

YUWEI CHENG

Chicago, IL | 872-225-7727 | yuweicheng@uchicago.edu | <https://yuwei-cheng.github.io/>

EDUCATION

PhD **University of Chicago** | Department of Statistics GPA 3.8/4.0, **2021-2026, Feb** (Expected)
Best Poster Award (15 out of 116) at Midwest Machine Learning Symposium 2024
Spring Quarter Consulting Cup Winner (Best team in the Statistical Consulting Program)
BS **National University of Singapore** | Department of Statistics GPA 4.9/5.0, 2016-2020
BA **National University of Singapore** | Department of Economics GPA 4.9/5.0, 2016-2020
Dean's List Recipient (GPA ranked in the top 5% for 4 consecutive years in both departments)
Science and Technology Scholarship (Full scholarship covering tuition and living expenses for 4 years)

RESEARCH INTEREST

AI for Social Good, Preference Alignment, Online Learning, Reinforcement Learning, Algorithmic Game Theory

INDUSTRIAL EXPERIENCE

Data Scientist Research Intern at Trust and Safety Team of **YouTube, Google** Jun 16, 2025 – Sep 19, 2025

- Researched and designed evaluation dimensions and metrics for AI-generated metric report summaries, contributing to the automation of the metric reporting process.
- Developed a Python library implementing the proposed LLM-based rater with prompting strategies that evoke chain-of-thought reasoning for improved interpretability and diagnostic insights.
- Conducted comprehensive sensitivity analyses to evaluate metric fidelity and collected human-written summary samples to align automated ratings with human preferences.
- Generated scorecards for AI-produced summaries and provided product improvement recommendations based on the evaluation results.

PUBLICATIONS

*indicates co-first author

[W3] Yi, B., Liu, Q., **Cheng, Y.**, & Xu, H. (2025+). Escaping Model Collapse via Synthetic Data Verification: Near-term Improvements and Long-term Convergence. Under Review

[W2] Wang, Q., **Cheng, Y.**, Song, A., Keedy, S., Berman, M., & Feamster, N. (2025+). Can LLMs Address Mental Health Questions? A Comparison with Human Therapists. Under Review

[W1] Wang, G, Cheng, **Y.**, & Xu, H. (2025+). Exponential Convergence of Probabilistic Bisection Algorithm with Noisy Labels. Under Review

[P5] **Cheng, Y.**, Zhao, Z., & Xu, H. (2025). Learning Personalized Ad Impact via Contextual Reinforcement Learning under Delayed Rewards. NeurIPS 2025: The Thirty-Ninth Annual Conference on Neural Information Processing Systems

[P4] Yao, F*, **Cheng, Y.***, Wei, Er., & Xu, H. (2025). Single-Agent Poisoning Attacks Suffice to Ruin Multi-Agent Learning. ICLR 2025: Proc. 13th International Conference on Learning Representations, 2025.

[P3] **Cheng, Y.**, Yao, F, Liu, X., & Xu, H. (2025). Learning from Imperfect Human Feedback: a Tale from Corruption-Robust Dueling. ICLR 2025: Proc. 13th International Conference on Learning Representations, 2025.

[P2] Quaye, S. E. D., **Cheng, Y.**, Tan, R. K. J., Koo, J. R., Prem, K., Teo, A. K. J., & Cook, A. R. (2023). Application of the network scale-up method to estimate the sizes of key populations for HIV in Singapore using online surveys. *African Journal of Reproduction and Gynaecological Endoscopy*, 26(3), e25973.

[P1] **Cheng, Y.**, et al. (2022). Estimates of Japanese encephalitis mortality and morbidity: a systematic review and modeling analysis. *PLOS Neglected Tropical Diseases*, 16(5), e0010361.

SKILLS & SERVICES

- Programming:** Python, shell scripting, R
- Data Analysis:** Bayesian hierarchical modelling, Generalized linear models, Gradient boosting, SVM, principal component analysis, neural network, reinforcement learning, sentiment analysis, large language models
- Teaching Experience:**
- Instructor for Elementary Statistics with 45+ students Winter 2025
 - Teaching assistant for Applied Regression Analysis Fall 2024
 - Teaching assistant for Introduction to Data Science Fall, Winter 2021-2024
- Professional Services:**
- Reviewer for ICML, NeurIPS 2025
 - Reviewer for The American Statistician 2024
 - Reviewer for BMJ Global Health 2023

TALKS

- Towards Development and Assessment of Evaluation Metrics for AI Generated Metric Trend Summary**
- Google PhD Summit Poster Session (July, 2025)
- Learning Personalized Ad Impact via Contextual Reinforcement Learning under Delayed Rewards**
- Salesforce AI Research Future Forum (May, 2025)
 - University of Chicago, research talk at Department of Computer Science Theory Lunch (May, 2025)
- Intrinsic Efficiency-Robustness Trade-offs in Modern Online Learning Algorithms**
- National University of Singapore, Department of Computer Science (April, 2025)
 - Nanyang Technological University, Agent Mediated Intelligence Research Group (April, 2025)