

TU VU

 tuvllms.github.io |  tuvu@vt.edu |  [Google Scholar](#) |  [@tuvllms](#) |  [@tuvllms](#)

APPOINTMENTS

Virginia Tech Fall 2024 — *present*
Assistant Professor, Computer Science
Research Interests: *large language models & natural language processing*

Google Spring 2025 — *present*
Faculty Researcher 2023 — 2024
Google DeepMind 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: 1st)

PROFESSIONAL EXPERIENCE

Google DeepMind Fall 2022 — Spring 2023
Research Intern *with* [Thang Luong](#) & [Quoc Le](#)

Google DeepMind Summer 2022
Research Intern *with* [Thibault Sellam](#) & [Elizabeth Clark](#)

Google DeepMind Winter 2021 — Spring 2022
Research Intern *with* [Noah Constant](#)

Google DeepMind Summer 2021 — Fall 2021
Research Intern *with* [Daniel Cer](#) & [Noah Constant](#)

Google DeepMind Winter 2020 — Spring 2021
Research Intern *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](#)

*: equal contribution. For my latest research papers, see my [Google Scholar](#) profile.

PRISM: Pushing the Frontier of Deep Think via Process Reward Model-Guided Inference
Rituraj Sharma*, Weiyuan Chen*, Noah Provenzano, **Tu Vu**
arXiv 2026

SealQA: Raising the Bar for Reasoning in Search-Augmented Language Models
Thinkh Pham, Nguyen Nguyen, Pratibha Zunjare, Weiyuan Chen, Yu-Min Tseng, **Tu Vu**
ICLR 2026

// Our benchmark dataset has been used by Google's Gemini, Qwen, DeepSeek, and Kimi

Efficient Model Development through Fine-tuning Transfer
Pin-Jie Lin, Rishab Balasubramanian, Fengyuan Liu, Nikhil Kandpal, **Tu Vu**
EMNLP 2025

// Oral presentation, rating 4.5/5 by area chairs

What Matters for Model Merging at Scale?
Prateek Yadav, **Tu Vu**, Jonathan Lai, Alexandra Chronopoulou, Manaal Faruqui, Mohit Bansal,
Tsendsuren Munkhdalai
TMLR 2025

Foundational Autoraters: Taming Large Language Models for Better Automatic Evaluation
Tu Vu*, Kalpesh Krishna*, Salaheddin Alzubi, Chris Tar, Manaal Faruqui, Yun-Hsuan Sung
EMNLP 2024

// The top-performing generative model on RewardBench as of July 15, 2024, trained only on publicly available data

Gemini: A Family of Highly Capable Multimodal Models
Google Gemini Team: Rohan Anil, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste
Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew Dai, Anja Hauth, and others including
Tu Vu
Technical report 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

FUNDING & RESEARCH SUPPORT

MGI Scaling Scholarship Award (\$4,000) PI: Tu Vu	2026
Sentient Labs (\$35,000) PI: Tu Vu	2026
Google Research Award (\$50,000) PI: Tu Vu	2025
Google DeepMind Research Award (\$30,000) PI: Tu Vu	2025
Amazon - VT Faculty Research Award (\$130,000) PI: Tu Vu	2025
Google AI Studio Tier 1 PI: Tu Vu	2025
VT New Faculty Mentoring Grant (\$1,500) PI: Tu Vu	2025
Adobe Research Gift (\$5,000) PI: Tu Vu	2024

ADVISING

PHD ADVISEES:

Weiyuan Chen, 1 st year PhD student	Fall 2025 — <i>present</i>
Yu-Min Tseng, 1 st year PhD student	Fall 2025 — <i>present</i>
Rishab Balasubramanian, 2 nd year PhD student	Fall 2024 — <i>present</i>
Thinh Pham, 2 nd year PhD student	Fall 2024 — <i>present</i>
Quyet Do, 2 nd year PhD student	Fall 2024 — <i>present</i>

Pin-Jie Lin, 2nd year PhD student

Fall 2024 — *present*

MASTERS ADVISEES:

