

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:

Glattpark (Opfikon) (X Kilometer und X Minuten von Ihrer Wohnadresse entfernt)

Jobbeschreibung:

AI/ML Engineer & MLOps (w/m/d)

Ihr Aufgabengebiet:

- Entwicklung und Industrialisierung von ML/GenAI-Lösungen: Übersetzung von Business-Roadmaps in Lösungsdesigns sowie Implementierung robuster Trainings-/Inferenz-Pipelines und Produktionsservices
- Mitgestaltung von Lösungsarchitekturen zusammen mit Data Scientists und Domain Pods (Batch/Echtzeit, APIs, RAG-Muster)
- Industrialisierung von Prototypen: automatisierte Tests, sichere Paketierung, Deployment/Rollback sowie Performance- und Kostenoptimierung
- Aufbau und Weiterentwicklung von MLOps-/LLMOps-Pipelines inkl. standardisierter CI/CD-Workflows und Modellpromotion durch Dev-Test-Prod
- Pflege von CI/CD/CT-Templates inkl. Modell-/Prompt-Paketierung, Validierungs-Gates und Promotion-Prozessen
- Betrieb von Model Registry/Versionierung sowie automatisierten Retraining-Triggern gemäss ADAO-Governance
- Bereitstellung reproduzierbarer Experimentierumgebungen sowie Verwaltung von Feature-Store-Mustern und Validierung
- Durchführung von Evaluationsprozessen für ML & GenAI (Regressionstests, A/B-Tests, Safety- und Mehrsprachigkeitsprüfungen)
- Erstellung konformer Modell-Dokumentationen sowie Umsetzung von Governance-, Sicherheits- und Datenschutzkontrollen
- Implementierung von Responsible-AI-/GenAI-Leitplanken wie Grounding, Human-in-the-loop-Freigaben und Audit-Logs
- Implementierung vollständiger Observability für Modelle und Services sowie Pflege von Runbooks und Alerting
- Definition von SLOs/SLIs sowie Monitoring/Alerting für Qualität, Drift, Latenz, Verfügbarkeit und Kosten
- Mitwirkung bei Incident-Resolution, Postmortems und kontinuierlichen Zuverlässigkeitsverbesserungen
- Bereitstellung wiederverwendbarer ML/GenAI-Komponenten sowie Coaching der Teams zu MLOps-/LLMOps-Best-Practices
- Aufbau wiederverwendbarer Bibliotheken und Referenzimplementierungen (z. B. RAG-Konnektoren, Evaluationstools)
- Coaching von Pod-Engineers durch Reviews, Pairing und Clinics sowie Mitwirkung an ADAO AI Academy Materialien

Was Sie ausmacht:

- Bachelor- oder Masterabschluss in Informatik, Data Science, Ingenieurwesen, Mathematik oder gleichwertige praktische Erfahrung
- Mindestens 3 Jahre Erfahrung in ML Engineering, MLOps, Software-/Data Engineering oder verwandten Rollen mit nachgewiesener End-to-End-Produktion von ML/GenAI-Lösungen
- Praktische Erfahrung mit CI/CD und DevOps für ML-Workloads sowie mit Infrastructure-as-Code (Terraform/Pulumi) und GitOps
- Erfahrung mit ML-Frameworks (scikit-learn, PyTorch, TensorFlow) und Serving-Mustern (REST/gRPC, Batch, Streaming) sowie MLOps-Tools (MLflow, Kubeflow, Azure ML, Vertex AI, SageMaker)
- Fundierte Python-Softwareentwicklungskenntnisse (Testing, APIs), Erfahrung mit Containern (Docker) und Cloud-Deployment-Mustern; GenAI/RAG-Erfahrung von Vorteil
- Vertrautheit mit Feature Stores, Experiment Tracking, Datenqualitätstests und Orchestrierungstools (Airflow/Dagster/Prefect)
- Starke Kommunikations- und Stakeholder-Management-Fähigkeiten, Ownership-Mentalität und Belastbarkeit mit strukturiertem Engineering-Ansatz
- Fähigkeit, ohne formale Autorität Einfluss zu nehmen, Standards über Hub und Pods hinweg zu fördern und sich kontinuierlich an neue KI/ML-Technologien anzupassen; Erfahrung in der Telekommunikationsbranche von Vorteil

Was diese Stelle ausmacht:

- Dynamisches und multinationales Arbeitsumfeld, in dem jede Stimme zählt
- Gelebte Vielfalt als Teil der Unternehmens-DNA
- Partnerschaft mit «Advance – Gender Equality in Business»
- Ausgezeichnet mit dem Swiss LGBTI Label und der Zertifizierung «Fair-ON-Pay Advanced»