

TRAN NGUYEN DUY BAO

15/02/2003 • Nha Be District, Ho Chi Minh City • trannguyenduybao152@gmail.com • +84 829 782 878

[Linkedin](#) • [Github](#) • [Portfolio](#)

AI Engineer with **2 years of experience** building and deploying ML solutions in Computer Vision, NLP, and Big Data. Proven ability to deliver real-world applications in healthcare, legal, and media domains. Strong in end-to-end development, from data engineering to model deployment.

TECHNICAL SKILLS

- **Languages & Frameworks:**
 - o **Languages:** Python, JavaScript
 - o **Frameworks:** Flask, FastAPI, Node.js, React.js
 - o **Databases:** SQL (PostgreSQL, MySQL), NoSQL (MongoDB, Elassandra)
- **AI / ML Libraries & Ecosystem**
 - o TensorFlow, PyTorch, Scikit-learn, XGBoost
 - o Hugging Face Transformers, LangChain, LangGraph, vLLM, Ollama
 - o YOLO (Ultralytics), DeepFace, face_recognition, OpenCV
 - o SpaCy, NLTK
 - o OCR (Tesseract, EasyOCR)
- **Tools & Platforms:** Git, Docker, GCP, Azure, OpenAI API, RabbitMQ, Chromium, Scrapy, Selenium,
- **Data Engineering:** NumPy, PySpark, Pandas, Matplotlib, Seaborn
- **Soft Skills:** Critical thinking, communication, research-driven, self-learning, scientific paper reading

WORKING EXPERIENCE

AI Engineer – Junior

Saigon, Vietnam

MCV Complex

Jul 2023 – Present

- **Built LLM healthcare chatbot** using Flask and vector DB (RAG pipeline) utilizing **AI Graph Agent**.
- **Developed FaceID event check-in** with YOLOv8 + DeepFace for spoofing detection.
- **Multimodal approach** for video + audio summarization.
- **Fine-tuned PhoBERT** for Vietnamese legal and sentiment classification, boosting classification accuracy on internal dataset.
- **Built full-stack AI tools** for data crawling and classification.

PROJECTS

Spaceone – Summarization & Sentiment Analyst

MCV Complex, 2023 - now

Technologies: Python, Kafka, GCP, AWS, Node.js, Spleeter, Whisper, OpenAI GPT-4, LangChain, Pinecone, PhoBERT

Context: Real-time system to summarize media content and classify viewer sentiment across platforms.

Responsibilities:

- Designed a two-module pipeline:
 - o Summarization: Used Spleeter to isolate vocals from video, transcribed with Whisper, then used structured prompts with GPT-4 via LangChain for summarization.
 - o Sentiment Analysis: Collected Vietnamese-language comments, fine-tuned **PhoBERT** for sentiment classification (positive, neutral, negative).
- Integrated the summarization module into a Kafka-based real-time streaming architecture.
- Designed a modular inference engine to toggle between local open-source models (**vLLM, Ollama**) and cloud-based API services for premium users, optimizing cost and performance.

Impact:

- Sentiment classifier reached F1-score of 0.84 on real-world, domain-specific data.
- Enabled non-technical staff to generate summaries and mood insights at scale.

Doctor Assistant – Multimodal Chatbot

MCV Complex, 2023 - now

Technologies: Python, LangChain, LangGraph, OpenAI GPT-4, Pinecone, Flask, Whisper, HuggingFace, Streamlit, PyPDF

Context: Developed an AI-powered assistant to help patients and doctors communicate with structured, context-aware dialogue across text, documents, and images.

Responsibilities:

- Built a LangChain-based conversational agent with memory and tool usage to simulate multi-turn medical conversations.
- Integrated Whisper for speech transcription and image/PDF support using PIL and PyPDF, allowing multimodal input.
- Embedded medical literature and uploaded files into a Pinecone vector store for **retrieval-augmented generation (RAG)**.
- Designed structured prompt templates to control LLM output (JSON format, FAQs, summaries).
- Built a frontend prototype in Streamlit for testing conversational flow and user experience.
- Deployed API backend using Flask with endpoints for chat, file upload, and retrieval.

Impact:

- Enabled multimodal document understanding (PDFs, prescriptions, images).
- System could answer patient queries, summarize uploaded files, and interpret symptoms using structured generation logic.

