Trung Nguyen

Austin, TX | email | linkedin | google scholar | homepage

Education

University of Texas at Austin *Aug 2022 - May 2027 (expected)*
PhD student in Computer Science

Hanoi University of Science and Technology *Sep 2015 – Aug 2020*
Engineer's degree in Mathematics and Informatics (Honors Program)

Relevant courses: Algorithms, Programming Languages, System Analysis & Design, Database, Computer Systems & Networks, Mathematical Statistics, Optimization Methods, Stochastic Models and Applications

Research projects

Behavioral economic preferences exhibited in large language models, intervention mechanisms

- This project explores behavioral biases in the economic decision-making of large language models, such as mental accounting, risk, intertemporal preferences
- This project has led to two papers under review.

Toward a Flexible Framework for Linear Representation Hypothesis

- Linear representation hypothesis (LRH) informally posits that highlevel concepts are encoded as linear directions in the representation spaces of LLMs.
- I proposed a framework to formalize LRH, and relate it to the Representation Engineer practice which focuses on designing, transforming, and manipulating LLM representations for applications such as probing, steering, and concept erasure. (link to the paper)

Active Learning with competitiveness guarantees

- My contributions: brainstormed ideas, conducted experiments.
- This project resulted in a paper (link) at the 28th International Conference on Artificial Intelligence and Statistics (AISTATS)

Unsupervised learning on 3D Point Clouds with Optimal Transport

- My contributions: brainstormed ideas with other researchers, designed and conducted experiments, and wrote and presented results.
- This project led to a paper (link) at the International Conference on Computer Vision (ICCV).

Recognizing the physical contact state of hands

- My contributions: modified an open-source tool to suit the need of data collection, checked with data annotators to ensure labeling quality, collected statistics of the proposed dataset, ran baseline experiments on the proposed dataset.
- This project led to a paper (link) at the Conference on Neural Information Processing Systems (NeurIPS).

Single-Click 3D Object Annotation on LiDAR Point Clouds

- My contributions: brainstormed ideas with other researchers, designed and conducted experiments, and wrote and presented results.
- This project led to a paper (link) at the NeurIPS Data-Centric AI Workshop.

Work Experience

Graduate Research Assistant *Jun 2024 – Present*
UT Austin

- Research topic: interpretability of large language models (LLMs), economic preferences exhibited in LLMs.
- Technical skills: Python, Pandas, Large language models, Git

Engineer *Jan – May 2022*
VinAI

- Worked in an engineering team to develop a web-based crowd-sourcing data annotation platform for object detection, object segmentation, skeleton detection, …

- Focused on developing the front end and experimenting with new features for efficient annotation, such as one-click annotation.
- Technical skills: React, Tensorflow JS, Git, Docker, Flask, CherryPy.

AI Resident *Jan 2020 - Dec 2021*
VinAI
- Brainstormed ideas with research scientists, designed and conducted experiments, and wrote and presented research papers.
- Technical skills: Deep Learning, Pytorch, Javascript/HTML/CSS, 3D Point Cloud, Object Detection.

Engineer Intern *Sep – Dec 2019*
VinBigData
- Worked with other engineers to detect and track vehicles coming in and out of construction sites.
- Worked with other engineers to detect people entering and exiting shopping malls.
- Checked with data annotators to ensure labeling quality.
- Trained deep neural networks using in-house data.
- Technical skills: Deep Learning, Object Detection, YoloV3, SORT, CenterNet.

Engineer Intern *Jun – Aug 2019*
Tri Nam group
- Created a small dataset from in-house data and trained a YoloV3 network to recognize car license plates.
- Technical skills: Deep Learning, Object Detection, YoloV3.

Part-time Web Developer *Dec 2018 – Feb 2019*
Miyatsu Vietnam
- Worked in an engineering team on a web project.
- Focused on developing the front end.
- Technical skills: Javascript/HTML/CSS.

Teaching Assistant *Aug 2022 – May 2024*
Computer Science department, University of Texas at Austin
- Discussed ideas and provided students with suggestions for homework and projects.
- Provided feedback for students' work.
- TA-ed courses: Predictive Machine Learning, Elements of Web Development, Algorithms: Techniques and Theory.
- Technical skills: PHP, Javascript/HTML/CSS, JQuery, SQL

Fellowships and awards

- Scholarship, National Program for the Development of Mathematics, Vietnam Ministry of Education and Training, 2017.
- Second prize, Vietnam Mathematical Olympiad for College Students, 2016.
- Scholarship, Hanoi University of Science and Technology, 2016.

Misc.

The convergence of Regula Falsi method
- I proved the convergence of a zeroth-order root-finding method (link)