Frontier of Artificial Network

Department of Data Science
香港城市大學
City University of Hong Kong

Towards a Mathematical Theory of Foundation Models

Invited Speaker
Hongkang Li
University of Pennsylvania

Date: June 25, 2026 (Thu)
Time: 21:00 (HKT)
Zoom Meeting: 801 137 0362

Biography

Hongkang Li is currently a postdoctoral researcher from the Department of Electrical and Systems Engineering of the University of Pennsylvania. He obtained his Ph.D. degree from the Department of Electrical, Computer, and Systems Engineering (ECSE) of Rensselaer Polytechnic Institute (RPI) in 2024. He obtained his Bachelor's degree from the Department of Electronic Engineering and Information Science of the University of Science and Technology of China in 2019. His research interests focus on the Mathematics of deep learning. He is a recipient of the 2025 IEEE Signal Processing Society Best Ph.D. Dissertation Award. He received the MLCommons ML and Systems Rising Star Award in 2025. He also received the Founders Award of Excellence and the Belsky Award of RPI in 2023.

Abstract

Large foundation models have demonstrated remarkable capabilities in language understanding, reasoning, content generation, and scientific discovery. In particular, they exhibit emerging learning paradigms beyond classical supervised learning, including learning from context during inference and generating data through iterative denoising. Despite their empirical success, the theoretical understanding of how these models learn, adapt, and generate remains limited. In this talk, I will present a series of theoretical studies aimed at establishing mathematical foundations for foundation models. Based on the feature learning and training dynamics analysis of nonlinear Transformers, first, I will introduce our recent theoretical works on how foundation models can acquire and manipulate task-relevant knowledge by In-Context Learning, Chain-of-Thought, and task arithmetic. I will then present how Transformer-based diffusion models can converge to the Bayes-optimal denoiser by gradient descent.