# 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/M. S: Computer Science + Mathematics *GPA:* ***3.9***
- Honors and Activities: Chancellor's Fellow, Taylor Scholar, APM Leader, **3x** WashU Hackathon (2023 Co-organizer), Computational Imaging Group, Multimodal Vision Laboratory, **VP** of WashU Robotics
- Relevant Coursework: ML II, Bayesian ML, RL, Data Mining, Data Structures and Algorithms, Rapid Prototyping, Computer Engineering, Linear Algebra, Optimization, Stochastic Processes, Discrete Mathematics

## *PROFESSIONAL EXPERIENCES & INVOLVEMENT*

### Head TA for CSE 240 Discrete Mathematics *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; provided support as secondary TA for CSE 412 and 247.

### 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** for improved search versatility; established access endpoints with **Fast-API** 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 Co-op *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 generate synthetic glucose fluctuation data based on cognitive function criteria using **TensorFlow**, streamlining more robust patient data analysis.

### Data Science Intern *May - July 2023*

Couch Biomedical Science *Remote, MO*

- Denoised and refined collected data by implementing a single-celled Deep-Count Autoencoder using **PyTorch**, updating databases with **PySpark**, **SQL**, and **Kubernetes** clusters for parallelization, increasing processing speeds by **200%.**
- Classified gene segments into clusters by creating SVM**,** Decision Trees, and Naïve Bayes models using **Sci-Kit, NumPy,** and **Pandas**; processed models on **Microsoft Azure**, enhancing the efficiency of the gene segmentation pipeline.

## *PERSONAL PROJECTS ( See Examples and More on my Portfolio )*

- **HalluAgent:** Developed a framework utilizing a LoRa fine-tuned small language models (openbmb/MiniCPM3-4B) to detect and correct hallucination patterns in the popular GPT-3.5. Created a mechanism in **Python** for assigning confidence scores to segmented LLM CoT by enabling SLM to call custom tools & Google's web search **API.** Leveraged confidence to tune GPT-3.5 using **HF** libraries, enhancing its trustworthiness; also generated a trajectory dataset for fine-tuning HalluAgent utilizing GPT-4.
- **Virtual Gym:** A online trainer that provides users feedback with dynamic workouts using CV; leveraged the **Open Weather** and Google's **Media-Pipe API** for suggesting routines & classifying movements with **80%** accuracy; servers hosted on **Flask.**
- **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, Numpy, PySpark, Pandas, Azure, SageMaker, Pinecone, CUDA
- **Backend:** Java, C++, Scala, Python, C, Node.js, Rest API, Fast API, SQL
- **Frontend:** PHP, React.js, JavaScript, AJAX, HTML, CSS, Swift, Apache
- **Others:** Jupyter**,** AWS, Docker, SSH, Git, Terminal, Bash, PowerShell, R, Matlab, Arduino