

RESEARCH SEMINAR

**Towards Scalable Human-Centric
Trustworthy Foundation Model Reasoning****Prof. Yi R. (May) FUNG**

Assistant Professor

Department of Computer Science and Engineering

Hong Kong University of Science and Technology

Hong Kong

Date : 29 Apr 2025 (Tue)**Time : 3:00 pm - 4:00 pm****Venue : N001****Abstract**

In recent years, multimodal large language models have achieved remarkable progress, excelling across diverse tasks and demonstrating impressive few-shot learning capabilities. However, ensuring these models align with principles of trustworthiness, robustness, and human-centric reasoning remains an open challenge. In this talk, we present a roadmap for enhancing foundation models' reasoning capabilities, with a focus on improving their knowledge boundary awareness and reasoning robustness through instruction tuning. Beyond textual modality, we introduce two benchmarks, VLM²-Bench and V²R-Bench, for evaluating large vision-language models (LVLMs)' ability to visually link matching cues and assessing the robustness of LVLMs to fundamental visual variations, respectively. We conclude with a discussion on scalable self-supervised learning approaches and emerging research directions that promise to make foundation models more interpretable, resilient, and aligned with human reasoning, paving the path towards AI agents with advanced reasoning capabilities that operate reliably in dynamic, real-world environments.

About the Speaker

Prof. Yi R. (May) Fung is an Assistant Professor at the Department of Computer Science and Engineering (CSE), Hong Kong University of Science and Technology (HKUST). She received her PhD from the University of Illinois, after which she spent time visiting MIT as a postdoctoral researcher. May drives cutting-edge research in the domain of human-centric trustworthy AI/NLP model reasoning, with cognitively grounded scalable alignment principles and a focus on advancing multimodal knowledge robustness mechanisms. In particular, she has published over 30 papers at top-tier machine learning venues along the topics of MLLM agentic frameworks, retrieval-augmented generation, and multi-lingual cross-culture situation understanding for diverse real-world applications (e.g., software, healthcare, business, education, media communication). Her stellar research has received much recognition internationally, including the ACL'24 Outstanding Paper Award, NAACL'24 Outstanding Paper Award, and NAACL'21 Best Demo Paper Award. In addition, she serves on the Organizing Committee for IJCAI, as Area Chair for NeurIPS/ACL/EMNLP/ACL-RR, and as Program Chair for ACM Multimedia System (MMSys). She leads a young, energetic, and growing research lab, with a number of students awarded or nominated highly selective HKPFS/RedBird merit fellowships.