☑ allencao@berkeley.edu **(**510) 935-3270

https://www.linkedin.com/in/allen-l-cao

https://github.com/AllenCaoo

EDUCATION

University of California, Berkeley

Shttps://allencaoo.github.io

B.A. Computer Science, GPA: 3.835/4.0

o Coursework: Data Structures, Algorithms, Computer Architecture, Database Systems, Operating Systems, Compilers, Computer Security, Internet Architecture, Artificial Intelligence, Machine Learning, Data Science, DeFi, Circuits, Linear Algebra, Discrete Math, Probability • Teaching Assistant (TA): CS188 (Intro to AI), CS61B (Data Structures)

llen Cao

SKILLS

Programming Languages: Python, Java, C/C++, C#, TypeScript/JavaScript, HTML/CSS, Rust, SQL, Scala, Bash, Go, OCaml

Frameworks: Apache Spark, LangChain, Node, React, Next, Flask, GraphQL, MySQL, PostgreSQL, PyTorch, TensorFlow, OpenCV Tools: AWS (Lambda, CloudFormation, S3, DynamoDB, Athena/Glue, API Gateway, RDS), Git, Sentry, MongoDB, Redis, Docker, Pinecone

WORK EXPERIENCE

Legalyze.ai

Software Engineer

- o Fine-tuning LLMs to create lawyer assistant for question-answering of PDFs, external source search and citing, and streamlined case filing.
- o Developing API using FastAPI, backend AI with LangChain, monitoring with Sentry, payment with Stripe, frontend with Next.js, and integrations with case management systems.
- Accelerating semantic search (\sim 10 ms) during question-answering by efficiently storing OpenAI text embeddings in Pinecone vector database.

Amazon Fashion

Software Development Engineer Intern

- o Reduced week-long data management tasks to mere seconds by building an efficient database catalog for seamless data search, modification, and migration.
- o Achieved a 97% decrease in data retrieval latency by designing a downstream layer of database partitions with Apache Spark, resulting in predictable latencies for 100k daily API requests.
- o Developed a scalable REST API using API Gateway, integrated with a React frontend and Lambda, Athena/Glue backend to serve thousands of Amazon Fashion content managers and data engineers daily.
- Constructed deployable cloud infrastructure using AWS CloudFormation/CDK and integrated across 3 development stages.

Baxter Healthcare

Software Development Engineer Intern

- o Fully obtained FDA compliance by incorporating an interrupt service routine into medical device systems to monitor device damage, collect data, and generate logs to an end-user database.
- o Upgraded internal API by developing a logging queue and integrating throughout codebase to resolve thread-safety concerns and optimize multi-threading.
- o Constructed a streamlined Python machine learning workflow that includes an ETL pipeline feeding into a damage detection classifier, resulting in a model with 99% accuracy on test data.

UC Berkeley EECS, Barsky Lab

Machine Learning/Computer Vision Researcher

- o Developing autonomous vision system for fall detection in crowded (occluded) multi-person scenes using Spatio-Temporal Graph Convolutional Networks (ST-GCNs).
- Using CUDA PyTorch to develop parallel video processing algorithms such as object tracking and optical flow, increasing frame rate by 5 fold.

PROJECTS

Pathfind Visualizer

- https://allencaoo.github.io/Pathfind-Visualizer
- o Built an interactive visualization tool for pathfind and maze generation algorithms on a 2D grid.
- As course staff, released the project to over 500 students in UC Berkeley's data structures (CS61BL) course as a visual learning tool.
- o Algorithms include BFS, DFS, Dijkstra's, A*, Greedy Best First Search, Randomized Prim's, and Inverted Randomized Prim's.

tAl

- Built an Al-powered educational tool delivering personalized assistance to all educational levels and subjects.
- Features include feedback-based guizzes, conversational office hours, and grading interface.
- Utilized MongoDB to manage content and student feedback, Express is and AI text generation APIs for backend, and React for UIs.

Pintos OS

- Programmed a uniprocessor operating system to handle and manage interrupts, system calls, memory allocation, thread scheduling, extensible file system (Berkeley FFS), caching, and virtual memory.
- Implemented comprehensive thread and process synchronization using locks, semaphores, and monitors.

Gitlet

- Designed and implemented a version control system for local and remote repositories.
- o Built an organized hash-based storage system for commits, branches, and remote repositories.
- o Supports init, add, commit, rm, log, global-log, status, checkout, branch, rm-branch, reset, merge, add-remote, rm-remote, fetch, push, pull.

Seattle, WA May 2023 - August 2023

Remote

November 2023 - Present

Berkelev, CA

Deerfield, IL

June 2022 - August 2022

August 2022 - Present

MongoDB, Node, Express, React, Redux

React

Java

C, x86 Assembly

Berkeley, CA

August 2020 - May 2024