Tan Nguyen

tan.nguyen.atwork@gmail.com | (+84) 914-774-606 | Ho Chi Minh City, Vietnam
github.com/tangerine0000 | linkedin.com/in/tangerine0000

Education

VNUHCM, University of Science (HCMUS) 2020 - 2024
Bachelor of Science in Computer Science **GPA: 3.66 / 4.0** | Ho Chi Minh City, Vietnam

Technical Skills

Languages: TypeScript, JavaScript, Python, Rust, SQL, Go / Golang
Backend & Frameworks: NestJS, Node.js, FastAPI, React, Next.js, Express.js
Databases & Caching: PostgreSQL, DynamoDB, OpenSearch, Neo4j, Redis
Cloud & DevOps: AWS (Lambda, DynamoDB, EC2, S3), AWS CDK, Docker, Kubernetes, CI/CD, RabbitMQ, Git
Architecture: Microservices, Event-Driven Architecture, Serverless, State Machines, RESTful APIs, RBAC

Experience

Full-Stack Engineer - Innomize Oct. 2025 - Present

Full-Stack Engineer developing enterprise billing-operations platforms and core financial domain services.

- Owned the core Write-Off module end-to-end (React, NestJS, DynamoDB), replacing manual spreadsheet workflows with an automated processing engine supporting millions of invoices, appeals, and adjustments.
- Engineered backend state-machine services for multi-stage approval workflows, ensuring strict financial audit compliance via guarded state transitions and an automated eligibility matrix.
- Modeled DynamoDB schemas and transactional write patterns to support multi-tier adjustments and write-down calculations while guaranteeing strict ACID consistency across related records.
- Integrated OpenSearch to overcome NoSQL query bottlenecks, enabling sub-second multi-attribute filtering, full-text search, and cursor-based pagination across millions of financial records.
- Architected high-throughput reporting pipelines generating raw Excel extracts and template-driven PDFs with dynamic attachment merging and background worker execution.
- Built an event-driven collections email system supporting configurable audience filters, queued dispatch, scheduled retries, and dynamic attachment routing.
- Designed and published a shared UI component library integrated with fine-grained RBAC, adopted across 5+ enterprise modules and reducing frontend turnaround time by ˜30%.
- Accelerated engineering velocity and codebase modernization by establishing AI-assisted coding workflows with Claude Code and Antigravity.

Software Engineer Contractor - DK Lab Nov. 2024 - Mar. 2025

Backend Engineer responsible for high-performance API services and deployment pipelines.

- Developed high-throughput backend API services in **Rust** for the platform's v2 core release, reducing memory footprint and improving request latency.
- Built automated CI/CD deployment pipelines targeting containerized AWS EC2 environments, shortening release lead time and establishing reproducible build steps.

Engineering Projects

Serverless Document & Resume Parsing Microservice 2025 - 2026

Python 3.12, AWS Lambda, AWS CDK, FastAPI, Pydantic, S3, Docker, Pytest | Cloud Project

- Architected an event-driven serverless document parsing service on AWS Lambda (ARM64) fronted by API Gateway and AWS Cognito JWT authorizers, achieving zero LLM API costs with sub-second execution.
- Built deterministic parsing pipelines using FastAPI and Pydantic v2 for PDF/DOCX files, resolving complex layouts (two-column sidebars, tables, Mojibake repair) with 98.5%+ F1 accuracy.
- Automated multi-environment IaC with AWS CDK (TypeScript) featuring zero-downtime rollbacks; maintained an 861-test automated suite (pytest) and regression benchmarking across 1,300+ documents.

AI-Foundations - Distributed LLM Orchestration Framework 2026

Python, AsyncIO, Pydantic, OpenAI API, Anthropic API | Open-Source Project

- Architected an asynchronous orchestration layer decoupling multiple model providers behind an abstract client interface with automatic failover, retries, and load balancing.
- Enforced deterministic JSON schemas via Pydantic; implemented real-time token tracking and telemetry monitoring request latencies.
- Built concurrent A/B benchmarking suites utilizing Python `asyncio` to evaluate model throughput, TTFT (Time-to-First-Token), and system performance.

Publications & Research

The Explanation Method for Vietnamese Visual Question-Answering Model 2025

Tan Nguyen, Linh Tran-Thi-Phuong* (*Co-first Authors), Tung Le-Thanh, Huy Nguyen-Tien*
Journal on Information and Communications Technology: Research, Development and Application (MIC)
DOI: 10.32913/mic-ict-research-vn.v2025.n1.1338

- Published peer-reviewed research (Co-first Author) proposing an eXplainable AI (XAI) methodology for multimodal VQA models; implemented modular, maintainable PyTorch training and evaluation codebases.