Hiep Tan To

Ho Chi Minh City – Viet Nam
+84 919703340 • tanhiep2012003@gmail.com

Bachelor's student in Computer Science at Ho Chi Minh University of Science, specializing in Artificial Intelligence and Machine Learning. Hands-on experience with deep learning, NLP, and generative AI through research projects, publications, and applied system development. Strong foundation in data engineering, algorithms, and model deployment.

Education

Master of Computer Science — In Progress
University of Science - VNUHCM — *2025*

Bachelor of Technology with Honors — 9.41/10.0
Computer Science and Engineering, University of Science - VNUHCM — *2021-2025*

Thesis — 10.0/10.0
Generative AI-Based Virtual KOL System — *2025*

Language Skills — IELTS 6.0, VSTEP B2

Internships and Experience

Associate Machine Learning Engineer
DeepCDR Team, OPSWAT — *Jul. 2025 - Present*

- Developed AI for file type detection, optimizing accuracy from 85% to 92%.
- Designed and trained classification models on large-scale file datasets, improving detection accuracy.
- Evaluated model performance using precision, recall, and F1-score metrics across various file types.
- Collaborated with cybersecurity experts to align ML models with threat detection requirements.
- Optimized model inference for integration into production pipeline with minimal latency.
- Conducted research and benchmarking on traditional and deep learning approaches for file analysis.

Contractor Software Engineering
Integration Team, Treasure Data — *Apr 2024 –Jun 2025*

Built and optimized data pipelines and connectors for ETL workflows (Embulk, REST APIs).
Developed an AI document summarization agent, improving summarization speed by 30%.
Collaborated with cross-functional teams to design scalable and reliable data integration.
Enhanced CLI manual testing tools to support QA processes effectively.

Undergraduate Thesis: Generative AI-Based Virtual KOL System
Guide: Dr. Le Trung Nghia, Assoc. Prof. Tran Minh Triet, University Of Science — *Sept. 2025*

Developed a generative AI system merging identity with varied styles, preserving semantics.
Built a modular, extensible architecture for easy integration of new algorithms and pipelines.

Undergraduate Research
Software Engineering Laboratory, University of Science — *Sept. 2023 – Present*

Researched Generative AI techniques such as Stable Diffusion, GANs, and RAG.
Contributed to successful publications, enhancing understanding of adversarial robustness and model defense strategies for NLP systems.

Technical Skills

- **Machine Learning / Deep Learning:** PyTorch, TensorFlow, HuggingFace Transformers, Scikit-learn, OpenCV
- **Generative AI:** Stable Diffusion, GANs, Retrieval-Augmented Generation (RAG), LangChain
- **Data Engineering / MLOps:** Embulk, Airflow, SQL, MongoDB, AWS (S3, EC2, Lambda)
- **Programming:** Python, C++, Java
- **Other Tools:** Streamlit, Docker, Git, Overleaf
- **Soft Skills:** Strong presentation, collaboration, and problem-solving

Scholastic Achievements

- Awarded **Outstanding Student in Scientific Research Achievement** (2023)
- Honored in the **Top 100 Outstanding Students of Vietnam National University** (2023)
- Ranked **1st Best Student of Regular Program, Faculty of IT** (2021)
- Selected among **Top 5 Students** of the Faculty of IT (2021)
- Won **Third Prize** – Provincial Competition for Excellent Students in Informatics (2020)

Key Academic Projects

Question Answering System with Transformer Models (Python, PyTorch, Hugging Face, Streamlit)
Role: Researcher | Course Project *2024*
Built a QA system using BERT and RoBERTa fine-tuned on SQuAD, Zalo Challenge, and custom Vietnamese datasets.
Collected, cleaned, and merged 26k+ QA samples; conducted EDA to ensure balanced distributions.
Integrated Google Custom Search, improving query response time by 25%.
Explored RAG for knowledge-intensive QA tasks with over 100 research papers.
GitHub: Question-Answering

Adversarial Robustness in NLP Models (Python, PyTorch, Hugging Face, TextAttack)
Role: Researcher | Course Project *2024*
Explored FGSM, PGD, C&W, TextFooler adversarial attacks on text classification models.
Implemented adversarial training strategy, boosting accuracy from 0% to 95.24% on adversarial test cases.
Conducted experiments on AG News dataset, achieving baseline accuracy of 75.12% (F1-score 74.58).
GitHub: Adversarial-Attacks

Budget-aware Road Semantic Segmentation in Unseen Foggy Scenes
Team Size: 3 *Aug. 2023*
Proposed efficient methods to improve autonomous driving in foggy conditions.
Published and presented at **RIVF'23 International Conference**.
Paper link: IEEE Xplore

Streamlining Virtual KOL Generation Through Modular Generative AI Architecture
Team Size: 2 *Aug. 2025*
Proposed a scalable generative AI system for Virtual KOL development (GenKOL).
Accepted at **ACM MultiMedia 25**.

Additional Certificates

- Completed **FPT AI Compass**
- Completed **AWS Academy Cloud Foundations**
- Completed **Intro to Machine Learning (Kaggle)**
- Completed **Unsupervised Learning, Recommenders, Reinforcement Learning (Coursera)**
- Completed **Advanced Learning Algorithms (Coursera)**
- Completed **Machine Learning Explainability (Kaggle)**
- Completed **Machine Learning Specialization (Coursera)**
- Completed **Ultimate AWS Certified AI Practitioner AIF-C01**
- Completed **AWS Certified Developer Associate**