Xunye (Yeager) Tian

Address: 60 A'Beckett Street, Melbourne, VIC 3000, Australia Date of Birth: 18 November 2000 Phone: +61 433 772927 Email: xunye.tian@unimelb.edu.au or xunyetian.ml@gmail.com Home Page: <u>https://yeager20001118.github.io/</u>

RESEARCH OBJECTIVE

My research is driven by a commitment to strengthening the statistical foundations of modern AI through innovative hypothesis testing frameworks and trustworthy machine learning methods. Currently, I focus on bridging statistical theory with practical AI challenges through:

- Developing novel two-sample testing approaches to distinguish between distributions with theoretical guarantees
- Advancing independence testing methodologies for semi-supervised representation learning and machine unlearning
- Creating testing-based detection algorithms for machine-generated content to preserve information integrity
- Exploring conformal prediction techniques for reliable uncertainty estimation in high-stakes domains

As AI continues to transform decision processes across critical sectors, I maintain that statistical soundness and trustworthiness aren't optional features but fundamental requirements.

EDUCATION BACKGROUNDS

Doctor of Philosophy - Artificial Intelligence

- **Research Focus:** Hypothesis Testing, Machine-Generated Text Detection
- Supervisor: Dr. Feng Liu, Dr. Liuhua Peng and Dr. Tingru Cui

Master of Data Science - Hypothesis Testing

- **Grade:** High Distinction (A in US Grade, 1st Class in UK Grade)
- Core Subjects: Statistical Machine Learning, Multivariate Statistics, Natural Language Processing, Cluster and Cloud Computing, Computer Vision (audit)
- Research Focus: Optimization in Deep Kernel Two-sample Testing
- Supervisor: Dr. Feng Liu

Bachelor of Science - Data Science

- Core Subjects: Machine Learning, Linear Statistical Models, Modern Applied Statistics, Applied Data Science
- **Relevant Subjects:** Design of Algorithm, Database System, Engineering Mathematics, and Web Information Technologies

RESEARCH PUBLICATIONS

(^ Co-first author, * Corresponding author)

1. **Xunye Tian,** Liuhua Peng, Zhijian Zhou, Mingming Gong, Arthur Gretton, Feng Liu*. A Unified Data Representation Learning for Non-parametric Two-sample Testing.

Key contributions: We propose a high-performed two-sample testing framework RL-TST, which addresses existing drawbacks of the mainstream pipelines. RL-TST can both extract information discarded from data-splitting process and learn discriminative representations to detect complex distribution patterns.

2. Xunye Tian, Feng Liu*. Take a Close Look at the Optimization of Deep Kernels for Non-parametric Two-Sample Tests.

Key contributions: We identify a problem existing in the deep kernel optimization procedure that directly maximizing the optimization objective will reduce the test power of deep kernel. We provide two viable techniques from non-smooth optimization and Pareto optimization to simultaneously maximize the objective and test power, leading to a more stable performance of deep kernel optimization.

Accepted by UAI 2025

The University of Melbourne

The University of Melbourne

The University of Melbourne

July 2018 - March 2022

October 2023 - Present

March 2022 - July 2023

Accepted by ADC 2023

3. Zhijian Zhou, Liuhua Peng, **Xunye Tian**, Feng Liu*. A Kernel Distribution Closeness Testing.

Key contributions: We propose a Norm-Adaptive Maximum Mean Discrepancy (NAMMD), a powerful new distributional discrepancy measure that scales MMD by the RKHS norms of distributions, enabling more informative distributional closeness testing across complex data domains. This framework achieves significantly higher statistical power than conventional MMD approaches while maintaining bounded type-I error rates across diverse applications including synthetic noise and real-world image analysis.

4. Zhijian Zhou, Liuhua Peng, **Xunye Tian**, Feng Liu*. Anchor-based Maximum Discrepancy for Relative Similarity Testing.

Key contributions: We tackled the challenge of relative similarity testing by developing an Anchor-based Maximum Discrepancy (AMD) approach that simultaneously learns both the optimal hypothesis and kernel, eliminating the need for manual hypothesis specification. We implemented a two-phase framework where we first estimate the AMD over diverse kernel spaces to identify potential hypotheses, then assess their statistical significance using a unified testing mechanism with theoretically-derived thresholds, demonstrating superior performance through extensive validation on benchmark datasets.

5. Zhijian Zhou[^], **Xunye Tian**[^], Liuhua Peng, Chao Lei, Antonin Schrab, Danica J. Sutherland, Feng Liu^{*}. Learning Diverse Kernels for Aggregated Two-sample and Independence Testing.

Key contributions: We identify a phenomenon that kernel aggregation does not always guarantee an increasing test power, as kernels might convey overlapping information that are redundant. We propose a method Diverse U-statistic Aggregation with Learned kernels (DUAL), which introduce a notion of diversity into aggregated multivariate U-statistics and learn a pool of both diverse and powerful kernels. As a result, DUAL provides a general power enhancement for both two-sample and independence tests.

HONOURS & AWARDS

- Melbourne Research Scholarship, UoM, 2023.
- Melbourne Welcome Grant, UoM, 2023.

ACADEMIC SERVICES

Conference Reviewer:

• ICML, NeurIPS, ICLR, AJCAI, etc.

Journal Reviewer:

• NEUNET, TMLR, etc.

TEACHING EXPERIENCES

- Teaching Assistant for MAST90139: Statistical Modelling for Data Science (2025 S1)
- Teaching Assistant for COMP90051: Statistical Machine Learning (2024 S1 & S2 & 2025 S1)

WORKING EXPERIENCES

SenseTime, Shanghai, China (Web Development Intern: July 2021 – March 2022)

- Utilized Java Springboot to develop a designed AI Platform to provide image recognition, text analysis, and traditional classification services and a feedback Platform to manage inquiries, complaints, and suggestions
- Collaborated with cross-functional teams to build Springboot websites based on specific project requirements, tracked project progress using Confluence and Jira, and configured Docker Compose files for product launch
- Created and updated the Mybatis database, extracted the database with SQL and mapper, and operated Docker and K8s for container management

TECHNICAL SKILLS

- Machine Learning: PyTorch, Python
- Software Development: Java, Docker, Javascript, HTML, CSS, Git

Submitted to NeurIPS 2025

Submitted to NeurIPS 2025