

Kai Breese

925-786-5550 | kai.breeze@me.com | [linkedin.com/in/kailbreese](https://www.linkedin.com/in/kailbreese) | github.com/ninjakaib

EDUCATION

University of California San Diego

Bachelor of Science in Data Science

La Jolla, CA

Sep. 2021 – March 2025

EXPERIENCE

Machine Learning Engineer Intern

June 2024 – Present

XiFin, Inc.

San Diego, CA

- Created IDE coding assistant using open-source LLMs and AWS SageMaker, implementing context-aware chat and code suggestions, inline editing, and real-time autocomplete while ensuring data privacy
- Created automated insurance appeal letter system using knowledge graphs and multimodal LLMs, reducing manual processing time for denial management team
- Improved insurance pricing data pipeline, achieving 90x throughput with multiprocessing and custom C backends
- Redesigned pricing database to reduce data duplication by 75% with normalization and optimized data parsing
- Built no-code deployment tool for the ML team to easily deploy models to AWS utilizing custom Docker containers
- Improved chatbot with streaming vision capabilities to reduce incorrect answer rate from 50% to 15%
- Created new RAG pipeline facilitating agentic retrieval and visual understanding of documents
- Designed AI SQL generator with context-guided outputs enabling data teams to write complex, accurate queries

Lead Developer

Mar. 2024 – July 2024

Clayton Health, LLC

San Diego, CA

- Built custom e-commerce platform for Shopify using Liquid driving \$100K in sales within first 6 months
- Implemented automated email marketing campaigns leading to 8% conversion rate
- Worked directly with CEO to rapidly bring new features from concept to production

Sales Consultant

Oct. 2023 – Apr. 2024

Verizon

San Diego, CA

- Ranked top 5 in San Diego district generating an average of over \$100/hr in company profits
- Created long term relationships with high-value business clients to drive repeat sales

Computer Science Tutor

Mar. 2023 – Feb. 2024

TheCoderSchool

San Diego, CA

- Led one-on-one and classroom based sessions, teaching students in ways they can easily understand and relate to
- Helped students create over 100 unique projects including games, websites, and assignments

PROJECTS

Machine Learning Hardware Accelerator Chip

Sept. 2024 – Present

- Leading 4-person capstone team developing a hardware accelerator achieving 80-95% energy reduction in floating-point operations while maintaining 99% accuracy in ML applications
- Created new bfloat16 multiplier units achieving 50% transistor count reduction compared to float8 designs
- Designed systolic array architecture with pipelining for parallelized matrix multiplication with high throughput
- Designed specialized SwiGLU hardware activation units optimized for transformer models
- Rapidly learned and applied computer architecture and VLSI design concepts to create energy-efficient ASIC

AI Healthcare Assistant | *Spigot API, Java, Maven, TravisCI, Git*

June 2023 – Oct. 2023

- Developed a conversational AI phone call service before similar products from OpenAI came to market, achieving 80% API cost reduction over comparable solutions
- Designed SQL relational database with nutritional information scraped from the web
- Fine-tuned LLMs and deployed with AWS for real-time streaming and high-user load capacity

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript, C/C++, Excel

Frameworks: FastAPI, Flask, Gradio, Svelte, Nginx, React, Liquid

Developer Tools: Git, Docker, Anaconda, Yosys, Jupyter, Jira, VS Code, ChatGPT

Libraries: pandas, NumPy, PyTorch, matplotlib, sklearn, transformers, langchain, Cython, PyRTL

Cloud: AWS Sagemaker, Lambda, S3, RDS, ECR, EC2, API Gateway, Oracle DB