

Kim Reuter

Data Scientist - Natural Language Processing

✉ kim.reuter@example.de

☎ +49 6221 9012 3456

📍 Heidelberg, Deutschland

🌐 kimreuter.dev

🌐 linkedin.com/in/kimreuter

📄 github.com/kimreuter



PROFIL

NLP-fokussierter Data Scientist mit 6 Jahren Erfahrung bei Bayer AG und BioNTech SE in klinischer Sprachverarbeitung. Fine-Tuning von 8B Open-Modellen mit QLoRA + DPO, RAG-Stacks mit pgvector + LlamaIndex, Hallucination-Reduktion 44%. Co-Autor eines EMNLP 2025 Papers zu Long-Context-Retrieval.

BERUFSERFAHRUNG

Data Scientist - NLP

Bayer AG

Pharmacovigilance-NLP im Drug-Safety-Team

📅 08/2022 - heute 📍 Heidelberg, Deutschland

- Fine-Tuning eines 8B Open-Modells mit QLoRA + DPO für klinische Berichts-Triage, Resolution-Rate von 61% auf 79%
- RAG-Stack auf pgvector + LlamaIndex mit Re-Ranking, Hallucination-Rate in blindem Eval um 44% reduziert
- Veröffentlichung eines EMNLP 2025 Papers zu Long-Context-Retrieval für Multi-Turn-Support-Workflows
- Aufbau einer HELM + Inspect-basierten Eval-Suite mit 180 Tasks für 5 Produkt-Teams
- Co-Pflege der Hugging-Face-Modell-Karten unter EU-AI-Act-Vorbereitung

Junior Data Scientist

BioNTech SE

Clinical-Note-NLP im R&D-Team

📅 10/2019 - 07/2022 📍 Mainz, Deutschland

- Training eines BioBERT + LoRA-Modells für klinische Entitäts-Extraktion (F1 0,91) auf 1,2 Mio. Records
- Aufbau einer Snakemake + dbt-Pipeline für 14 Genomics-NLP-Studien
- Co-Autor eines ACL 2022 Workshop Papers zu Faithfulness in klinischer Zusammenfassung

PROJEKTE

EMNLP 2025 - Long-Context-Retrieval

📅 06/2025 - 09/2025

Co-Autor des Hauptkonferenz-Papers zu Long-Context-RAG in Multi-Turn-Support

Bayer Drug-Safety-RAG

📅 01/2024 - 05/2024

Aufbau eines Pharmacovigilance-RAG mit pgvector, Hallucination-Rate -44%

AUSBILDUNG

M.Sc. Computational Linguistics

Universität Heidelberg

📅 10/2017 - 09/2019

📍 Heidelberg, Deutschland

Computerlinguistik • 1,1

B.Sc. Computerlinguistik

Universität Stuttgart

📅 10/2014 - 09/2017

📍 Stuttgart, Deutschland

Linguistik & Informatik • GPA: 1,3

FÄHIGKEITEN

Python (PyTorch, JAX)

Hugging Face Transformers

LoRA, QLoRA, DPO

RAG (pgvector, LlamaIndex)

HELM, Im-eval, Inspect

MLflow 2.x & KServe

Ray Train & SageMaker

spaCy & scispaCy

ZERTIFIKATE

DeepLearning.AI Generative AI with LLMs

📅 07/2025



NVIDIA DLI - Building LLM Applications with RAG

📅 01/2025



Hugging Face Reinforcement Learning Course

📅 11/2023



SPRACHEN

Deutsch	● ● ● ○ ○
Englisch	● ● ● ● ●
Französisch	● ● ● ● ●

STÄRKEN

Eval-Disziplin

Baue jede LLM-Pipeline mit HELM + Inspect inkl. Faithfulness- und Safety-Tasks

Domain-Tiefe

6 Jahre Pharma- und Biotech-NLP mit klinischen Terminologien (SNOMED, MedDRA, ICD-10)

Wissenschaftliche Strahlkraft

EMNLP 2025 Paper plus ACL 2022 Workshop Paper als Tiefen-Signal