



Sumer Kohli

 /in/sumerkohli |  sumer.kohli@berkeley.edu |  @firebolt55439

 Saratoga, CA 95070 |  (408) 621-6422

EDUCATION

University of California, Berkeley

B.S., Electrical Engineering & Computer Sciences

GPA: 4.00

Aug 2018 – May 2022

- **Distinctions:** 2021 Outstanding GSI Award; CalHacks Prize Winner (4x); Dean's List (3x); HackMIT Prize Winner (2x)
- **Organizations:** Cal Launchpad (AI/ML); DevOps @ Berkeley (VP of Tech); Data Science Society of Berkeley
- **Relevant Coursework:** CS170 (*Algorithms*); CS162 (*OS & Systems*); CS189 (*Machine Learning*); CS161 (*Cybersecurity*); CS186 (*Databases*); CS188 (*Artificial Intelligence*); CS61C (*Computer Architecture*); CS61B (*Data Structures & Algorithms*); EECS127 (*Optimization Models*); EECS126 (*Probability*); EECS16A/B (*Electronic Systems*); Math 53 (*Multivariate Calculus*)


RESEARCH

Berkeley RISELab

Undergraduate Research Assistant – Advisor: Prof. Randy Katz

Berkeley, CA

Aug 2020 – present

- Researching runtime optimizations for distributed serverless applications with the aim of running optimized code from the very start ("hot-start"), thereby greatly reducing latency and cost of serverless workloads.
- Designed and iterated upon software to execute research strategies, evaluate findings, and visualize results at a rapid pace.
- Interpreted and presented results every weekly meeting to advance research goals and improve collective understanding.
- Published a paper with initial findings in HotOS '21: From Warm to Hot Starts: Leveraging Runtimes for the Serverless Era . Joao Carreira, **Sumer Kohli**, Rodrigo Bruno, and Pedro Fonseca. 2021. From Warm to Hot Starts: Leveraging Runtimes for the Serverless Era. In *Workshop on Hot Topics in Operating Systems (HotOS 21)*. <https://doi.org/10.1145/3458336.3465305>
- Working towards a submission to OSDI '21 to present findings on the potential of intelligent checkpoint scheduling to provide secular improvements for serverless applications employing a checkpoint-restore mechanism.

Group Matching Research

Undergraduate Teaching Assistant, later Research Assistant – Advisor: Prof. Gireeja Ranade

Berkeley, CA

Jun 2020 – present

- Led development of group matching software that has been used during COVID semesters by classes totalling over 5,000 students. Has been used in CS 61A, 61B, 88, 169, 189; EECS 16A, 16B.
- We are analyzing the efficacy of the software, explaining its algorithmic underpinnings; and incorporating deep RL to improve group match quality over time.
- We are planning to publish our findings to a CS education-focused conference early next year; I will share first authorship.
- Won Berkeley's 2021 Outstanding GSI Award along with my collaborator for spearheading the development of this project.

Berkeley RISELab

Undergraduate Research Assistant – Advisor: Prof. Alvin Cheung

Berkeley, CA

Oct 2021 – present

- Focused on leveraging verified lifting, a breakthrough in program analysis, to improve and generalize program synthesis.
- Helping build a framework to leverage these techniques for a variety of domain-specific languages, offering low-to-zero-cost language abstractions by synthesizing performant and formally correct equivalent programs.

WORK EXPERIENCE

Citadel LLC

Software Engineer Intern

New York, NY

Jun - Aug 2021

- Engineered a Kafka trade pipeline in Java for regulatory reporting that parses, transforms, and transports up to 6B trades/day.
- Built a Java library and accompanying write-behind cache to replay misprocessed Kafka messages, critical for error handling.
- Rigorously tested pipelines and replay library due to zero industry error tolerance for missing trades, and deployed to production.

University of California, Berkeley

Teaching Assistant for EECS 16A (Fall '19, '20), 16B (Spring '20), and CS 61B (Spring, Fall '21)

Berkeley, CA

Aug 2019 – present

- Led development of group matching software that has been used during COVID semesters by classes totalling over 5,000 students. I am co-authoring a research paper on its efficacy to publish to a conference. Won the 2021 Outstanding GSI Award.
- Jointly led the core infrastructure team to support 1,000+ students and 40+ staff, designing and writing software as needed.
- Taught discussion sections, labs, and office hours, and was rated markedly above (4.81/5) the course staff average (4.64/5).

Microsoft Inc.*Software Engineer Intern*

Sunnyvale, CA

Jun - Aug 2020

- Designed, developed, and deployed a new customer-facing Azure Communications service using C#/ASP.NET with my team, and a fully-featured UI using React/TypeScript (further details under NDA). Won the 2020 Garage Team Hero award.
- Implemented a C# backend for automatic ML-based captioning for the Windows Photo app with 300M+ yearly users.

Lawrence Livermore National Laboratory*Computational Scholar Intern*

Livermore, CA

Jun - Aug 2019

- Researched and developed a Python-based key-escrow server on AWS and Docker to enable Full Disk Encryption (FDE) on the Lab's 3,500+ Macs, greatly improving operational security in response to escalating state-sponsored cyberattacks.
- Programmed a client-side service in Swift to enforce FileVault enablement on the 3,500+ employees, ensuring compatibility with YubiKey-based multi-factor authentication (MFA) while enabling instant roll-out of critical settings updates.
- Integrated and documented a REST-based API to enable authenticated access to user, machine, and recovery key data.

Nutanix Inc.*Software Engineer Intern*

San Jose, CA

Jun - Aug 2015, Jun - Aug 2017

- Developed a performant Python-based backend to process and store over 1 million product telemetry data points a day.
- Built a fully-featured web interface to efficiently tabulate and visualize gigabytes of product telemetry in near real-time.
- Implemented reliable logging of core processes in C++, preventing potential catastrophic data loss during cluster imaging.

PROJECTS**DreamRL**

Jan 2019

A CNN-VAE → MDN-RNN → CMA-ES TensorFlow model co-developed with my Cal Launchpad team; achieved state-of-the-art performance on the CarRacing-v0 OpenAI gym.

 $\frac{d}{dx}$ it!

Oct 2018

A web application that instantaneously takes partial derivatives of arbitrary mathematical functions; supports extremely complex input, and computes everything client-side using Scheme and JavaScript (far faster than WolframAlpha!).

Cobra Compiler

Sep 2017

An LLVM-based optimizing compiler for my very own object-oriented, Turing-complete programming language; code generation is fully C/C++ ABI-compatible and supports multithreading and custom operators.

Sumer's OS

Dec 2016

A Unix-like mini-operating system with scheduling and multitasking support written in C; able to run code in user land and provides a (mostly) POSIX-compliant environment as well as a basic libc runtime.

SCE Chess

May 2016

A blazingly-fast C++ chess engine rated around 2045 ELO, or able to beat over 98% of chess players; employs alpha-beta pruning and a bitboard representation to evaluate nearly 2 million positions per second.

SKILLS

Languages	C/C++, Python, Java, JavaScript/Node.js, TypeScript, Go, Swift, Objective-C, C#, Shell, Rust, Wolfram, R
Technologies	AWS, Docker, Heroku; MongoDB, MySQL, PostgreSQL; React, AngularJS, Vue.js
AI/ML	TensorFlow, PyTorch, Keras; SVM, Random Forests, LASSO; CNN, LSTM, GAN