

John Carl Santos

Bulacan, Philippines · johncarsantos30@gmail.com · github.com/JohnCarl-30 · linkedin.com/in/john-carl-santos · 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)

TECHNICAL SKILLS

Languages: PHP, JavaScript, TypeScript, Python, Java, Swift

Frameworks: Laravel, ReactJS, FastAPI, Flask, Express, SwiftUI

AI / ML: RAG Pipelines, LLM APIs (Claude, Gemini), Vector DBs (Pinecone, pgvector), LangChain, LangGraph, TensorFlow, PyTorch, NumPy, Pandas

Databases & Tools: PostgreSQL, MySQL, Supabase, Firebase, SQLAlchemy, Git, Docker, REST APIs, Postman, GitHub Actions, Render, Vercel, DigitalOcean

EXPERIENCE

AI Engineer — SOFI AI Tech Solutions

Present

- Build and curate training datasets for LLM and machine learning models — sourcing, cleaning, deduplication, and formatting raw data into training-ready sets.
- Define annotation and labeling standards, then review outputs for consistency and quality before data enters training runs.
- Construct evaluation sets used to measure model output quality and catch regressions across iterations.

Backend Engineer Intern — FlyRank AI

2 Months

- Built a Retrieval-Augmented Generation pipeline with vector search; integrated LLM APIs for document analysis and Q&A.
- Implemented JWT authentication and role-based authorization.
- Containerized a multi-service backend using Docker Compose.

Software Engineering Intern — Alphaexplora

3 Months

- Developed and integrated RESTful APIs, contributing backend features and improving system functionality.
- Debugged and optimized application performance in a production-like environment within an agile team.

PyTorch Data Scientist — Campus Student Program

Ongoing

- Built and trained deep learning models in PyTorch (tensors, autograd, custom architectures); evaluated ML pipelines for classification and regression.

PROJECTS

StudyAI — RAG-Powered Document-to-Flashcard Platform

FastAPI, pgvector, Supabase, React, Redis, Celery

- Built an async RAG pipeline for document ingestion and semantic retrieval; benchmarked sub-2s embedding latency (p95 $\approx$ 2.17s) and $\sim$ 50 chunks/sec throughput.
- Delivered end-to-end processing in $\sim$ 14–71s, with OCR-heavy documents adding up to $\sim$ 60s extraction time.

Resumae — AI Resume Analyzer

Next.js, TypeScript, Express, Zod, Tailwind

- Built a full-stack resume analyzer with AI extraction, real-time scoring, and job-tailored suggestions using GPT-4o-mini and vector embeddings.
- Architected a pnpm monorepo (28,500+ lines) with MVVM separation and 201 test cases across 61 files; deployed via Vercel, Cloudflare Containers, and GitHub Actions CI/CD.

CiviReport — Barangay Complaint Management System

FastAPI, PostgreSQL, Redis, JWT, Android

- Built a complaint management system for barangay governance, letting residents file and track complaints via an Android app.
- Added real-time status notifications with Server-Sent Events and Redis-backed OTP auth over a RESTful FastAPI backend.

CERTIFICATIONS & AWARDS

Oracle OCI 2025 Generative AI Professional · Oracle OCI 2025 AI Foundations Associate (1Z0-1122-25) · MongoDB Building RAG Applications · AWS Machine Learning Foundations · AWS Introduction to Cloud 101 · Databricks Generative AI Fundamentals · 3rd Place, Java Problem Solving Hackathon