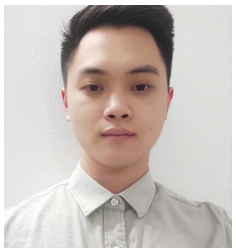

DOAN DUC THANG

Address: Ha Dong, Ha Noi
Phone: 0978 851 058
Email: ducthang8558@gmail.com
Github: https://github.com/DoanDucThang0805

SUMMARY

Graduated in Electronics and Telecommunications Engineering with a strong passion for AI. Experienced in building and training models, fine-tuning LLMs, and developing RAG systems. Focused on advancing my career in AI and transforming innovative AI technologies into practical, impactful real-world products.

EDUCATION

Posts and Telecommunications Institute of Technology **Aug 2020 - May 2025**
Degree of Engineer – Electronics and Telecommunications
GPA: 2.94

WORK EXPERIENCE

FPT TELECOM **June 2024 - Dec 2024**
Intern AI Engineer

Project: RAG Chatbot
Description: This project focuses on developing an advanced chatbot for efficient document or information retrieval. The knowledge base of chatbot is the combination with unstructured data (.doc, PDF, ...) and structured data (SQL)
Responsibilities:

- Preprocessing documents (.pdf, .doc, etc.), performing text chunking and embedding, and storing them in Milvus vector DB.
- Building an information retrieval system, including:
  - Processing queries and prompt optimization.
  - Developing tools for LLM function calling.
  - Implementing similarity search for unstructured data in Milvus DB.
  - Converting queries into SQL and executing them in SQLite.

Tech stack: Python | Langchain | LangGraph | Streamlit | Milvus DB | SQL Lite | Gemini LLM

Project: Camera Object Tracking
Description: This project focuses on developing a high-performance, real-time vehicle detection and tracking system (car, bus, truck, motorbike, bicycle) for surveillance applications.
Responsibilities:

- Collecting, preprocessing, and annotating data
- Training the YOLOv8 model
- Converting and quantizing the model to TensorRT
- Implementing the DeepSort tracking algorithm
- Deploying the system on a Jetson Nano (Docker)

Tech stack: Python | OpenCV | YoLo | DeepSort | TensorRT | Docker

---

ICNLAB PTIT
Student Researcher

Project: Tomato Plant Pest and Disease Diagnosis Application
Description: This research was accepted and presented at the REV ECIT 2024 National Conference, focusing on the development of a tomato plant disease diagnosis solution based on DCNN with an LLM integrated to enhance application functionality.
Responsibilities:

- Collecting and preprocessing image datasets.
- Fine-tuning the MobileNetV3 model model and quantization model.
- Deployment:
  - Building APIs for model deployment.
  - Integrating LLaMA LLM and optimizing load balancing with Redis and Celery.

Tech stack: Python | OpenCV | Onnx | Llama LLM | FastAPI | Streamlit | Redis | Celery
Link paper: https://drive.google.com/file/d/1pni-BasHE8u92Wr0EwI1Oo5V6irCNQCZ/view

SKILLS

Technical Skills:

- **AI Skills:** Computer Vision, LLMs, RAG-Chatbot, MLOps, Docker
- **Programming Languages:** C++, Python
- **Framework:** Scikit-learn, Tensorflow, Pytorch, Langchain, LangGraph, Onnx, TensorRT

English Communication: Intermediate

CERTIFICATIONS

- B1 Aptis English 2024
- Certificate of Scientific Research 2024
- Certificate course in AI, BIG DATA by SAMSUNG 2023