

VU VAN THAI

Phone: +84 865-735-024 ·

Email: thaivuit19120@gmail.com

LinkedIn: <https://www.linkedin.com/in/thaivv190120/>

SUMMARY

AI Engineer with more 3 years of experience in developing and deploying AI solutions, specializing in using TensorFlow and PyTorch to build NLP and Computer Vision models, and more recently utilizing LLM for chatbot, recommendation, and search tasks with local models (Qwen, Llama,...) and Cloud model. Experienced in deploying AI models through ONNX and developing APIs for AI Backend. Always eager to explore and learn new technologies, contribute ideas, and devise solutions to encountered challenges.

Languages: Python, PHP, JS

Frameworks/Platforms used:

- OpenAI, Hugging face, vLLM, SGLang.
- Pytorch, Scikit K-learn, Keras, Tensorflow.
- OpenCV, Pillow, Numpy, Pandas.
- Docker, MIFlow.
- LangChain, LlamaIndex, LangGraph
- Ray serve, Triton server, Flask, Django, Fastapi.
- Serving LLM: vLLM, SGLang

The technology of AI:

- Gen AI:
 - Model: GPT-4o, Llama 3.2, ...
 - Prompting, RAG, Knowledge Graph, GraphRAG
 - Database: Milvus, Neo4j, MongoDB
 - AI Agent, LangGraph
- Fundamental knowledge of Machine Learning: Supervised/Unsupervised Learning, Cost function, Gradient Descent, Overfitting, Regularization, Cross Validation, ...
- Basic Machine Learning methods: Linear Regression, Logistic Regression, K Nearest Neighbors, Neural Network, Backpropagation, SVM, ...
- Knowledge about popular CNN (Convolutional Neural Network) architectures: VGG, Inception, ResNet, MobileNet, DenseNet.
- Object detection, image segmentation models: YOLO, U-Net, U2Net, Deeplabv3.
- NLP: RNN, LSTM, Transformer, Bert

PROFESSIONAL EXPERIENCE

BRAVO

June 2025 - Present

AI Engineer

Project: Chatbot AI Agent

- Description: Build an API service, solution for chatbot tasks such as QA on documents, Search Agent.
- Team Size: 2
- Skills: Python, vLLM, SgLang, LangGraph, Qwen 3, Local model LLM
- Domain: LLM

March 2024 - June 2025

AIC (FPT AI CENTER)

AI Engineer

Project: Chatbot Team

- Description: Build an API service for chatbot in MsTeam
- Team Size: 5
- Skills: Python, GPT- 4o, OpenAPI, LangGraph, Milvus
- Domain: LLM

Project: Extract Information

- Description: Build an API service for extracting structured information from unstructured text.
- Team Size: 2
- Skills: Python, GPT-3.5, 4o, OpenAPI
- Domain: LLM

Project: Object Counting Web Interface with LandingAI

- Description: Develop counting model using LandingAI, a web interface to display and manage object counting results.
- Team Size: 2
- Skills: Python, LandingAI, Javascript.
- Domain: CV

Project: Object Healthcare Information Extraction with RAG and Knowledge Graphs

- Description: Develop an API service for extracting healthcare-related information from unstructured text, leveraging RAG (Retrieval-Augmented Generation) and GraphRAG techniques to build and query knowledge graphs. Integrate Milvus for vector search and Neo4j for managing relationships between extracted entities.
- Team Size: 2
- Skills: Python, Milvus, Neo4j, LLM Model, OpenAPI, Knowledge Graphs, GraphRAG
- Domain: LLM

Kaopiz Solution
AI Engineer

March 2022 - March 2024

Project: ChatBot

- Description: Create a bot to respond to information related to the provided documents
- Team Size: 2
- Skills: Python, LangChain, LLM model, OpenAPI
- Domain: NLP

Project: Count Wiring

- Description: Count the number of wires in the picture
- Team Size: 1
- Skills: Python, YOLOvx, ONNX
- Domain: Computer Vision

Project: Detect Number

- Description: Detect the sequence of numbers in the image
- Team Size: 1
- Skills: Python, YOLOvx, ONNX, U2Net
- Domain: Computer Vision

Project: Facial analysis

- Description: Detect individuals in the video and recognize and classify them based on the provided database
- Team Size: 2
- Skills: Python, YOLOvxFace, ONNX, InsightFace
- Domain: Computer Vision

Project: Character Classification

- Description: Classify handwritten Japanese characters, classify correct and incorrect character writing database
- Team Size: 2
- Skills: Python, Mobilenetv3, Annoy search
- Domain: Computer Vision

Project: Hand Game

- Description: Touch the ball to get scores and complete the mission
- Team Size: 3
- Skills: Python, OpenCV, Mediapipe hand
- Domain: Computer Vision

Personal**Project: Fraud detection - MLOPS competition**

- Description: Implement the model, write an API that handles multiple client responses
- Team Size: 2
- Skills: Python, FastAPI, Ray serve, Nginx, Redis, Docker, Triton server
- Domain: MLOPS

Project: OCR - OCR vietnamese handwriting competition

- Description: Read handwritten Vietnamese
- Team Size: 2
- Skills: Python, Vietocr
- Domain: OCR

Project: Photo Editing App

- Description: Developing a simple photo editing application with an additional image enhancement and image inpainting task
- Team Size: 1
- Skills: Python, Pyqt5, LaMa, ESRGan
- Domain: Computer Vision

EDUCATION

[May 2023] HaNoi University of Science and Technologies (Vietnam) Bachelor's degree