LE THANH THIEN

(+84) 96 194 13 29 | thienlt.7@gmail.com | thanhthienle.github.io

EDUCATION

Hanoi University of Science and Technology, Hanoi — Aug 2019 – Sep 2023

Bachelor of Science in Computer Science

- **Thesis** — ***Parameter-efficient Continual Relation Extraction***: combine generative replay with a novel multi-objective optimization method for competitive Continual Relation Extraction **(GPA 4.0)**
- **Achievements** — Talent Scholarship 2019 (merit-based, value of 1,300 USD)
- **Modules** — Artificial Intelligence (**GPA 4.0**), Computer Architecture **(GPA 4.0),** Data Science (**GPA 4.0**), Evolutionary Computing **(GPA 4.0)**, Information Retrieval **(GPA 4.0)**, Linear Algebra (**GPA 4.0**), Machine Learning and Data Mining (**GPA 4.0**), Programming Techniques (**GPA 4.0**)
- **Projects** — ***Volatility Prediction for Vietnam Stock Market***: benchmark ARCH, GARCH, SVM, and Multi-Layer Perceptron on VNINDEX and VN30INDEX
  Data Platform for Financial Market: build big data pipeline for the financial market
  Bird Classification: train ResNet models, resulting in accuracy of over 90%
  Image Captioning: build and train *Show, Attend and Tell* model using CNNs and LSTMs

VNU.HCM High School for the Gifted, Ho Chi Minh City — Aug 2016 – May 2019

*High School Diploma (**GPA 9.7**)*

- **Achievements** — Odon Vallet Scholarship 2019, from Rencontres Du Vietnam

EXPERIENCE

VinAI Research, Hanoi | **AI Research Resident** — Feb 2023 – Present

Top 20 global company for leading AI research in 2022

Supervisor: Assoc. Prof. Thien Huu Nguyen (University of Oregon)

- Large Language Models' trustworthy, safety, reasoning - 1 published paper
- Continual learning in information extraction - 4 published papers
- Explored cross-lingual information extraction
- Collected and preprocessed data to train Large Language Model for Vietnamese language
- Trained and benchmarked PhoWhisper, state-of-the-art Speech Recognition model for Vietnamese language

VND Credit, Hanoi | **AI Engineer** — Jul 2022 – Feb 2023

P2P lending platform

- Wrote, trained, and deployed API for a job-matching model as part of a job recommendation system
- Gathered and standardized data on job opportunities for training AI models extracting useful information from job description, achieving precision and recall of over 90%
- Collaborated with 6 engineers and 3 interns to build graph neural networks for credit scoring from mobile-usage data

HUST Data Science Lab (BKAI), Hanoi | **Research Student** — Apr 2021 – Aug 2023

Affiliated with NAVER – most-used search portal in South Korea

- Researched deep generative models (e.g. GANs, Diffusion Models) and theoretical generalization for deep neural networks, led by Assoc. Prof. Than Quang Khoat
- Conducted project "Deep learning for imbalanced data", designing architectures and training paradigms for deep neural networks to enhance performances on skewed and long-tailed data

PUBLICATIONS

Asterisk (*) denotes equal contribution

Realistic Evaluation of Toxicity in Large Language Models

Tinh Son Luong*, *Thanh-Thien Le**, Linh Van Ngo and Thien Huu Nguyen, *ACL 2024*

ToVo: Toxicity Taxonomy via Voting

Tinh Son Luong*, *Thanh-Thien Le**, Thang Viet Doan*, Linh Ngo Van, Thien Huu Nguyen and Diep Thi-Ngoc Nguyen, *under review*

Continual Relation Extraction via Sequential Multi-Task Learning

*Thanh-Thien Le**, Manh Nguyen*, Tung Thanh Nguyen*, Linh Van Ngo and Thien Huu Nguyen, *AAAI 2024*

SharpSeq: Empowering Continual Event Detection through Sharpness-Aware Sequential-task Learning
*Thanh-Thien Le**, Viet Dao*, Linh Van Nguyen*, Thi-Nhung Nguyen, Linh Van Ngo and Thien Huu Nguyen, *NAACL 2024*

PhoWhisper: Automatic Speech Recognition for Vietnamese
Thanh-Thien Le, Linh The Nguyen, Dat Quoc Nguyen, *Tiny Paper @ ICLR 2024*

Lifelong Event Detection via Optimal Transport
Viet Dao*, Van-Cuong Pham*, Quyen Tran*, *Thanh-Thien Le*, Linh Ngo Van, Thien Huu, *EMNLP 2024*

Few-Shot, No Problem: Descriptive Continual Relation Extraction
Thanh Nguyen*, Anh Duc Le*, Quyen Tran*, *Thanh-Thien Le**, Linh Ngo Van, Thien Huu, *under review*

Adaptive Prompting for Continual Relation Extraction: A Within-Task Variance Perspective
Minh Le, Tien Ngoc Luu, An Nguyen The, *Thanh-Thien Le*, Thien Trang Nguyen Vu, Tung Thanh Nguyen, Linh Van Ngo and Thien Huu Nguyen, *under review*

ACHIEVEMENTS

Physics Unlimited Explorer Competition (PUEC) | **Honorable Mention** 2018
International Physics competition originated from Princeton University
- Track: Statistical Mechanics

Vietnam Physics Olympiad (VPhO) | **Second Prize** 2019
Most prestigious Physics competition for high school students in Vietnam.
- Top < 0.05% of nationwide high-school students in 2018-2019 school year

SOICT Hackathon 2023 | **Finalist** 2023
AI Development competition for nationwide students
- Track: Spoken Language Understanding

SKILLS

- **Programming** Python (NumPy, pandas, Matplotlib, PyTorch, PySpark, Hugging Face Transformers), C++
- **Machine Learning** Linear Regression, K-Means, SVMs, KNN, Probabilistic Models, DNNs, GNNs, LLMs
- **Mathematics** Calculus, Linear Algebra, Probability and Statistics, Optimization, Discrete Mathematics
- **Languages** English (IELTS 8.0), Vietnamese (Native)

REFERENCES

Assoc. Prof. Thien Huu Nguyen thienn@uoregon.edu | https://ix.cs.uoregon.edu/~thien/
Linh Van Ngo, Ph.D. linhnv@soict.hust.edu.vn | https://users.soict.hust.edu.vn/linhnv