Daniel (Haolan) Zuo

EDUCATION

M.S. in *Computer Science*, **Yale University**, New Haven, CT — May 2024
B.S. in *Computer Science* & ***B.S.*** in *Finance*, **Sichuan University**, Chengdu, China — Jun 2023

PROFESSIONAL EXPERIENCE

Meituan — Beijing, China
Machine Learning Engineer Intern — May 2023 – Aug 2023

- Part of a 3-person MLE team in Mobike division, driving the algorithmic optimization of shared e-bike logistics
- **Lifted profit by 7%** in end-to-end pipeline, by pioneering graph neural networks for **bike station recommendations**
- Created a **data ETL pipeline** with **PySpark** and **SQL** to transform spatio-temporal tabular data into a graph structure, yielding a large-scale heterogenous graph dataset (300K nodes 20M edges) ready for comprehensive network analysis
- Achieved **72% Acc@500** in component-specific testing, by building scalable GNN models using Tensorflow
- **Deployed** the model as a service using **Docker** and **Flask**, into Java heuristic-search order system via RESTful APIs
- Facilitated cross-team collaboration by authoring technical blogs, leading code reviews and sharing best practices

Johns Hopkins University - Center for Bioengineering Innovation and Design — Baltimore, MD
Machine Learning Engineer Intern — Jun 2022 – Sep 2022

- Developed iOS and watchOS apps to collect and transmit sensor data **using Swift** with a bioengineering team
- Established a **data pipeline** to structure and export timestamped sensor data from mobile **HealthKit** to a local server
- Achieved **84% accuracy** in classifying patient activity by building LSTM in Keras for a spine deformity clinical trial
- **Deployed** the model as a service on **a local server** with Flask for batch inference, enhancing **data privacy**

Yale Biotech Club – Bexorg Inc. — New Haven, CT
Data Scientist Externship — Oct 2023 – Nov 2023

- Enhanced high-frequency signal data integrity via **data cleansing**, feature selection, and downsampling using **Pandas**
- **Uncovered key causal relationships** among controllable and measured variables **using scikit-learn** for random forest, mutual information, and Bayesian networks, delivering **visual reports** to optimize brain cultivation experiment

SELECTED PROJECTS

Automated Fact-Checking System Using Knowledge Graph-Enhanced RAG — Jul 2024 – Present

- Improved retrieval and fact-checking quality by **innovating** a two-stage knowledge graph-enhanced RAG technique
- Boosted adaptability of retrieve-and-verify system via **joint finetuning** of the retriever and pretrained language model
- Increased transparency in decision-making by providing reference-supported justifications alongside predicted veracity, making the system more reliable for real-world use

Scalable Web Chatbot for Proprietary Data Question Answering with Voice Interaction — May 2024 – June 2024

- Delivered a seamless question-answering experience based on user proprietary data by developing a full-stack web chatbot, integrating Retrieval-Augmented Generation (RAG) with **LangChain** and **Chroma vector databases**
- Enhanced user engagement through intuitive voice interaction capabilities by integrating TTS and STT **cloud services**
- Ensured efficient, scalable performance by engineering **serverless deployment** using Docker in IBM Cloud

PUBLICATIONS

- J. Chen, **H. Zuo**, et al. *Graph Foundation Model for Expressive Structural Encoding*. **In Submission, ICLR 2025**
- W. Qiu, H. Chu, S. Wang, **H. Zuo**, et al. *Learning High-Order Relationships of Brain Regions.* **ICML 2024**
- **H. Zuo**. *A Fully Convolutional Denoising Auto-Encoder with 2.5 D Convolutional Classifier*. **IEEE TOCS 2022**

SKILLSET

Machine Learning: Python, PyTorch, Keras, TensorFlow, scikit-learn, HuggingFace (Transformers), XGBoost, ChatGPT;
MLOps: Git, Docker, Kubernetes, Flask, FastAPI, RESTful APIs, Gradio, Streamlit, LangChain, LlamaIndex, W&B, Tensorboard, Pytest, Prometheus, Grafana; Cloud Services: IBM Cloud, AWS, Google Cloud, Azure, Heroku;
Data Visualization: Jupyter, Matplotlib, Seaborn, Tableau, Plotly; Data ETL: Pandas, NumPy, PySpark, SQL, Hive