Pengfei Zhao

https://pzdkn.github.io/

Email: p.zhao.6174@gmail.com
Mobile: +49-176-418-397-05
Github: https://github.com/pzdkn

EXPERIENCE

- **Merantix Momentum** — Berlin, Germany
  Machine Learning Engineer — *April 2022 - Present*
  - **RAG with LLM**: I developed the backend for a **retrieval augmented generation** application that included a vector database, **LLM** prompt engineering for generation, and experimentation on improving the semantic search using reranking and query expansion methods.
  - **Legal AI**: To help lawyers in the validation process of NDAs, I developed and deployed a classifier for identifying relevant parts of NDAs based on **sentence transformers.**
  - **Document AI**: I benchmarked Document AI models such as **LayoutLM** or **Donut** for finding a viable internal solution for doing ML on PDFs
  - **MLOps**: For my work, I use Terraform, Kubernetes, Docker and GCP, for deployment. The ML part is most often packaged into a Fast API backend and deployed via Kubernetes on GCP.

- **P3** — Stuttgart, Germany
  Data Scientist (WS) — *June 2021 - Sep 2021*
  - **Semantic Search**: To improve the search results of a public tender platform. I integrated a semantic search functionality into an existing web application. I further performed unsupervised finetuning of the sentence embeddings to domain-specific language of tenders using DeCLUTR

- **Ubiquitous Knowledge Processing Lab** — Darmstadt, Germany
  Student Research Assistant — *Oct 2019 to May 2020*
  - **Language Model Benchmarking**: UKP is a world-class research lab known for their work on sentence transformer or adapters. The project I worked at had the goal to identify pro and contra arguments inside a large document corpus. I benchmarked models to find the optimal trade-off between performance and resource-efficiency.

- **Squirrel-core** — Berlin, Germany
  Open Source Contributor — *April 2022 - Present*
  - **Python**: Squirrel-core is a **dataloading library** optimized for deep learning and machine learning over **cloud storage**. Its main perks include efficient storage access via sharding, parallel loading and streaming. I am one of the two main maintaner of the library, resolving issues, crafting RFCs, and implementing new features.

EDUCATION

- **Technical University Darmstadt** — Darmstadt, Germany
  Master of Science in Autonomous Systems; Grade 1.2 (best: 1.0) — *Oct. 2018 – Jan. 2022*
  - **Master Thesis**: Estimating high quality gradients is crucial for **reinforcement learning** agents to learn fast. Our study focused on utilizing a concept called **Measure Valued Derivatives** (MVD) for gradient estimation. We combined the MVD with exiting gradient estimation techniques to yield a novel estimator significantly lowering the gradient variance relative to exitsting estimators.
    Courses: Robotics, Sensor Technologies, Computer Vision, Machine Learning, Deep Learning

- **University Stuttgart** — Stuttgart, Germany
  Bachelor of Engineering in Engineering Cybernetics. — *Oct. 2013 – May. 2017*
  Courses: System Dynamics, Nonlinear Dynamics, Introduction to Control Theory

PROJECTS

- **Reimplementation Of Neural Radiance Fields**: A **Jax** reimplementation of the original **NERF** paper

- **Implementation of Merkle Tree**: A **Rust** implementation of Merkle Trees.

LANGUAGES

- **German**: Native   **English**: Proficient   **Chinese**: B2   **French**: A2

PROGRAMMING SKILLS

- **Languages**: Python, Rust, C++   **Technologies**: GCP, Kubernetes, Docker, PyTorch, Jax

AWARDS

- **Recipient of the Deutschlandstipendium:** The Deutschlandstipendium is a scholarscip provided by the German goverment for high-achieving students.