

## EDUCATION

<b>The University of Melbourne</b> Ph.D. in Engineering and IT (expected) Department: Faculty of Engineering and Information Technology Thesis title: Data-adaptive Non-parametric Hypothesis Testing Thesis advisor: Feng Liu, Liuhua Peng	2024-Now
<b>Nanjing University</b> MSc (Eng) in Computer Science and Technology Department: School of Artificial Intelligence Thesis title: On the Exploration of Local Significant Differences for Two-Sample Test Thesis advisor: Wei Gao	2021-2024
<b>Dalian University of Technology</b> BSc (Eng) in Transportation Engineering Department: Department of Transportation and Logistics Thesis title: Signalized Intersection Optimization Based on Machine Learning Thesis advisor: Shaopeng Zhong	2017-2021

## RESEARCH INTERESTS

My research interests focus on making hypothesis testing usable for modern machine learning problems:

- **Single-example testing for probabilistic ML outputs.** Most classical statistical tests are designed for batch samples, while many ML applications require decisions on individual examples with probabilistic outputs. This motivates my research on single-example testing methods that provide per-example statistical significance and theoretical false-alarm control.
- **Data-adaptive, learnable testing beyond rigid hypothesis specification.** Conventional hypothesis testing requires specifying formal hypotheses in advance and then designing a test statistic, which often forces ML problems into restrictive statistical formulations and limits practical adoption. My research aims to build learnable testing frameworks in which both the test statistic and the testing procedure are constructed from data, enabling rigorous evaluation of a broader class of questions.
- **Certified LLM safety and evaluation.** Modern LLM safety and evaluation often rely on empirical pipelines, often without rigorous finite-sample guarantees, particularly in .under adaptive and label-efficient settings. My **current research** focuses on **certified and cost-efficient LLM evaluation** and **certified detection of training data contamination**, aiming to develop theory-driven methods with finite-sample validity, and provable error control for LLM evaluation and safety assessment.

## PUBLICATIONS (indicates equal contribution)\*

1. **Z. Zhou\***, X. Tian\*, L. Peng, C. Lei, A. Schrab, D. J. Sutherland, and F. Liu. DUAL: Learning Diverse Kernels for Aggregated Two-sample and Independence Testing. In NeurIPS, 2025 [**CORE: A\***]
2. **Z. Zhou**, L. Peng, X. Tian, and F. Liu. Anchor-based Maximum Discrepancy for Relative Similarity Testing. In NeurIPS, 2025 [**CORE: A\***]
3. X. Tian, L. Peng, **Z. Zhou**, M. Gong, A. Gretton, and F. Liu. A Unified Data Representation Learning for Non-parametric Two-sample Testing. In UAI, 2025 [**CORE: A**]
4. F. Xu, **Z. Zhou**, J. Ni, and W. Gao. Interpretation with baseline shapley value for feature groups on tree models. Frontiers of Computer Science, 2025.
5. **Z. Zhou**, J. Ni, J. Yao, and W. Gao. On the exploration of local significant differences for two-sample test. In NeurIPS, 2023 [**CORE: A\***]

6. S. Zhong, Y. Gong, **Z. Zhou**, R. Cheng, F. Xiao. Active learning for multi-objective optimal road congestion pricing considering negative land use effect. Transportation Research Part C, 2021.
7. **Z. Zhou\***, X. Tian\*, J. Zhang\*, Z. Ye, Y. Guo, D. Zhang, L. Peng, F. Liu. USAD: Uncertainty-aware Statistical Adversarial Detection. [Under review at ICML]
8. **Z. Zhou**, L. Peng, X. Tian, M. Gong, and F. Liu. Do You Want to Know if Two Distributions Are Close to Each Other? Testing the Closeness with Statistical Significance. [Under review at ICML]
9. X. Tian\*, **Z. Zhou\***, L. Peng and F. Liu. LOTTERY: Learning from Reference-Only Samples in Two-Sample Testing under Size Asymmetry. [Under review at ICML]
10. X. Jiang\*, **Z. Zhou\***, X. Tian\*, Y. Song and F. Liu. LOTTERY: Learning from Reference-Only Samples in Two-Sample Testing under Size Asymmetry. [Under review at ICML]
11. Z. Zheng, **Z. Zhou**, W. Gao and Z. Zhou. Variance-Normalized Mean Embedding Discrepancy for Two-Sample Test. [Under review at JMLR]

## RESEARCH EXPERIENCE

- Graduate Researcher @ the University of Melbourne 2024-Now  
Supervised by Feng Liu, Liuhua Peng  
Trustworthy Machine Learning and Reasoning (TMLR) Group  
Research on Data-adaptive Non-parametric Hypothesis Testing
- Graduate Researcher @ Nanjing University 2021-2024  
Supervised by Wei Gao  
Learning and Mining from Data (LAMDA) Group  
Research on Interpretable Two-Sample Testing

## WORKING EXPERIENCE

- The University of Melbourne** 2026.03 - Now  
Tutor of MAST20006: Probability for Statistics Average hours per week: 6
- Alibaba Cloud Computing Co. Ltd** 2023.06 - 2023.09  
Machine Learning Engineer (Internship) Average hours per week: 40
- Detected product anomalies using temporal distribution shifts in post-sales customer requests.
  - Analyzing the dialogue between customer service representatives and customers using large language models (LLMs) and prompt engineering, and constructing rule-based algorithms for anomaly filtering.
- Nanjing University** 2023.02 - 2023.07  
Teaching Assistant for Game Theory and Applications Average hours per week: 20

## AWARDS

- Conference on Neural Information Processing Systems (Neurips), Top Reviewer 2025
- The Melbourne Research Scholarship 2024.06
- Master's Academic Scholarship 2021.09 - 2024.06
- National Transportation Science and Technology Contest (Class A), Second Prize 2021.07
- China Association for Traffic Education Research, Outstanding Award 2020.11

## PROFESSIONAL SERVICE

- Reviewer: International Conference on Machine Learning (ICML), 2024-2026
- Reviewer: International Conference on Learning Representations (ICLR), 2024-2026
- Reviewer: Neural Information Processing Systems (NeurIPS), 2024-2025
- Reviewer: International Conference on Artificial Intelligence and Statistics (AISTATS), 2024-2025
- Reviewer: IEEE Transactions on Knowledge and Data Engineering (TKDE)