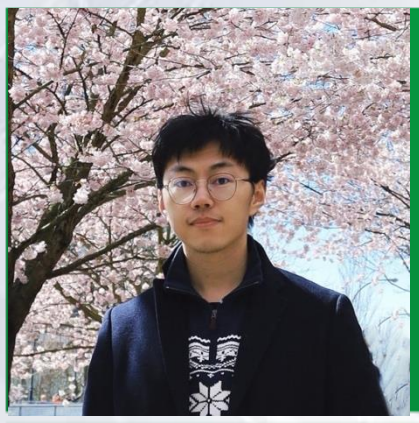
香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

TMLR Young Scientist SEMINAR

2026 SERIES

Trustworthy Machine Learning and Reasoning Group



Dr. Yi Ren

Postdoctoral Researcher,
OATML Group,
Oxford University

Date: 30 March 2026 (Monday)

Time: 19:00 – 20:30 (HKT)

Meeting: https://hkbu.zoom.us/j/6603117755

Learning Dynamics of Deep Learning – Force Analysis of Deep Neural Networks

ABSTRACT

Deep neural networks exhibit rich yet structured dynamics during training. In this talk, we will discuss a theoretical framework for understanding these mechanisms through a perspective akin to force-analysis in high-school physics. In this view, neural network training can be interpreted as a dynamical system where samples exert forces on each other. These interactions naturally decompose into two components: an interaction Kernel determined by the sample similarity in model's gradient space; and a gradient Force, describing the strength of learning signals.

This framework provides a unified lens for analyzing modern training paradigms. In particular, it offers new insights into phenomena observed in LLM fine-tuning (e.g., SFT, DPO, GRPO), including repetition, forgetting, hallucination, and exploration collapse. More broadly, learning dynamics and force analysis offer a principled foundation for understanding training behavior and building more reliable and interpretable AI systems.

BIOGRAPHY

Yi Ren is currently a postdoc researcher at OATML group, University of Oxford, led by Prof. Yarin Gal. His research focuses on deep learning theory, particularly learning dynamics and its applications in modern deep learning systems. His recent work examines the training dynamics of large language model fine-tuning, self-improvement, and the mechanisms behind simplicity bias. He proposed a force-analysis framework interpreting neural network training as sample interactions, unifying paradigms like knowledge distillation, RLHF, and DPO. His work has been published at top-tier conferences, such as ICLR and NeurIPS, and received the ICLR 2025 Outstanding Paper Award as well as the Best Paper Award at the 2nd AI for Math Workshop (ICML 2025).

ENQUIRY

Email: bhanml@comp.hkbu.edu.hk