Tu Vu

tuvu@vt.edu
https://tuvllms.github.io

Appointments

Virginia Tech August 2024 —
Assistant Professor, Computer Science
Research Interests: *natural language processing & machine learning*

Google DeepMind August 2023 — *present*
Research Scientist

Education

University of Massachusetts, Amherst 2016 — 2023
M.S/Ph.D. in Computer Science
Advisor: Mohit Iyyer 2018 — 2023
Thesis committee: Mohit Iyyer, Subhransu Maji, Hamed Zamani, Thang Luong, Colin Raffel

Vietnam National University, Hanoi 2009 — 2013
B.S. Honors Program in Computer Science
Highest distinction (class rank: 1/100)

Professional Experience

Google DeepMind Fall 2022 — Spring 2023
Student Researcher *with* Thang Luong & Quoc Le

Google DeepMind Summer 2022
Research Intern *with* Thibault Sellam & Elizabeth Clark

Google DeepMind Winter 2021 — Spring 2022
Student Researcher *with* Noah Constant

Google DeepMind Summer 2021 — Fall 2021
Research Intern & Student Researcher *with* Daniel Cer & Noah Constant

Google DeepMind Winter 2020 — Spring 2021
Student Researcher *with* Thang Luong & Quoc Le

Google DeepMind Summer 2020
Research Intern *with* Grady Simon & Zi Yang & Nan Hua

Microsoft Research Summer 2019
Research Intern *with* Tong Wang & Tsendsuren Munkhdalai & Adam Trischler

Selected Preprints & Publications

For an up-to-date list of my research papers, please see my Google Scholar profile.

Gemini: A Family of Highly Capable Multimodal Models
Gemini Team, Google: 1300+ authors including **Tu Vu**
arXiv preprint 2023
// Google AI Blog

FreshLLMs: Refreshing Large Language Models with Search Engine Augmentation
Tu Vu, Mohit Iyyer, Xuezhi Wang, Noah Constant, Jerry Wei, Jason Wei, Chris Tar, Yun-Hsuan Sung, Denny Zhou, Quoc Le, and Thang Luong
ACL 2024 Findings
// Our dataset and method have inspired or been used for the development of Google's Gemini, Perplexity.AI's Online LLMs, You.com, and Contextual AI's RAG 2.0

The Flan Collection: Designing Data and Methods for Effective Instruction Tuning
Shayne Longpre, Le Hou, **Tu Vu**, Albert Webson, Hyung Won Chung, Yi Tay, Denny Zhou, Quoc Le, Barret Zoph, Jason Wei, and Adam Roberts
ICML 2023
// Google Research Blog

Mixture-of-experts meets instruction tuning: A winning combination for large language models
Sheng Shen, Le Hou, Yanqi Zhou, Nan Du, Shayne Longpre, Jason Wei, Hyung Won Chung, Barret Zoph, William Fedus, Xinyun Chen, **Tu Vu**, Yuexin Wu, Wuyang Chen, Albert Webson, Yunxuan Li, Vincent Zhao, Hongkun Yu, Kurt Keutzer, Trevor Darrell, and Denny Zhou
ICLR 2024

SPoT: Better Frozen Model Adaptation through Soft Prompt Transfer
Tu Vu, Brian Lester, Noah Constant, Rami Al-Rfou, and Daniel Cer
ACL 2022

Overcoming Catastrophic Forgetting in Zero-Shot Cross-Lingual Generation
Tu Vu, Aditya Barua, Brian Lester, Daniel Cer, Mohit Iyyer, and Noah Constant
EMNLP 2022

STraTA: Self-Training with Task Augmentation for Better Few-shot Learning
Tu Vu, Thang Luong, Quoc Le, Grady Simon, and Mohit Iyyer
EMNLP 2021

Exploring and Predicting Transferability across NLP Tasks
Tu Vu, Tong Wang, Tsendsuren Munkhdalai, Alessandro Sordoni, Adam Trischler, Andrew Mattarella-Micke, Subhransu Maji, and Mohit Iyyer
EMNLP 2020

Advising

PhD advisees:

Quyet Do, incoming PhD student at **Virginia Tech** — Fall 2024 —

Thinh Pham, incoming PhD student at **Virginia Tech** — Fall 2024 —

Rishab Balasubramanian, incoming PhD student at **Virginia Tech** — Fall 2024 —

Pin-Jie (Linus) Lin, incoming PhD student at **Virginia Tech** — Fall 2024 —

Others:

Prateek Yadav, Research Intern at **Google Gemini** — Summer 2024

Simeng (Shirley) Han, Student Researcher at **Google DeepMind** — Summer 2024

Dheeraj Mekala, PhD student at **UCSD** — Spring & Summer 2022

Recent Invited Talks

Efficient Adaptation of Large Language Models — November 2023
Graph Neural Networks Reading Group, **Google**

Effective and Efficient Transfer Learning — Spring 2023
in the Era of Large Language Models
Faculty job talk

Overcoming Catastrophic Forgetting — October 2022
in Zero-Shot Cross-Lingual Generation
Parameter Efficient Tuning Methods Sync, **Google**

Transfer Learning with Large-scale Language Models — August 2022
Lecture at **VietAI**

The Appeal of Parameter-efficient Transfer Learning — June 2022
Natural Language Accelerated Team, **Google**

SPoT: Better Frozen Model Adaptation through Soft Prompt Transfer — December 2021
Parameter Efficient Tuning Methods Sync, **Google**

Academic Service

Area Chair for ACL 2024, EMNLP 2024

Program Committee/Reviewer for NeurIPS, COLM, ACL, EMNLP, NAACL, COLING, CoNLL, INLG

Selected Media

FreshLLMs: ZDNET — 2023

The Flan Collection: Google Research Blog — 2023

SPoT: Headlines of Google AI's Natural Language Accelerated Newsletter — Q1, 2022

Selected Awards & Honors & Funding

Google Student Researcherships — 2020 — 2023

UMass Amherst Graduate Assistantships — 2016 — 2023

Honda Y-E-S Award for young engineers and scientists, Vietnam — 2013
// in the top 10 nationally

Outstanding Academic and Co-curricular Achievements, Vietnam National University — 2013

Prominent Young Figure Award, Vietnam National University — 2010 & 2012

First Runner-up Prize, International Programming Contest, Japan — 2011
// ranked 2nd among 64 teams internationally

Outstanding Young Talent of the Capital City, Vietnam 2010
(in the top 100 most outstanding young talents selected from a wide range of fields)

Champion Prize, National Mathematical Olympiad, Vietnam 2010
// ranked 1^{st} among more than 600 contestants nationally

A number of prizes in National/International Olympiads 2009 — 2013
(in both Mathematics and Informatics)

A number of academic scholarships for undergraduate students 2009 — 2013

Patents

*: original inventor

Frozen Model Adaptation Through Soft Prompt Transfer
Tu Vu, Daniel Cer, Noah Constant, Brian Lester, Rami Al-Rfou
U.S. Patent Application, 17/863,840

Task Augmentation and Self-training for Improved Few-shot Learning
Thang Luong, **Tu Vu***, Quoc Le, Grady Simon
U.S. Patent Application, 17/826,690