Ruqi Bai

Email:bairuqi@gmail.com | Telephone: +1 (765)714-5086
Website: http://ruqibai.netlify.app | Linkedin: https://www.linkedin.com/in/ruqi-bai/

SUMMARY

Ph.D. candidate at Purdue University specializing in both predictive AI and generative AI with expertise in deep machine learning models such as Transformers, CNNs, VAEs, GANs. Authored multiple research contributions, including a spotlight publication (first author and sole student author) at ICLR and poster presentations at NeurIPS and ICLR. Prior to the Ph.D., as a Senior Site Reliability Engineer at Baidu, developed the first machine learning-based system for abnormality detection and clustering, solving large-scale, real-world challenges. Now equipped with cutting-edge AI expertise and a strong research foundation, ready to return to industry to tackle impactful, real-world problems through innovative AI solutions.

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

Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China — Aug. 2012 — Jun. 2016
B.S. in Applied Physics

RESEARCH INTERESTS

Data Centric AI: datasets and benchmarks for ML applications;
Robust and Trustworthy AI: causal machine learning; (federated) domain generalization; counterfactual fairness;
Machine Learning Applications: transformer (large language model); anomaly detection;

PUBLICATIONS

* *denotes equal contribution.*

Data Centric AI:

[1] Benchmarking Algorithms for Federated Domain Generalization
Ruqi Bai, Saurabh Bagchi, David I. Inouye
International Conference on Learning Representations (ICLR) (**Spotlight, 5% acceptance**), 2024

Robust and Trustworthy AI:

[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)

Robust and Trustworthy AI:

[6] Fewshot Counterfactual Matching in Improving Domain Generalization
Ruqi Bai, Yao Ji, Zeyu Zhou, David I. Inouye

[7] FedLOE: Federated Domain Generalization via Local Overfitting and Refitting
Ruqi Bai, David I. Inouye

WORKING PROJECTS

Deep Generative Models and Data Centric AI:
[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: Constructed a 139.6 million frame dataset focusing on complex 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

Senior Software Engineer **Baidu Inc.** Beijing, Jan. 2016 — Jun. 2019

- 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 1 minute.
- Contributed to the Baidu Search System Assurance during the 2019 Spring Festival Gala, managing billions of page views in one minute.

SKILL

ML models: Transformers, CNNs, VAEs, GANs;
Python packages: PyTorch, NumPy, SciPy, matplotlib, scikit-learn, Pandas, WandB
Cloud Computing: Hadoop, Spark, Kubernetes
Databases: MySQL,MongoDB: HBase, Redis
Other tools: LaTeX, Linux, Git, Shell

CORE COURSES AT PURDUE

Probability and Statistics	**Machine Learning**	**Optimization and Algorithm**
ECE 60000 Random Variables	ECE 57000 Artificial Intellengence	MGMT 69000 Stoch Iter Methods
MA 51100 Linear Algebra	ECE 59500 Machine Learning	MA 59800 Convex Optimization
CS 59200 Information Theory	ECE 69500 Generative Models	ECE 60800 Comp Models & Methods
ECE 69500 Probabilistic Causal Inference	CS 59200 Probabilistic Machine Learning	
	ECE 66200 Pattn Recog & Decis Proc	

TEACHING EXPERIENCE

Teaching Assistant **Purdue University**
ECE 57000 Artificial Intelligence
ECE 50024 Machine Learning
ECE 69500 Big Data for Reliability and Security