Nguyen Anh Quan

Ho Chi Minh City, Vietnam | (+84) 369 423 476 | anhquan200304@gmail.com | in LinkedIn | Github

Technical Skills & Frameworks

- **Frameworks:** FastAPI, LangChain, Flask, Streamlit
- **Cloud & Data:** GCP, AWS, Azure
- **AI/ML:** Fine-tuning, Hugging Face, LoRA, Prompt Engineering
- **Vector & Databases:** Pinecone, PostgreSQL, MongoDB
- **MLOps:** Docker, Git, CI/CD
- **Analytics:** GA4, GTM, Facebook Pixel, Tableau, Power BI
- **Programming:** Python, SQL, TypeScript

Work & Experience

- **GFI Group**   *Dec 2024 - Present*
  AI Engineer   Ho Chi Minh City
  ◦ Designed and maintained scalable end-to-end ETL pipelines on GCP, reducing data processing time by **40%** for both batch and real-time workflows using Cloud Composer, external APIs, and cloud storage.
  ◦ Optimized BigQuery performance and cost efficiency, achieving **30%** cost reduction through SQL tuning, partitioning, clustering, and best practices in data modeling.
  ◦ Managed GCS-based data lake architecture and implemented HTTPS Load Balancing for Cloud Run services to ensure high availability, autoscaling, and reliable global routing.
  ◦ Designed and deployed AI chat systems using LangChain and AWS, including intent classification, memory integration, and dynamic response generation, improving engagement and reducing failure rate by **20%**.
  ◦ Enhanced AI agent performance through prompt engineering, cloud-based inference optimization, and systematic evaluation to ensure accurate, context-aware interactions.
  ◦ Researched and applied cutting-edge LLM and vector database technologies to build scalable, production-ready AI solutions in collaboration with cross-functional teams.

- **RE.LOVED - Sustainable Fashion**   *Jun 2024 - Jan 2025*
  Data Engineer   Ho Chi Minh City
  ◦ Optimized GTM setup for **90%** data accuracy (GA4, Facebook Pixel).
  ◦ Built scalable data architecture for marketing, sales, and operations data.

- **TDTU NLP-KD LAB**   *Jan 2024 - Present*
  AI Researcher   Ho Chi Minh City
  ◦ Developed educational RAG assistant with **80%** retrieval accuracy, and built chatbot systems (50+ weekly users) using OpenAI API, LangChain, and FastAPI.
  ◦ Fine-tuned Vietnamese NLP models using LoRA and transformer architectures, achieving strong performance in sentiment analysis and text classification tasks.
  ◦ Contributed to NLP research community by reviewing state-of-the-art papers, experimenting with small model efficiency, and presenting at academic conferences.
  ◦ Independently coded and deployed a full-stack AI project for MindWell, a mental health platform, handling both frontend and backend development single-handedly

- **Lotus Company**   *Jun 2023 - Dec 2023*
  Software Engineer   Ho Chi Minh City
  ◦ Developed full-stack app with AI chatbot, reducing support workload by **35%**.
  ◦ Set up CI/CD pipelines and cloud workflows for fast AI feature iteration.

Education

- **TON DUC THANG UNIVERSITY**   *Expected 2026*
  Bachelor of Computer Science   Ho Chi Minh City, Vietnam
  ◦ **Honors:** 3rd Place, mAInia AI Hackathon (2025); Honorable Mention, OAI HCMC 2025; 3rd Prize, Student Scientific Research.
  ◦ **Coursework:** Machine Learning, AI, NLP, Cloud Computing.

Language

- English (Advanced)