Yuanchun Shen

+49 157 5806 5470 | y.c.shen@tum.de | linkedin.com/in/yuanchunshen | github.com/happen2me

Education

Technical University of Munich — Munich, Germany
Master of Computer Science, Grade 1.6 — *Nov 2020 – Feb 2024*

Beijing University of Aeronautics and Astronautics — Beijing, China
Bachelor of Software Engineering, GPA 3.6 — *Sep 2015 – Jun 2020*

Experience

Machine Learning Working Student *(Part-Time)* — Apr 2023 – Dec 2023
Infineon Technologies — *Munich, Germany*

- Developed a **recommendation system** for informed project management with small internal datasets.
- Compared **12 regression models** for predicting numerous monetary costs, staff costs, and project duration.
- **Containerized** the inference endpoint, front-, and back-end; deployed them on OpenShift.

Data Scientist Intern *(Full-Time)* — May 2021 - Oct 2021
Amazon Web Service — *Shenzhen, China*

- Developed an **end-to-end knowledge-based dialogue system**. Partially accepted as official AWS samples.
- Constructed a **graph database** from text on Neptune; developed **natural-language query models** with BERT.
- Implemented a **continuous monitoring system** that monitors code update, data update, and performance drift.
- Enabled 1% **inference data collection** and automatically created **labelling task** for model improvement.
- Developed **MLOps pipelines** that automatically **retrain and deploy** the model with AWS Pipelines; the pipelines are triggered by the monitoring system.

Android Developer Intern *(Part-Time)* — Sep 2018 - Mar 2019
Institute of Automation, Chinese Academy of Science — *Beijing, China*

- Prototyped ADAS android application, integrated object detection algorithms written in OpenCV with JNI.

Projects

Graph-Retrieval-Augmented Language Generation | *NLP, LLM, Multi-Modality* — Nov 2022 – Nov 2023

- Proposed **graph-language alignment modules** to enable LLMs to **comprehend graph modality**.
- Created **GraphextQA**, a **new benchmark dataset** for evaluating graph-enhanced language models.
- Pretrained the graph alignment modules with **distant supervision** on Wikidata graphs and associated texts.
- **Enhanced generation accuracy** by integrating retrieved graph modality knowledge with proposed methods.

SRTK: Subgraph Retrieval Toolkit | *NLP, Contrastive Learning, Retrieval* — Mar 2023 – May 2023

- Developed a **Python library** for graph retrieval from Wikidata with natural questions; **published** on PyPI.
- Implemented **semantic-based** retrieval by iteratively querying graph database and comparing semantic relevance
- **Pipelined** preprocessing, training, retrieving, validation, and visualization into **reusable** subcommands.

Command Recognition for Surgical Robots | *Intention Detection, Sequence Labelling* — Nov 2021 - Mar 2022

- Implemented commands and numbers recognition with JointBERT to assist surgeon operations;
- Improved it by newly collecting and annotating datasets, along with text augmentation.

Distillation-based Domain Adaptation | *Model Compression, Transfer Learning* — Jan 2020 - May 2020

- Compressed **closed-source** trained vision models by **knowledge distillation** on reconstructed pseudo-images.
- Adapted the model to a target domain by leveraging **GAN** to extract **domain-invariant** intermediate features.

Skills

Programming: **Python**, Java, C/C++, SQL, Go, JavaScript
Technologies: **Deep Learning (NLP** & **CV)**, **MLOps**, **Huggingface**, **transformers**, PEFT, Quantization, Qdrant, Redis, AWS, Linux, Git, Docker
Certificates: **AWS Certified Machine Learning (Speciality)**, AWS Solution Architect (Associate)
Language: English (C1), German (B1), Mandarin (native), Japanese (N1)