

Yuxuan Zhang (Wayne)

linkedin.com/in/yxzwayne

github.com/yxzwayne

yxzwayne@gmail.com

EXPERIENCE

Machine Learning Intern

Mar. 2024 – Current

Stealth

Remote

- Built and validated an in-house reranker model using open-source **cross-encoders**, achieving proprietary system accuracy with lower cost and **200ms** reduction in search response time.

Machine Learning Intern

Sep. 2023 – Current

Salk Institute for Biological Studies

San Diego, CA

- Trained custom protein **variational autoencoders** and **Transformers** in **Keras** on 6 million protease sequences for mutation analysis using latent space tracking.
- Engineered a **PCA**-based tool for cryo-EM data selection to complement clustering analysis in cryo-DRGN, uncovering new protein structures by surpassing traditional K-means and GMM clustering limitations.

Software Engineer

Dec. 2021 – Jul. 2022

Awesung Inc.

Cranbury, NJ

- Optimized logistics operations using previously unincorporated third-party data, devised operational plans by analyzing data with **Pandas** and **Matplotlib**, reduced Item-Not-Received complaints by **15%** in two months.
- Developed a logistics management application using **React.js**, **Django**, and **MySQL**.
- Optimized **MySQL** database schema and queries, reducing serialization time by **20%** and **RestAPI** response times by **50%**, improving performance and user experience.

Software Engineer Intern

Oct. 2020 – Dec. 2021

AILaw

Remote

- Shipped new features and unit tests for a legal document management app in **Vue.js** weekly.
- Revitalized a customer performance reporting service in **Python**, enhancing efficiency by **20%** through query optimization, re-engaging a dozen clients within two months.

PROJECTS

Medilora — Large language models, Fine-tuning

- Led a team of 4 to fine-tune OpenHermes-2.5-Mistral-7B with Q-LoRA on 300M medical text tokens.
- Implemented distributed training with **Axolotl**, inference with Llama-cpp-python and evaluation with **LM-Eval**.
- Improved **PubMedQA** and **MedQA** over **20%** and matching state-of-the-art 70B open models on **MMLU-Medical** with **0.05%** of data size.

Open-ended Document Classification — Large language models, Classification

- Implemented the recursive summarization method in 2109.10862, generated summary embeddings for ArXiv papers using OpenAI ChatGPT API and generated open-ended class labels with RoBERTa model.

CompassX Platform — Software Engineering, Education

- Founding developer of CompassX at UCSD, led full-stack development using **Flask**, **JavaScript** and **PostgreSQL**, scaled data schema and backend up to support multi-college expansion.

EDUCATION

University of California San Diego

La Jolla, CA

M.S. Data Science. GPA: 3.5

Sep. 2022 - Present

- Courses and Involvements** Attention/Diffusion Reading Group, RL/LLM seminar, Search and Optimization, Learning Algorithms, Deep Learning, Scalable Data Systems, Data-driven Text Mining, Statistical Modeling
- Teaching Assistant** DSC 30: Data Structures & Algorithms

University of Washington Seattle

Seattle, WA

B.S. Informatics. GPA: 3.5

Sep. 2017 - Jun. 2021

TECHNICAL SKILLS

Programming: Python, Java, SQL, JavaScript, R

Computing Libraries: Keras, Jax/Haiku, PyTorch, XGBoost, Pandas, NumPy, Matplotlib

Frameworks: Apache Spark, Dask, React, Django, Node.js, Flask, JUnit

Tools: Git, Docker, Linux, AWS, Vim, Jupyter