Lê Ngọc Tú

AI Engineer

+84 969 691 930 · lntu2003.dev@gmail.com
Phan Thiết, Việt Nam · lengoctu.com
github.com/Le-Ngoc-Tu · huggingface.co/LeNgocTu

KỸ NĂNG

AI / ML
LLM, RAG, AI Agent, Fine-tuning (LoRA), Embedding, TTS, Computer Vision (Face, Object)

Ngôn ngữ lập trình
Python, JavaScript, TypeScript, SQL

Backend
FastAPI, Node.js, REST API, WebSocket

Frontend
ReactJS, NextJS, Vite

Cơ sở dữ liệu
MySQL, MariaDB, Vector Database (Qdrant)

DevOps & Tools
Git, Docker, Linux, Nginx, Ollama

HỌC VẤN

Cử nhân Công nghệ Thông tin
Đại học Phan Thiết
09/2021 – 01/2025 · GPA 3.77 / 4.0

GIẢI THƯỞNG & NGHIÊN CỨU

Road to AI · Stage 2
Top 10, vào Vòng Chung kết Demo Day
Hạng mục AI Solo với dự án FinWhale.

02/2025
Phát hiện gian lận thi cử bằng mô hình YOLO
Đăng trên Tạp chí Giáo dục Việt Nam. Phát hiện gian lận theo thời gian thực, độ chính xác 85%.

11/2024
Giải thưởng NCKH Sinh viên Euréka lần XXVI
Do Đại học Quốc gia TP.HCM trao cho đóng góp nghiên cứu xuất sắc trong lĩnh vực AI và Machine Learning.

GIỚI THIỆU

AI Engineer gần 2 năm kinh nghiệm đưa AI vào sản phẩm thật: thị giác máy tính, LLM, RAG, TTS và web full-stack. Đã triển khai nhiều giải pháp end-to-end cho chính phủ và doanh nghiệp tại Việt Nam.

KINH NGHIỆM LÀM VIỆC

AI Engineer
12/2024 – Hiện tại

Công ty TNHH Giải pháp Phần mềm Nguyễn Luân (NLTECH) · Phan Thiết

- Nghiên cứu, fine-tune và huấn luyện các mô hình AI cho tiếng Việt (embedding, LLM persona, TTS); công bố mô hình trên HuggingFace.
- Xây dựng pipeline RAG & LLM để tự động hóa truy xuất thông tin và sinh câu trả lời.
- Phát triển nhận diện khuôn mặt, phát hiện đối tượng và tích hợp vào backend đưa vào sản xuất.
- Thiết kế REST API, mô hình hóa cơ sở dữ liệu và phát triển giao diện với ReactJS.

DỰ ÁN TIÊU BIỂU

ViEn-Matcha — TTS song ngữ Việt – Anh siêu nhẹ, chạy được trên CPU
Matcha-TTS (flow matching), Vocos, Whisper, ECAPA, UTMOS, CTC Alignment, ONNX Runtime
Huấn luyện từ đầu trên bộ dữ liệu tự xây dựng và lọc kỹ, cho ra 6 giọng đọc tự nhiên, xử lý tốt câu trộn Việt – Anh. Lỗi phát âm chỉ **1,9% CER**, nhanh gấp **11 lần** thời gian thực trên GPU phổ thông và **5 lần** trên CPU.

NLTECH Camera AI — Giám sát giao thông thời gian thực
YOLO26n (TensorRT), ByteTrack, OCR biển số (ONNX), NVDEC, ZeroMQ, MariaDB, WebSocket
Biến camera thường thành hệ thống giám sát thông minh: theo dõi người, xe, đọc biển số và cảnh báo xe vi phạm tức thì qua Telegram.

Mô hình TTS song ngữ Việt – Nhật (Irodori-Sen)
Text-to-Speech, Fine-tuning, Deep Learning
Giọng đọc AI nói tự nhiên cả tiếng Việt lẫn tiếng Nhật.
HuggingFace: LeNgocTu/Irodori-Sen-TTS-v2

Công cụ dịch tài liệu PDF
Python, LLM APIs, PyMuPDF
Dịch hàng loạt tài liệu PDF dày mà vẫn giữ nguyên bố cục gốc.

FinWhale — AI Agent phân tích tài chính · Top 10 Road to AI
AI Agent, LLM, Pandas
Hỏi bằng ngôn ngữ thường, AI tự viết truy vấn Pandas và trả kết quả trên giao diện kiểu Bloomberg Terminal.
GitHub: github.com/Le-Ngoc-Tu/finwhale

Bộ cấu hình AI Agent hỗ trợ lập trình
Claude Code, Antigravity CLI, MCP, Skills
Bộ luật và công cụ giúp AI viết code đúng ý, đúng phạm vi, không làm thừa.

EduViz AI — Trực quan hóa Toán & Vật lí THPT
Next.js 16, React 19, TypeScript, JSXGraph, Three.js, Gemini 2.5 Flash, Strands Agents, SSE
Chụp đề là có lời giải gợi mở và đồ thị 2D/3D tương tác, phản hồi chỉ 1.5–3.5s thay vì hơn 60s.
Demo: eduviz.lengoctu.com · GitHub: github.com/Le-Ngoc-Tu/eduviz

Kiosk AI hỗ trợ thủ tục hành chính (AAAS)
FastAPI, ReactJS, LlamaIndex, Qdrant, OmniVoice, LLMs
Người dân hỏi bằng giọng nói, kiosk đa agent trả lời thủ tục chính xác nhờ RAG.

Hệ thống AI luyện thi IELTS
FastAPI, ReactJS, Strands Agents, Qdrant, Groq API, Ollama
Gia sư AI luyện đủ 4 kỹ năng, tự chấm điểm và chỉ ra cách cải thiện.

Gemma persona & Embedding tiếng Việt
LLM, Embedding, Fine-tuning (LoRA), GGUF, RAG
LLM tiếng Việt chạy mượt trên máy yếu, kèm embedding 300M cho tìm kiếm ngữ nghĩa.
HF: LeNgocTu/gemma-4-E2B-vietnam-personas-gguf · LeNgocTu/vietnamese-embeddinggemma-300m