

JOHN CARL SANTOS

Bulacan, Philippines (open to relocation: Singapore) · johncarlsantos30@gmail.com · github.com/JohnCarl-30
· linkedin.com/in/santosjohnncarl · dyeyc.dev

Education

Philippine Christian University

Expected 2027

BS Computer Science — Consistent Dean's Lister, GWA 1.15 (Top of Class)

Coursework: Data Structures & Algorithms, Software Engineering, Linear Algebra, Cloud Computing (AWS)

Experience

AI Engineer — SOFI AI Tech Solutions

Present

- Build client LLM document and email automation, internal ticketing tools, and training and eval datasets.
- Expanded automated coverage from **0 to 254 tests gated in CI**, including 23 browser tests and 24 admin-only routes.

AI Engineer Apprentice — Webpuppies

Sep 2026 – Present

- Cost analysis for a multilingual AI phone caller in Singapore: a self-built Pipecat + Twilio + Claude pipeline at **\$0.08–0.11/min vs \$0.14–0.18/min** for Vapi, Retell, ElevenLabs or OpenAI Realtime.
- Audited open-source speech and language models for licence and language coverage: about **half the popular “open” voice models bar commercial use**, and none commercially usable speaks Indonesian.
- Prototyped member-scoped agent tools with capped refunds and canary leak checks (**94 tests**).

Backend Engineer Intern — FlyRank AI

2 Months

- Built a RAG pipeline for document Q&A, JWT auth with roles, and a Docker Compose multi-service backend.

Software Engineering Intern — Alphaexplora

3 Months

REST APIs and performance debugging in an agile team.

Projects

Relaydesk — AI Support Agent for Small Businesses *Next.js, TypeScript, LangGraph, Zod, Langfuse, MCP*

- Right article first on **50/60** labeled questions; held back hybrid retrieval (hit@1 **30/36 to 34/36**) when precision fell, and chunking when recall fell **0.098**, past a 0.05 gate.
- Opens a human ticket when the help center can't answer: **zero unnecessary tickets** (precision 1.0) on 23 labeled questions, including a prompt-injection attempt and Taglish.
- Multi-tenant help centers, self-serve signup, reply co-pilot, MCP server, redacted Langfuse traces; **141 tests**.

autocards.app — AI Flashcards from PDFs (live)

FastAPI, Celery, pgvector, Next.js, Redis

- **20 real users**. Turns PDFs and images into flashcards with OpenAI, schedules reviews with SM-2, and answers questions over each user's documents through RAG chat.
- Async Celery ingestion on Railway, documents in Cloudflare R2; embedding p95 $\approx 2.2s$ at ~ 50 chunks/s.

Resumae — AI Resume Analyzer and Builder (live) *Next.js, NestJS, TypeScript, Vercel AI SDK, Postgres*

- Scores resumes against a job description and builds new ones from 4 templates; **600+ tests**, PostHog funnel.

directory-pipeline — Company Directory to Search Index *Python, Temporal, OpenSearch, PostgreSQL*

- Crawls a company directory into a deduplicated search index on crash-safe Temporal workflows; deterministic extraction recovers **232 of 234 fields with zero incorrect values**.

Technical Skills

Languages & frameworks: TypeScript, Python, SQL, Next.js, React, NestJS, Node.js, FastAPI, Zod

AI: Claude and OpenAI APIs, Vercel AI SDK, LangGraph, MCP, hybrid-search RAG, pgvector, Langfuse

Infrastructure & testing: PostgreSQL, Redis, Temporal, Docker, AWS, Vercel, GitHub Actions, Playwright

Certifications & Awards

Claude Certified Developer (Anthropic, 2026) · Oracle OCI GenAI Professional (2025) · 3rd Place, Java Hackathon