andycbruce@gmail.com
www.andycbruce.com
github.com/andychenbruce

Andrew Bruce

6241 Placid Ct, San Jose, CA 95135
+1 (408) 207-3992

I am available to work full time from June 17, 2024 to September 20, 2024.

Work Experience

May 2024 - Present

Graduate Researcher - *University of California Santa Cruz, California*

- Ported OpenMM, a molecular dynamics simulation library used by thousands of scientists, to run with 16 bit floating point precision on CUDA. backend of GPU to speed up the simulation of atoms.
- Currently training many different neural net models on predicting the energy fields of a protein. If successful could be used to massively speed up the simulation of proteins.

July 2023 - March 2024

Machine Learning Engineer - *Suzhou Benqio Technology Ltd., Suzhou, China*

- Quantized LLMs to GGUF files for more efficient inference on the GPU, reducing memory usage by 4x and allowing inference of a 33 billion parameter model to run with under 24GB of ram.
- Very efficiently retrained LLM's to small custom datasets using LoRA matricies on just a single GPU.
- Designed and wrote a custom HTTP server implementing REST APIs. Backend application allowed conversations with any GGUF formatted LLM using Llama.cpp and Candle, and being able to read information from documents using sentence embeddings and the Qdrant vector database, and storing conversations using SQLite.
- Created a frontend in Webassembly for real time conversation and uploading PDF documents to the LLM.
- Designed and wrote a data pipeline for extracting text from pdfs, OCR-ing embedded images using Tesseract, and extraction of relevant information using sentence transformers for an insurance company to save employees time on finding viable insurance contracts. The entire pipeline had be done locally, without connection to any of the internet, to prevent sensitive data from leaking.

October 2022 - June 2023

Embedded and Frontend Developer - *PinPoint AVL, Santa Cruz, California*

- Created device for tracking busses and counting capacity using a Raspberry Pi Pico, Arduino, a bus driver dashboard, powered by a battery.
- Implemented inter-device communication using I2C, and debugging over the UART protocols.
- Sent realtime coordinates to the cell network using LTE chips.
- Utilized custom multi-core scheduling in the Raspberry Pi to balance workload between logging to disk, LTE communication, and I2C to the Arduino.
- Crosscompiled C++ using Cmake to ARM AArch64. Implemented backend REST API in C#.
- Implemented frontend Google Maps integration in TypeScript to track the real time location and capacity of UCSC bus routes.
- Created many Python scripts for processing data in GTFS feeds and logged data in CSVs.

July 2019 - June 2022

Mobile Game Developer - *AbiTalk, Inc., San Jose, California*

- Developed educational mobile games for kids learning English and speech impediments.
- Implemented GUI, sound design, and sending statistics to the backend to track progress.
- Worked with Apple ecosystem, Swift programming, music licensing, and publishing to the IOS App Store

Education

2024-Present

University of California Santa Cruz - *Computer Science MS*

I will complete the MS in June of 2025 through UCSC's Contiguous Pathways master's degree

2022-2024

University of California Santa Cruz - *Computer Science BS*

3.93 GPA, with a GPA of 4.0 in my CS classes.

Skills

C, C++, Rust, Python, JavaScript, TypeScript, React, SQL, SQLite, PyTorch, CUDA kernels, OpenGL, WebGL, GLSL, WebAssembly,
Bash, Make, CMake, Git, Embedded Programming, Linux Devops, Linux Device Drivers, Linux Kernel Modules, OSDev
Machine Learning, Computer Vision, OpenCV
Full-stack, Agile, Scrum
Postscript, LaTeX