Diabetes Prediction with Tabular Data Classification

MCV Complex, 2025

Technologies: Python, Pandas, Scikit-learn, OCR, SVR, ADASYN, RobustScaler, Matplotlib, Seaborn

Context: Built a regression-based ML pipeline to estimate blood sugar levels from questionnaire responses and basic physiological metrics, enabling low-cost, non-invasive diabetes screening.

Responsibilities:

- Collected and cleaned medical survey data from Google Sheets and OCR survey form
- Preprocessed categorical and binary features with one-hot encoding; normalized continuous variables using **RobustScaler**.
- Handled skewed glucose distribution using **log transformation**:
 $y' = \ln(1 + y)$, reversed by $y = e^{y'} - 1$
- Applied **ADASYN oversampling** to balance class distribution between diabetic and non-diabetic samples.
- Trained 10+ models (Linear, Ridge, Lasso, SVR, RF, AdaBoost, MLP, etc.), tuned via **RandomizedSearchCV**.
- Visualized results using residual plots, prediction-vs-actual charts, and correlation heatmaps.

Impact:

- **Support Vector Regression (SVR)** achieved best performance:
RMSE: ~0.6–0.7, $R^2 > 0.85$, and ~5% average prediction error.
- Enabled early-stage diabetes risk classification without lab tests.
- Demonstrated scalability and interpretability for real-world healthcare screening tools.

Spooface – Face ID with Spoofing Detection

MCV Complex, 2024

Technologies: Python, YOLOv8, Ultralytics, OpenCV, DeepFace, face_recognition, Pandas

Context: Developed for live event check-in systems requiring real-time face verification and anti-spoofing.

Responsibilities:

- Fine-tuned a YOLOv8 detection model to classify real vs. spoofed facial inputs (printed images, digital screens, masks).
- Created custom dataset with synthetic and real-world spoofing scenarios using OpenCV for pre-processing.
- Integrated DeepFace and face_recognition to verify identity post-liveness detection.
- Built real-time inference pipeline using OpenCV and Ultralytics, optimizing for latency and FPS.
- Deployed prototype in live event scenarios with real-world lighting and camera variation.

Impact:

- Achieved 92% spoof detection accuracy, reducing unauthorized check-ins by over 85%.
- Enhanced event security while maintaining sub-1s detection latency for smooth user experience.

Medical Abnormality Detection from Chest X-rays

Ton Duc Thang University, 2023

Technologies: Python, YOLOv8, Ultralytics, OpenCV

Responsibilities:

- Curated and cleaned open-source datasets (NIH ChestX-ray, Shenzhen) for abnormality detection.
- Used YOLOv8 to fine-tune a 6-class object detector on annotated X-ray scans.
- Applied image preprocessing (resizing, histogram equalization) and Albumentations for heavy augmentation (elastic distortion, CLAHE, gamma correction).
- Visualized bounding boxes and confidence maps to assist in radiology decision support.

Impact:

- Achieved **70.2% mean Average Precision (mAP)** and **60.4% recall** on held-out test set.
- Contributed annotated dataset and model checkpoints to university medical AI lab for further research.

Stock Forecasting Using LSTM and Statistical Analysis

Ton Duc Thang University, 2022

Technologies: Python, TensorFlow, Pandas, Matplotlib, Seaborn, Selenium, Scrapy, OLS

Context: Personal research project to understand predictive modeling on time series data from financial markets.

Responsibilities:

- Built a web scraping pipeline with Selenium and Scrapy to collect multi-source financial time-series data (stock prices, earnings reports, indicators).
- Engineered features like rolling averages, volatility bands, and lag windows using Pandas.
- Applied OLS regression to interpret basic trend and correlation factors.
- Designed and trained an **LSTM** model to predict next-step closing price; included **dropout** and early stopping to prevent overfitting.
- Conducted visual EDA using Seaborn and Matplotlib to understand outliers and long-term trend shifts.

Impact:

- LSTM model achieved **mean squared error within 5%** of real closing prices.
- Discovered anomaly clusters around earnings dates, used as input for future risk modeling.

EDUCATION

Ton Duc Thang University – B.Sc. in Computer Science (2025)

GPA: 8.0 / 10 | English: C1 | Top 10 – Recursion 2022

REFERENCES

Dr. Nguyen Trang Thao - Former PhD at TDTU Lab - [ORCID](#) - thao.nguyentrang@vlu.edu.vn

Mr. Huynh Hai Huynh - CTO at MCV Group - haihuynh.huynh@mcvnetworks.com