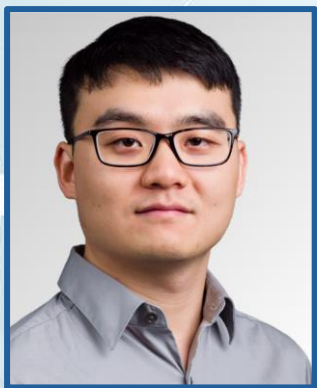


# Towards Theoretical Foundations of Foundation Models: Feature Learning, Optimization Dynamics, and Generalization



**Invited Speaker**

**Shuai Zhang**

*New Jersey Institute of Technology*

**Date:** June 18, 2026 (Thu)

**Time:** 10:00am (HKT)

**Zoom Meeting:** 801 137 0362

## Biography

Shuai Zhang has been an Assistant Professor in the Department of Data Science at the New Jersey Institute of Technology (NJIT) since 2023. He received the B.E. degree in Electrical Engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2016, and the Ph.D. degree in Electrical Engineering from Rensselaer Polytechnic Institute (RPI), Troy, NY, USA, in 2021. From 2022 to 2023, he was a Postdoctoral Research Associate at RPI. His research focuses on the theoretical foundations of modern machine learning, including self-supervised learning, parameter-efficient transfer learning, and emerging large language model architectures. He has served as an Area Chair for leading machine learning conferences, including ICLR and NeurIPS. His work has been selected for oral presentation at top-tier machine learning venues, including ICML 2023 and ICLR 2025.

## Abstract

Foundation models have rapidly become a central paradigm in modern AI, driving breakthroughs across language, vision, and multimodal learning. Despite their empirical success, many fundamental questions remain, including why these large-scale neural networks learn effective representations and why they generalize well despite extreme overparameterization. This talk discusses recent theoretical progress toward understanding foundation models within the feature learning framework. In this framework, data are modeled as a mixture of latent features, and learning dynamics guide models toward particular solutions through implicit preferences that lead to strong generalization. Specifically, the talk covers two topics: contrastive learning and recent state-space models (e.g., Mamba). Key results provide theoretical insights into how feature learning emerges during self-supervised pretraining, how data imbalance shapes the representations learned by large models, and how pruning-based methods can improve the learning of minority features. For state-space models, the talk analyzes selective gating mechanisms that enable adaptive information filtering and efficient long-sequence modeling, and examines how token ordering influences the model's ability to capture sequential dependencies.