Yuxuan Zhang (Wayne)

yxzwayne.com | yxzwayne@gmail.com

github.com/yxzwayne

linkedin.com/in/yxzwayne

EXPERIENCE

Machine Learning Intern
Mar. 2024 – Current

trufflepig.ai — *Remote*

- Implemented document ingestion, **FAISS** vector search and reranking with both Cohere Rerank API and BGE cross-encoders on PDFs from finance, patent law and academia domain.
- Evaluating hit rates and mean reciprocal ranks from BGE-m3 and MiniLM based cross-encoders by orchestrating GPT-4 and Hermes-2-Pro-Llama3 to generate over 10k synthetic queries and evaluate rerank outputs.

Machine Learning Intern
Sep. 2023 – Current

Salk Institute for Biological Studies — *San Diego, CA*

- Training protein **Transformers** and **variational autoencoders** in **Keras** on multiple sequence alignments from over 6 million protease and integrase sequences, applying sparse autoencoders for epistasis analysis.
- Developed 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.

PROJECTS

Medilora — Large language models, Fine-tuning
Oct. 2023 - Dec. 2023

- Fine-tuned OpenHermes-2.5-Mistral-7B with Q-LoRA using **Axolotl** on 300 million medical text tokens.
- Improved **PubMedQA** and **MedQA** evaluation scores by over **20%**, matching the state-of-the-art 70B **Meditron** on **MMLU-Medical** with **0.05%** of the data size.

Open-ended Document Classification — Large language models, Classification
May 2023 - Jun. 2023

- Implemented the recursive summarization method in 2109.10862 by generating summaries for ArXiv papers using ChatGPT API and generating open-ended class labels with **RoBERTa**.

CompassX Platform — Software Engineering
202

- Founding developer of CompassX at UCSD, led full-stack development using **Flask**, **JavaScript**, and **PostgreSQL**, scaling data schema and backend 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, Maths of Deep Learning, Scalable Data Systems, 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, PyTorch, XGBoost, Pandas, NumPy, Matplotlib
Frameworks: React, Django, Node.js, Flask, JUnit
Tools: Git, Docker, Linux, AWS, Vim, Jupyter