HAOLIN ZOU

+1-929-216-9678 | kyrinzou@gmail.com | Linkedin | Google Scholar
New York, NY 10025, United States

Research Expertise

Research Interest: I have a broad interest in the theoretical foundations of machine learning and AI, primarily using high-dimensional statistical theory both to analyze existing methods and to design new, more reliable algorithms.
Core areas: proportional high dimensional asymptotics, risk estimation and cross validation methods, machine unlearning, data evaluation and interpretability.

Education

- **Columbia University** *Sep 2020 - May 2026*
  Ph.D in Statistics New York, USA
  - Advisor: Arain Maleki, Victor de la Peña.
  - Advanced Coursework: High-Dim Statistics, Decoupling and Self Normalization, Probability (A+), Stats Inference.
- **Columbia University** *Sep 2018 - Dec 2019*
  Master in Actuarial Science New York, USA
- **Peking University** *Sep 2014 - Jul 2018*
  Bachelor of Applied Mathematics, Bachelor of Economics (double major) Beijing, China
  - Strong foundation in mathematical analysis, computational methods, quantitative finance and economics.

Publications

- **Key Publications**
  - Pandey, A., Auddy, A., **Zou, H.**, Maleki, A. and Kulkarni, S. (2026). Gaussian certified unlearning in high dimensions: A hypothesis testing approach. *ICLR2026*. **[Selected for oral presentation, 1.18%]**
  - **Zou, H.** et al. (2025). Certified Machine Unlearning Under High Dimensional Regime. Accepted for publication in *Journal of Machine Learning Research (JMLR)*.
  - **Zou, H.** et al. (2025). Theoretical Analysis of Leave-one-out Cross Validation for Non-differentiable Penalties under High-dimensional Settings. *AISTATS 2025*, PMLR 258:4033-4041. **[Acceptance rate: 31%]**
  - Auddy, A., **Zou, H.**, Rahnama Rad, K. and Maleki, A. (2024) Approximate Leave-one-out Cross Validation for Regression with L1 Regularizers. *IEEE Transactions on Information Theory*, 70(11):8040–8071.
  - Auddy, A., **Zou, H.**, Rahnama Rad, K. and Maleki, A. (2024) Approximate Leave-one-out Cross Validation for Regression with L1 Regularizers. *AISTATS 2024*, PMLR 238:2377-2385. **[Selected for oral presentation, 9%]**
- **Additional Publications**
  - **Zou, H.** et al. (2025). Newfluence: Boosting Model Interpretability and Understanding in High Dimensions. *ICML 2025, Workshop: Assessing World Models: Methods and Metrics for Evaluating Understanding.*
  - Baydil, B., de la Peña, V., **Zou, H.** and Yao, H. (2025). Unbiased estimation of the Gini coefficient. *Statistics and Probability Letters*, 222:110376.
  - de la Peña, V., Gzyl, H., Mayoral, S., **Zou, H.** and Alemayehu, D. (2024). Prediction and estimation of random variables with infinite mean or variance. *Communications in Statistics - Theory and Methods*, 54(1):1-15.
- **Preprints and Working Papers**
  - Tong, H., Ghosh, S., **Zou, H.** and Maleki, A. (2026). Imperfect Influence, Preserved Rankings: A Theory of TRAK for Data Attribution. Under review by *ICML 2026*.
  - **Zou, H.** et al. (2025). A Scalable Formula for the Moments of a Family of Self-Normalized Statistics. Submitted to *Statistics and Probability Letters*.
  - **Zou, H.** et al. (2025). Error Analysis of K-fold Cross Validation Under High-dimensional Settings. *Working Paper.*

Presentations

- **Minghui Yu Memorial Conference 2025** *Apr 2025*
  Presentation on approximate data removal
- **AISTATS 2024** *May 2024*
  Paper S.1 Selected for oral presentation in Oral Session 9 ("Statistics")
- **Minghui Yu Memorial Conference 2024** *Apr 2024*
  Presentation on approximate leave-one-out cross validation
- **INFORMS 2022** *Oct 2022*
  Session chair ("Heavi-tailedness, Dependence and Robustness"), presentation on the bias of Gini coefficient
- **Columbia Statistics Seminar** *Nov 2024*
  Presentation on high dimensional statistics

Teaching Experience

- **Instructor** *Jan 2024 - May 2024*
  Columbia University
  - Weekly recitation for Stat Inference and Modeling.
- **Co-instructor** *Apr 2024*
  Short Course on Decoupling and Self-normalized Inequalities, Georgia Institute of Technology
  - Co-instructed with Prof. Victor de la Peña on the application of decoupling and self-normalization on bandit and sorting problems.
- **Teaching Associate** *Sep 2020 - Now*
  Columbia University
  - Core MS course: Probability Theory (2 semesters), Statistical Inference (4 semesters)
  - Advanced methods: Generalized Linear Models (1 semester)
  - Undergraduate course: Introduction to Statistics (4 semesters)

Industry Experience

- **Voleon Capital Management** *May 2025 - Aug 2025*
  Quantitative Research Intern
  - Conducted quantitative research in bond pricing and comparison among algorithms.

Services

- **Peer Reviewing**
  - Annals of Applied Statistics
  - Journal of the Royal Statistical Society
  - IEEE Transactions on Information Theory
  - Conference on Neural Information Processing Systems (NeurIPS)

Awards and Distinctions

- **Second Prize** *Sep 2017*
  China Contemporary Undergraduate Mathematical Contest in Modeling (CUMCM)
- **First Class Scholarship** *Sep 2017*
  Yizheng Alumni Scholarship
- **First Prize** *Oct 2014*
  China Undergraduate Contest in Physics

Additional Information

Coding and Computing: Python (proficient), R, MS Office, STATA, STAN (advanced)
Languages: English (proficient), Chinese (native), Japanese (proficient), Latin (basic).

Referees

- **Arian Maleki, Professor, Department of Statistics, Columbia University**
  Email: mm4338@columbia.edu
- **Victor de la Peña, Professor, Department of Statistics, Columbia University**
  Email: vhd1@columbia.edu
- **Kamiar Rahnama Rad, Associate Professor, Baruch College, CUNY**
  Email: kamiar.rad@baruch.cuny.edu