Ruqi Bai

Seeking opportunities in ML research, focusing on robust and causality-inspired AI.
Email:bai116@purdue.edu | Telephone: +1 (765)714-5086
Website: http://ruqibai.netlify.app | Linkedin: https://www.linkedin.com/in/ruqi-bai/

SUMMARY

A Ph.D. candidate in Probabilistic and Understandable Machine Learning with extensive experience in developing scalable AI and distributed systems at Purdue. Skilled in machine learning research, system optimization, and experimental development using Python and industry-standard tools. Demonstrates strong collaboration and communication abilities, effectively working with diverse teams to solve complex ML challenges.

EDUCATION

Purdue University, West Lafayette, Indiana   Aug. 2019 — May. 2025 (Expected)
Ph.D. in Probabilistic and Understandable Machine Learning Lab, ECE
Advisor: David I. Inouye

RESEARCH INTERESTS

Robust and Trustworthy AI: Causal Machine Learning; Domain Generalization; Counterfactual Fairness; Distribution Matching
Data-Centric AI Counterfactual Fine-tuning; Datasets and Benchmarks for ML Applications
Deep Generative Models: VAE; Transformer

PUBLICATIONS

* *denotes equal contribution.*
[1] Benchmarking Algorithms for Federated Domain Generalization
Ruqi Bai, Saurabh Bagchi, David I. Inouye
International Conference on Learning Representations (ICLR) (**Spotlight, 5% acceptance**), 2024
[2] Towards Characterizing Domain Counterfactuals for Invertible Latent Causal Models
Ruqi Bai*, Zeyu Zhou*, Sean Kulinski*, Murat Kocaoglu, David I. Inouye
International Conference on Learning Representations (ICLR), 2024.
[3] Counterfactual Fairness by Combining Factual and Counterfactual Predictions
Zeyu Zhou, Tianci Liu, **Ruqi Bai**, Jing Gao, Murat Kocaoglu, David I. Inouye
Neural Information Processing Systems (NeurIPS), 2024.
[4] Improving Practical Counterfactual Fairness with Limited Causal Knowledge
Zeyu Zhou, **Ruqi Bai**, and David I. Inouye
ICLR Workshop on Navigating and Addressing Data Problems for Foundation Models, 2024
[5] Towards Characterizing Domain Counterfactuals for Invertible Latent Causal Models
Sean Kulinski, Zeyu Zhou, **Ruqi Bai**, Murat Kocaoglu, David I. Inouye
NeurIPS Workshop on Causal Representation Learning, 2023

TECHNICAL REPORTS (UNDER REVIEW)

[6] Fewshot Counterfactual Matching in Improving Domain Generalization **Ruqi Bai**, Yao Ji, Zeyu Zhou, David I. Inouye
Short intro: We proved a counterfactual matching constraint to improve model robustness in domain generalization.
[7] FedLOE: Federated Domain Generalization via Local Overfitting and Refitting
Ruqi Bai, David I. Inouye
Short intro: We demonstrate that reducing communication frequency can effectively enhance domain generalization in the Federated learning (FedDG) performance, surpassing their centralized counterparts.

WORKING PROJECTS

[8] StarCraft Motion: A Multiple Object Tracking Dataset with Ground Truth Intent and Complex Behavior
Ruqi Bai, James Z. Hare, Nicholas R. Waytowich, David I. Inouye
Short intro: We constructed a 139.6 million frame dataset focusing on adversarial behaviors for Multiple Object Tracking (MOT), and built a transformer-based model for effective tracking prediction.
[9] Counterfactual Pairing for Robustness in Large Language Models, A data-centric approach.
Ruqi Bai, David I. Inouye
Short intro: Domain counterfactual sentence pairs can robustify pretrained LLM models, reducing hallucinations.

WORKING EXPERIENCE

Baidu, Inc Jan. 2016 — Jun. 2019
Senior Site Reliability Engineer

- Led the design and development of Baidu Phoenix Nest's first AI distributed tracing and failure location system, reducing MTTR from 45 minutes to 17 minutes and achieving millions in cost savings annually.
- Assisted in building Baidu Phoenix Nest's large-scale tracing infrastructure, enabling distributed log splicing across thousands of servers within one minute.
- Contributed to the Baidu Search System Assurance during the 2019 Spring Festival Gala, managing billions of page views in one minute.

SKILL

Python packages: PyTorch, numPy, sciPy, matplotlib, scikit-learn, pandas, wandb
Systems: operating systems, distributed systems
Cloud Computing: cloud computing
Other tools: linux, LaTeX, Git

TEACHING EXPERIENCE

Purdue University
Teaching Assistant

- ECE 57000 Artificial Intelligence
- ECE 50024 Machine Learning
- ECE 69500 Big Data for Reliability and Security