Le Van Hoang — AI Engineer Intern

(+84) 3 282 81624 • 22520465@gm.uit.edu.vn • hoanglvuit • hoanglvuit

Summary

An AI/Machine Learning student with a strong foundation in **Deep Learning (DL)**, **Machine Learning (ML)** and has experience conducting research on adversarial examples. I am eager to gain hands-on experience through an internship in a real-world production environment. I am highly motivated to learn, grow, and contribute to impactful projects.

Education

University of Information Technology (UIT) — **HCMC, Vietnam**
Bachelor of Computer Science — *9/2022–1/2026 (expected)*
GPA: 8.95/10.0 (current)

Technical Skills

Languages: Python, HTML, CSS

ML & Frameworks: PyTorch, scikit-learn, Hugging Face, FastAPI

Tools: Git, Docker, CI/CD, Linux, Jenkins

Skill: MLOps

Projects

Chatbot about Vietnam criminal law QA **Solo Project** **05/2025 – 06/2025 [GitHub] [Website]**:
- Developing a Vietnamese Criminal Law QA Chatbot using RAG architecture.
- Developed the backend with FastAPI and the frontend using Tailwind CSS and Vite.
- Deployed the frontend on Vercel (free-tier) and the backend on VPS
- Implemented CI/CD automation using Jenkins to build and push Docker images, SSH into the VPS, pull images, and run containers. Full deployment workflow is documented on GitHub

Legal Document Retrieval - Top 3 SoICT **Team Size: 4** **10/2025 – 12/2025 [GitHub]**:
- Lead team and served as main developer, developed legal document retrieval system by integrating a Bi-Encoder model for retrieval and Cross-Encoder models for ranking, enabling effective retrieval from text-based queries.
- Addressed the limitation of having only question-answering datasets by fine-tuning the Bi-Encoder with Multi-Negative Ranking Loss and enhancing Cross-Encoder training by applying negative mining techniques.

Drought Prediction Using Meteorological & Soil Data **Team Size: 5** **03/2025 – 05/2025 [GitHub]**:
- Responsible for the research and design of an AI architecture to forecast drought severity 7 weeks ahead, using 175 days of historical weather data and regional soil features.
- Engineered temporal features (trends, seasonality, lags) and evaluated multiple sequence models (RNN, GRU, LSTM, Transformer) to learn time-aware embeddings.

Adversarial Patch Attacks on Sign Classifiers **Solo Project** **03/2025 – 05/2025 [GitHub]**:
- Train a simple CNN to classify traffic signs, achieving 95% accuracy and 90% F1 score
- Applied the CamoPatch technique to deceive a self-trained sign classifier, reducing its accuracy from 95% to 1%.
- Enhanced the genetic algorithm by customizing the individual update mechanism for real-world constraints.

Certifications

Machine Learning Specialization and Deep Learning Specialization : Coursera - Andrew Ng (2024)

Applications of AI for Anomaly Detection: NVIDIA (2025)

TOEIC: Listening & Reading: 765 (2024), Speaking & Writing: 280 (2024)

Research and Competitions

SoICT Hackathon 2024: Top 3 - Legal Document Retrieval

Home Credit - Credit Risk Model Stability, Kaggle Competition: Top 80/3856 - Silver medal

Optimizing Legal Document Retrieval in Vietnamese with Semi-Hard Negative Mining: Submitted to ICCCI 2025 (first author - accepted)

Prompt Manipulation for Targeted Adversarial Object Generation in Stable Diffusion: (first author - Manuscript ready for submission)