Hieu Tran

michael.tranhieu@gmail.com | fatebreaker.github.io

Research Interests

My primary research interest is enabling LLMs to understand, generate, and reason about human language to enhance their ability to perform complex cognitive and language-related tasks. Recently, I have expanded my work to advanced topics such as Preference-based Learning, Retrieval-Augmented Generation (RAG), and Test Time Computing, with a particular emphasis on improving the reasoning capabilities, factual accuracy, and efficiency of language models in both general and domain-specific applications.

Education

University of Massachusetts, Amherst (UMass) Aug. 2022 - now
Computer Science PhD student *Massachusetts, USA*
- CPA: 3.92/4.0

Hanoi University of Science and Technology (HUST) Aug. 2014 - Aug. 2019
Engineer's Degree in Computer Science *Hanoi, Vietnam*
- CPA: 3.6/4.0 (top 1.3%), graduated with Excellent Degree

Bien Hoa High School for The Gifted Sep. 2011 - Sep. 2014
Specialized in Informatics *Hanam, Vietnam*

Research Experience

Computer Science PhD student Aug 2022 - now
BioNLP lab (https://bio-nlp.org) *Massachusetts, USA*
- Advisor: Prof. Hong Yu (google scholar)
- Research Topic: Large Language Models for Medical

Research Intern May 2024 - Aug 2024
United Imaging Intelligence (http://www.uii-ai.com) *Boston, USA*
- Supervisor: Dr. Weijing Huang (google scholar)
- Working on fact-checking of LLMs responses

AI Research Resident Nov 2019 - Jun 2022
VinAI Research (www.vinai.io) *Hanoi, Vietnam*
- Supervisor: Prof. Thien Huu Nguyen (google scholar)
- Research Topic: Information Extraction

Applied Rotation Program May 2021 - Sep 2021
VinAI Research *Hanoi, Vietnam*
- Supervisor: Dr. Duy Tin Vo (google scholar), Dr. Dat Quoc Nguyen (google scholar)
- Working on knowledge-based question answering and machine translation problem

Publications

RARE: Retrieval-Augmented Reasoning Enhancement for Large Language Models 2024
Hieu Tran, *Zonghai Yao , Zhichao Yang, Hong Yu*
Under review

SemiHVision: Enhancing Medical Multimodal Models with a Semi-Human Annotated Dataset and Fine-Tuned Instruction Generation 2024
Junda Wang, Yujan Ting, Eric Z. Chen, ***Hieu Tran****, Hong Yu, Weijing Huang, Terrence Chen*
Under review

LEAF: Learning and Evaluation Augmented by Fact-Checking to Improve Factualness in Large Language Models 2024
Hieu Tran*, Junda Wang, Yujan Ting, Weijing Huang, Terrence Chen*
Under review

BioInstruct: Instruction Tuning of Large Language Models for Biomedical Natural Language Processing 2024
Hieu Tran*, Zhichao Yang, Zonghai Yao, Hong Yu*
Journal of the American Medical Informatics Association, ocae122

Readme: Bridging medical jargon and lay understanding for patient education through data-centric nlp 2023
Zonghai Yao, Nandyala Siddharth Kantu, Guanghao Wei, ***Hieu Tran***, , Zhangqi Duan, Sunjae Kwon, Zhichao Yang, Hong Yu
Proceedings of EMNLP 2024 (Findings)

A Vietnamese-English Neural Machine Translation System 2022
Thien Hai Nguyen, Tuan-Duy H. Nguyen, Duy Phung, Duy Tran-Cong Nguyen, ***Hieu Minh Tran****, Manh Luong, Tin Duy Vo, Hung Hai Bui, Dinh Phung, Dat Quoc Nguyen*
Proceedings of InterSpeech 2022 - Show & Tell Demonstrations

Exploiting Document Structures and Cluster Consistencies for Event Coreference Resolution 2021
Hieu Tran*, Duy Phung, Thien Huu Nguyen*
Proceedings of ACL-IJCNLP 2021 (oral presentation)

Learning Cross-lingual Representations for Event Coreference Resolution with Multi-view Alignment and Optimal Transport 2021
Duy Phung, ***Hieu Tran****, Thien Huu Nguyen*
Proceedings of the first Workshop on Multilingual Representation Learning (MRL 2021) at EMNLP 2021

Vietnamese Speech-based Question Answering over Car Manuals 2021
Tin Duy Vo, Manh Tien Luong, Duong Minh Le, ***Hieu Tran****, Nhan Tri Do, Duy Nguyen, Thien Hai Nguyen, Hung Hai Bui, Dat Quoc Nguyen, Dinh Quoc Phung*
Proceedings of Data-Centric AI workshop NeurIPS 2021

The Dots Have Their Values: Exploiting the Node-Edge Connection in Graph-based Neural Models for Document-level Relation Extraction 2020
Hieu Tran*, Minh Trung Nguyen, Thien Huu Nguyen*
Proceedings of EMNLP 2020 (Findings)

Honors and Awards

The Excellence Scholarship - Level A *2015*
School of Information and Communication Technology, HUST
Each semester top 5% of students with greatest academic performance are awarded

National Excellent Student Award *2014*
Vietnam Ministry of Education and Training
Third prize in Informatics subject, National Excellent High School Student Contest

Activities

Technical Talk *August 2021*
AI Day 2021: Empowering Innovations – website: https://www.vinai.io/aiday2021
Present at AI Day 2021 - the event which welcomed technical talks from top researchers around the world and attracted over 20,000 views online

Invited speaker *October 2021*
Public Natural Language Processing workshop organized by VinAI Reseasrch.

Technical Skills

Programming: Python (proficient), Java, MATLAB (Familiar)

Libraries: Pytorch, TensorFlow, Numpy, Matplotlib, Transformers, etc.

Developer Tools: Git, Docker

Certificates

Deep Learning Specialization by deeplearning.ai
- Neural Networks and Deep Learning
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
- Structuring Machine Learning Projects
- Convolutional Neural Networks
- Sequence Models

Languages

- **Vietnamese**: Native

- **English**: Fluent
  IELTS: Overall 7.0, Listening 7.5, Reading 8.5, Writing 6.5, Speaking 6.0