Kavana Venkatesh

Blacksburg, VA - 24060 | 857-869-9358

kavanav@vt.edu | https://github.com/KavanaVenkatesh | https://www.linkedin.com/in/kavanavenkatesh/

EDUCATION

Ph.D. Computer Science and Applications, Virginia Tech, Blacksburg, VA**,** GPA: 4.00 ***2024* – 2028 (expected)**
Advisor: Dr. Pinar Yanardag
Research Focus: Interpretable and High-Fidelity Generative AI at the Intersection of Natural Language Processing and Computer Vision, with an Emphasis on Leveraging Large Language Models and Diffusion Models.

MS in Data Science, Northeastern University, Boston, GPA: 3.75 ***2019 - 2021***
Related Courses: Machine Learning, Deep Learning, Visualization, Linear Algebra

B.E in Electrical and Electronics Engineering, SJCE, India, GPA: 9.27/10.00 ***2014 - 2018***

RESEARCH

Graduate Research Assistant, *Virginia Tech,* Blacksburg, VA ***Aug 2024 - Present***
- Developing novel frameworks for high-fidelity domain-specific text-to-image generative tasks at the intersection of NLP and vision, by leveraging advancements in Large Language Models and Diffusion Models
- Researching methodologies to enhance domain-specific image generation capabilities of Diffusion Models by augmenting them with the instructional following capabilities of LLMs and innovative prompt engineering strategies
- Exploring sophisticated techniques to enhance interpretability of complex agentic frameworks for coherent and efficient task completion.

Research Assistant, *SJCE,* Mysuru, India (Advisor: Dr. Neethi M) ***Aug 2017 – Apr 2018***
- Studied the effects of signal variation on the performance and longevity of electrical components by conducting extensive experiments in carefully simulated environments.
- Designed novel frameworks that synergize machine learning and high voltage systems to automate and enhance the safety monitoring of complex, large-scale electrical circuits.

Summer Research Fellow, *IIT,* Hyderabad, India (Advisor: Dr. Sumohana Channappayya) ***May – Aug 2017***
- Optimized hybrid solar-wind power generators by collecting large streaming data from numerical relays, using Deep Neural Networks. Boosted output voltage by 27%, beating existing benchmarks.
- Collaborated with chemists and computer scientists to develop cutting-edge machine learning and vision methodologies to identify precise drug molecular structures with minimal side effects.
- Contributed to novel drug discovery efforts by synthesizing over 100K high-quality synthetic data samples with GANs.

INDUSTRY EXPERIENCE

iLink Digital (Remote) ***Jan 2024 – Aug 2024***
Data Scientist (Generative AI)
- Developed a scalable *RAG* system for a banking client, to process over ***15M+*** *documents in under 0.5s response times*, leveraging *Llama 2, CosmosDB, Nvidia TensorRT*, and *Triton Server* for real-time, high-accuracy decision-making.
- Significantly improved RAG system performance using hybrid search algorithms and reranking techniques, boosting in context relevance while reducing hallucinations. Rigorously validated through *LLMEval* and *RAGAs*.
- Designed and launched a transformative pet health dashboard with instant AI-driven insights and a *custom* ***GraphRAG*** *chatbot*, processing over ***6K+ documents*** via *Neo4j* for real-time pet history summaries and precise entity extraction.
- Implemented a robust live feedback loop with adaptive retraining, improving chatbot precision by 15% and recall by 12%, enhancing diagnostic accuracy and decision-making for veterinarians.

Fidelity Investments, Boston, MA
Data Scientist ***Feb 2022 – Oct 2023***
- Developed and deployed an end-to-end tool to extract, classify and parse complex financial tables from annual reports using NLP, OCR, computer vision (*YOLOv5, GNNs, Bert*) and classical ML, saving ***1000+ hours*** of manual work.
- Designed a *Streamlit* application for managing ***half a trillion*** assets containing millions of portfolios using efficient *recommendation systems* and anomaly detection, thereby generating ***$850M*** *in revenue.*
- Fine-tuned, and deployed ***open-source LLM*** apps on large custom text data for a variety of internal use cases (code translation, NLG, Chatbots, summarization) using techniques like ***PEFT***, ***text-generation-inference, RAG, Agents.***

- Managed and trained a team of ***10 graduate interns*** to research Large LLM use cases, develop novel solutions considering global compliance requirements, risk aversion techniques, and presented to senior management.
- Researched emerging trends in AI as a Board Member at the ***AI Council at Fidelity*** to identify and automate complex business processes. Quantified product success using custom KPIs, dashboards and presented to stakeholders.

