Thang Le Viet

AI Engineer

✉ levietthang0512@outlook.com | 📞 (+84) 989394637 | 🔗 thanglv | 🐙 toreleon | 📍 Ho Chi Minh City, Viet Nam

ABOUT ME

I'm an AI researcher with a strong background in computer science and a deep interest in history, innovation, and solving complex problems. As an INTJ-A personality type, I naturally thrive in analytical environments, where I enjoy exploring deep theories and historical perspectives that fuel creative thinking.

With a startup spirit and firsthand experience as a co-founder, I bring a unique blend of technical expertise, strategic insight, and a passion for bringing complex ideas into real-world solutions. My strengths lie in analytical problem-solving, long-term vision, and the ability to bridge theory with impactful execution.

WORKING EXPERIENCE

Co-founder and Vibe Coding Engineer at ▲ *Smart Hirink*   *Jun 2024 — Present*

- **Smart Hirink Platform:** Developed a comprehensive platform automating candidate screening, interviewing, and evaluation processes, enhancing talent acquisition.
- **Conversational AI:** Built an end-to-end conversational AI solution integrating open-source TTS, STT, and LLM technologies for low-latency AI-driven interviews.
- **Multi-agent UX Research:** Conducted research to transform traditional SaaS products into multi-agent platforms, preparing for next-gen software interaction.
- **Achievement:** Successfully acquired the first client in March 2025.

AI Engineer III at 🗔 *FPT AI Center - AI4SE Laboratory*   *Jan 2024 — Present*

- **Agent Platform Development:** Developed a multi-agent system platform for automated code analysis, code review, and bug localization.
- **Mainframe Code Summarization:** Designed algorithms for clustering and summarizing mainframe codebases, improving documentation and comprehension.
- **Code Knowledge Graph:** Researched software knowledge graphs for enhanced information retrieval and question-answering in code repositories.
- **xMainframe Fine-Tuning:** Utilized LoRA for efficient fine-tuning of the xMainframe model, optimizing its mainframe-specific capabilities.
- **Mainframe Studio Backend:** Built backend APIs enabling detailed mainframe code analysis for modernization efforts.

AI Research Engineer at 🔲 *Resonance Technology*   *Apr 2022 — Oct 2023*

- **Prompt Engineering:** Boosted AI platform usage by 83.7% with innovative prompt techniques.
- **Keyword Clustering Optimization:** Developed high-performance SERP-based clustering algorithms, achieving 40% speed improvements.
- **LLM Prompt Automation:** Created automated GPT-driven workflows for prompt optimization.
- **Model Fine-Tuning:** Applied LoRA and QLoRA for efficient fine-tuning of language models.
- **Keyword Intent Classification:** Developed intent classification systems using semantic analysis and SERP features.

Data Scientist at 🟡 *Be Group*   *Jul 2021 — Apr 2022*

- **Search Optimization:** Improved search API accuracy, increasing top recommendation click rates from 55% to 70%.
- **Recommendation Engine:** Implemented personalized Bayesian ranking systems, increasing recommendation clicks by 8%.
- **Support Automation:** Developed a Rasa-based conversational AI platform, significantly automating customer support.
- **Semantic Search:** Built a FastText-based semantic search engine, improving contextual search effectiveness.

RESEARCH EXPERIENCE

Research Assistant at DS@UIT Laboratory, *University of Information Technology*   Feb 2020 — Feb 2022

- **COVIDROP Dataset:** Created the first Vietnamese numerical reasoning dataset based on DROP, enhancing Vietnamese NLP research resources.

- **Annotation Tools:** Developed guidelines and tools for accurate data labeling.

- **Model Evaluation:** Implemented and evaluated NAQANet and NumNet models to assess dataset usability.

SKILLS

Languages & Tools:	Python, Cython, C/C++, Docker, Git, Copilot, Cursor, FastAPI, WebSocket, Next.js
Data Science Libraries:	NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-learn, PyTorch, XGBoost, LightGBM, OR-Tools, PySpark, Kafka, MLflow, Wandb
NLP & GenAI:	OpenAI API, Gemini API, Ollama, Huggingface, LangChain, LangGraph, Autogen
Cloud Platforms:	Google Cloud Platform, Microsoft Azure
Research:	Natural Language Processing, Large Language Models, Deep Learning, Operations Research, Evolutionary Algorithms, Reinforcement Learning
Languages:	Vietnamese (Native), English (Advanced)

EDUCATION

Master of Science (M.Sc.) in Computer Science, *University of Information Technology*   Sep 2023 – Present
GPA: 3.7/4.0

Bachelor of Science (B.Sc.) in Data Science, *University of Information Technology*   Sep 2018 – Jan 2022
GPA: 3.6/4.0 (Top 1 in Major)
Thesis: "Numerical Reasoning Reading Comprehension on Vietnamese COVID-19 News: Task, Corpus, and Challenges"

PUBLICATIONS

Van Nguyen, K., Le, T.V. Do, T.PP. Numerical reasoning reading comprehension on Vietnamese COVID-19 news: task, corpus, and challenges. Neural Comput Applic 36, 14053–14073 (2024). https://doi.org/10.1007/s00521-024-09744-5