

Education

University of California, San Diego | San Diego, CA

SEP 2024 - JUN 2026

Intended M.S. in Computer Science

University of California, Berkeley | Berkeley, CA

AUG 2020 - MAY 2024

B.A. in Computer Science & B.A. in Applied Math GPA: 3.90

Relevant Coursework: Artificial Intelligence, Machine Learning, Computer Security, Special Topics in Deep Learning, Game Design & Development, Data Structures, Principles of Data Science, Complex Analysis, Machine Structures, Efficient Algorithms, Real Analysis, Abstract Algebra, Numerical Analysis, Compiler Design, Operating Systems, Probability and Random Processes

Professional Experience

Machine Learning Engineer Intern | Accretional

JUN 2024 -

- Design and implement **RAG vector** embedding model to ground LLM with respect to cloud infrastructure
- Create command-line tool in Golang for indexing and vectorization of files, **semantic & lexical search** of keywords
- Fully construct VSCode application in Svelte and TypeScript to help users easily implement and deploy APIs in less than 5 seconds

Research Intern | Hao AI Lab @ UCSD

APR 2024 -

- Develop cutting-edge LLM monitoring system, utilizing **SAEs** and new sequence-level early exit strategies
- Configure new CLLM models to be compatible with **qLoRA fine-tuning** on Jacobi trajectories, decreasing memory usage over 3x

Software Engineer Intern | Hinge Health

MAY 2023 - AUG 2023

- Overhauled and integrated a tedious (100+ hour) manual user data reset workflow process
- Designed 5+ endpoints and event consumers in TypeScript and Ruby on Rails, restructuring and building new GraphQL queries
- Discussed and presented solutions for over 10 UI/UX features and bugs in React with PMs & engineers
- Administered and piloted 5+ standup & sprint planning meetings using Jira, monitoring bug intake
- Touched 8 different repos with Docker, linking many of them with RabbitMQ event emitting, deployed in Kubernetes

Full Stack Intern | Cisco Meraki

MAY 2022 - AUG 2022

- Spearheaded a full stack project to create uplink ping stat graphs, helping **500,000 customers** visualize wireless trends over time
- Collected ping stats data from over 4 million wireless internet nodes, using Scala scripts to build a grabber and Jenkins to test
- Conducted experiments with global LT & SQL queries across 600 shards to mitigate the size of rapidly-growing tables
- Introduced 3 multi-functional endpoints in Ruby on Rails using backend data collectors

Software Engineer Intern | Breakout Mentors

JUN 2021 - AUG 2021

- Amplified Python scripts, reducing number of calls to pSQL databases by 50%, improving efficiency of mapping algorithms by 10x
- Revised React code to minimize frontend API calls and used caching to store data on 100% of administrative webpages

Projects & Research

MusicCLLM, an adaptation of Consistency LLMs for music generation

JUN 2024 -

Convolutional Neural Network implementation in Python

MAR 2024 - APR 2024

- Engineered a basic CNN, including construction of layers, activations, and NN models, using NumPy, PyTorch, & TensorFlow

Insite, a React, NestJS & TypeScript app that connects high schoolers to professionals

JUN 2023 - SEP 2023

- Utilized AWS S3 buckets for file storage, assembled 40+ REST API endpoints for tasks like search, connections, and profile pages
- Built custom Docker containers and PSQL databases, integrating them with CRA frontend & backend repos with networks

Regulus, a Golang secure file sharing system

MAR 2023 - MAY 2023

- Employed over 5 cryptographic tools and libraries like HashKDF, PBKDF, UUID, HMACs for authentication & access revocation

NumC, a C++ version of numpy, optimized 40x using loop unrolling, SimD, OpenMP, caches

OCT 2021 - DEC 2021

Gitlet, a Java-based git-like version control system

FEB 2021 - MAR 2021

Skills

Python, SQL, JavaScript, TypeScript, React, PyTorch, TensorFlow, Java, Golang, Kubernetes, AWS, Ruby on Rails, Scala, Docker, C, C#, C++, RabbitMQ, GraphQL, HTML, CSS, Figma, Jenkins, PHP, Unity, Git / GitLab / GitHub, Unix, Jira, JupyterLab