NGUYEN THANH TAI

AI / Software Engineer



Hanoi, Vietnam · n.t.tai435@gmail.com · 0356542240
linkedin.com/in/se-nttai · github.com/Tai-TZ · portfolio-eight-indol-12.vercel.app

SUMMARY

Software engineer with 2+ years of experience building web applications (SaaS products, an ERP system, AI chat platforms) with Angular, React/Next.js, TypeScript and Python. Trained in AI engineering through VinUniversity's AI20K Talent Program, now a lab coach and engineer there. Design, build and deploy agent and RAG systems end to end (LangGraph, hybrid retrieval, guardrails, evaluation, FastAPI, React), working with AI coding agents under my direction.

TECHNICAL SKILLS

Languages	Python, TypeScript, JavaScript, SQL
AI / LLM engineering	LangGraph, LangChain, RAG (BM25 + vector hybrid search, RRF, reranking), structured outputs, guardrails, prompt-injection checks, model routing and failover, evaluation sets; Claude, Gemini and OpenAI APIs; PyTorch, OpenCV, NumPy, Pandas
Frontend	Angular, React, Next.js, TanStack, React Native (Expo), Tailwind CSS, Angular Material, Kendo UI, Ant Design, D3.js, three.js / React Three Fiber
Backend & data	FastAPI, Pydantic, SQLAlchemy, Alembic, PostgreSQL (row-level security, full-text search), Prisma, SQLite, REST, Server-Sent Events, WebSockets
DevOps & tooling	Docker, GitHub Actions, Google Cloud Run, Vercel, Render, pytest, Vitest, Playwright, Git, GitHub, Sourcetree, Atlassian tools, Claude Code, Cursor

EXPERIENCE

Lab Coach & Engineer
Sep 2026 – Present

VinUniversity · Hanoi · Full-time

- Contribute to VLearn, the learning platform of VinUniversity's AI20K program: features for learners and instructors in lab sessions (real-time support, class management, in-class activities).
- Built an internal tool that turns lecture slides into lesson videos and packages course content, cutting preparation time per course day.
- Researched personalization for GSM within VRIC's scope: studied the problem and evaluated candidate approaches.
- Support learners in lab sessions as a teaching assistant, and help lecturers run classes and prepare materials.

Independent Contractor
Dec 2025 – Mar 2026

Freelance (client: CDIMEX) · Ho Chi Minh City · Remote

- Main contractor for CDIMEX on BOOKMEDI, an online store for imported foreign-language books; negotiated scope, specs and deliverables and signed the agreement on behalf of the team.
- Led a cross-functional team (milestones, task assignment, progress) as the main client contact for progress and change requests.
- Designed the architecture and set up the codebase from scratch (structure, coding standards, workflow); reviewed code and guided team members.

Software Engineer
Oct 2024 – Jan 2026

LIZAI Inc. · Vietnam · On-site · Full-time

- Built chatbot features for the company's AI chat platform, from the chat interface to its integration with backend services.
- Reimplemented paid features as custom components with Angular 17+, React, Next.js and Python.
- Worked with the team on core business logic, performance, UX and complex technical issues.

Project Coordinator & Developer
Mar 2025 – Jun 2025

Freelance · Ho Chi Minh City · Remote

- Coordinated an e-commerce platform project end to end as liaison between client stakeholders and engineers.
- Turned business requirements into functional specs and prioritized backlogs; planned timelines, assigned and tracked tasks across frontend and backend teams; managed risks, check-ins and scope changes.
- Reviewed deliverables and contributed hands-on to frontend development and e-commerce API integration.

Developer May 2023 – Aug 2024

3P Solutions Limited · Ho Chi Minh City · Full-time

- Developed the frontend of a web-based ERP system with Angular 16 and TypeScript: business modules and data-heavy screens (forms, tables, dashboards).
- Built reusable components and shared services (including customized Kendo UI), integrated them end to end with backend APIs, and fixed performance issues and bugs.
- Trained and mentored interns through onboarding, code reviews and Angular best practices.

SELECTED PROJECTS

Edico – AI-assisted LaTeX editor for scientific papers May – Oct 2026

Bootcamp team of 3, system architect · Deployed to Google Cloud Run (2026) · github.com/Tai-TZ/edico-ai-paper-editor

- Architected an in-browser LaTeX editor where an LLM intent router dispatches to style, edit, structure and citation agents (FastAPI, LangGraph, SSE) and every AI change is an accept/reject diff; wrote most of the backend and frontend (React 19, TanStack Start, PostgreSQL).
- Co-built the LangGraph agents and guardrails (numeric-drift, length and prompt-injection checks); built a citation verifier (arXiv, Crossref, Semantic Scholar, OpenAlex), a 29-rule preflight and an AI "Defense Mode" mock viva.
- Delivered JWT + Google SSO auth, an admin console with LLM cost reports, the mobile editor UI, a cross-user isolation fix in the TeX Live compile sandbox, Docker, GitHub Actions CI and scripted Cloud Run deploys.

V-Game – Learn-by-playing game for building RAG agents Oct – Nov 2026

Solo · Live at vgame.ai20k.cloud · github.com/Tai-TZ/v-game

- Built a game where learners assemble RAG agents from blocks; a FastAPI engine compiles each into a LangGraph state graph, runs it on golden question sets, streams step traces over SSE and grades deterministically (no LLM judge).
- Serves real hybrid retrieval from a 512 MB free-tier host with no model in memory: 24 chunking variants, e5 embeddings and reranker scores precomputed offline; BM25 + vector search with RRF; replay cache, Gemini fallback and rate limits.
- Shipped a React 19, TypeScript and three.js front end (3D hub, accessible workbench, projector mode) with CI bundle budgets, a strict CSP, and Vitest, pytest and Playwright + axe tests.

Talent Hub – Admissions and training-quality platform with AI screening Oct 2026

Solo · Local prototype with synthetic data · github.com/Tai-TZ/talent-hub

- Built a multi-tenant platform (FastAPI, async SQLAlchemy, PostgreSQL 16, Next.js 16) with tenants isolated by forced row-level security and a test that fails if any tenant table lacks a policy.
- Designed human-in-the-loop AI screening: evidence quotes are kept only if found verbatim, PII never reaches the model, prompt injection is flagged, and four-eyes approval keeps every status change human.
- Raised bulk triage on 20,000 synthetic applications from ~95 to ~2,000 per second (process pool, batched inserts) and cut analytics p95 latency from 1.8 s to 49 ms with identical outputs.

HOICON – Scam and document helper app for elderly parents Nov 2026

Team lead (team VCV), Codex Community Hackathon – Hanoi · Hackathon prototype · github.com/Tai-TZ/hoicon

- Led the team and was sole committer of the shipped code: an Expo/React Native + FastAPI iPhone app where parents check scam messages, prescriptions and documents by photo or voice and hear a spoken verdict.
- Built a multi-provider LLM pipeline on LangChain structured outputs: Claude Haiku/Sonnet routing, staggered Gemini failover, and code-enforced rules so every high-risk result alerts the adult child in the app and on Telegram; 63 pytest tests with fake LLMs.

EDUCATION

VinUniversity – AI Engineering, AI20K Talent Program 2026

Van Lang University – B.Eng. in Software Engineering 2020 – 2024

GPA 3.08/4

AWARDS & SCHOLARSHIPS

- Student scientific research: rice leaf disease recognition and classification system.
- 50% university scholarship.
- 50% scholarship for a Master of IT in Taiwan.
- AI20K Talent Program, VinUniversity.