Arezoo Rajabi

Email: arezoorajabi2@icloud.com
LinkedIn: https://linkedin.com/in/arezoo-rajabi
HomePage: https://arezoo-rajabi.github.io

Senior AI/ML Engineer
Technical leader specializing in machine learning verification and validation, statistical modeling, AI evaluation, multimodal AI, and production ML systems. 10+ years of experience developing and evaluating machine learning systems, including 3+ years building and validating LLM/GenAI applications. Experienced in leading technical solution design, developing comprehensive evaluation frameworks and automated pipelines, defining performance metrics, curating large-scale datasets, and conducting statistical analysis, benchmarking, annotation, and error analysis to identify edge-case failures and improve system reliability. PhD in Computer Science with research and publications in trustworthy AI, adversarial machine learning, privacy, and robust ML.

AREAS OF EXPERTISE

Model Validation & Verification, ML Evaluation Pipeline Design & Benchmarking (LLM/RAG/Agentic AI Systems), Statistical Modeling, AI Safety, Trustworthy AI, LLM post-training, Inference Optimization

SKILLS

Programming:	Python, SQL, R, MATLAB
ML Systems & Frameworks:	PyTorch, Hugging Face Transformers, Fine-Tuning, JAX, MLflow
LLM / Generative AI:	LLMs, Agentic AI, RAG, Ranking, Embeddings, Prompt Optimization
ML Evaluation & Validation:	Statistical Analysis, Experimental Design, Hypothesis Testing, Error Analysis, Failure-Mode Analysis, Edge-Case Testing, Benchmarking, Regression Testing, A/B Testing, KPI & Metric Design
Data & Statistical Modeling:	Statistical Modeling, Dataset Creation, Data Extraction, Data Annotation, Feature Engineering, Classification, Regression, Forecasting, Anomaly Detection
Data & Engineering:	PySpark, GCP, Hadoop, MLflow, Git/GitHub, CI/CD
LLM Training & Evaluation:	LLM quantization (LoRA, qLORA), LLM post-training (RLHF, DPO), LLM Evaluation Harness, LLM Benchmarking, RAGAS, DeepEval, LLM-as-a-Judge for Synthetic Data Generation, Evaluation and Verification, Prompt Engineering and Evaluation, Robustness Evaluation, Trustworthy AI

WORK EXPERIENCE

Senior ML/AI Engineer — Dec 2022 - Present
Wells Fargo, — *SF, CA, USA*

- **Enterprise LLM Chatbot – 27,000+ users, $6M ROI**
  - Led the technical team in designing and developing comprehensive LLM evaluation pipelines to assess response quality, factuality, relevance, and hallucinations.
  - Created benchmark datasets to compare models and prompts and identify failure modes.
  - Developed automated prompt generation to improve response quality and reduce hallucinations
  - Implemented offline and human-in-the-loop evaluation to guide model improvements.
  - Established quality metrics and production monitoring to detect performance degradation.
- **Multimodal AI for Suspicious Behavior & Policy Violation Detection – $1.5M+ ROI**
  - Led technical solution design for a production AI system leveraging vision-language models to detect safety risks and policy violations.
  - Built video data, annotation, and segmentation pipelines for model evaluation and deployment.

  - Reduced 80% of inference/token costs through intelligent frame selection and automated video segmentation.
  - Improved model grounding and detection precision through context-aware metadata and prompting.
- **Model Validation and Verification**
  - Designed end-to-end evaluation and validation frameworks for production ML and AI systems, defining performance metrics, validation criteria, test datasets, and failure-analysis methodologies for anomaly detection, forecasting, conversational AI, and multimodal models.
  - Developed data extraction, curation, annotation, benchmarking, and statistical analysis pipelines to evaluate model performance, reliability, robustness, and edge-case behavior.
  - Conducted quantitative model validation through statistical and error analysis, threshold optimization, regression testing, and comparative experimentation to identify failure modes and guide model improvements.

Postdoctoral Scholar Feb 2021 - Dec 2022
NSL Lab, University of Washington *Seattle, WA, USA*

- Multi-Domain Trojan Detection: Proposed a multi-domain Trojan detection framework across text, image, and audio models, achieving 85% detection accuracy through extensive benchmarking under adversarial settings.
- Privacy-Preserving RL Algorithm: Developed and evaluated privacy-preserving reinforcement learning algorithms, conducting extensive experiments to quantify privacy-utility tradeoffs under adversarial settings.
- Federated Learning for Trojan Prevention: Proposed a federated learning approach to counteract Trojan samples during the training phase.
- Led privacy and adversarial-robustness research projects, guiding research direction and mentoring master's students research

EDUCATION

Ph.D. in Computer Science, Oregon State University 2014 - 2021

Master of Computer Engineering, Sharif University of Technology 2011 - 2013

Bachelor of Computer Science, Sharif University of Technology 2005 - 2010

SELECTED PUBLICATIONS

1. A. Rajabi, S. Asokraj, F. Jiang, L. Niu, B. Ramasubramanian, J. Ritcey, R. Poovendran, MDTD: A Multi-Domain Trojan Detector for Deep Neural Networks, ACM Conference on Computer and Communications Security (ACM CCS), Sep. 2023.

2. J. Jia, Z. Yuan, D. Sahabandu, L. Niu, **A. Rajabi**, B. Ramasubramanian, B. Li, R. Poovendran, FLGAME: A Game-theoretic Defense against Backdoor Attacks In Federated Learning, Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS), Sep 2023 (https://neurips.cc/virtual/2023/poster/70499).

3. A. Rajabi, D. Sahabandu, L. Niu, B. Ramasubramanian, R. Poovendran, LDL: A Defense for Label-Based Membership Inference Attacks, ACM Asia Conference on Computer and Communications Security (AsiaCCS), July 2023 (**7% acceptance rate**).

4. A. Rajabi, B. Ramasubramanian, A. Marruf, R. Poovendran, Privacy Preserving Reinforcement Learning Beyond Expectation, Accepted in 61st IEEE Conference on Decision and Control, 2022.

5. A. Rajabi, M. Abbasi, R. B. Bobba, K. Tajik, Adversarial Images Against Super-Resolution Convolutional Neural Networks for Free, Privacy Enhancing Technology Symposium (PETS), 2022.

PROFESSIONAL SERVICES

- Adjunct Faculty at University of Washington 2022- 2024
- Organizer at The Trojan Detection Challenge (LLM Edition),NeurIPS 2023
- Organizer at Trojan Detection Challenge, NeurIPS 2022