

REN HWA (ZACK) LOW

Singapore Citizen | +65 8777 1534 | zack.low.dev@gmail.com
renhwa.com | github.com/sunraymoonbeam | linkedin.com/in/ren-hwa-low-855080224

Summary

Software engineer specializing in AI, with experience across multiple domains: agentic systems, computer vision, generative AI and speech processing. Strong programming fundamentals and full-stack experience. I work from first principles: I take the time to understand the problem deeply and how the ideal solution might fit into the bigger picture. I'm comfortable owning ambiguous problems end to end, from discovery with stakeholders to evaluation and deployment.

Work Experience

AI Engineer – Carro

August 2025 – Present

- Led development of Carro's ticketing agents for **60,000+** monthly tickets across 3 countries. Worked with operations teams to turn their processes into agentic workflows spanning the customer journey (triage, inventory creation, appointments).
- Developed an internal multi-agent orchestration platform that mirrors Carro's departments, automating their workflows such as data retrieval and reporting, governed by role-based access controls for multiple users and roles.
- Reduced LLM costs by **80%** through harness engineering and intelligent model routing.
- Designed an LLM-as-judge evaluation pipeline that validates agent changes before release, reducing SOP deviations by **30%**.
- Automated marketing outreach (customer engagement, follow-up) on n8n, driving **30%** revenue growth for SG workshops.

AI Engineer – AI Singapore

June 2024 – April 2025

- Delivered a GenAI MVP for a gaming industry project, generating 3D character animations from text.
- Built end-to-end ML pipelines in PyTorch to train and deploy diffusion models for text-to-motion generation.
- Auto-labeled training data with multimodal video LLMs, improving text-to-motion metrics by **20%** over baseline.
- Optimized GitLab CI/CD pipelines, reducing testing and deployment time by **80%**.

Research Assistant – National Institute of Education

October 2023 – March 2024

- Partnered with professors to build a speaker diarization system that measures student participation and learning outcomes.
- Fine-tuned NVIDIA NeMo ASR models on self-annotated transcripts, beating open-source baselines by **30%**.

Machine Learning Engineer Intern – Singapore Department of Statistics

May 2023 – August 2023

- Prototyped NLP tools for several departments (classification, Q&A, report generation) using LangChain and GPT-3.5.

Database Administrator Intern – Tesseract.io

July 2022 – February 2023

- Maintained production PostgreSQL databases with PL/pgSQL and Python, migrating **300,000+** records for 10+ law firms.

Education

B.E. Computer Science, Nanyang Technological University, Singapore

August 2020 – June 2024

Projects

Kanta – collaborative event camera with face-based photo sorting | [GitHub](#)

April 2025 – May 2025

- Guests upload photos to one shared album; face detection and unsupervised clustering group them by person.
- FastAPI backend with Azure Blob Storage and PostgreSQL on Azure Container Apps; clustering runs on AWS Lambda.

Echo – personal LLM digital clone | [GitHub](#)

April 2025 – May 2025

- Chat clone fine-tuned on 5 years of personal Telegram messages to reply in its owner's texting style.
- Fine-tuned LLaMA 3 and Gemma 3 with Unsloth; served through FastAPI microservices with data in AWS S3.

Technical Skills

Languages: Python, TypeScript, SQL

Backend & DevOps: Git, GitLab CI/CD, FastAPI, Django, Docker, n8n

AI & Agents: LangChain, LangGraph, Google ADK, Hermes Agent, LiteLLM, Langfuse, PyTorch

Cloud & Infrastructure: Terraform, GCP (Compute Engine, Cloud Run, BigQuery), Azure (Blob Storage, Container Apps)

Interests

Rock climbing, guitar, coffee brewing, cooking