Kanishk Jain

Postdoctoral Research Scientist in AI | Ph.D. in Physics

www.kanishkjain.com | linkedin.com/in/kanishkbjain | github.com/kanishkbjain | kanishkbjain@gmail.com | (512)200-6091
scholar.google.com/citations?user=XrnzkXQAAAAJ

EDUCATION

Emory University (Atlanta, GA) | Ph.D. Physics (Focus: Applied ML) 2017 – 2024
Indian Institute of Science Education and Research | B.S., M.S. Physics 2012 – 2017

PROFESSIONAL EXPERIENCE

Emory University: Atlanta, GA Jan 2025 – current
Postdoc in Physics and AI

Evaluating LLM input representation robustness

Perturbation Design: Developed a controlled evaluation framework to isolate how tokenization perturbations affect LLM reasoning and code-generation performance across MMLU, GSM8K, GSM8K-Python, and HumanEval.
Test-time Scaling: Designed retokenization, typo, and sampling-temperature experiments to compare representation-induced performance shifts against standard stochastic-generation baselines.
Eval Harness: Built empirical pass@ analysis tooling to quantify robustness, variance, and solution diversity under prompt-level perturbations.

Emory University: Atlanta, GA May 2018 – Oct 2024
Doctoral Researcher

End-to-end development of predictive models for high-dimensional, multi-modal time-series data.

Time Series Forecasting: Designed and trained a multi-timescale predictive model using Recurrent Neural Networks (RNNs) to quantify latent dynamics and predict state transitions at both short and long time horizons.
Scalable Data Engineering: Built a production-grade pipeline using SQL queries to ingest and process time-series datasets exceeding ~10 million data points.
Anomaly Detection: Leveraged Kalman Filters for state estimation, achieving 98% accuracy in identifying and imputing missing data in noisy signal streams using Autoencoders.
Generative Modeling: Developed dynamic representations of behavioral states using Variational Autoencoders (VAEs), enabling the detection of latent patterns (anomalies) invisible to traditional statistical methods.
Computer Vision: Built a high-fidelity pose estimation system using Convolutional Neural Networks (CNNs) with custom heuristics to track complex objects in video data.
Dimensionality Reduction: Implemented PCA, t-SNE, and UMAP to compress multi-modal high-dimensional data into interpretable latent manifolds for downstream classification.
Causal Analysis: Used logistic regression and decision tree classifiers to quantify latent patterns in animal behavior that reflect exposure to early-life stress in mice, using multi-level bootstrapping to calculate statistical upregulation in stress across animals.

University of Texas at Austin: Austin, TX Summer of 2015 & 2016
Summer Researcher

Monte-Carlo Simulations: Created Python-based Monte-Carlo simulations to model stochastic spread dynamics on structured populations; identified key predictors of system persistence.
Algorithm Development: Devised a recursive algorithm (O(n) complexity) to identify and classify fingering patterns formed in biological cell membranes.
Automated Classification: Developed automated algorithms to classify force separation data from Atomic Force Microscopes, replacing manual workflow and leading to a Nature publication.

PUBLICATIONS

1. **Jain, Kanishk;** Day, Matthew; Can, Tankut; (2026); Emergent retokenization symmetry in large language models: phenomenology and applications. arXiv:2606.15521
2. Kaur, Rajpreet; **Jain, Kanishk**; Berman, Gordon; (2026); Using timescale as a state coordinate reveals the metastable geometry of behavior. *bioRxiv*, 2026.05.25.727718, 2026.
3. **Jain, Kanishk**; Day, Matthew; Can, Tankut; (2026); Generating output diversity from prompt re-tokenization. **ICLR** 2026; Scientific methods for Deep Learning Workshop
4. Winner Taniel; Rosenberg Michael; **Jain, Kanishk**; Kesar Trisha; Ting Lena; Berman Gordon; (2023); Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. PLOS Computational Biology 19 (10)
5. **Jain, Kanishk**; Berman, Gordon; (2019); Opening the black box of social behavior. Nature Neuroscience 22 (12), 1947-1948.