# YU GUI

https://yugjerry.github.io/ yugui@wharton.upenn.edu

## ACADEMIC POSITION

**The Wharton School, University of Pennsylvania, U.S.A.** *July 2025 - present*
*Postdoctoral Researcher, Department of Statistics and Data Science*
- Advisors: Dylan Small and Zhimei Ren
  - Member of *Wharton Causal Data Science Lab, Wharton AI & Analytics Initiative*
  - Member of *National Studies on Air Pollution and Health Group, Harvard T.H. Chan School of Public Health*

## EDUCATION

**University of Chicago, U.S.A.** *Sep 2020 - Jun 2025*
*PhD in Statistics*
- Advisors: Rina Foygel Barber and Cong Ma
- Thesis: *Statistical learning and inference in weakly specified settings: shifted distributions and unlabeled data*
- Dissertation committee: Rina Foygel Barber, Cong Ma, Rebecca Willett
- IMS Lawrence D. Brown Ph.D. Award
- William Rainey Harper Dissertation Fellowship[1]

**University of Science and Technology of China (USTC)** *Sep 2016 - Jun 2020*
*School of the Gifted Young, Bachelor of Science*

**Research intern at Harvard** *Jun 2019 - Nov 2019*
- Advisor: Professor Jun S. Liu

## RESEARCH INTERESTS

- **Reliable, adaptive, and *human-in-the-loop* statistical inference**
  - Adaptive statistical methods for causal inference that incorporate exploratory data analysis while preserving statistical validity when hypotheses and inference depend on the same data
  - Distribution-free uncertainty quantification for black-box models, including reliable and human-in-the-loop deployment of AI foundation models with statistical guarantees
- **Robust uncertainty quantification under heterogeneity and weak specification**
  - Robust uncertainty quantification when the observed data are not representative of target populations—due to distribution shift, missingness, or time-to-event censoring
  - Doubly robust methods for causal mediation analysis with time-varying covariates, time-to-event censoring, and recurrent outcomes
- **Multi-modal learning and applications in single-cell spatial multi-omics**
  - Theoretical foundations for self-supervised representation learning with multi-modal and unlabeled data
  - Information-theoretic disentanglement and cross-modal experimental design with applications to single-cell spatial multi-omics data

[1]This award, one of the University of Chicago's highest honors, is given in recognition of a PhD candidate's record of achievement and professional promise.

## HONORS AND SELECTED AWARDS

- Rising Stars in Statistics and Data Science, Cornell University 2026
- IMS Lawrence D. Brown Ph.D. Award, *Institute of Mathematical Statistics* 2026
- IMS ICSDS Junior Researcher Travel Award (declined due to visa renewal) 2025
- William Rainey Harper Dissertation Fellowship 2024
- IMS Hannan Graduate Student Award 2023
- NeurIPS Scholar Award 2023
- IMS International Conference on Statistics and Data Science (ICSDS) Travel Award 2023
- Winner of the Consulting Cup, Department of Statistics, University of Chicago 2022
- National Scholarships 2019
- National Scholarships 2018

