Daniel Chen

San Diego, CA | (858) 396-4613 | dac030@ucsd.edu | github.com/dan9574 | danielchen.app

EDUCATION

University of California San Diego **Expected Spring 2028**
B.S. in Mathematics–Computer Science | GPA: 3.915/4.000 *San Diego, CA*

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript/JavaScript, Rust, Swift, Java, C/C++
Machine Learning: PyTorch, Transformers, PEFT, TRL, QLoRA, bitsandbytes, vLLM, scikit-learn, Pandas
RAG & Backend: FastAPI, SQLAlchemy/Alembic, PostgreSQL, pgvector, pytest, Supabase, Spring Boot
Tools: Git, GitHub Actions, Linux, React, Tauri 2, SwiftUI

EXPERIENCE

Altheros Capital — Cara AI **Sep 2026 – Present**
Backend Lead, Part-time (Software Engineer, Sep–Oct 2026) *Remote*

- Promoted to Backend Lead for the AI/RAG backend of an AI companion app (FastAPI, PostgreSQL, pgvector) within three weeks; assign backend tasks and review pull requests.
- Designed and shipped the RAG ingestion pipeline that syncs journals, profiles, and evicted chat history into a pgvector store: deterministic chunk serialization, SHA-256 content dedup for idempotent retries, and a pending-delete queue with idempotent sweeps for eventual consistency.
- Enabled source attribution by returning doc_id from the retrieval service and resolving each chunk to its user-scoped journal or conversation; kept evicted chat turns retrievable via fixed-window batching.
- Tiered retrieval timeouts by responsive/thoughtful mode from benchmarks showing cross-encoder rerank latency grows linearly with top_k, verified by mutation testing; 6 merged PRs across 2 repos with ~70 tests.

Xiamen Limayao Network Technology Co., Ltd. **Jun 2026 – Sep 2026**
AI/ML R&D Intern *Beijing, China*

- Processed 6.17M raw product records into 66,429 training/validation samples via category-aware cleaning, deduplication, Max-Min diversity sampling, and hard negatives; a two-stage Doubao/DeepSeek audit corrected 4,173 mislabeled records.
- Fine-tuned Qwen3-4B with QLoRA SFT, raising core 1:N accuracy from 58.78% to 88.49% and candidate-set compliance from 78.59% to 99.98%.
- Deployed Qwen3.5-2B as the production model with 2.33x vLLM throughput (45.89 vs. 19.73 samples/s) while staying within 0.17 pp of Qwen3-4B accuracy.
- Designed candidate-order augmentation and paired evaluations, cutting shuffled-order degradation from 3.03 to 1.25 pp; error analysis prioritized data expansion over additional epochs.

Beijing Mapvision Technology Co., Ltd. **Jun 2025 – Aug 2025**
Software Engineering Intern *Beijing, China*

- Normalized ~440K brand/store records into ~400K canonical brands (tokenization, rule-based matching, LLM-assisted checks); built experimental multilingual BERT sentiment and NER pipelines for Xiaohongshu content.
- Built an internal full-stack platform with Spring Boot, JPA, PostgreSQL, and React, including REST APIs, cookie-session authentication, and user-level authorization.

PROJECTS

Hope — Cross-Platform Time Tracker **2025 – Present**
Open-Source Project | Tauri 2, Rust, React, TypeScript, Swift, SQLite, Supabase | GitHub | Demo

- Rebuilt an Electron app as a Tauri 2 desktop app (macOS/Windows) plus native iPhone and Apple Watch apps in Swift (~23K lines, 150 unit tests), with Rust as the single data layer behind typed commands.
- Designed an offline-first sync protocol on Supabase (UUID keys, last-writer-wins soft deletes, server-sequence cursors, Postgres conflict triggers, row-level security) and a WidgetKit watch-face complication.