

Tuyêñ Vū

AI Research Engineer

📞 0896023142
✉️ tuyenvt455@gmail.com
🌐 github.com/SuZeAI

Professional Summary

- Responsible and passionate AI Research Engineer with strong expertise in **Machine Learning, Deep Learning, and Large Language Models (LLMs)**. Proactive in self-learning, exploring cutting-edge research, and applying innovative solutions to real-world problems. Experienced in designing and deploying production-level AI systems, building multi-agent frameworks, and leading research projects. Committed to continuous growth and aspiring to contribute as a Tech Lead in AI Research, leveraging technical skills and leadership to drive impactful outcomes for the organization.

Education

2021 – 2026 **IT - PTIT - Posts and Telecommunications Institute of Technology, Hanoi**

- GPA: 3.34/4.0
- I served as an advisor for the PAYT Club and was a member of both IEC & Naver Lab.

Experience

May 2025 – **AI Research Engineer, HBLAB**
Now

- **Extractor Date:** Developed a **Japanese medical-date extraction agent** using the **OpenAI API**, achieving **87% accuracy** through domain-specific data processing and evaluation.
- **GenQuestion (Testcase Generation Agent):** Built **automated test and QA question generation agents** based on project specifications, significantly reducing QA workload; evaluated **9/10 by QA Leads**.
- **AI-Project (Multi-Role Agent System):** Developed an **AI-driven project management system** with agents acting as PM, Tester, and other roles; integrated with **GitLab, Google Drive, and additional tools via MCP** to automate project workflows.
- **Agent Dependency:** Researched and developed an agent to **update source code dependency information** when versions change, constructing **code graphs** and applying **context engineering techniques**.

Dec 2023 – **AI Research Engineer, Viettel Cyber Security**
May 2025

- **Intelligent Code Vulnerability Detection:** Conducted problem surveys, reviewed relevant research, collected and analyzed datasets, and built models using **AST, CFG, and DFG graphs**. Leveraged **PyTorch, PyTorch Lightning, Scikit-learn, Torch Geometric**, configured via **Hydra**, and tested with **Pytest**.
- **Chatbot Development & Model Finetuning:** Researched **Knowledge Distillation (KD)** from large models such as GPT and Gemini; explored synthetic data generation, model construction, and benchmarking. Finetuned **LLaMA 7B, 8B, 17B** on **A100 GPUs** (SFT, DPO, RLHF), including experiments with **Data Parallel (DP) training**.
- **MLOps Experience (Airflow, MLflow, MinIO):** Built scalable **LLM training pipelines** using **Airflow**, tracked experiments via **MLflow**, and managed data/model storage with **MinIO**.
- **Customer Status Alert System (Production):** Performed data analysis, cleaning, and **SHAP-based feature filtering**; trained predictive models and **deployed the system to production**.
- **LLM Training Configuration (DeepSpeed, Unsloth):** Customized training workflows using **DeepSpeed** and **Unsloth**, streamlining processes into single YAML configuration files for more efficient **LLM training**.
- **Other Experiences:** Self-Learning & Problem-Solving; Pretrained Language Models (PLMs); Viettel AIHub Competition.

Sep 2023 – **AI Assistant, PTIT IEC Lab & Naver Lab**

Now

- **Malware Detection in Android Applications:** Co-led a project leveraging **Graph Neural Networks (GNNs)** with fine-tuned **Word2Vec embeddings** to model function names in call graphs; exploited shared naming conventions in malware and achieved **98.8% F1-score**. ([link github](#))
- **Graph Fusion Detection in Smart Contracts:** Researched **vulnerabilities in smart contracts** with a focus on error classification; developed **graph-based models using batch training** to capture feature interactions. Paper currently under review at a **Q1 Journal ICT express**. ([link github](#))
- **Utilization of Advanced Frameworks:** Applied **PyTorch, Torch Geometric, Deep Graph Library (DGL), Google Colab**, and **Kaggle** for model training, evaluation, and analysis.
- **Other Experiences:** Participation in Knowledge Seminars, NLP Knowledge (CS224N, CS224U); Support for GCN-Computer Vision Project.

Sep 2021 – **Personal Projects**

Now

- **ADM Permissions 2024 Project: Malware detection in APK files** leveraging **BERT & RoBERTa** for feature extraction; moved beyond traditional graph-based approaches to create features suitable for conventional model recognition. ([link github](#))

- **Contrastive Learning for ADM:** Applied **contrastive learning** + **Graph Neural Networks (GNNs)** for classification; achieved **98% accuracy.** ([link](#) [github](#))
- **Agent Gen Slide Template (LLM Pipeline):** Developed an **LLM-based pipeline** with **agent planning** and template generation; planned enhancements include image retrieval. ([link](#) [github](#))
- **Text Normalization for TTS:** Fine-tuned **Qwen models** to normalize dates/numbers for text-to-speech; incorporated preprocessing and postprocessing techniques; achieved **80% accuracy.** ([link](#) [github](#))
- **Job Recommendation System:** Analyzed LinkedIn dataset to extract features and build a vector database using **Sentence Transformer L6v2 embeddings** + **FAISS** for efficient retrieval and recommendation. ([FE](#), [BE](#))
- **CrewAI Contributor:** Contributed to the CrewAI project via [pull request #3856](#); improved **execution stability** and added **crew streaming support.**
- **Other Projects:** NER with BERT; Comment Classification (Mamba); AI Tetris Project; Hand-gesture Controlled Game.

Achievements

Research & Publications (Highlighted)

- **Q1 Journal Submission — GraphFusionVulDetect:** Second author of a manuscript under review at **ICT Express (Q1).** ([link](#))
- **FAIR Publication — SlideGen Paper:** Successfully published at **FAIR.**
- **Android Malware Paper:** Published at PTIT.

AI Competitions (Top Rankings)

- **Olympic AI Northern Region:** **Top 1 at PTIT • Top 5 Northern Region • Top 16 Vietnam.**
- **3rd Prize — AI Contest PTIT 2025.**
- Prize winner — AI Contest PTIT 2022.
- Selected among the **Top 100 at Thucchien.ai.**

Academic Achievements

- **PTIT Scholarship** for first-semester GPA of **3.6/4 (Excellent).**
- Recipient of the **Australia & Vietnam Scholarship.**

Certifications & Leadership

- Samsung Python for AI Certification (2022).
- Google "Crash Course on Python" Certification.
- Certified organizing committee member (JAIST & PTIT).
- **Mentor** for Samsung for AI program (PTIT × Samsung).

Mathematics Awards (Earlier Years)

- Encouragement Prize in Mathematics at the district level (7th grade).
- 1st Prize in Mathematics at the district level (8th grade).

- 2nd Prize in Mathematics at the provincial level (12th grade).

Skills

- **Programming:** Python, C, C++, Java
- **Machine Learning & Deep Learning:** PyTorch, TensorFlow, Keras, Scikit-learn; DeepSpeed, Unslloth; Hugging Face Transformers
- **LLM & NLP:** LLaMA, Mistral, RoBERTa, Sentence-BERT; Text Generation, RAG, Seq2Seq, Contextual Understanding
- **Data Science & Visualization:** NumPy, Pandas, Matplotlib, Seaborn; Jupyter Notebook
- **Agent Frameworks:** crewAI, ADK, LangChain
- **DevOps & MLOps:** Git, Linux, Docker; Airflow, MLflow, MinIO, Mysql
- **Languages:** English, Vietnamese