

# Facheng Yu

[fachengy@uw.edu](mailto:fachengy@uw.edu) • (206) 677-6512 • Seattle, WA • <https://fachengyu.github.io/>

## EDUCATION

**University of Washington, Seattle**

*Ph.D. in Statistics (Expected 2028)*

*M.S. in Statistics (2025)*

Advisor: Zaid Harchaoui

**Research Interests:** Machine Learning, Causal Inference, Core AI, and AI Domain Applications

**Coursework:** Statistical Inference, Statistical Learning, Advanced Probability, Convex Optimization, Regression Methods

## SKILLS

**Programming:** Python (NumPy, PyTorch, scikit-learn, Pandas), R, C++, SQL.

**Techniques:** Machine Learning, Causal Inference, Optimization, High-Dimensional Statistics.

**Tools:** Git, LaTeX.

## RESEARCH EXPERIENCE

**University of Washington, Seattle**

*Graduate Student Researcher*

Advisors: Zaid Harchaoui and Alex Luedtke

**Project: LLM-based Privacy-Preserving Authorship Obfuscation**

Sep. 2025–Present

- Developed a pipeline in PyTorch for authorship obfuscation in Chinese with fine-grained control over stylistic features such as usage of sarcasm, word lengths, sentence lengths, frequency of function words.
- Evaluated the LLM performance with different decoding strategies such as greedy, top-k, top-p, and mixed strategies etc.

**Project: The Benefit of Balancing with Known Marginals for Multimodality**

Sep. 2025–Present

- Generalized a multi-CLIP loss for n modalities as one step of balancing using known marginal distributions of each modality.
- Proposed an iterative algorithm that reduces the estimation variance of the multi-CLIP minimizer and performed experiments on three modalities (text, video, audio) in PyTorch.

**Project: Stochastic Gradients under Nuisances**

Jun. 2024–Dec. 2025

- Investigated stochastic gradient algorithms for optimizing objectives with unknown nuisance parameters, with applications in high-dimensional inference and causal learning.
- Applied **Neyman orthogonality** to mitigate the bias introduced from nuisance estimation, enabling more reliable gradient-based optimization.
- Designed and implemented an approximately **orthogonalized gradient oracle** to debias learning for non-orthogonal objectives.
- Developed **OSGD** (Python library) to implement orthogonalized stochastic gradients for partially linear models (PLM), achieving improved risk consistency (~0.001) with ~0.1 nuisance error and ~0.5 operator error compared to standard SGD.

**Renmin University of China, Beijing, China**

*Undergraduate Student Researcher*

Advisor: Yuqian Zhang

**Project: Data Integration Using Covariate Summaries from External Sources**

May 2023–Jun. 2024

- Designed robust mean outcome estimators under both homogeneous and heterogeneous data settings.
- Extended our integration methods to estimate treatment effects for improved **generalizability** and **transportability**.
- Conducted extensive simulations in R, demonstrating strong robustness to model misspecification and improved accuracy over baseline methods (e.g., DoubleML).

## PUBLICATIONS

[1] F. Yu, R. Mehta, A. Luedtke, and Z. Harchaoui, “[Stochastic Gradients under Nuisances](#)”, *NeurIPS 2025*.

[2] F. Yu and Y. Zhang, “[Data Integration Using Covariate Summaries from External Sources](#)”, *arXiv:2411.15691*.

## SCHOLARSHIPS AND AWARDS

- Institute for Foundations of Data Science (IFDS) Scholarship, 2024.
- Third Prize in Asia and Pacific Mathematical Contest in Modeling, 2022.
- Excellent Student Scholarship, Wuhan University, 2020, 2021, 2022.