

Varun Venna

973-517-0645 | vvenna@g.ucla.edu | linkedin.com/in/varun-venna | varunv22.github.io/

EDUCATION

University of California, Los Angeles

Los Angeles, CA

Bachelor of Science in Computer Science

Sep. 2022 – June 2026

- Relevant Coursework: Data Structures and Algorithms, Software Construction, Operating Systems, Computer Systems Security, Machine Learning, Computer Systems Architecture, Recommender Systems

EXPERIENCE

Software Engineering Intern

June 2025 – Sep 2025

Amazon

Seattle, WA

- Built a full-stack module to surface accounting metadata and USL lines in internal finance tools, enabling Amazon teams to view and track their accounting events
- Used AWS Athena to paginate 1M+ USL records, reducing load times by 60%
- Cut DynamoDB query latency by 40% via custom GSI and elastic search for high-volume metadata lookups
- Fixed concurrency bugs from batch parallelism, restoring consistent event status propagation

Software Engineer/Co-Founder

June 2024 – October 2024

Swyft (Acquired by Duffl)

Los Angeles, CA

- Developed a mobile app for UCLA students to order groceries using Expo, React Native, and WooCommerce API
- Integrated user accounts with WooCommerce for seamless order tracking and real-time inventory updates
- Implemented payment gateways and optimized API calls with local caching for faster load times
- Implemented push notifications to alert users about order status updates, ensuring timely communication throughout the delivery process

Software Engineering Intern

March 2024 – July 2024

Brev.dev (Acquired by NVIDIA)

San Francisco, CA

- Tested and fixed over 30 Jupyter Notebooks for LLMs and multi-modal models including Mistral 7B, NVIDIA's NeMo Framework, TensorRT-LLM, LLaVa, and the StreamingLLM framework
- Upgraded the CLI by doing a large overhaul and strengthening/removing commands as necessary
- Integrated a cloud service named Crusoe into the console through multiple functions that allow users to create, delete, stop, start, and retrieve information about an instance
- Created a Golang script capable of creating and terminating Windows/Linux instances via Microsoft Azure

Software Engineering Intern

Oct. 2023 – Feb. 2024

Avolta

Toronto, ON

- Established and maintained routing framework with Flask and SQL for a mobile vehicle monitoring app while enhancing security with a bcrypt hashing function and parametrized queries
- Collaborated in a team to develop functionality and security for password change/reset, data accesses, login, register user, etc.

PROJECTS

Affina | *Express, HTML/CSS, React Native, Expo, MongoDB, Node, Postman*

- Built a comprehensive mental health support mobile app with AI-powered therapy chat and mood tracking
- Integrated Anthropic AI for personalized therapy conversations with crisis detection, implementing message chunking for natural conversation flow
- Developed mood tracking calendar with audio diary entries, allowing users to record voice notes and track emotional patterns over time with visual mood indicators
- Published app to iOS App Store with 150+ downloads, implementing secure authentication, profanity filtering, and age verification

TECHNICAL SKILLS

Languages: Java, Python, C/C++, SQL, JavaScript, HTML/CSS, GoLang, Lisp

Technologies: React, Node.js, Flask, Azure, AWS, Crusoe, VastAI, Postman, Swagger, Docker, MongoDB, Redux, Shell, Git, OpenCV, Triton, Expo, CUDA

Libraries: Pandas, NumPy, Matplotlib, TensorFlow, TensorRT, Ollama, OpenAI, LLaVa, Seaborn