
Cade Brown

Agentic performance optimization | Compilers & DSLs | GPU numerical libraries

brown.cade@gmail.com · linkedin.com/in/cade-brown · github.com/cadebrown · cade.io

Selected Experience

NVIDIA | Senior CUDA Math Libraries Engineer | 2023–Present

- Optimize cuFFT performance for Blackwell, Rubin, Feynman, and future GPU architectures.
- Developed an AI-agent framework for CUDA kernel generation, profiling, and autotuning.
- Designed AI-agent harnesses and tools (MCPs, CLIs, ...) to accelerate development.
- Developed CUDA Tile integration and autotuning across SIMT, cuTile Python, NVCC, and LTO IR.
- Optimized ML inference in PyTorch using internal research techniques.

Innovative Computing Laboratory @ UTK | Research Assistant (HPC) | 2019–2023

- Developed MAGMA and SLATE GPU kernels; optimized MAGMA BLAS for AMD GPUs.
- Designed SABATH for reproducible scientific ML following FAIR principles.

Qardian Labs | Machine Learning Consultant | 2020

- Trained and evaluated an ML model for heart-disease identification and anomaly detection, with a web interface for risk assessment.

Oak Ridge National Laboratory | Research Intern (HPC) | 2017–2018

- Built FractalExplorer, a distributed GPU fractal visualizer for the SimpleSummit Jetson cluster.

Selected Publications

Brown et al. Design, Optimization, and Benchmarking of Dense Linear Algebra Algorithms on AMD GPUs. IEEE HPEC, 2020.

Luszczek & Brown. Surrogate ML/AI Model Benchmarking for FAIR Principles' Conformance. IEEE HPEC, 2022.

Selected Projects

mycc: C-subset compiler using Bison, Flex, and LLVM, with an optional ORC JIT path.

kscript: Dynamic language implemented in C, with arrays, linear algebra, FFTs, and a WASM REPL.

Skills

Languages: C/C++, CUDA, Python, Rust, TypeScript.

Machine Learning: PyTorch, JAX, Triton, cuTile Python, LLVM/MLIR, JIT/LTO.

Agentic Engineering: Claude, Codex, Cursor, OpenCode, Pi, MCPs, harnesses, CLIs.

Education

Bachelor's degree, Computer Science | University of Tennessee, Knoxville | 2023