## PUBLICATIONS AND MANUSCRIPTS

### Publications

- **Yu Gui**, Rina Foygel Barber, and Cong Ma.
  Distributionally robust risk evaluation with an isotonic constraint.
  *Information and Inference: A Journal of the IMA*, 2026.
  IMS Lawrence D. Brown Ph.D. Award.
- **Yu Gui***, Ying Jin*, Yash Nair*, and Zhimei Ren*.
  ACS: An interactive framework for conformal selection.
  *Journal of the American Statistical Association*, 2026.
- **Yu Gui**, Cong Ma, and Zongming Ma.
  Multi-modal contrastive learning adapts to intrinsic dimensions of shared latent variables.
  *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- **Yu Gui***, Rohan Hore*, Zhimei Ren*, and Rina Foygel Barber.
  Conformalized survival analysis with adaptive cut-offs.
  *Biometrika*, 2024.
- **Yu Gui***, Ying Jin*, and Zhimei Ren*.
  Conformal alignment: knowing when to trust foundation models with guarantees.
  *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.
- **Yu Gui**, Rina Foygel Barber, and Cong Ma.
  Conformalized matrix completion.
  *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
  IMS International Conference on Statistics and Data Science (ICSDS) Travel Award, 2023.
  NeurIPS Scholar Award, 2023.
- Xiaofan Zhou, Baiting Chen, **Yu Gui**, and Lu Cheng.
  Conformal Prediction: A Data Perspective.
  *ACM Computing Surveys*, 2025.
- **Yu Gui**.
  ADAGES: adaptive aggregation with stability for distributed feature selection.
  (Undergraduate work) *ACM-IMS Foundations of Data Science Conference*, 2020.

- Xin Xing*, **Yu Gui***, Chenguang Dai, and Jun S Liu.
  Neural Gaussian Mirror for controlled feature selection in neural networks.
  (Undergraduate work) *International Conference on Machine Learning and Applications*, 2020.

### Under Review/Preprints

- **Yu Gui**, Dylan Small, and Zhimei Ren.
  Adaptive discovery of effect modification in matched observational studies.
  *Journal of the American Statistical Association*, under review, 2026.
- **Yu Gui**, Rina Foygel Barber, and Cong Ma.
  The cost of enforcing long-run coverage: a phase transition in online conformal prediction. 2026.
- **Yu Gui**, Cong Ma, and Zongming Ma.
  IndiSeek learns information-guided disentangled representations. 2026.
- **Yu Gui** and Bahar Taşkesen.
  Statistical equilibrium of optimistic beliefs.
  *Games and Economic Behavior*, under review, 2025+.
- **Yu Gui**, Cong Ma, and Yiqiao Zhong.
  Unraveling projection heads in contrastive learning: insights from expansion and shrinkage.
  *Revise & Resubmit at Journal of the American Statistical Association*, 2025+.
  IMS Hannan Graduate Student Award.

Google Scholar (* indicates equal contributions or alphabetical ordering.)

## PROFESSIONAL SERVICES

**Reviewer for journals**: The Annals of Statistics, Journal of the Royal Statistical Society Series B: Statistical Methodology, Journal of the American Statistical Association, Operations Research, Journal of Computational and Graphical Statistics, IEEE Transactions on Information Theory, Transactions on Machine Learning Research, ACM Transactions on Knowledge Discovery from Data, Electronic Journal of Statistics, Statistica Sinica, Communications in Statistics–Theory and Methods, Statistics in Medicine.

**Reviewer for conferences**: NeurIPS, ICML, ICLR, AISTATS, AAAI.

**Invited discussion**: Discussant for "Multiple Testing of Linear Forms for Noisy Matrix Completion" by Wanteng Ma, Lilun Du, Dong Xia, and Ming Yuan. *International Seminar on Selective Inference.* Apr 2024.

**Mentorship**: Research mentoring, mentee: Elena (Yutong) Li, University of Chicago; University of Chicago International Allies Program; Physical Sciences Division Peer Mentorship Program.

**Student consultant**: University of Chicago Statistics Consulting Program. *Winner of the Consulting Cup, 2022*: in collaboration with Pritzker School of Molecular Engineering.

