

Nguyen Manh Tu

+ (84) 397888035 | tunm17421@gmail.com | [linkedin.com/in/tunm17421/](https://www.linkedin.com/in/tunm17421/) | <https://tunm-dev.vercel.app>

Summary

Backend Software Engineer with 2+ years of experience in Java/Spring Boot and AWS, currently specializing in AI Engineering to build RAG systems, semantic search, and real-time AI platforms using FastAPI, Qdrant, and OpenAI.

Experience

Software Engineer - FPT Software

May 2024 – Mar 2026

Java, Spring Boot, PostgreSQL, AWS, Docker, GraphQL, Testing

- Backend Development:** Developed backend APIs for HR systems using Java, Spring Boot, and PostgreSQL, improving system efficiency by 20%.
- Cloud Infrastructure:** Implemented batch jobs and asynchronous processing with AWS SQS, utilizing Docker and LocalStack to optimize local development/testing.
- Leadership:** Led a 4-5 member backend sub-team, coordinating tasks and performing code reviews to boost team productivity by 30% during the lead's absence.

Projects

VCareer — AI Career Development Platform

May 2026 – Present

Next.js 16 & React 19, TypeScript, Tailwind CSS v4, Drizzle ORM & Neon PostgreSQL, Upstash Redis, Cloudflare R2, OpenAI GPT / Google Gemini Live, LiveKit

- Core Architecture:** Built a full-stack AI career platform using **Next.js 16 (App Router)**, **TypeScript**, and **Drizzle ORM**, leveraging **Server Components** and **Actions** to optimize performance and SEO.
- AI & Real-time Interview:** Developed a real-time AI interview module using **WebRTC (LiveKit)**, integrating **OpenAI Realtime** and **Gemini Live** with a **Three.js** 3D TalkingHead for immersive interactions.
- Data Processing & Storage:** Engineered asynchronous scoring workflows and CV analysis logic using **Cloudflare R2** for storage and **Zod** for schema-strict parsing of PDF/DOCX resumes.
- Recruitment Features:** Implemented end-to-end recruitment workflows, including an **ATS-friendly CV builder**, CV-JD matching algorithms, and automated job crawling/application tracking systems.

Scholar AI — AI-Augmented Research Workspace

Apr 2026 – June 2026

Next.js, React, FastAPI, PostgreSQL, Qdrant Cloud, OpenAI, Supabase, Render, Vercel, LangSmith

- RAG Architecture:** Designed a high-performance **RAG workflow** using **hybrid retrieval** (Dense Embeddings & BM25) and **hierarchical chunking** to improve semantic search accuracy and context preservation.
- AI Reliability & Trust:** Implemented **citation-aware chat** and **LLM guardrails** (cosine similarity thresholds, scope-enforcing prompts) to minimize hallucinations and ensure evidence-based responses.
- System Evaluation:** Built a comprehensive evaluation pipeline using **LangSmith**, **tracking key metrics** such as **hit rate**, **citation accuracy**, and **faithfulness** to iteratively optimize RAG performance.
- Full-stack Integration:** Developed the workspace using **Next.js**, **FastAPI**, and **Qdrant Cloud** (Vector DB), delivering a seamless experience for academic paper synthesis and writing.

Education

FPT University

Bachelor's Degree, Information Technology

Hoa Lac, Ha Noi

2019 – 2024

Languages

English (Independent)

Vietnamese (Native)

Korean (Independent)