

# Hieu Tran

(+84) 935 957 250 | [tthieu22.06@gmail.com](mailto:tthieu22.06@gmail.com) | [linkedin.com/in/hieutt0](https://linkedin.com/in/hieutt0) | [heraclex12.github.io](https://heraclex12.github.io)

## EXPERIENCE

---

### Machine Learning Engineer

June. 2023 – Present

*BLACKBOX AI*

*Remote (Canada)*

- Research Code Chat & Code Completion
- Scored **86.5%** on HumanEval, surpassing GPT-4 (August) which achieved 84%

### Lecturer & Reseach Collaborator

May. 2021 – Present

*VietAI-NTI (founded by Minh-Thang Luong)*

*Ho Chi Minh, Vietnam*

- Taught Advanced NLP Course in Spring & Winter 2023
- Built English-Vietnamese machine translation systems which outperform previous SOTA by 4.5% (BLEU). Having **7K+** downloads/month
- Released the first Vietnamese generative models, achieving SOTA in Summarization and Biomedical benchmarks. Having **8K+** downloads/month

### Senior AI Engineer

August. 2023 – October. 2023

*HCLTech*

*Ho Chi Minh, Vietnam*

- Initialized an efficient pipeline for fine-tuning large language models on AWS to facilitate Test Plan/Cucumber Generation.

### AI Research Engineer

August. 2021 – August. 2023

*M\_Service (MoMo)*

*Ho Chi Minh, Vietnam*

- Build Intent-based/ChatGPT-based AI chatbots for Customer Support on MoMo e-wallet platform
- Deploy the training pipeline and chatbots on different environments using Kubernetes, GCP
- Effectively train large language models (e.g., BERT, GPT-like models with billion params) on multiple GPUs

### Undergraduate Research Student

September. 2020 – February. 2022

*AISIA Research Lab*

*Ho Chi Minh, Vietnam*

- Explored the efficiency of Transformer-based models for Vietnamese Punctuation Prediction

### AI Software Engineering Intern

April. 2019 – August. 2021

*Zalo Group - VNG Corporation*

*Ho Chi Minh, Vietnam*

- Constructed RDF-triplestore Knowledge Bases for Kiki - first Vietnamese AI virtual assistant
- Proposed a lite and state-of-the-art Vietnamese Spelling Corrector. [Try it!](#)

## EDUCATION

---

### VNUHCM - University of Science

Ho Chi Minh, Vietnam

*Bachelor of Computer Science, GPA: 8.3/10*

*Aug. 2017 – Jul. 2021*

## PROJECTS

---

### Vietnamese Alpaca-LoRA | *Pytorch, Transformers, DeepSpeed, PEFT*

- Finetuned large language models (e.g., LLaMa, BLOOM) on Vietnamese instruction-following dataset using DeepSpeed and Low-Rank Adaptation technique

### HackEng | *Python, Flask, Tensorflow, Git*

- Built a standardized English test website utilizing T5 model to automatically generate reading comprehension questions and correct answers from reliable text sources

### NLP2SPARQL | *Python, Pytorch, Transformers, Fairseq*

- Employed Transformer-based Encoder-Decoder models (e.g., BERT2BERT) to translate between natural language and SPARQL query language on Question Answering over Knowledge Graphs

## HONORS AND AWARDS

---

**Outstanding Undergraduate Thesis Award**, *Undergraduate Thesis*, VNUHCM - University of Science

**First Prize**, in *MindX Grand Hackathon 2020*, MindX Technology School

**Second Place**, in *Apkathon 2019*, Aptech Computer Education

## TECHNICAL SKILLS

---

**Languages:** Python, Kotlin, SQL, SPARQL

**Developer Tools:** Git, Docker, Google Cloud Platform, MongoDB, PostgreSQL, Kubernetes, ElasticSearch, Linux

**Libraries:** Pytorch, Tensorflow, Transformers, DeepSpeed, PEFT, scikit-learn, Pandas, Numpy, Flask

## ACTIVITIES

---

**Unlocking the potential of GenAI for coding**, Google Devfest HCMC 2023, 100+ audiences

**Explore the Power of AI - Career Trends in This Industry**, USAID WISE Online Event 2023, 500+ audiences

**Shaping Your NLP Career Path in 2024**, VietAI Online Event 2023, 150+ audiences

**Challenges and Opportunities When Working In AI Industry (2022)**, Van Lang University, 100+ audiences

**MTet: Multi-domain Translation for English and Vietnamese**, talk at Google I/O Extended Hanoi 2022

**How MoMo Builds an AI Chatbot**, published on MoMo Technology Blog (Medium)

Participated in **BigScience Research Workshop**, mainly organized by HuggingFace

## PUBLICATIONS

---

Long Phan\*, Tai Dang\*, **Hieu Tran\***, Trieu H. Trinh\*, Vy Phan, Lam D. Chau, Minh-Thang Luong. 2023. Enriching Biomedical Knowledge for Low-resource Language Through Large-Scale Translation. In *EACL 2023*

Teven et al. 2022. BLOOM: A 176B-parameter Open-access Multilingual Language Model. In *ArXiv Preprint*

Laurençon et al. 2022. The BigScience ROOTS Corpus: A 1.6TB Composite Multilingual Dataset. In *Proceedings of NeurIPS 2022 Track Datasets and Benchmarks*

Chinh Ngo\*, Trieu H. Trinh\*, Long Phan\*, **Hieu Tran\***, Hieu Nguyen, Tai Dang, Minh Nguyen, Minh-Thang Luong. 2022. MTet: Multi-domain Translation for English and Vietnamese. In *ArXiv Preprint*

Long Phan, **Hieu Tran**, Hieu Nguyen and Trieu H. Trinh. 2022. ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation. In *Proceedings of NAACL SRW 2022*

Hieu Nguyen, Long Phan, James T. Anibal, Alec Peltekian and **Hieu Tran**. 2021. VieSum: How Robust Are Transformer-based Models on Vietnamese Summarization?. *ArXiv Preprint*

**Hieu Tran**, Long Phan, James T. Anibal, Binh T. Nguyen and Truong-Son Nguyen. 2021. SPBERT: An Efficient Pre-training BERT on SPARQL Queries for Question Answering over Knowledge Graphs. In *Proceedings of ICONIP 2021*, Acceptance rate ~20%

Long Phan, **Hieu Tran**, Daniel Le, Hieu Nguyen, James T. Anibal, Alec Peltekian and Yanfang Ye. 2021. CoTextT: Multi-task Learning with Code-Text Transformer. In *Proceedings of NLP4Prog @ ACL-IJCNLP 2021*

**Hieu Tran**, Cuong Viet Dinh, Long Phan and Son Nguyen. 2021. Hierarchical Transformer Encoders for Vietnamese Spelling Correction. In *Proceedings of IEA/AIE 2021*

**Hieu Tran**, Cuong Viet Dinh, Quang Hong Pham and Binh T. Nguyen. 2021. An Efficient Transformer-Based Model for Vietnamese Punctuation Prediction. In *Proceedings of IEA/AIE 2021*

Long Phan, James T. Anibal, **Hieu Tran**, Shaurya Chanana, Erol Bahadroglu, Alec Peltekian and Grégoire Altan-Bonnet. 2021. SciFive: a text-to-text transformer model for biomedical literature. *ArXiv Preprint*

**Google Scholar:** <https://scholar.google.com/citations?hl=en&user=atONCyEAAAAJ>