

Charles Phu

U.S.Citizen | 510-458-1787 | charlesphu18@gmail.com | [linkedin.com/in/charles--phu](https://www.linkedin.com/in/charles--phu) | github.com/charlesphu

EXPERIENCE

Supermicro

July 2025 – Current

Systems and Software Engineer

- Perform bring-up, validation, and software integration of NVIDIA **GB200**, **GB300**, and **Vera Rubin** AI platforms
- Built internal web platform that automates L10, L11, and L12 validation testing across rack-scale GPU clusters, providing centralized monitoring, execution, and reporting
 - Built using React JS, MongoDB, Fast Api, Python, Bash, Docker
- Validated NVIDIA Mission Control software and cluster management workflows on GB200 AI infrastructure
- Validated GB200 and GB300 racks for NVIDIA L11 PVP (Partner Validation Playbook)
- Developed Python and Bash automation to streamline rack deployment, diagnostics, validation, and log collection across multi-node GPU clusters
- Executed MLperf AI training / inference on GB300 using **Kubernetes** and **Docker**
- Supported onsite deployment L11 and L12 testing projects for **OpenAI**, **Tesla**, **Lambda**, **Firmus**, and **Global AI**

Headstarter AI

July 2024 – September 2024

Software Engineering Fellow

- Collaborated in a team of 4 to develop an interactive customer support agent using Next.js, integrated a custom RAG pipeline using **OpenAI** and **Pinecone** that responds based on a company's knowledge base
- Engineered a SaaS product that generates dynamic flashcards based on any topic using the **Llama 3.1 LLM**
- Implemented a web scraper to extract data from Rate my Professor and upsert to a **Pinecone** index, integrated with a **RAG** pipeline using **LangChain** and **OpenAI GPT-4o** to get up-to-date and relevant answers to user queries

EDUCATION

University of California, Santa Cruz

Santa Cruz, CA

Bachelor of Science in Computer Science, GPA: 3.75

Relevant Coursework: Applied Machine Learning: Deep Learning, Artificial Intelligence, Data Structures and Algorithms, Principles of Computer Systems Design, Computational Models, Database Systems

SKILLS

Programming: Python, Bash, C/C++, JavaScript, SQL, HTML/CSS, Lua/Luau

AI Infrastructure: Linux, NVIDIA GB200, NVIDIA GB300, Vera Rubin, CUDA, NCCL, MPI, InfiniBand, DCGM

Frameworks & Databases: React, Next.js, FastAPI, MongoDB, Tailwind CSS, Pandas, PyTorch, Docker, Kubernetes

Tools: Git, SSH, Bash Scripting, JSON, YAML, Regex

PROJECTS

[Roblox 2v2 team space game](#) | *ROBLOX studio, Lua, Scripting, Modeling* | roblox.com/games/8715055706/Solar-Strife

- Team-based strategic PVP sci-fi game featuring fully armed spacecraft with gyroscopic movement
- Custom-modeled planes, ships, bases, and turrets with fully functional ship and player combat system

[SpeedFill](#) | *Javascript, Next.js, Supabase, Vercel, Gemini, Figma* | <https://speed-fill.vercel.app/>

- **Product Owner and Manager**
- Developed a website to analyze how well resumes match up with job descriptions, generate cover letters, and receive AI-feedback
- Led the implementation of authentication, data storage, and UI components while optimizing backend processing and user experience

AI Form Detector | *YoloV8, Transfer Learning, Computer Vision*

- Created an AI model to validate squat, bench, and deadlift form by using YoloV8 achieving 85% accuracy in classification

LEADERSHIP AND PROFESSIONAL DEVELOPMENT

Upper Division Tutor: Principles of Computer System Design

September 2024 – July 2024

University of California, Santa Cruz

- Provided one-on-one support to students, resolving programming issues and debugging code to improve assignment outcomes
- Assisted students with problem-solving and code debugging in programming assignments, ensuring efficient and effective solutions