

Daniel (Haolan) Zuo

EDUCATION

<i>M.S. in Computer Science, Yale University</i> , New Haven, CT	May 2024
<i>B.S. in Computer Science & B.S. in Finance, Sichuan University</i> , Chengdu, China	Jun 2023

PROFESSIONAL EXPERIENCE

Meituan	Beijing, China
<i>Machine Learning Engineer Intern</i>	May 2023 – Aug 2023
<ul style="list-style-type: none">Part of a 3-person MLE team in Mobike division, driving the algorithmic optimization of shared e-bike logisticsIncreased overall profit by 7% in end-to-end pre-launch A/B testing, by pioneering scalable graph neural networks for bike station recommendations using Tensorflow, achieving 72% Recall@500 in component-specific testingCreated a data ETL pipeline with PySpark and SQL to transform spatio-temporal tabular data into a graph structure, yielding a large-scale heterogeneous graph dataset (300K nodes 20M edges) ready for comprehensive network analysisDeployed the model as a service using Docker and Flask, into Java heuristic-search order system via RESTful APIsFacilitated 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
<i>Machine Learning Engineer Intern</i>	Jun 2022 – Sep 2022
<ul style="list-style-type: none">Developed iOS and watchOS apps to collect and transmit sensor data using Swift with a bioengineering teamEstablished a data pipeline to structure and export timestamped sensor data from mobile HealthKit to a local serverAchieved 84% accuracy in classifying patient activity by building LSTM in Keras for a spine deformity clinical trialDeployed 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
<i>Data Scientist Externship</i>	Oct 2023 – Nov 2023
<ul style="list-style-type: none">Enhanced high-frequency signal data integrity via data cleansing, feature selection, and downsampling using PandasUncovered 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

<i>Automated Fact-Checking System Using Knowledge Graph-Enhanced RAG</i>	Jul 2024 – Present
<ul style="list-style-type: none">Improved retrieval and fact-checking quality by innovating a two-stage knowledge graph-enhanced RAG techniqueBoosted adaptability of retrieve-and-verify system via joint finetuning of the retriever and pretrained language modelIncreased transparency in decision-making by providing reference-supported justifications alongside predicted veracity, making the system more reliable for real-world use	
<i>Scalable Web Chatbot for Proprietary Data Question Answering with Voice Interaction</i>	May 2024 – June 2024
<ul style="list-style-type: none">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 databasesEnhanced user engagement through intuitive voice interaction capabilities by integrating TTS and STT cloud servicesEnsured 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