

XIANGRUI (CINDY) KONG

(217) 819-6849 | kxrcindy@gmail.com

EDUCATION

Purdue University

Ph.D. in Mathematics

Department of Mathematics

West Lafayette, IN

Aug 2024–Aug 2028 (Expected)

Carnegie Mellon University

M.S. in Information Systems Management – Business Intelligence and Data Analytics

Heinz College of Information Systems and Public Policy

Pittsburgh, PA

Aug 2022–Dec 2023

Shanghai University of Finance and Economics

B.S. in Data Science

School of Statistics and Management

Shanghai, China

Sep 2018–Jul 2022

WORKING PAPERS & PUBLICATIONS

- [1] **Xiangrui Kong**, Yueqi Wang, Haoyang Zheng, Weijian Luo, and Guang Lin. Ultra Fast PDE Solving via Physics Guided Few-step Diffusion. *NeurIPS*, 2026. [paper]
- [2] Haoyang Zheng, Xinyang Liu, **Xiangrui Kong**, Nan Jiang, Zheyuan Hu, Weijian Luo, Wei Deng, and Guang Lin. Ultra-Fast Language Generation via Discrete Diffusion Divergence Instruct. *International Conference on Learning Representations (ICLR)*, 2026. [paper]
- [3] **Xiangrui Kong**, Haoyang Zheng, and Guang Lin. BLADE: Bayesian Langevin Active Discovery with Replica Exchange for Identification of Complex Systems. *Machine Learning: Science and Technology*, 2026. [paper]
- [4] Song Wei, **Xiangrui Kong**, Sarah A. Huestis-Mitchell, Shixiang Zhu, Yao Xie, Alinson Santos Xavier, and Feng Qiu. Detecting Electricity Service Equity Issues with Transfer Counterfactual Learning on Large-Scale Outage Datasets. *Causal Representation Learning Workshop at NeurIPS*, 2023. [paper]

RESEARCH & WORK EXPERIENCE

Purdue University, West Lafayette, IN

Research Assistant; Advisor: Prof. Guang Lin

Aug 2024 – Present

- Developed Phys-Instruct, a **scientific machine learning** framework for distilling multi-step **diffusion models** into physics-guided one/few-step PDE solvers, achieving orders-of-magnitude faster inference and over **8× lower PDE error**.
- Co-developed DiDi-Instruct, a **Generative AI** framework for accelerating **diffusion language models** through distribution matching and integral KL-divergence minimization, enabling high-quality few-step generation with up to **64× acceleration**.
- Developed reward-based post-training methods for **text-to-image generation** and **multimodal generative models**, integrating reward optimization with **one-step discrete diffusion** for efficient generation.
- Developed **Bayesian machine learning** methods combining **MCMC sampling**, replica exchange, and active learning for uncertainty-aware system discovery, reducing measurement requirements by up to 60%.

Carnegie Mellon University, Pittsburgh, PA

Research Assistant; Advisor: Prof. Shixiang Woody Zhu

Jan 2023 – Dec 2023

- Developed a spatio-temporal **machine learning** framework for **power systems** outage prediction under extreme weather using **neural differential equations**.
- Modeled power failure and restoration dynamics with latent processes to enable **uncertainty quantification** and interpretable forecasting of large-scale power outages.
- Developed a **causal inference** framework with **transfer learning** to study energy equity and quantify the impact of demographic factors on power outage duration.

Carnegie Mellon University, Pittsburgh, PA

Open Science Graduate Research Consultant

Jan 2023 – July 2023

- Independently held consultations for a wide range of groups, including librarians, researchers from across disciplines, students, and other units on campus to provide data consultations on their projects.
- Conducted sentiment analysis on group interviews about open science framework using NLP techniques.

ByteDance, Shanghai, China

Intern, Search Ads Data Science, TikTok Monetization Product

Feb 2022 – July 2022

- **Reports and insights:** Conducted reports on advertiser scoring model, spark ads profit lift space; effectively segmented advertisers with multidimensional Indicators and came up with efficient toggle rate to indicate sending efficiency.

- **Experiment and attribution analysis:** Dived into and implemented statistical root cause analysis tools to locate structural and quality reason of data fluctuation, significantly helped in analyzing cost per action experiments.
- **Data construction:** Created and managed 4 dashboards (around 15 interfaces) with different functions at the search toggle ads business launching stage for inner use.
- **Cross-functional and cross-cultural communication:** Collaborated with local and foreign PM & RD in daily work on conducting analysis and data construction.

United Automotive Electronic Systems, Co., Ltd., Shanghai, China

Summer 2020

Data Science Intern, TEF3 Division

- **Statistical process control:** Designed production monitoring dashboards and control charts with Oracle to reproduce the quality control system; helped engineers to entrust the package software and understand the underlying statistical principles.
- **Tableau:** Mastered Tableau software and assisted in company-wide Tableau workshop and training; designed monitoring dashboard for new product line after communicating with cross functional group engineers; continuously updated essential charts to help engineers grasp production line issues in real-time.

PRESENTATIONS

Ultra Fast PDE Solving via Physics Guided Few-step Diffusion	[video]
Crunch Seminar, Brown University, 2026	
Langevin-Assisted Bayesian Active Learning for Physical Discovery	[poster]
IMSI Workshop on UQ and Trustworthy AI Algorithms for Complex Systems and Social Good, 2025	
Large-Scale Data Analysis for Fairness in Grid Resilience Enhancement	[poster]
CMU Energy Week, 2023	
Compartmental Modeling for Power Outage Prediction using Neural Stochastic Differential Equations	[poster]
YinzOR Conference, 2023	

PROFESSIONAL SERVICE

Conference & Workshop Reviewing	
International Conference on Machine Learning (ICML)	2026
Advances in Neural Information Processing Systems (NeurIPS)	2026
KDD AI4Sciences Track	2026
SIGKDD International Workshop on Mining and Learning from Time Series	2023, 2024
Journal Reviewing	
Journal of Agricultural, Biological, and Environmental Statistics	2023
IEEE Transactions on Power Delivery	2024

SKILLS

Programming Languages: Python, R, Java | **Machine Learning:** Scikit-learn, PyTorch
Database and Big Data: SQL, Scala Spark | **Data Visualization:** Tableau, SPSS, Stata
Statistical Analysis: A/B Testing, Regression, Clustering, PCA, Attribution Analysis