TONY (KHANG) CAO

916-891-7031 tonykcao@gmail.com linkedin.com/in/tonykcao/ github.com/tonykcao

EDUCATION

Swarthmore College Expected May 2025
Bachelor of Arts in Computer Science and Applied Math (GPA: 3.76 / 4.00) *Swarthmore, PA*

- *Relevant Coursework*: Operating Systems, Parallel & Distributed Systems, Randomized Algorithms, Machine Learning, Statistical Methods II, Probability, Stochastics, Differential Geometry, Calculus III, Differential Equations, Linear Algebra

EXPERIENCE

Research Assistant Jun 2024 – Aug 2024
Dive Into Systems, Swarthmore College Research — GitHub *Swarthmore, PA*

- Enhanced 5 chapters of open-source computer systems textbook used at over 60 institutions by developing interactive exercises using jQuery and HTML, improving engagement and learning outcomes for CS undergraduates
- Created 2 core algorithms for simulating assembly and C code syntaxes with JavaScript and D3.js, enriching educational content with programmatically generated problems with answers validation
- Refined algorithms to enable control of output distribution, and redesigned UI to increase interactivity and educational value

Software Engineer Intern Jun 2023 – Aug 2023
GMD Protocol (DeFi Startup) *Ho Chi Minh City, Vietnam*

- Built Python-based chatbot that enables users to place leveraged trades instantly on decentralized exchanges, securing $500,000 in external funding by demonstrating the product's potential to investors
- Developed internal price monitoring and command-line tools for portfolio management, streamlining traders' workflows and eliminating manual input, reducing transaction times from minutes to seconds in volatile markets
- Launched a secure MongoDB database with security measures including salt and pepper hashing, password protection, and killswitch to safeguard private user data
- Resolved critical price-fetching issue on launch day, ensuring a seamless and timely product release

PROJECTS

Hypercubic Peer-to-Peer Network | *C* — GitHub

- Authored hypercubic P2P chat network optimized for high scalability and tested final models on mock network of 256 nodes
- Designed topology maintenance mechanisms and messaging protocol to achieve logarithmic scaling on logical jumps
- Profiled network performance using Python, producing insights on performance of the topology in a final report

Music Anti-Recommender | *Python,TypeScript* — GitHub

- Developed robust backend for a React web-app that uses cosine similarity to deliver song "anti-recommendations"
- Built a scalable pipeline with Python futures to efficiently mine and validate dataset of 10,000 songs, ensuring data quality
- Enabled real-time lookups via Spotify's API, incorporating debouncing and rate limiter to maintain API compliance and boost responsiveness

Game of Life Optimization | *C* — GitHub

- Refactored Game of Life simulation with SIMD and threading, reducing runtime significantly by up to 120 times
- Increased efficiency with SSE and AVX vector instructions and bitmasks, achieving a 25x reduction in runtime
- Utilized gprof tracing information to improve cache and stack behavior with function inlining, attaining 40% speed up

PRNG Behavior Analysis | *C++, R*

- Examined long-term behavior of mt19937 number generator in C++ by processing 9 TB of data using NIST test suite
- Performed analysis over 360,000 blocks of 25 MB generated data, compared collected p-values against theoretical distribution of bits ratio and oscillation frequency
- Visualized random behavior through graphics, presented findings on reliability of mt19937 as a non-cryptographic PRNG

US Econometrics and Immigration Data Analysis | *R*

- Utilized fixed effects models and Principal Component Analysis (PCA) to analyze 20 econometric variables related to social security as potential predictors for immigration rates across all US states
- Performed comprehensive exploratory data analysis (EDA) to clean and preprocess the dataset, successfully reducing the number of variables by 50%, enhancing model efficiency and interpretability

TECHNICAL SKILLS

Languages: C, C++, Python, JavaScript, R, Rust
Technologies: CUDA, MPI, jQuery, D3.js, TensorFlow, MongoDB
Concepts: Parallel Programming, Performance Optimization, Cache Memory, Vector Instructions, Compilers, UI/UX, Data Processing and Visualization, Statistical Analysis