NGUYEN VAN TU

AI Engineer Intern

0964511270 nvtu2305@gmail.com github.com/Tuprott991 linkedin.com/in/nguyen-van-tu

Education

VNUHCM - University of Science **2022 - 2026**

Bachelor of Information of Technology — Major: Information Technology *Ho Chi Minh City*

- **Courseworks**: Machine Learning, Multivariate Data Analysis, AI Fundamentals, Data Structures & Algorithms, Databases (SQL), Calculus, Linear Algebra, Probability & Statistics, Digital Image & Video Processing, IoT.
- **Foreign language:** IELTS 6.0

Honors & Awards

- **Champions** of Web3 & AI Ideathon (2025 - National Hackathon) (among 450+ teams)
- **Finalist** in AI Challenge HCMC 2024
- **Champions** of Line Follower Robot competition HCMUS (F-RACE) 2024
- **Consolation Prize** in The National Youth Informatics Competition 2021
- **Awarded** for outstanding contributions to Youth Union and Student Association activities in 2024
  - *Role: **Vice Head** of M.A.T Communications Committee, University of Science – VNUHCM*

Publication & Research

Tu et al. **An Interactive System For Visual Data Retrieval From Multimodal Input**. *The International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (IUKM 2025).*

- Work under Dr.Dang Bui from Sep to Nov 2024 to develop a conversational and multimodal video event search event.
- Research and apply AI models like CLIP, Whisper, PaddleOCR, TransnetV2, and **GPT-4o API**, allowing users to **semantically retrieve** visual data using natural language, image, and voice.

Projects

Educhain | 5 members | *github.com/Tuprott991/Educhain-AI* **Jan – Mar 2025**

- **Role**: Team lead / AI Engineer
- **Description:** Built a personalized learning platform integrating **LLMs** and **prompt engineering** to build agents for **lightRAG-chatbot**, **RAG-based** quiz & study guide generation, **user knowledge profiling**, and **file processing**.
- Utilized **FastAPI** for fast and efficient request handling backend. Leveraged **LangChain** to build agents with optimized retrieval and LLM-database interactions.
- Implemented file understanding with **Azure Document Intelligence** to extract structured insights, and used **Azure Speech** for voice-based interaction and transcription.
- Integrated **lightRAG** to embed and retrieve personalized content from user-uploaded documents.
- Managed all data with **PostgreSQL**, leveraging **pgvector** for vector storage and **Apache AGE** for knowledge graphs
- Fine-tuned **Qwen2.5-7B** using the **LoRA** method and deployed it with **vLLM** for optimized inference.
- **Techs:** Python, FastAPI, LLMs, LangChain, Azure, PostgreSQL, LoRA, vLLM, lightRAG, Generative AI, ReactJS

Multimodal Video Retrieval | 5 members | *github.com/Tuprott991/AIthena-C* **Aug – Oct 2024**

- **Role:** Team lead / AI Engineer
- **Description:** Developed a **AI-Driven semantic search** videos event system based on natural language, scene, voice, OCR, and other metadata.
- Reduced search latency by **30%** through optimized keyframe extraction using **OpenCV** and **TransNetV2** model.
- Applied **CLIP ViT L/14** and **BLIP-2** for text and image embedding generation, enabling efficient vector search with **FAISS**. Leveraged **Jaro-Winkler** algorithm and **PaddleOCR** for text in image retrieval
- Enhanced retrieval performance with **multimodal inputs** (text, voice, prompts, objects); integrated **GPT-4o** for query refinement and visual question answering, and employed **Whisper** for accurate real-time speech-to-text conversion.
- **Techs:** Python, Flask, CLIP, HuggingFace, NLP, Numpy, Transformer, OpenAI, Semantic Search, ReactJS

BoneDiseaseVQA | *github.com/Tuprott991/BoneDiseaseVQA-2* **March - April 2025**

- **Description:** I developed and trained a multimodal transformer model (natural language & image) for bone disease visual question answering, which achieved **90.50%** accuracy for validate dataset
- Implemented a novel semi-open architecture leveraging the **viHealthBERT** (language model for medical) and **Vision Transformer** to extract features and apply **cross-multimodal attention network** mechanism for feature fusion stage. Using a **learnable answer embedding** for the Transformer decoder increases precision by approximately **12%** compared to the closed architecture.
- Leveraged **Gradio** for web interface and **GPT-4** to enhance the diagnose information
- **Techs:** Python, Pytorch, BERT, Deep Learning, Transformers, Huggingface, torchvision, Gradio, OpenAI

Vision Language Object Tracking | *github.com/Tuprott991/Object-tracking-Natural-Language* **Dec 2024**

- **Description:** Designed an object tracking pipeline that integrates natural language inputs for enhanced object identification and tracking across video frames. Leveraged vision-language models for understanding scene context and tracking targets based on user queries.
- Improved object tracking precision by 20% with a custom-trained **YOLOv9** model on a Vietnamese vehicle dataset.
- Implemented a hybrid approach combining candidates matching with **CLIP** for semantic query understanding and **DeepSORT** for robust multi-object tracking.
- **Techs:** Python, Ultralytics, YOLO, CLIP, DeepSORT, Googletrans

Technical Skills

- **Languages**: Python (Proficient), C++ (Good)
- **Libraries/Frameworks**: FastAPI, LangChain, LLMs, PyTorch, Transformers, Scikit-learn, Pandas, Matplotlib, Milvus
- **Developer Tools & Technology**: Azure, Git, Docker, PostgreSQL

Certificates

Medical Image Processing Workshop - *Masayuki Fukuzawa, Dr Eng. Faculty of Information and Human Sciences, Kyoto Institute of Technology*

- Participated in a workshop focusing on medical image processing techniques.