**EIKE**

STEFFEN

KOHLMEYER



**M**

**A**

**C**

**H**

**I**

**N**

**E**

**L**

**E**

**A**

**R**

**N**

**I**

**N**

**G**

**E**

**N**

**G**

**I**

**N**

**E**

**E**

**R**

**Python**

-

years

4

**SQL**

-

4

years

**scikit-learn**

- 4 years

**numpy / pandas**

- 4 years

**Tensorflow**

- 3 years

**Pytorch**

- 2 years

**Git**

- 2 years

**Docker -**

years

2

**Linux**

2 years

-

**MS Azure**

- 2 year

**AWS**

-

1.5

years

**deepspeed**

- 1 year

**DBT**

- 1 year

**S**

**K**

**I**

**L**

**L**

**S**

+49 151 67204928

Trausnitzstr. 8, 81671, Munich

ekohlmeyer21@gmail.com

https://github.com/EikeKohl

Passionate team player experienced in applied Machine Learning,

particularly NLP. My experience spans the entire MLOps lifecycle, and I

am experienced in the application of Large Language Models. My high

intrinsic motivation, skillset, and solution-oriented work style make me

an excellent fit for your position.

**P**

**R**

**O**

**F**

**I**

**L**

**E**

**E**

**X**

**P**

**E**

**R**

**I**

**E**

**N**

**C**

**E**

SCAILEX GmbH - Munich, Germany

Atruvia AG - Munich, Germany

12/2021

- 08/2023

12

/2020

- 11/2021

**MACHINE LEARNING ENGINEER**

I am responsible for the complete MLOps lifecycle including training.

In this position, I was able to develop my skillset as a machine

learning practitioner as well as a software engineer. Please refer to

the "Projects" section for relevant projects and technologies used.

**DATA SCIENTIST**

As a Data Scientist at an IT Service Provider for banks, I developed

DL / ML applications in a banking context. Please refer to the

"Projects" section for relevant projects and technologies used.

HSBC - Düsseldorf, Germany

06

/2019

- 08/2020

**WORKING STUDENT - ASSET, LIABILITY**

**AND CAPITAL MANAGEMENT**

https://shorturl.at/aJK49

KPMG AG - Zurich, Switzerland

10

/2018

- 12/2018

**AUDITOR BANKS & ASSET MANAGEMENT**

KPMG Société coopérative - Luxembourg, Luxembourg

- 08/2018

02

/2018

**TRAINEE AUDIT ASSET MANAGEMENT**

**E**

**D**

**U**

**C**

**A**

**T**

**I**

**O**

**N**

Hochschule Düsseldorf

2019 - 2020

**BUSINESS ANALYTICS | M.SC.**

Fachhochschule Dortmund

2014 - 2018

**FINANCE, ACCOUNTING,**

**CONTROLLING AND**

**TAXES | B.SC.**

**P R O J E C T S**

# AI LAWYER (WIP)

AWS - Huggingface - Pytorch - DeepSpeed - OpenAI

Research and prototyping for an instruction-based, German LLM using RLHF, capable of:

1.Understanding legal cases,

2.Identifying applicable legal contexts,

3.Autonomously asking questions to obtain information relevant to the claim.

# INFORMATION RETRIEVAL (LLM)

LangChain - OpenAI - Huggingface - AWS - Streamlit

I have developed multiple Information Retrieval pipelines that utilize a vectorized knowledge database to enhance LLM prompts (e.g. ChatGPT) with contextual information, resulting in personalized and fact-based chatbot interactions. The services have been deployed on AWS, utilizing terraform Infrastructure as Code to ensure that authentication, load balancing, and other essential features are implemented effectively. Public reference: https://github.com/EikeKohl/paperqa-web-app

# INFORMATION EXTRACTION (NER & OCR)

Sagemaker - Huggingface - Pytorch - MLflow -AWS Lambda

Training, evaluation, deployment, and monitoring of information extraction models. In the context of the legal domain, this involves Named Entity Recognition (NER) of pertinent details from client, court, and insurance correspondence. To achieve this, I utilized transformers-based architectures and a BiLSTM-CRF ensemble model approach. Additionally, my team and I developed an Optical Character Recognition model to extract information from images sent by customers, employing a YOLOv5 + TrOCR model ensemble. Experiments for model training were conducted on Sagemaker and tracked in MLflow.

# DOCUMENT CLASSIFICATION

Sagemaker - Huggingface - Pytorch - MLflow -AWS Lambda

As a Machine Learning Engineer, I've trained, evaluated, and deployed document classification models for SCAILEX and its SaaS platform handling 500,000+ annual documents from various sources. Using SWEM and transformers-based approaches, we achieved a 60% automation rate with 95% accuracy. Model training experiments were conducted on Sagemaker and tracked in MLflow for efficient development.

# OTHER

AWS - DBT - SQL - CI/CD - Package Registry - Huggingface Pytorch - pydantic - CLI - Jupyter - Tensorflow - HDFS -

Docker

Usage of **Dynamic Time Warping** to reconstruct reference time series based on real-life time series data

Implementation of **Interpretable ML** approaches

Automatic **Speech Recognition** in banking context

Bosch **Hand Pose Detection** Hackathon

Developed internal "Dr.Prepper" **data preparation pipeline tool** for NLP

Implemented ML **data mart** **on Amazon Redshift** using DBT