Owen Hochwald

US Citizen Local to Bay Area ✉ owenhochwald@gmail.com 💼 owenhochwald.me 🔗 /in/oHoch ⭘ /owenHochwald

Education

University of British Columbia   Vancouver, Canada
BS Computer Science **GPA 4.33/4.33**, Dean's Scholar.   **Expected Graduation: May 2027**

Experience

Stealth Startup   Remote
Backend & AI Engineering Intern   June 2025 - Present
- Built and optimized a **high-performance caching layer** using a custom hash table-based database in **Kotlin + Spring Boot**, enabling exceptionally fast read/write access for **real-time applications**.
- Led the design of a **distributed processing system** to chunk, route, and orchestrate tasks to LLM inference services, improving system throughput and scalability.
- Developed a **RAG-based bias classification pipeline** leveraging **Pinecone** for semantic search and vector similarity, achieving over 90% classification accuracy on unstructured content.

UBC Visual Cognition Lab   Vancouver, Canada
Coding Team Full Stack Intern   April 2025 - Present
- Spearheaded integration of a **RAG pipeline chatbot** and **custom object detection model** into a research platform.
- Delivered full-stack web features using **React.js, Redux, MongoDB, AWS S3, and Express.js**.
- Authored **developer onboarding guides** and **ran internal tech talks**, reducing onboarding time by **70%.**

Insomniac Design, Inc   Washington, D.C.
Backend Machine Learning Engineer Intern   *June 2023 – Aug 2023*
- Built a **FastAPI-based microservice** around a **Lbl2Vec NLP model**, automating sentiment classification with a **>90% accuracy** benchmark.
- Established internal **CI/CD workflows** and standardized **GitFlow** branching strategy.

Technical Projects

AdmitifyAI: College Admissions Prediction & Matchmaking   NextJS, PostgreSQL, AI Integration
- Launched full-stack SaaS product with **200+ active users**; integrated **LLM-driven recommendation** engines for personalized college application targeting.
- Led product design, devops, and growth strategy while managing a **5-person agile team**.

RecMind: Real-Time Article Recommender   Go, Python, RabbitMQ, LangChain, Redis, AWS
- Architected a **distributed, event-driven system** using **RabbitMQ** and **Go** to process high-throughput interaction data in real-time.
- Embedded **semantic vector search** via **LangChain + Pinecone**, generating intelligent recommendations from unstructured data.
- Deployed and monitored infrastructure using **Docker, EC2, Prometheus, and Grafana** for end-to-end observability

Shakespeare GPT: Generative Language Model   Python, PyTorch
- Developed a **GPT-style transformer model** with character-level tokenization trained on Shakespeare's corpus, achieving **4x loss reduction** over bi-gram baseline
- Demonstrated understanding of **sequence modeling, embedding layers, and loss function tuning**.

Skills & Interests

Programming Languages: Python, TypeScript, JavaScript, Kotlin, SQL, C++, HTML/CSS
Frameworks & Libraries: React, Redux, FastAPI, PyTorch, LangChain, Next.js
DevOps & Cloud: Docker, AWS (EC2, S3), DigitalOcean, Vercel, Git, CI/CD, Prometheus, Grafana
Infra & Databases: PostgreSQL, MongoDB, Supabase, Redis, Pinecone, Kafka
Concepts: REST APIs, microservices, LLM integration, semantic search, RAG, real-time systems, cloud infrastructure, backend architecture, agile development