Rituraj Sharma, 1st year MS student

Spring 2025 — *present*

Noah Provenzano, 2nd year MS student

Spring 2025 — *present*

UNDERGRAD ADVISEES:

Mikaela Cankosyan, senior student

Spring 2026 — *present*

Jaydon Bingham, senior student

Spring 2026 — *present*

Christian Calvo, senior student

Spring 2026 — *present*

Nguyen Nguyen, junior student

Spring 2025 — *present*

OTHERS:

Zhenting Qi, Student Researcher @ **Google**

Summer — Fall 2025

Prateek Yadav, Research Intern @ **Google DeepMind**

Summer 2024 — Spring 2025

Simeng Han, Student Researcher @ **Google DeepMind**

Summer 2024 — Spring 2025

Salaheddin Alzubi, Masters student @ **UMass Amherst**

Fall 2022 — Spring 2023

Dheeraj Mekala, PhD student @ **UCSD**

Spring — Summer 2022

TEACHING

CS 6804: Frontier AI Systems

Spring 2026

CS 4804: Introduction to AI

Fall 2025

CS 5624: Natural Language Processing

Spring 2025

RECENT INVITED TALKS & LECTURES

Efficient Model Development through Fine-tuning Transfer
Lightning talk @ **The Amazon-Virginia Tech AI Workshop**

October 2025

Attention Mechanisms and Transformers
Two lectures @ **The New Turing Institute**

August 2025

SealQA: Raising the Bar for Reasoning in Search-Augmented Language Models
Open AGI Symposium @ UC Berkeley, hosted by **Sentient**

August 2025

Multimodal LLMs & Test-time Scaling
Lecture @ **The New Turing Institute**

June 2025

Efficient Model Development in the Era of Large Language Models
Qualcomm

November 2024

Efficient Model Development in the Era of Large Language Models
Mila / McGill NLP Seminar

October 2024

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

November 2023

Effective and Efficient Transfer Learning in the Era of Large Language Models Faculty job talk	Spring 2023
Overcoming Catastrophic Forgetting in Zero-Shot Cross-Lingual Generation Parameter Efficient Tuning Methods Sync, Google	October 2022
Transfer Learning with Large-scale Language Models Lecture @ The New Turing Institute	August 2022
The Appeal of Parameter-efficient Transfer Learning Natural Language Accelerated Team, Google	June 2022
SPoT: Better Frozen Model Adaptation through Soft Prompt Transfer Parameter Efficient Tuning Methods Sync, Google	December 2021

ACADEMIC SERVICE

Area Chair for ACL 2026, 2025, 2024; EMNLP 2025, 2024, NAACL 2025; COLING 2025

Session Chair (Machine Learning for NLP) @ EMNLP 2024

Program Committee/Reviewer for NATURE 2025, NEURIPS 2025, 2024; ICLR 2026, 2025; ICML 2026, 2025; TL4NLP@NEURIPS 2022; MCDL@ICLR 2025; COLM 2026, 2025, 2024; AAAI 2023; ACL 2022, 2021, 2020, 2019; EMNLP 2022, 2021; NAACL 2022, 2021; COLING 2020; CoNLL 2019; INLG 2020, 2019

SELECTED MEDIA

Capability Transfer	2026
Amazon Faculty Award	2025
GEMINI: Google AI Blog	2023
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

Google Student Researchships	2020 — 2023
UMass Amherst Graduate Assistantships	2016 — 2023
Honda Y-E-S Award for young engineers and scientists, Vietnam <i>// in the top 10 nationally</i>	2013
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 <i>// ranked 2nd among 64 teams internationally</i>	2011

Outstanding Young Talent of the Capital City, Vietnam <i>// in the top 100 most outstanding young talents selected from a wide range of fields</i>	2010
Champion Prize, National Mathematical Olympiad, Vietnam <i>// ranked 1st among more than 600 contestants nationally</i>	2010
A number of prizes in National/International Olympiads (in both Mathematics and Informatics)	2009 — 2013
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