



Friedrich Cloud-Data-Steinhuber

MLOps Data Engineer

friedrich.steinhuber@example.de +49 6221 8901 2345

Heidelberg, Deutschland

friedrich-steinhuber.dev

linkedin.com/in/friedrich-steinhuber

github.com/friedrich-steinhuber

PROFIL

MLOps Data Engineer mit M.Sc. Data Science (TUM, Note 1,2) und 6 Jahren Erfahrung in produktiven ML-Pipelines bei DeepL SE und Aleph Alpha. Spezialisiert auf Feature-Engineering-Pipelines, Feature-Store-Architektur und LLM-Inference-Infra. Co-verantwortlich fuer Trainings-Daten-Pipelines im PB-Bereich fuer 24 produktive Modelle.

BERUFSERFAHRUNG

Senior MLOps Data Engineer

06/2022 - heute

Aleph Alpha

Heidelberg, Deutschland

MLOps-Team fuer LLM-Trainings- und Inference-Pipelines

- Aufbau einer LLM-Trainings-Daten-Pipeline mit 2,4 PB taeglichem Volumen auf Iceberg + Spark + Databricks
- Architektur eines Feature Stores mit 480 produktiven Features fuer 24 ML-Modelle, Latenz unter 28 ms
- Reduktion der GPU-Trainings-Kosten um 1,4 Mio. EUR p.a. durch Spot-Instance-Scheduling und Checkpoint-Resumption
- Implementierung eines Data-Versioning-Workflows mit DVC + lakeFS, Reproduzierbarkeit-Score 100%
- Mentoring von 2 Engineers zur Promotion und 4 ADRs pro Quartal zur Plattform-Architektur

MLOps Data Engineer

10/2018 - 05/2022

DeepL SE

Koeln, Deutschland

MLOps-Team fuer NMT-Trainings-Pipelines

- Aufbau einer Trainings-Daten-Pipeline fuer 280 Sprach-Paare auf Kubeflow + MLflow
- Migration von Airflow-1.x-Pipelines auf Dagster-Assets, MTTR von 4,2 h auf 22 min reduziert
- Aufbau eines Feature-Stores fuer 84 NLP-Features mit Feast + Redis
- Co-Autor der DeepL-Trainings-Daten-Quality-Policy mit 184 Daten-Qualitaets-Regeln

AUSBILDUNG

M.Sc. Data Science

10/2016 - 09/2018

Technische Universitaet Muenchen (TUM)

Muenchen, Deutschland

Data Science

1,2

B.Sc. Informatik

10/2013 - 09/2016

Technische Universitaet Muenchen (TUM)

Muenchen, Deutschland

Informatik

GPA: 1,4

FÄHIGKEITEN

MLflow + DVC + Weights & Biases

Apache Spark + Databricks

Feature Stores (Feast, Tecton)

Kubeflow + Vertex AI +

SageMaker

Python (pytorch, scikit-learn)

Apache Airflow + Dagster

Vector-DBs (Pinecone, Weaviate)

Terraform + Kubernetes + Helm

ZERTIFIKATE

Databricks Machine Learning Professional

10/2024

AWS Certified Machine Learning Specialty

07/2024

Google Cloud Professional Machine Learning Engineer

04/2024

MLflow Certified Practitioner

11/2023

SPRACHEN

Deutsch

Muttersprache

Englisch

C2

Franzoesisch

B2

PROJEKTE

feast-databricks-adapter (Open Source)

06/2022 - heute

Feast-Adapter fuer Databricks Feature Store, 1.640 GitHub-Stars, im offiziellen Feast-Plugin-Catalog

Masterarbeit: Feature Store Patterns fuer LLM-Trainings-Daten

06/2018 - 12/2018

TUM, Note 1,1, mit DFKI Saarbruecken kofinanziert

PUBLIKATIONEN

Feature Stores at PB Scale for LLM Training

06/2024

Co-Autor der Aleph-Alpha-Feature-Store-Story, im Databricks-Marketing-Material referenziert

STÄRKEN

ML-Pipeline-Disziplin

Pflegt MLflow-Experiment-Tracking pro Modell-Variante mit deterministischen Daten- und Code-Hashes

Feature-Engineering

Aufbau eines zentralen Feature Stores mit 480 produktiven Features fuer 24 ML-Modelle

Cost-bewusstes Training

Reduktion der GPU-Trainings-Kosten um 1,4 Mio. EUR p.a. durch Spot-Instance-Scheduling