

Gabryel Soh

Available for Penultimate Internships from Jan 2026 - Aug 2026

+1 341 258 2522 | gabryelsoh@berkeley.edu | linkedin.com/in/gabryel-soh | github.com/gable-github

EDUCATION

University of California, Berkeley

M.Sc. Technological Entrepreneurship | Immersive Program in Entrepreneurship

Berkeley, California

Expected Apr 2027

Singapore University of Technology and Design

B.Eng. Computer Science (Honors with Distinction)

Singapore

Expected Apr 2027

- **STEP Scholar** | Awarded to top 5% of cohort for academic excellence
- Relevant Coursework: Software Engineering, Machine Learning, Data Structures and Algorithms, Databases and Distributed Systems, Operating Systems, Machine Learning Operations

EXPERIENCE

MetaBob (*AI codebase analysis company at NEC X*)

San Francisco, California

Software Engineering Intern

May 2025 – Aug 2025

- **Built an end-to-end rule enforcement pipeline** that converts natural language contribution guides into structured rulesets using LLMs and NLP models, **reducing manual rule creation time by >80%**.
- Integrated rule enforcement into client CI workflows with **Python, regex, AST matching**, and **GNNs via Celery tasks**, **reducing code review cycles by >70%** across client developer teams

WebAI (*Series A AI Orchestration Startup, \$60M+ raised, \$700M valuation*)

Austin, Texas

Software Engineering Intern

Feb 2025 – May 2025

- Led a team at **Open Project @ UC Berkeley** to build an **AI-powered video summarization pipeline** using **OpenCV, OpenAI Whisper, and HuggingFace Transformers**, serving **1,000+ end users** in law, finance, and professional sectors.
- **Reduced content review time by 10×**, converting **5-hour keynotes into 5-minute topic-based summaries** with keyframes extracted in **<10 minutes** using **EAST** and **pHash** for visual deduplication.

OrangeKloud Technology Inc. (*NASDAQ: ORKT*)

Singapore

Software Engineering Intern

Aug 2024 – Dec 2024

- Leveraged **MetaGPT, LangChain, PostgreSQL/PGVector** to develop a **Multi-Agent RAG App Generation Pipeline** capable of generating production frontend and backend code in **<10 minutes** with **full API and DB integrations**.
- Deployed the pipeline to **2,500+ enterprise customers** globally with **YAML, Docker and Ansible** across **Azure VM, Blob Storage and PostgreSQL** instances.
- **Reduced hallucinations by >95%** using **HITL-integrated RAG** which enabled dynamic UI referencing through a 100+ template knowledge base. **Increased LLM output token limits by 8×** using recursive algorithms and **drove code generation time down to <4 minutes**.
- **Designed PostgreSQL schema and efficient query patterns** to handle **10K+ records with low latency**; implemented **FastAPI** endpoints and **WebSocket** channels for real-time pipeline control, **reducing user wait times by 70%**

Korya Co. Pte Ltd

Singapore

Software Engineer and Co-Founder

Sept 2021 – Sept 2022

- **Co-founded Korya Co.** and led the development of its **Shopify-based eCommerce platform (Javascript, HTML, CSS)**, driving a **3.2×** customer growth and **4-figure monthly revenue** increase by **onboarding 11+ distributors**.

PROJECTS

Planned Parenthood: Serverless NLP for Patient Feedback Analysis

Boston, Massachusetts

- Led a team at **Code for Good @ UC Berkeley** to build an **ML sentiment analysis tool**, **reducing manual review by 90%** and scaling patient feedback analysis **across all Planned Parenthood clinics in Massachusetts**.
- Utilized **Python, HuggingFace Transformers and Trainers**; integrated the pipeline into **AWS Lamda and Power Automate** for scalable usage across Planned Parenthood.

Ascenda: Full-Stack Credit Card Points Exchange Application

Singapore

- **Led and completed a MVP development project for Ascenda**, applying the Unified Process to deliver a functional full-stack prototype through a complete SDLC.
- Employed **Node.js, Express.js, React, MongoDB** stack with **Socket.io** for **webhook event-driven communication** leveraging **MongoDB's change streams**.
- Incorporated **ChakraUI with React** for the frontend interface; Used **custom hooks for data handling and frontend logic**.

WayFare: Full-Stack P2P Tour Hosting and Discovery Android App

Singapore

- Utilized **Java and SpringBoot** to develop an app that allowed **>1000 users** to book or host tours with short-form content.
- Used **JWT Tokens, Spring Boot Security, Spring Web, Spring Data MongoDB** to implement an account-based system and secure user data
- Employed **Material Design 3 components** to create a seamless UI/UX experience for users.

TECHNICAL SKILLS

Programming Languages: Java, Python, JavaScript / Node.js

Backend Frameworks and Databases: Springboot, Django, FastAPI, Node.js/Express.js, Maven, React, Django, PostgreSQL / PGVector, MongoDB, Redis, Celery, MetaGPT, Langchain, HTML, CSS

Tools: Docker, Azure Cloud Storage, AWS EC2, Git / Github, Jira

Licenses & Certifications: Meta Back-End Development, Meta Django Web Framework, Meta Databases for Back-End Development