Xu Yuanjian(徐元健)

phone: (+86) 17637303503 · email: xuyuanjian233@gmail.com · homepage

Education

Hong Kong University of Science and Technology, Ph.D. in Fintech (Candidate) 2023.09–Present
Peking University, Master in Computer Science 2020.09–2023.06
Nankai University, Bachelor in Computer Science 2014.09–2018.06

A Brief Introduction

I am a Ph.D. candidate in Computer Science at the Hong Kong University of Science and Technology, Guangzhou campus. I am supervised by Prof. Guang Zhang from HKUST-GZ, and Prof. Zhong Li from MSRA. My research focuses on **Data Centric Artificial Intelligence**, particularly on understanding the mechanisms of **large language models (LLMs)** and exploring their applications in domains such as finance.

I have published at venues including **ACL 2023**, **ACL 2025**, and **ACM ICAIF 2025**, with additional manuscripts under review at **ICLR 2026**, **ACL Rolling Review**, and **ICASSP 2026**. I also serve as a reviewer for leading conferences such as **NeurIPS**, **ICLR**, and **AAAI**.

Research Publications[1]

Language Model Data Selection

This problem is fundamentally *distributionally robust optimization*. We first identified hard-to-learn samples, then showed that from the LLM's perspective, *sample difficulty, domain partitioning, and human intuition differ significantly*—both papers received positive reviews at ACL. We developed general methods: $\mathbf{D}^3$ explains why training order matters, while our annealing framework addresses sample selection during annealing. We also investigate tokenization strategy impacts.

- **Yuanjian Xu**, Qi An, and Zaiqing Nie. "Hard Sample Aware Prompt-Tuning." **In Proceedings of ACL 2023** *(CCF A, CORE A*)*.
- **Yuanjian Xu**, Jianing Hao, Guang Zhang, et al. "DoGraph: Towards Domain-Aware LLM Training with Graph-Guided Weighting." **Under review at ACL 2026** *(CCF A, CORE A*; All reviews are positive)*.
- **Yuanjian Xu**, et al. "D^3: Dynamic Directional Graph-Constrained Data Scheduling for LLM Training." **Target: ICML 2026** *(CCF A, CORE A*; In collaboration with MSRA)*.
- **Yuanjian Xu**, et al. "Towards Efficient LLMs Annealing with Principled Sample Selection." **Target: ICML 2026** *(CCF A, CORE A*; In collaboration with MSRA)*.
- **Yuanjian Xu**, et al. "Information Aggregation under Model Capacity Constraints." **Target: ICML 2026** *(CCF A, CORE A*)*.

General Sequence Model Capabilities

Push-forward models, *Neural SDEs*, and *autoregressive generative models* are three paradigms of generative sequence models, differing in their randomness injection mechanisms. In this line of research, we investigate the expressive power limitations of Neural SDEs.

- **Yuanjian Xu**, Jianing Hao, Guang Zhang. "State-Aware Neural Stochastic Differential Equations for Multi-Modal Dynamics." **Under review at ICLR 2026** *(CCF A, CORE A*)*.
- **Yuanjian Xu**, Jianing Hao, Guang Zhang. "Mitigating Discretization Bias in Neural SDEs via Inference-Time Dropout." **Under review at ICLR 2026** *(CCF A, CORE A*)*.
- **Yuanjian Xu**, Yuan Shuai, and Guang Zhang. "Hermite Discriminator for Neural SDEs." **Under review at ICASSP 2026** *(CCF B, CORE A; All reviews are positive)*.
- **Yuanjian Xu**, et al. "Neural Causal Process." **Target: ICML 2026** *(CCF A, CORE A*)*.

[1] • Published; • Under review (high acceptance probability, all reviews are positive); • Under review; • Ongoing

Generative Applications

I investigate practical applications of generative models, particularly in finance, focusing on aligning LLMs with domain-specific knowledge and benchmarking their reasoning capabilities.

- **Yuanjian Xu**, Jianing Hao, and Guang Zhang. "FinRipple: Aligning Large Language Models with Financial Market for Event Ripple Effect Awareness." **In Proceedings of ACL 2025** *(CCF A, CORE A*).*
- **Yuanjian Xu**, Jianing Hao, and Guang Zhang. "LENS: Large Pre-trained Transformer for Exploring Financial Time Series Regularities." **In Proceedings of ACM ICAIF 2025** *(Leading conference for AI in finance).*
- Jianing Hao*, Yuhe Wu, **Yuanjian Xu***, Guang Zhang. "BizCompass: Benchmarking the Reasoning Capabilities of LLMs in Business Knowledge and Applications." **Under review at ACL 2026** *(CCF A, CORE A*; All reviews are positive; *Equal contribution).*

Research & Teaching Experience

Microsoft Research Asia (MSRA), Supervised by Dr. Zhong Li

Tsinghua University, Institute for AI Industry Research (AIR), Supervised by Prof. Zaiqing Nie

Hong Kong University of Science and Technology, Teaching Assistant, Advanced Statistics

INTERNSHIP EXPERIENCE

HuaTai Securities, Finance Engineering Department 2021.1-2022.3

Skills

Machine Learning: I am well-versed in modern generative models, LLMs, and machine learning.

Mathematical Foundation: I have a relative solid mathematical foundation. During my Ph.D. studies, I passed the PQE (the Ph.D. qualifying exam) tests in *optimization*, *advanced probability theory*, *advanced statistics*, and *stochastic analysis*.

Financial Theory: I also have exposure to financial theory, having passed *FRM Level 1* and studied *advanced microeconomics* and *advanced asset pricing*.

Honors & Awards

- Award for Excellent Academic Excellence, Peking University 2021 Certificate No.: H2021000170320
- Air Star Plan, Tsinghua University, Institute for AI Industry Research (AIR) 2021
- Full Ph.D. Scholarship, Hong Kong University of Science and Technology 2023–Present