

Murong Yue

🏠 4703 Forestdale Drive, Fairfax, VA 22032 ✉ myue@gmu.edu ☎ (703) 975-8878 🌐 personal webpage

EDUCATION

George Mason University, Fairfax, VA

Ph.D. in Computer Science (Advised by Prof. Ziyu Yao)

2022 – Present

University of Southern California, Los Angeles, CA

M.S. in Electrical and Computing Engineering

2018 – 2020

Beijing Jiaotong University, Beijing, China

B.E. in Electrical and Computing Engineering

2014 – 2018

PUBLICATION

Murong Yue, Ziyu Yao, “Efficient but Vulnerable: Benchmarking and Defending LLM Batch Prompting Attack”, **Preprint**[paper]

Murong Yue, Wenlin Yao, Haitao Mi, Dian Yu, Ziyu Yao, Dong Yu, “DOTS: Learning to Reason Dynamically in LLMs via Optimal Reasoning Trajectories Search”, **ICLR 2025**[paper]

Murong Yue, Wijdane Mifdal, Yixuan Zhang, Jennifer Suh, Ziyu Yao, “MathVC: An LLM-Simulated Multi-Character Virtual Classroom for Mathematics Education”, **AAAI 2025 AI4EDU Workshop**[paper]

Murong Yue, Jie Zhao, Min Zhang, Liang Du, Ziyu Yao, “Large Language Model Cascades with Mixture of Thoughts Representations for Cost-efficient Reasoning”, **ICLR 2024**[paper]

Min Zhang, Jiangeng He, Shuo Lei, **Murong Yue**, Linhan Wang, Chang-Tien Lu, “Can LLM find the green circle? Investigation and Human-guided tool manipulation for compositional generalization”, **ICASSP 2024**. [paper]

Binfeng Xu, Xukun Liu, Hua Shen, Zeyu Han, Yuhao Li, **Murong Yue**, Zhiyuan Peng, Yuchen Liu, Ziyu Yao, Dongkuan Xu, “Gentopia: A Collaborative Platform for Tool-Augmented LLMs”, **EMNLP 2023 Demo**. [paper]

EXPERIENCE

George Mason University

Graduate Research Assistance

Fairfax, VA

2023 - present

- Explore the interactive large language model systems for simulating students in a virtual classroom [paper].
- Developing cost-efficient LLM cascades to reduce computational expenses by 40% while maintaining performance on multiple reasoning tasks.
- Participated in the construction of the Gentopia [paper], which enabled us to create a context/distribution shift specialized to some target tasks easily.

Graduate Teaching Assistance

2022 - 2023

- Teaching Assistant for CS110 Essentials of Computer Science and CS112 Introduction to Computer Programming

Tencent America, AI labs

Research Intern

Bellevue, WA

2024 Summer

- Responsible for enhancing the reasoning capability of large language models in the post-training process. Developed DOTS [paper] methods in making LLM learn to reason dynamically via optimal reasoning trajectories search.

Alibaba Group, TmallGenie AI labs

Machine Learning Engineer

Beijing, China

2020 - 2022

- Responsible for vocal search in voice assistant. Designed a cross-modal retrieval method based on contrastive learning connecting the text and music to improve music tagging accuracy and enable zero-shot capabilities.
- Responsible for dialect classification in voice assistant. Designed a dialect classification system with an early classification module based on curriculum learning.

Pingan Technology, AI Labs

Research Intern

Beijing, China

2019 Summer

- Trained the TDNN neural network to extract the features of speakers and designed a speaker verification method based on the extracted speaker features. The error rate decreased from 13.7% to 8.4%.

PROFESSIONAL SERVICE

Organizing Committee

MASC-SLL 2023 - 10th Mid-Atlantic Student Colloquium on Speech, Language and Learning

Reviewer

NeurIPS'23, NeurIPS'24, ICLR'25, AISTATS'25, ARR'24, ARR'25, ICML'25

INVITED TALKS

LLM agent: build your own personal assistant, Invited talk at George Mason University Event for Hand-on Generative AI. Host: Amadra Shehu. Feb, 2024

Large Language Model Cascades with Mixture of Thoughts Representations for Cost-efficient Reasoning, Invited talk at Microsoft Semantic Machines. Host: Hao Fang. Oct. 2023

HONOR AND AWARD

2024 **Outstanding PhD Student Award**

2024 Recognition for Service on Hangs-On Generative AI Tutorial

2016 BJTU Study Scholarship