

Yuxuan Zhang (Wayne)

http://www.yxzwayne.com

github.com/yzwayne | linkedin.com/in/yzwayne

yzwayne@gmail.com | 206-557-8944

EXPERIENCE

Machine Learning Intern

Sep. 2023 – Current

Lyumkis Lab, Salk Institute for Biological Studies

San Diego, CA

- Led research on testing the generative capacity of VAEs, containerized model training programs in Docker.
- Developed latent-space tracking methods on cryo-DRGN output data, leading to new methodology research and streamlining cryo-EM data interpretation.
- Collaborating with other researchers, developing data-driven methods to explore new research directions from existing cryo-EM datasets and artifacts.

Teaching Assistant

Apr. – Aug. 2023

Hacıoğlu Data Science Institute, UC San Diego

San Diego, CA

- Head TA for DSC 30: Data Structures & Algorithms, managed and supervised students in Java programming assignments, led discussion sessions of 20-30 students to review and clarify students' queries.

Software Engineer

Dec. 2021 – Jul. 2022

Awesung Inc.

Cranbury, NJ

- Shipped features weekly, working full-stack with React.js and Django for the internal logistics management app.
- Reduced data serialization time by over 20% and achieved a 50% reduction in API response times.
- Spearheaded the eBay Verification initiative, implementing an ETL pipeline to leverage untapped data, decreasing Item-Not-Received complaints by over 15% in two months, significantly improving operational reliability and customer trust.

Software Engineer Intern

Oct. 2020 – Dec. 2021

AILaw

Remote

- Shipped new features and unit tests in Vue.js for a legal workflow management app.
- Restored and optimized a defunct daily reporting script, reducing runtime by over 20% and enabling the recovery of a dozen clients within two months.

PROJECTS

Medilora — Large language models, Fine-tuning

- Led Medilora project, fine-tuned large language models with Q-LoRA on 300M medical text tokens.
- Set up model+data parallel training, inference and evaluation experiments in remote linux servers.
- Achieved comparable medical benchmark performance with 0.05% of the training data required by larger medical models, demonstrated effectiveness of small pretrained models.

Open-ended Document Classification — Large language models, Classification

- Adapted 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.

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

University of Washington Seattle

Seattle, WA

B.S. Informatics. GPA: 3.5

Sep. 2017 - Jun. 2021

TECHNICAL SKILLS

Programming: Python, Java, SQL (Postgres), JavaScript, R

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

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

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