbeitsatmosphäre
- Mentoring, kontinuierliches Lernen und enger Austausch mit führenden Forschungsteams