JOHNNY HUANG

San Mateo, CA, 94403 | (650)-278-6570 | h.johnny@wustl.edu | Portfolio | Linkedin | GitHub

EDUCATION

St. Louis, MO

Washington University in St. Louis

August 2022- May 2026

- B. S: Computer Science + Mathematics — *GPA: **3.9***
- Honors and Activities: Chancellor's Fellow, Taylor Scholar (APM Mentor), **3x** WashU Hackathon (2023 Co-organizer), Computational Imaging Group, Multimodal Vision Laboratory, **VP** of WashU Robotics, **VP** of First-Generation Investors Club
- Relevant Coursework: Machine learning, Bayesian ML, Reinforcement Learning, Data Mining, Data Structures and Algorithms, Computer Engineering, Convex Optimization

PROFESSIONAL EXPERIENCES & INVOLVEMENT

Head TA for CSE 412: Intro to AI

Jan 2023 - Current

McKelvey School of Engineering — *St. Louis, MO*

- Managing a team of **30+** TAs to assist over **1500+** students across multiple terms; regularly attended lab sections and hosting bi-weekly recitations; assigning grading and proctoring exams.

Large Language Models Intern

May - July 2024

Rad AI — *San Francisco, CA*

- Led the development of a full-stack **RAG chain system** using **Langchain** and **Pinecone** vector database for optimized search versatility; established access endpoints with **FastAPI** and deployed through **Docker** on **AWS Sagemaker**.
- Fine-tuned Gemma2-7b for function-calling leveraging **Hugging Face's PEFT** and **LoRa** finetuning which increased training speed by **600%**; optimizing model performance through sharding; utilized the **Pydantic** framework for data validation.

Machine Learning Intern

Jan - May 2024

Mallinckrodt Institute of Radiology — *St. Louis, MO*

- Developed a modified U-Net using **PyTorch, CUDA** and **Caffe** for fMRI segmentation**;** applied YOLO for anomaly detection, integrating a PnP-FISTA pipeline for improving runtime speed; algorithm successfully predicts **93%** of critical regions.
- Constructed a deep cGAN to streamline more robust synthetic CGM v.s. Cognitive function data generation using **TensorFlow.**

Software Engineering Lead

September 2023 - Feb 2024

WashU Robotics, MATE ROV — *St. Louis, MO*

- Spearheaded robot sensory systems processing efficiency speed through integrating the AutoViz and Gazebo API in **C++**; optimizing scripts compilation time in poolside controller by **150%** using ROS and **C** on **Linux**.

Data Science Intern

May - July 2023

Couch Biomedical Science — *St. Louis, MO*

- Denoised and refined collected data by implementing a single-celled Deep-Count Autoencoder using **PyTorch**; updating databases with **PySpark** and **MySQL** on **Kubernetes** clusters for parallelization, increasing processing speeds by **500%.**
- Clustered and classified DNA sequence into gene segments by developing SVM and Agglomerative models using **Sci-Kit, NumPy,** and **Pandas**; processed models on **Azure**; visualized results and presented to wet lab.

PERSONAL PROJECTS (Please See More on my Portfolio!)

- **HalluAgent:** Developed a framework for utilizing LoRA tuned SLMs to detect and correct hallucination patterns in GPT-3.5; leveraged SLMs as agents to call various functional **API's** to evaluate LLM response; retrained GPT-3.5 using **HF** libraries with results; generated **2000** robust trajectory training data with GPT-4 for agent tuning.
- **Petrichor:** A mental health website aiming to match users with their perfect therapist; implementing user login, menus, calendars; data stored with **MySQL** on **Linux**; hosted on **AWS**; coded mainly using **HTML/CSS, PHP,** and **JS.**

SKILLS

- **ML:** PyTorch, TensorFlow, HuggingFace, Pydantic, Langchain, PySpark, Azure, SageMaker, Pinecone, CUDA
- **Backend:** Java, C++, Scala, Python, C, NodeJS, Express, FastAPI, MySQL, NoSQL
- **Frontend:** PHP, React, JavaScript, HTML, CSS, Swift, Apache
- **Others:** Jupyter**,** AWS, MongoDB, Docker, Kubernetes, Git, Bash, PowerShell, R, Matlab