

# Khôi Lê

✉ minhkhoile2001nsg@gmail.com    🌐 tokisakikurumi2001.github.io    🎓 Google Scholar

## Research interests

My research interests are in large language models and their applications, including efficient data sampling, cross-lingual transfer learning and low-resource adaptation.

## Education

**University of Technology, VNUHCM**

*Bachelor of Computer Science (Advanced Program)*

**HCMC, Vietnam**

*Aug 2019 – Nov 2023*

- **GPA:** 3.81/4.00 (Rank: 3/343)
- **Thesis:** LAMPAT: Low-rank Adaptation for Multilingual Paraphrasing using Adversarial Training
- **Thesis Advisor:** [Assoc. Prof. Quan Thanh Tho](#)

## Work experience

**AI Resident**

*VinAI Research*

**HCMC, Vietnam**

*Jul 2022 – Jul 2025*

- **Research Topics:** Multilingual pretrained language models, low-resource adaptation

## Publications

**UniBridge: A Unified Approach to Cross-Lingual Transfer Learning for Low-Resource Languages**

May 2024

*Khôi Lê\**, Trinh Pham\*, and Anh-Tuan Luu

[Proceedings of Association for Computational Linguistics \(ACL\) 2024](#)

**LAMPAT: Low-Rank Adaption for Multilingual Paraphrasing Using Adversarial Training**

Dec 2023

*Khôi Lê\**, Trinh Pham\*, Tho Quan and Anh-Tuan Luu

[Proceedings of Association for the Advancement of Artificial Intelligence \(AAAI\) 2024](#)

## Projects

**Cross-lingual transfer learning**

[Github](#) [🔗](#)

- Develop a system for low-resource languages modeling that leverages the knowledge encoded in multilingual pre-trained language models.
- The system can strategically find a suitable vocabulary size for each language and initialize the new embedding from both semantic and lexical knowledge.
- The system can incorporate different languages at inference time to fully benefit from the knowledge shared between group of similar languages.

**Unsupervised multilingual paraphrasing system**

[Github](#) [🔗](#)

- Develop a system that could paraphrase sentences in multi-language settings.
- The system is developed without the aid of parallel corpora, only take advantages of the rich and resourceful monolingual corpora.
- The system enhances the quality of paraphrase by using virtual label training - a branch of adversarial training - to generate paraphrase with various linguistic features.

## Skills

**Programming:** Python, JavaScript, Ruby

**Language:** Vietnamese, English (IELTS 7.5 : Reading: 8.0, Listening: 8.0, Writing: 7.5, Speaking: 6.5)

**Technologies:** Pytorch, Huggingface, Numpy, Pandas, FastAPI

## Reference

---

- [Dr. Quan Thanh Tho](#)  
Associate Professor  
Dean, Faculty of Computer Science and Engineering  
Ho Chi Minh City University of Technology (HCMUT)  
Email: qttho@hcmut.edu.vn
- [Dr. Luu Anh Tuan](#)  
Assistant Professor  
College of Computing and Data Science  
Nanyang Technological University (NTU)  
Email: anhtuan.luu@ntu.edu.sg