Tran Ngoc Tuong

(+84)987649725 — tuongtrangoc@gmail.com — tuongtrangoc — tuongtrangoc

SUMMARY

I am a Senior AI Engineer/ Data Scientist in the field of machine learning. My research interests include the Machine Learning/Deep Learning/AI algorithms and applying software engineering principles to real-world applications.

EDUCATION

Ha Noi University of Science and Technology **Aug 2015 – Aug 2020**
Applied Mathematics and Informatics of Engineering

Minors: Mathematics ; Informatics

CPA: 3.11/4

Certifications

Certificated ML/DL course sponsored by Viet Nam Institute for Advanced Study in Mathematics

Certificated Kubernetes for the Absolute Beginners-Hands-on by Udemy

MLOps — Machine Learning Operations Specicalization by Coursera

AI Agent Course - HuggingFace.

TECHNICAL SKILLS

Hard Skills

OS	Debian, Ubuntu, CentOS
DMS	PostgreSQL, MongoDB, MySQL, Milvus
Cloud	GCP, Microsoft Azure
Languages	Python, Bash
Version Control	Git, DVC
DevOps/BE	MinIO, Redis, Nginx, Flask API, CI/CD, Github Actions, Docker, Kubernetes
Frameworks	Tensorflow, Pytorch, Transformers, TF-serving, LLMDeploy, Langchain, LlamaIndex, Huggingface

Soft Skills

– Good English communication skills (TOEIC - 650)
– Have the ability to quickly learn new technology
– Able to work independently as well as in a team
– Working under high pressure

EXPERIENCE

OmiGroup **Sept 2025 — Present**
AI Engineer/ AI Researcher

Project: OmiKG - Knowledge Graph-based Root Cause Analysis for Functional Medicine (5 members)

Role: AI Engineer/ Researcher

Description: A method for constructing causal knowledge graphs for root cause analysis of chronic diseases in functional medicine.

Technologies: Knowledge Graph, UMLS Entity Normalization, Relation Extraction, LLMs, FastAPI, Prompt Engineering.

Responsibility:

- Researched and analyzed existing approaches for chronic disease modeling using biomedical knowledge graphs.
- Studied and applied UMLS entity normalization techniques for mapping clinical and biomedical terms.
- Evaluated the effectiveness of the proposed method through case studies and qualitative analysis.
- Assisted in drafting and reviewing research documentation and technical reports.
- Led the design and formulation of the evaluation methodology for the research paper submitted to AMIA (Top-tier in Health / Medical Informatics).

AI consultant

Description: Participate in presale projects in the field of digital healthcare as a Solution and AI Consultant.

AI Projects:

- **AI-OCR**: Design and propose an OCR approach to extract structured data from prescription documents (FAX, Medication record, Patient information) in the prescription ordering workflow.
- **AI Agent-CallBot**: Build a PoC for an AI Agent integrated with the EMR/EHR system to assist patients in adjusting, scheduling, and canceling bookings.
- **AI RAG-ChatBot**: Build a PoC for an RAG ChatBot integrated with the EMR/EHR system to assist patients by providing guidance on how to use and access their medical records.

KALAPA, SJC **Sept 2024 — Sept 2025**

Data Scientist/AI Engineer/Technical Lead

Project: eKYC - Say good bye to digital identity fraud (4 members)

Role: Data Scientist/AI Engineer

Description: Maintain, optimize, and improve AI models and system.

Technologies: FastAPI, Kubernetes, Unittests, Locust, TF-serving, model quantization.

Responsibility:

- Improve quality check models to prevent fraud on identification cards and passports.
- Optimize system, design load test, high available test, and unit test for service APIs.
- Develop a face-antispoofing module to prevent spoofing in the Liveness Detection API.

Project: Agentic Document Extraction (3 members)

Role: Technical Lead

Description: Build an end-to-end system to extract information from an unstructured document (such as financial reports, bank statements, etc.).

Technologies: Opensource-VLM (Intern-VL, Qwen-VLM, LLMDeploy), Google Gemini, Prompting Techniques, FastAPI, Kubernetes, MySQL, MinIO, and Nginx.

Responsibility:

- Research algorithms and design System Architecture solutions for system.
- Develop a solution for system deployment and integration.
- Optimize system, design load test, high available test, and unit test for API services.

Eastgate-Software **Jun 2021 – Sept 2024**

Middle AI Engineer

Project: ChatGPT - AI Scoring System (3 members)

Description: Evaluate companies' annual reports to score list of defined criteria.

Role: AI Engineer

Technologies: Azure ChatGPT services, prompting techniques, fine-tuning model, RAG with LLMs, Machine Learning regression models, Milvus Vector Database, Mongo Database.

Responsibility:

- Build end-to-end an RAG with LLMs for exact and semantic search.
- Build end-to-end a Machine Learning model to score for defined criteria.

