

SAM JOHNSON

AI/ML Researcher and Engineer

sj110.pages.iu.edu || github.com/sej2020 || sejohnson2020@outlook.com || (812) 341-7037

EDUCATION

Luddy School of Informatics, Indiana University - Bloomington, IN

Bachelor of Science in Data Science, May 2024

Master of Science in Data Science, May 2025

Major GPA: 4.00/4.00

GPA: 4.00/4.00

RESEARCH

Greedy-Advantage-Aware RLHF

Blog post on LessWrong

- Engineered modification of the LLM fine-tuning RLHF loop that combats reward-hacking behavior by deterring agents from generating disproportionately high-reward tokens during training
- Used second-order network analysis methods to detect the presence of a sharp parameter topology in reward-hacking agents—detecting the co-occurrence of these model properties is a novel result

Diffusion Models for Carbon Reduction in Time Series

Academic Paper: Targeting 51st International Conference on Very Large Data Bases

- Designed a time series data storage system from a score-based diffusion model for conditional data generation to reduce the carbon footprint of conventional time series data storage and processing methods
- Integrating our model into a popular database with a low-friction interface for model training and querying

Indirect Prompt Injection Attacks on Web Navigation Agents

Academic Paper and Demo: Targeting Empirical Methods in Natural Language Processing Conference 2025

- Investigating significant security vulnerability in LLM-integrated web navigation applications—indirect prompt injection—a method by which a malicious actor could control the actions of a remote web navigation agent
- Demonstrated this attack in presentation of an optimized adversarial trigger embedded in a hidden malware link

Are They What They Claim: A Comprehensive Study of Ordinary Linear Regression Among the Top Machine Learning Libraries in Python

Paper presented at 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

- Authored a comprehensive survey of current implementations of the Least Squares method in popular Python libraries (TensorFlow, PyTorch, scikit-learn, MXNet) to give users actionable information about state of ML
- Conducted original experiments to analyze the runtime across platforms, space requirement, performance over big data, and strength of model implementation of these popular libraries

TECHNICAL SKILLS

Programming Languages: Python (*adv*), R (*int*), SQL (*int*), Ruby (*int*), C (*beg*)

Machine Learning Software: PyTorch, TensorFlow, HuggingFace Transformers, scikit-learn, pandas, NumPy

Dev Tools: Docker, Git, Google Cloud Platform

WORK EXPERIENCE

Sustainable Computing Research, Indiana University – Bloomington, Indiana

January 2024 – Present

Graduate Research Assistant

- End-to-end execution of AI research projects for green applications, including the diffusion model database and an LSTM interpolation model for a non-profit collaborator, improving data fidelity by upwards of 20%

Groundwork Innovations, Inc. – Zionsville, IN

May 2023 – January 2024

Software and Data Engineering Intern

- Developed, tested, and maintained new features of a web application using Ruby on Rails framework to support CRM and lead qualification startup with client-base of 150+ contractors in home-improvement industry
- Established several data analysis pipelines, including an exploration aimed to increase client lead conversion by 20% and a prediction of new pricing structure expected to increase company ARR by upwards of 6%

PROFESSIONAL DEVELOPMENT:

ETH Zurich's Computer Architecture Course, BlueDot Impact's AI Safety Fundamentals Course, UC Berkeley's Supervised Program for Alignment Research