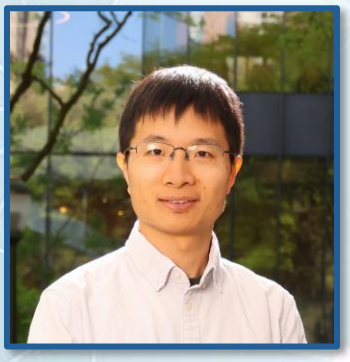
Frontier of Artificial Network

Department of Data Science
香港城市大學
City University of Hong Kong

The ICML 2023 Ranking Experiment: Empirical Performance and Analysis of the Isotonic Mechanism



Invited Speaker
Weijie Su
University of Pennsylvania

Date: October 13, 2025 (Monday)
Time: 20:30 - 21:30 (Hong Kong Time)
Zoom Meeting: 960 4908 9823

Biography

Weijie Su is an Associate Professor in the Wharton Statistics and Data Science Department and, by courtesy, in the Departments of Biostatistics, Epidemiology and Informatics, Computer Science, and Mathematics, at the University of Pennsylvania. He is a co-director of Penn Research in Machine Learning (PRiML) Center. His research interests span statistical foundations of generative AI, privacy-preserving machine learning, high-dimensional statistics, and optimization. He serves as an associate editor of Journal of the American Statistical Association, Journal of Machine Learning Research, Annals of Applied Statistics, Harvard Data Science Review, Foundations and Trends in Statistics, Operations Research, and Journal of the Operations Research Society of China, and he is currently on the ICML 2026 Organizing Committee as the Scientific Integrity Chair. His work has been recognized with several awards, such as the Stanford Anderson Dissertation Award, NSF CAREER Award, Sloan Research Fellowship, IMS Peter Hall Prize, SIAM Early Career Prize in Data Science, ASA Noether Early Career Award, ICBS Frontiers of Science Award in Mathematics, IMS Medallion Lectureship, and Outstanding Young Talent Award in the 2025 China Annual Review of Mathematics. He is a Fellow of the IMS.

Abstract

The rapid growth of submissions to ML/AI conferences such as NeurIPS and ICML---for example, NeurIPS received 2,425 submissions in 2016 compared to a projected 25,000 submissions in 2025, with an annual growth rate of 30%---has led to a substantial decline in peer review quality in ML/AI. To address this issue, in 2023 we conducted an experiment at ICML, asking authors with multiple submissions to rank their papers based on perceived quality. In total, we received 1,342 rankings covering 2,592 submissions. The experiment was subsequently conducted at ICML 2024 and 2025, and more recently at NeurIPS 2025. We present empirical analyses of how author-provided rankings could be leveraged to improve peer review processes using the Isotonic Mechanism, which calibrates raw review scores using author-provided rankings. Our analysis demonstrates that these ranking-calibrated scores outperform raw review scores in estimating the ground truth "expected review scores" in terms of both squared and absolute errors.