



THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

PHD STUDENT SEMINAR

**Scalable Labeling with LLMs: Towards Causally-Aware
Factor Discovery**

By

Miss Lifang ZHENG

Abstract

Labeling large-scale unstructured data remains a fundamental challenge, as manual annotation is costly and existing automated approaches often rely on large language models (LLMs) that are difficult to deploy at scale. While LLMs possess strong reasoning capabilities, their direct use for labeling raises concerns of resource efficiency and interpretability. This seminar will first introduce recent approaches using LLMs for label prediction, including factor-extraction-based methods such as Bayesian Concept Bottleneck Models and CHiLL, which focus on improving explainability and transparency. Building on this line of research, our work extends these factor extraction methods to address LLM scalability challenges and explores potential directions for future study.

Date : 15 May 2025, Thursday

Time : 4:00pm

Venue : Room 5510 (Lifts 25/26)

All are Welcome!