Data Science Co-Op ***Jan – Jun 2021***

- Forecasted stock prices for efficient portfolio management using advanced multi-variate time-series models resulting in returns *beating the market*. Conducted A/B tests for feature selection and built interactive real-time dashboards.
- Built a large-scale distributed collaborative filtering-based pipeline for similarity mapping of ***billions of financial records*** using advanced *SQL* and *PySpark*. Tuned hyperparameters and set up auto monitoring-validation ML systems.
- Simulated beta of assets by training and tuning an *XGBoost* regressor model on data with thousands of features using *Map-Reduce*. Speeded up execution by ***15 times*** by reducing dimensionality and saved ***40h/month.***

PUBLICATIONS

Context Canvas: Enhancing Text-to-Image Diffusion Models with Knowledge Graph-Based RAG ***Dec 2024***

- We introduce a novel training-free approach to significantly enhance fidelity and contextual accuracy of T2I models
- We propose a Self-Correction and an editing framework to generate high-fidelity images with minimal user prompts.

Project Page: *https://context-canvas.github.io/* | ***arXiv preprint :*** *https://arxiv.org/abs/2412.09614*

FluxSpace: Disentangled Semantic Editing in Rectified Flow Transformers ***Dec 2024***

- We introduce a domain-agnostic image editing method leveraging a representation space with the ability to control semantics of images generated by rectified flow transformers, such as Flux.

Project Page: *https://fluxspace.github.io/* | ***arXiv preprint:*** *https://arxiv.org/abs/2412.09611*

Fault Analysis and Predictive maintenance of 3-phase induction motors using Machine Learning ***Aug 2018***

- First Author Conference Paper presenting pre-fault detection of 6 types of electrical faults
- Best paper award at ICEECCOT- 2018 and the paper is published in IEEE Xplore digital library (link)

INVITED TALKS

University of Massachusetts at Amherst ***Mar 2024***
[Talk] Subject - Crafting Balance: The Dual-Edged Sword of Open-Source LLMs in the Quest for Responsible AI

Analytics Vidhya, *DataHour Virtual Global Webinar* (5K+ registrations from 35+ countries) ***Oct 2023***
[Webinar] Subject – Harnessing the Power of LLMs: A Deep Dive into Practical Solutions

Northeastern University, Boston

[Workshop] Subject – Generative AI (Two-part Workshop attended by 150+ students)

[Panel Event] Subject – Data Science Career Paths (In Association with Women in Data Science Worldwide) ***Apr 2023***

[Workshop] Subject – Machine Learning in Production: Best Practices ***Mar 2023***

TEACHING

Virginia Tech, *Graduate Teaching Assistant* ***Aug 2024 – Present***

- Introduction to AI: CS4804

Northeastern University, *Graduate Teaching Assistant*

- Programming with Data: DS2000 ***(Fall 2020, 2021)***

TECHNICAL SKILLS

Languages/Databases: Python, R, C, MySQL, SQL Server, PostgreSQL, XPath, Snowflake, RStudio
Frameworks/Libraries: TensorFlow, PyTorch, Transformers, Sklearn, NumPy, Pandas, Hugging Face, Diffusers
Large Language Models: Fine-Tuning, Prompt Engineering, VectorDatabases, Quantization, Agents, RAG, Evaluation
Machine Learning: Regression, Classification, Clustering, Anomaly Detection, Dimensionality Reduction
Computer Vision: Diffusion Models, CNN, GNN, Text-to-Image Generation, Segmentation, Object Detection
NLP: OCR, Bert, PDF Table Extraction, Text Classification, Bert, Transformers, Text Generation, QA Systems
Data Visualization: Matplotlib, Seaborn, ggplot2, Tableau, MS Excel, RShiny, Plotly Dash, PowerBI
Deployment: Streamlit, Jenkins, Airflow, Amazon Sagemaker, FastAPI, Docker, vLLM, TGI, GGML, GPTQ, Azure
Big data Processing: MapReduce, Scala, AWS (EC2, EMR, S3), PySpark, Airflow, Koalas, Amazon Redshift

HONORS AND ACHIEVEMENTS

- Ambassador, Women in Data Science (Greater Boston) ***Feb 2024 – Present***
- Mentor, Women in Big Data ***May – Sep 2024***
- National Merit Scholarship (MHRD, Govt. of India) – Awarded for Four Years of B.E ***2014 – 2018***
- Pre-University (12th class) State Rank 3 (Major Percentile: 99) ***2014***
- Department Rank 2 (Out of 150 Students), Bachelor of Engineering, EEE ***2018***