## TEACHING EXPERIENCE

**Teaching assistant, University of Chicago**

- Financial Statistics: Time Series, Forecasting, Mean Reversion, and High Frequency Data — Winter 2025
- Statistical Models and Methods — Autumn 2024
- Linear Models and Experimental Design — Spring 2024
- Multiple Testing, Modern Inference, and Replicability — Autumn 2023
- Introduction to Mathematical Probability — Spring 2022, Autumn 2022

- Statistical Methods and Applications — Winter 2021, Spring 2021, Autumn 2021

**Teaching assistant, USTC**

- Differential Equations — Spring 2020
- Multivariate Mathematical Analysis — Spring 2019
- Linear Algebra — Autumn 2019

## SOFTWARE

- **ada-subgroup-selection**, developer. R implementation of Gui, Small, and Ren (2026).
- **adacs**, contributor. Python implementation of Gui et al (2025).
- **conformal-alignment**, developer. Python implementation of Gui, Jin, and Ren (2024).
- **iso-DRL**, developer. Python implementation of Gui, Barber, and Ma (2024).
- **conf-mc**, developer. Python implementation of Gui, Barber, and Ma (2023).

## INVITED AND CONTRIBUTED PRESENTATIONS

- Adaptive discovery of effect modification in matched observational studies
  - *ASA Section on Statistical Learning and Data Science 2026: Inference and Intelligence. New York.* Scheduled Nov 2026.
  - (Poster) *Rising Stars in Statistics and Data Science. Cornell. Ithaca, NY.* Sep 2026.
  - *IMS 2026 New Researchers Conference. UMass Amherst, MA.* Jul 2026.
- Distributionally robust risk evaluation with an isotonic constraint
  - *IMS Annual Meeting 2026. Salzburg, Austria.* (Canceled due to visa renewal) Jul 2026.
  - *The New England Statistics Symposium (NESS) 2026. From Data to Discovery: Statistical Insight in the Age of AI. University of Connecticut, Storrs.* May 2026.
  - *13th International Conference on Multiple Comparison Procedures. Philadelphia.* Aug 2025.
  - *2025 Joint Conference on Statistics and Data Science (JCSDS).* Jul 2025.
  - *Seminar at Department of Statistics, University of Wisconsin-Madison.* May 2025.
  - *Professor Genevera Allen's group meeting. Columbia University (virtual).* Mar 2025.
  - *2024 International Conference on Frontiers of Data Science. Hangzhou, China.* Jul 2024.
- Conformal alignment: selective deployment of foundation models with statistical guarantees
  - (Poster) *The Thirty-eighth Annual Conference on Neural Information Processing Systems. Vancouver.* Dec 2024.
  - *Workshop on Statistical Foundations of LLMs and Foundation Models. Vancouver.* Dec 2024.
- Conformalized matrix completion.
  - *The 7th International Conference on Econometrics and Statistics. Beijing, China.* Jul 2024.
  - (Poster) *The Thirty-seventh Annual Conference on Neural Information Processing Systems. New Orleans, LA.* Dec 2023.
  - *Data Science Seminar at Wayne State University.* Mar 2024.
  - *2023 IMS International Conference on Statistics and Data Science. Lisbon, Portugal.* Dec 2023.

  - *Professors Yuxin Chen and Yuting Wei's group meeting. University of Pennsylvania (virtual).* Oct 2023.
- Multi-modal contrastive learning adapts to intrinsic dimensions of shared latent variables.
  - (Poster) *The Thirty-ninth Annual Conference on Neural Information Processing Systems. San Diego, CA.* Dec 2025.
- Unraveling projection heads in contrastive learning: insights from expansion and shrinkage.
  - *2024 IMS International Conference on Statistics and Data Science. Nice, France.* Dec 2024.
  - (Poster) *Statistical Foundations of Data Science and their Applications. Princeton, NJ.* May 2023.
- Conformalized survival analysis with adaptive cutoffs.
  - *ICML Workshop on Distribution-free Uncertainty Quantification, joint with Rohan Hore.* Jul 2022.
- Adaptive aggregation with stability for distributed feature selection.
  - *ACM-IMS Foundations of Data Science Conference, Seattle (virtual).* Oct 2020.
- Neural Gaussian Mirror for controlled feature selection in neural networks.
  - *19th International Conference on Machine Learning and Applications.* Dec 2020.
  - *Statistical Learning and Data Mining Lab, Pennsylvania State University.* Oct 2019.