Andrew Bruce

+1(408)207-3992 | andy@andycbruce.com | LinkedIn | GitHub | Portfolio

Experience

University of California Santa Cruz, CA

Graduate Researcher, Teaching Assistant *Dec 2023 – Present*

- Ported OpenMM, a molecular dynamics simulation library used by thousands of scientists, to run with 16-bit floating point precision CUDA increasing simulation speed.
- Designed and trained graph neural networks for predicting the energy surface of coarse-grained proteins, massively speeding up the simulation of protein interactions.
- Implemented sharding algorithms for running simulations over hundreds of A100 GPUs in multiple compute clusters

Suzhou Benqio Technology Ltd. Suzhou, China

Machine Learning Engineer *July 2023 – March 2024*

- Designed and wrote a machine learning data pipeline for automatically filtering insurance contracts for MSIG insurance.
- Saved hundreds of employees' time to find the most viable contracts.
- Implemented the entire pipeline locally to prevent sensitive client data from leaking.
- Saved thousands of dollars in deployment by quantizing LLama to lower precision floating point for more efficient inference, reducing memory usage 4x and allowing inference of a 33 billion parameter LLM to run with under 24GB of VRAM in just a single GPU.
- Very efficiently fine-tuned LLMs to small custom datasets using LoRA training.
- Designed and wrote a custom HTTP server implementing REST APIs handling internal company requests.
- Used sentence embeddings and Qdrant as a vector database to index and recall relevant parts of documents for use by LLMs.

Pinpoint AVL Santa Cruz, California

Software Engineer *October 2022 – June 2023*

- I made GPS tracking for TAPS and Santa Cruz Metro, which tens of thousands of students use for transportation to class.
- I integrated Google Maps to track individual metro bus routes' real-time location and capacity, allowing students and faculty to plan their trips more efficiently.
- Designed and built a device mounted on buses for tracking location and capacity.
- Created data processing pipelines for GTFS feeds and GPS data to optimize transportation routes and detect bottlenecks.
- I used Raspberry Pi Picos, Arduinos, and a custom bus driver dashboard connected to the bus power supply with inter-device communication using I2C and UART.
- Utilized custom multi-core scheduling in the Raspberry Pi to balance workload between logging to disk, LTE communication, and I2C.
- I wrote embedded AArch64 applications and implemented the backend REST API.

AbiTalk, Inc. San Jose, California

Software Engineer *July 2019 – June 2022*

- Developed educational mobile games for kids learning English and speech impediments used by hundreds of school districts.
- Implemented GUI, sound design, and statistics gathering for data-driven teaching.
- I worked with the Apple ecosystem, Swift programming, music licensing, and publishing for the Apple App Store.

Skills

Languages: C/C++, Python, CUDA, Javascript/Typescript, Rust, Go, SQL, React, RISC-V Assembly, Haskell, Agda, LaTeX, Postscript, GLSL

Technologies: PyTorch, OpenCV, SQLite, PostgreSQL, Webassembly, WebGL, Git, Linux Kernel Modules

Education

University of California, Santa Cruz Santa Cruz, California

B.S ***GPA: 3.97/4.00*** *Sep 2022 – Jun 2024*