Project: Document Intelligence - Table Structure Recognition (TSR) (2 members)

Role: AI Engineer

Description: Extract information tables from PDF/images in hematology reports.

Technologies: Table Transformer, Tkinter, fine-tuning model, table generation, Object Detection.

Responsibility:

- Research and propose technical solutions for customer.
- Training and fine-tuning model.
- Build a Windows application.
- Documentation of the proposal.

Accuracy: approximately 91%

Project: Document Intelligence - Key Information Extraction (KIE) (2 members)

Role: AI Engineer

Description: Extract key information from PDF/images from reports on medical and veterinary medicine.

Technologies: LayoutXLM, Machine Learning Classifier models, AWS DIP services, Semantic Entity Recognition, Relation Extraction, Docker.

Responsibility:

- Research and propose technical solutions for customer.
- Implement and optimize models for the relationship between key and value pairs using an ML model, with the aim of achieving superior performance compared to the Deep Learning model.
  Repository: https://github.com/tuongtranngoc/Language-independent-Entity-Linking
- Build a Docker application.
- Documentation of the proposal.

Accuracy: approximately 91%

Project: Computer Vision - Vehicle Identification Number Recognition (VIN) (2 members)

Role: AI Engineer

Description: Build end-to-end a system to detect and recognize Vehicle Identification Number.

Technologies: PaddleOCR, Object Detection, Object Recognition, MySQL.

Responsibility:

- Optimizing the OCR detection model: instead of relying on previous object detection algorithms to enhance the model using the highly advanced Differential Binarization (DB) algorithm.
- Optimizing the OCR recognition model: enhance the CRNN recognition algorithm by using a bag of tricks for CTC loss (Focal-CTC loss, A-CTC loss).

Accuracy: approximately 98%

Project: Data Science - Traffic flow prediction (3 members)

Role: Data Scientist

Description: Using data of features (weather, rainfall, wind speed, ...) from the websites Hong Kong government to predict the number of vehicles on road segments.

Technologies: Data Visualization, Time series models, Docker, Kafka.

Responsibility:

- Develop and optimize a time series model for real-time casting of the number of vehicles on road segments(up to 200 road segments/2s).
- Support to deploy the model to DEV, UAT, DR, and PROD environments.

Accuracy(MAE): approximately 15-30 vehicles/200 road segments.

Innovation Center – VNPT Information Technology Company **Sept 2019 - May 2021**
Junior AI Engineer

Project: Smart Vision - Action Recognition, Detect the elderly and children on surveillance camera and Weapons Recognition (2 members)

Description: Research, develop and improve AI models and system.

Technologies: Object detection (YOLO-series, CenterNet), Image Classification, Pose Estimation, Face Recognition, Pruning and Quantization model.

Responsibility:

- Research approaches and proposes solutions based on human pose estimation algorithms.
- Improve and optimize the model to increase the accuracy and run-time of the algorithm.
- Build Docker, FastAPI to integrate with Backend service.

ICOMM Media and Tech, Jsc **Feb 2019 - May 2019**
AI Developer Internship

Project: People Tagging - Data Crawling (1 members)

Description: Build a tool to crawl data from social webs automatically (Facebook, Google, ...).

Technologies: Selenium Chrome C#, JavaScript/CSS, MSSQL, ...

Project: Plate Recognition (2 members)

Description: Research and improve AI models in plate recognition.

Technologies: OCR, Image Classification, Object detection.

PROJECTS & SELF-STUDY EXPERIENCE

Agentic Document Extraction Portal **Aug 2024 – Present**
AI Software Engineer

Description: Build an end-to-end open-source production to extract information from an unstructured document

Technologies: Open-VLM (Intern-VL, Qwen-VLM, LLMDeploy), Google Gemini LLM, Prompting Techniques, FastAPI, Kubernetes, MySQL, MinIO, and Nginx.

Re-implementing Deep Learning/Machine Learning algorithms from scratch **Aug 2015 – Present**
AI Researcher

Re-implementing OCR algorithms (CRNN, Differential Binarization, ...).

Re-implementing object detection, segmentation, human pose estimation algorithms (YOLO-series, CenterNet, OpenPifpaf, ...).

Github profile: https://github.com/tuongtranngoc

AWARDS AND ACHIEVEMENTS

Top 18/30 Final Round - CODEWAR 2019 programming contest organized by FPT

– *Registered team name*: TTBKK60

– *Competition ranking*: Top 18/30

Best applicability project prize of scientific research contest for students - 2019

– *Advisor*: Assoc. Prof. Thi-Lan Le, International Research Institute MICA

– *Research topic*: Personal protective equipment (PPE) - Build a system to detect and monitor the personal protective equipment of workers at construction sites

– Certificate of Participation in the Student Scientific Research Competition

Encouragement Award in the District-level Mathematics Competition for Outstanding Students - 2009

– *Math score*: 18/20

– *Venue*: My Hung High School, My Loc, Nam Dinh