Hieu Tran

(+1) 240-462-6365 | tthieu22.06@gmail.com | linkedin.com/in/hieutt0 | heraclex12.github.io

EXPERIENCE

Senior Machine Learning Engineer
Jun. 2023 – Jul. 2024

BLACKBOX AI
Remote (Canada)

- BLACKBOX AI is an AI coding assistant powered by Code LLMs with over **6M users** and **2M downloads** on Visual Studio
- Research the best ranking pipeline for Code Search and Code Retrieval
- Trained CodeLlama-34B on synthetic instruction-following datasets, achieving **86.6%** on HumanEval, surpassing GPT-4 (Nov 2023) which scored 85.4% (See our report)

Lecturer & Reseach Collaborator
May. 2021 – Aug. 2024

VietAI
Ho Chi Minh, Vietnam

- Taught Foundations of Deep Learning, Advanced NLP, and LLMs courses to over **200 students**
- Worked directly with Dr. Minh-Thang Luong and Dr. Trieu H. Trinh (Google DeepMind) to build English-Vietnamese machine translation systems that outperform previous SOTA by 4.5% (BLEU). Having **10K+** downloads/month
- Released the first Vietnamese generative models, achieving SOTA in Summarization and Biomedical benchmarks. Having **8K+** downloads/month

AI Research Engineer
Aug. 2021 – Aug. 2023

M_Service (MoMo) (3rd unicorn in Vietnam)
Ho Chi Minh, Vietnam

- Built Intent-based/ChatGPT-based AI chatbots for Customer Support on MoMo #1 e-wallet platform in Vietnam
- Managed the training pipeline and chatbots on Kubernetes deployed on Google Cloud Platform
- Trained base language models that outperformed open-source Vietnamese models for our internal needs

Undergraduate Research Student
Sep. 2020 – Feb. 2022

AISIA Research Lab
Ho Chi Minh, Vietnam

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

AI Software Engineering Intern
Apr. 2019 – Aug. 2021

Zalo Group - VNG Corporation (1st unicorn in Vietnam)
Ho Chi Minh, Vietnam

- Constructed RDF-triplestore Knowledge Bases for Kiki - the first Vietnamese AI virtual assistant
- Proposed a light and state-of-the-art Vietnamese Spelling Corrector. Try it!

EDUCATION

University of Maryland
Maryland, USA

CS PhD Student in Natural Language Processing, Machine Translation
Aug. 2024 – Present

VNUHCM - University of Science
Ho Chi Minh, Vietnam

Bachelor of Computer Science
Aug. 2017 – Jul. 2021

- Achievements: Outstanding Undergraduate Thesis Award

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

Technical Skills

Languages: Python, NodeJS, 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

Talks and Activities

The Art of Generative AI, talk at Google Developer Student Clubs - IU 2024, 100+ audiences
Unlocking the potential of GenAI for coding, talk at 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, talk at 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

1. Phan Phuc*, **Hieu Tran***, Long Phan. 2024. Distillation Contrastive Decoding: Improving LLMs Reasoning with Contrastive Decoding and Distillation. *Preprint*

2. 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*

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

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

5. 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*

6. 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*

7. **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%

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

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

10. **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*

11. 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