JIAXUAN WANG

Tel: (734) 834-7996
E-mail: jiaxuan@umich.edu
Website: nathanwang000.github.io

SKILLS

Machine Learning: Large Language Model post training, Transformer architectures, Attention mechanisms, Direct Preference Optimization (DPO), Feature attribution, Model interpretability, Reinforcement learning, Expert iteration, Out of domain generalization, Non-convex optimization
Languages & Frameworks: Python, PyTorch, Lean 4, C++, Javascript, R, LangGraph, LangChain
Agents: Retrieval augmented generation (RAG), Reasoning and Acting agents (ReAct), Harnesses for maintaining long running multi-agentic systems, LLM as a Judge

EDUCATION

Ph.D. at the University of Michigan, Ann Arbor Sep 2017-April 2022

Computer Science and Engineering
Advisor: Jenna Wiens
GPA: 4.00 / 4.00
Research interests: Model interpretability; Time-series analysis; Non convex optimization; Out of distribution generalization; Reinforcement learning; Causal inference; Basketball analytics

Bachelors of Science in Engineering, Ann Arbor Sep. 2013 - Dec. 2016

Computer Science major and Mathematics minor
GPA: 3.96 / 4.00
Directed research: Computer vision; Basketball analytics

EMPLOYMENT

Senior AI research scientist @ GE Healthcare April 8 2024 - Present

- **Multi-Agent Systems for High-Stakes Clinical Care:** Designed, implemented, and deployed a ReAct-based multi-agent LLM system for perinatal care to reduce clinical burnout. Engineered RAG pipelines for hospital protocols and real-time patient summarization, accelerating clinical handoffs and allowing staff to maximize direct patient care. My work is showcased in the HLTH conference (2025 press release, multimodal agentic AI vision).
- **Inference & Throughput Optimization:** Optimized LLM inference call patterns to minimize end-to-end latency for real-time streaming patient data.
- **Evaluation & Reliability Guardrails:** Built automated evaluation pipelines using LLM-as-a-judge and validated against clinicians to ensure robust, deterministic performance in mission-critical environments. Implemented multi-turn guardrails, staleness checks, and a high-precision medical-context PII redaction pipeline.
- **Technical ML Leadership:** Led cross-functional data stewardship and AI governance initiatives, providing architectural design guidance for enterprise digital assistants and ML models safely deployed in live clinical settings.

Research scientist in machine learning @ Meta Jun. 6 2022 - April 1 2024

- **Adversarial Reinforcement Learning:** Co-developed and deployed Predictive Response Optimization (PRO), a large-scale reinforcement learning system used in production to dynamically optimize platform friction and deter scraping across Facebook and Instagram. *(*Published at USENIX Security 2025*)*.
- **Distribution Shift & Model Diagnostics:** Architected a rigorous feature attribution framework to monitor, debug, and correct real-time distribution shifts for the production RL agent. Ensured statistical stability and model robustness in a highly non-stationary, adversarial environment (technical report).

- **LLM Fine-Tuning & Synthetic Data (Duo-Llama):** Engineered 'Duo-Llama', an advanced fine-tuned LLM system leveraging machine translation to augment training data and drastically reduce human labeling costs. Outperformed the legacy production classifier on 4 core languages while successfully scaling capabilities to handle 13 new, low-resource languages.

Research Intern @ Microsoft Research Jun. 1 - Aug.21 2020
Mentor: Scott Lundberg
- **Advanced Model Interpretability:** Developed *Shapley Flow*, a novel graph-based feature attribution algorithm that mathematically unifies and resolves the theoretical limitations of three foundational interpretability methods. *(*published at AISTATS*)*.
- **Algorithmic Transparency:** Engineered a rigorous mathematical framework to trace the flow of feature importance through complex, non-linear machine learning architectures, providing a systematic approach to debug and validate black-box predictive models.

Software Engineering Intern, NLP group, Bloomberg L.P. (New York) Jun. 7 - Aug.19 2016
Mentors: Konstantine Arkoudas and Srivas Prasad
- **Financial NLP & Parsing:** Implemented and optimized a highly efficient C++ algorithm based on Probabilistic Context-Free Grammar (PCFG) for natural language parsing within the financial chart domain, extracting structured signals from unstructured natural language query to plot on the Bloomberg Terminal.

Research Assistant @ University of Michigan Oct. 2014 - Jan. 2016
Mentor: Jia Deng
- **Model Architecture Research:** Researched and developed rotation-equivariant neural networks, engineering specialized network architectures that mathematically encode spatial symmetries to improve generalization.
- **Data Engineering & Feature Extraction:** Built scalable feature extraction pipelines and managed large-scale, crowd-sourced dataset collection infrastructure using Amazon Mechanical Turk.

PUBLICATIONS (* denotes equal contribution)

- Predictive Response Optimization: Using Reinforcement Learning to Fight Online Social Network Abuse
Garrett Wilson, Geoffrey Goh, Yan Jiang, Ajay Gupta, **Jiaxuan Wang**, David Freeman, Francesco Dinuzzo
USENIX Security 2025

- Learning Concept Credible Models for Mitigating Shortcuts
Jiaxuan Wang, Sarah Jabbour, Maggie Makar, Jenna Wiens
Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS), 2022

- Shapley Flow: A Graph-based Approach to Interpreting Model Predictions
Jiaxuan Wang, Jenna Wiens, Scott Lundberg
Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS), 2021

- AdaSGD: Bridging the gap between SGD and Adam
Jiaxuan Wang, Jenna Wiens
arXiv preprint, 2020

- Relaxed Parameter Sharing: Effectively Modeling Time-Varying Relationships in Clinical Time-Series
Jeeheh Oh*, **Jiaxuan Wang***, Shengpu Tang, Michael Sjoding, Jenna Wiens
In Proceedings of the 4th Machine Learning for Healthcare Conference, 2019

- Learning Credible Models
Jiaxuan Wang, Jeeheh Oh, Haozhu Wang, Jenna Wiens
ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2018

- The Advantage of Doubling: A Deep Reinforcement Learning Approach to Studying the Double Team
Jiaxuan Wang*, Ian Fox*, Jonathan Skaza, Nick Linck, Satinder Singh, Jenna Wiens
MIT Sloan Sports Analytics Conference, 2018

- Learning to Exploit Invariances in Clinical Time-Series Data using Sequence Transformer Networks
Jeeheh Oh, **Jiaxuan Wang**, and Jenna Wiens
In Proceedings of the 4th Machine Learning for Healthcare Conference, 2018

- HICO: A Benchmark for Recognizing Human-Object Interactions in Images
Yu-Wei Chao, Zhan Wang, Yugeng He, **Jiaxuan Wang**, Jia Deng
International Conference on Computer Vision (ICCV) 2015