

Nguyen Quoc Vinh

District 2, Ho Chi Minh City, Vietnam | vinhquocnguyen.work@gmail.com | 0384832293

<https://www.linkedin.com/in/nguyen-quoc-vinh-nqv/> | <https://github.com/VinhAlth>

Career Objective

AI Engineer with 1 year of experience in developing AI applications and intelligent automation systems. Skilled in Python, AI workflows, backend development, and deploying AI solutions into real-world products. Passionate about Machine Learning, NLP, LLMs, and Generative AI, with a strong interest in building scalable AI systems and continuously expanding expertise in AI Engineering.

Work Experience

AI Engineer, Long Van System Solution 6/2025 – Present

- Developed AI-powered applications and automation workflows using Python, Java, and LLM technologies; also designed and built user interfaces, ensuring scalable and reliable solutions.
- Managed the full development lifecycle, from business analysis and technical design to implementation, deployment, and monitoring.

AI Research Intern, SIU AILAB – Artificial Intelligence Lab 6/2024 – 12/2024

- Explored LLMs and RAG architectures to enhance intelligent chat systems.
- Studied Transformer-based architectures and AI scientific research

Honors & Awards

Talent Award – AI Challenge Ho Chi Minh City 2025 11/2025

1st Prize Research and Top 5 Outstanding Students - University, Consolation Prize in EUREKA - National Research, Consolation Prize in the AI Challenge HCM 2024

1st Prize Research & 3rd Prize Startup – University, Consolation Prize – National Research, Finalist (AI Challenge HCM) 2023

Finalist in the AI Challenge Ho Chi Minh City 10/2022

Publications

An Optimized And Interactive Video Event Retrieval System With An Improved Temporal Algorithm. [Link] 4/2025

Video Retrieval System with Advanced Temporal Algorithm and Vision Language Models Integration (Accepted proceedings paper, to appear). 2026

Projects

AI Engineer: AI Assistant for Long Vân CRM Portal 2/2026 – 5/2026

Technologies Used: Python, LangGraph, LLMs, Knowledge Graph, Web Search, FastAPI, WebSocket, Prompt Engineering, Dynamic UI Generation.

Role: Analyzed business requirements, designed AI workflows and system architecture, developed backend systems, integrated AI features into the CRM portal, and collaborated with cross-functional teams to optimize internal development processes.

Description: Developed and deployed an AI assistant integrated directly into the Long Vân CRM Portal. The system supports context-aware conversations, technical discussion, technical documentation generation, dynamic UI generation, diagram and workflow creation, test case generation, blog/post writing, and AI-generated product images. Applied Knowledge Graph and web search mechanisms to improve contextual understanding and response quality.

Result: Production-deployed and used by more than 100 Long Vân employees, helping significantly reduce tool-switching time, accelerate technical analysis and documentation workflows, and improve overall internal productivity.

AI Engineer: Multi-Agent Customer Support System

11/2025 – 2/2026

Technologies Used: Python, LLMs, RAG, Multi-Agent, FastAPI, LangGraph, OpenAI, Vector Database, Prompt Engineering, MCP.

Role: Designed and developed multi-agent architecture including router, knowledge agent, task execution agent, and fallback agent. Built RAG pipeline with vector database, implemented tool-calling, and developed backend APIs for real-time customer support automation.

Key Contributions: Built multi-agent system with MCP-servers, implemented RAG using vector and non-vector knowledge sources, developed tool-using agents for API, database, and workflow automation.

Technologies

Technical Skills: Python, TensorFlow, PyTorch, Scikit-learn, Transformers, ML/DL, LLMs, RAG, RecSys, model optimization, Pandas, NumPy, Milvus, CI/CD, Docker, K8s, FastAPI, Linux, Ultralytics, Redis, PostgreSQL, MongoDB, Mlflow, n8n, MCP Sever, REST APIs.

Soft Skills

Skills: Learning mindset, presentation, business analysis, problem-solving, self-learning, research, teamwork, project leadership, analytical thinking, system thinking and read English documents.

Certifications

Foundation of machine learning and deep learning, SIU AILAB	2023
Machine Learning Specialization, Stanford University [Link]	2023
Neural Networks and Deep Learning, DeepLearning.AI [Link]	2025
Advanced Multi-Agent AI System, The Coursera	2026

Education

Computer Science - AI Specialization, Saigon International University 9/2022 – 3/2026

- GPA: 3.43/4.0 - Full scholarship (4 years)