

# Tuan-Anh VO

📍 Ho Chi Minh City ✉️ votuananh.frtk@gmail.com ☎️ +84357231917 🌐 in/tuananhvoai

## SUMMARY

Data scientist with an MSc in Natural Language Processing, now pivoting into LLMs & Agentic AI. Blending tech proficiency with sharp communication, I am driven to turn AI into real-world impact—enhancing learning, boosting productivity, and driving business success.

## EXPERIENCE

### Data Scientist | Deployment Strategist

FPT Software | Palantir Technologies

January 2023 – September 2024, Ho Chi Minh City, Vietnam

- Conducted **3 international onboarding trips** to different Airbus' client airlines to deliver **Skywise** — a Palantir's Foundry data framework for Aviation industry.
- Developed Machine Learning solution for **FlightOps prediction based on weather data**, successfully delivered to South African Airways.
- Builded and deployed **10+ analytic workflows** regarding Reliability, Safety, Flight Ops and sensor data. Cobuilted and instructed airlines' users with **an average 80% rate of implementation in their daily work**.
- Delivered **100+ hours of user trainings** for Skywise and workflow scoping.
- Looked over and coordinated with data engineers in **modeling and transforming a complex ecosystem of airlines data** from various sources before each trip.

### Data Analyst | Business Analyst

Wilmar International

May 2022 – January 2023, Ho Chi Minh City, Vietnam

- Designed workflow/data logic for new **multi-level Promotion features** of company's WDMS/SFA software solution.
- Delivered **30+ hours of user training** for international clients in retails in the rolling out process to new countries.
- Participated in a **transition to microservices architecture** for the software (previously monolithic).
- Chief BA for developing new mobile version** of the software.

### Intern NLP Researcher

LORIA

October 2020 – September 2021, Nancy, Grand Est, France

- Mastered different grammar formalism such as **Abstract Categorical Grammar (ACG)** and **Meaning Text Theory (MTT)** and applied to **analyse Vietnamese Phrasemes within lambda calculus framework**, and embedded into a formal, semantic approach to Natural Language Generation.
- Led a team of 4 conducted research on enhancing embedding performance on graph data, applying **hybrid approach to combine attention mechanism with auto-encoded meta-embedding**, with significant improvement on relation prediction tasks.

## PROJECT

### Economics Research Assistant

- Implements **agentic RAG techniques** to create a Research Assistant Bot that learned from relatively large corpus of academic economic literature, to autonomously **aid learning, critique writing, summarizing and complex reasoning** [*PydanticAI, LangGraph, Crawl4AI*]

### Finetuning vs In-context learning for Deepseek R1

- Based on a few highly technical texts on economics, build both a **finetuning pipeline using LoRA**, and a **contextual RAG with adding BM25 Retriever + Reranker** layers to comparatively study the performance of Deepseek R1. [*Langchain, unsloth, wandb*]

## EDUCATION

### MSc in Natural Language Processing

University of Lorraine • 2021 • Mention: Good

- Coursework:** Computer Science, Statistics, Data Science, Information Theory, Machine Learning/Deep Learning, Linguistics, NLP Techniques.

## CERTIFICATIONS

### Udacity Nanodegree: Artificial Intelligence

2024

- Agent theories, Constraint Propagation, Automated Planning, Heuristic Search, Adversarial Search, Optimization.

### Udacity Nanodegree: Data Engineering with Microsoft Azure

2022

- Databases, Data Modeling, Warehouse and ETL, Data lake and ApacheSpark, Data Pipeline.

## SKILLS

**LLMs Core:** Backpropagation, RNN, LSTM, BERT, Transformer, Tokenization, MoE, Distillation, RLHF

**LLMs Specialization:** RAG, Fine-Tuning, AI Crawling, Vector Databases, Contextual Embeddings, [*HuggingFace, Langchain, Crawl4AI*]

**LLM-powered Application:** Prompt Engineering, Autonomous Agent, Workflow, AgenticRAG [*PydanticAI, LangGraph, N8N*]

**Data Analytics & Machine Learning:** Wrangling, EDA, BI Analytics, Feature Engineering, ML/DL Modeling [*SQL, Pandas, NumPy, PySpark*]

**Data Engineering:** ETL, Star Schema, Warehouse & Datalake, SparkSQL, SparkStreaming, RDD [*Foundry, ApacheSpark, Databricks, Azure*]

**MLOps & LLMOps:** FastAPI, Model Packaging, *MLFlow*, Continuous Monitoring, Orchestration, Post Productionize

**Software Production & Agile Method:** Microservices, Docker, Kubernetes, CI/CD

**Working skills:** *User training, Public speaking, Technical writing*

**Language:** *English [Bilingual Proficiency], French [Limited Working Proficiency], Vietnamese [Native]*