

# Ruqi Bai

Email: [bairuqi@gmail.com](mailto: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, diffusion models, CNNs, VAEs, GANs, etc. 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 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

---

<b>Purdue University</b> , West Lafayette, Indiana	Aug. 2019 — Aug. 2025 (expected)
Ph.D. in Probabilistic and Understandable Machine Learning Lab, ECE	
Advisor: <a href="#">David I. Inouye</a>	
<b>Nanjing University of Posts and Telecommunications</b> , 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); diffusion 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 (SUBMITTED)

---

### **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

**Deep Generative Models:**

[8] StarCraft Motion: A Motion Forecasting Dataset with Ground Truth Intentions and Strategic 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.

**WORKING PROJECTS**

---

[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**

---

Site Reliability 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.

**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	

**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:** L<sup>A</sup>T<sub>E</sub>X, Linux, Git, Shell

**TEACHING EXPERIENCE**

---

Teaching Assistant

**Purdue University**

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