

**Kavana Venkatesh**  
Blacksburg, VA - 24060 | 857-869-9358  
[Email](#) | [GitHub](#) | [LinkedIn](#)

## EDUCATION

---

**Ph.D., Computer Science and Applications, Virginia Tech, Blacksburg, VA** *Aug 2024 – May 2029*

**Advisor:** Dr. Pinar Yanardag

**Research Focus:** Developing Interpretable and High-Fidelity Generative Models 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: 3.75/4.00** *Sep 2019-Dec 2021*

**Related Courses:** Machine Learning, Deep Learning, Large Scale Parallel Data Processing, Data Visualization, Linear Algebra and Probability, Algorithms

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

**Related Courses:** Advanced Multivariable Calculus, Numerical Methods of Computing, Probability and Statistics

## PROFESSIONAL EXPERIENCE

---

**Data Scientist (Generative AI), iLink Digital (Remote)** *Jan 2024 – Aug 2024*

**LLMs, Prompt Engineering, GraphRAG, Agentic Systems, TensorRT, Azure, Finetuning, Evaluation, Huggingface**

- Spearheaded the development and deployment of a highly scalable, low-latency **Retrieval-Augmented Generation (RAG)** system for a banking client, scaling to process over **15M+ documents with 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 by incorporating cutting-edge hybrid search algorithms and reranking techniques, leading to a substantial increase in context relevance and reduction in hallucinations. Rigorously validated through **LLMEval** and **RAGAs** evaluation frameworks.
- Designed and launched a transformative pet health information dashboard, empowering veterinary professionals with instant, AI-driven insights by deploying a **custom chatbot** using **GraphRAG**, processing over **6K+ documents** via **Neo4j**, and generating real-time pet history summaries with precise entity extraction capabilities.
- Implemented a robust live feedback loop and adaptive retraining system for continuous optimization of chatbot performance, resulting in a **15% improvement in precision and a 12% boost in recall** for the chatbot's entity extraction capabilities, contributing to more accurate diagnostic support and enhancing decision-making for veterinarians.

**Data Scientist, Fidelity Investments, Boston, MA** *Feb 2022 – Oct 2023*

**Generative AI, LLMs, NLP, SQL, Deep Learning, Machine Learning, Technical Writing, Thought Leadership**

- Developed and deployed an end-to-end scalable tool to extract, classify and recognize structure of complex financial tables from annual reports using NLP, OCR, computer vision (**YOLOv5, Graph Neural Networks, Bert**) and classical ML techniques, saving **1000+ hours** of manual work. Marketed across the firm using technical articles, product presentations.
- 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**.
- Validated, 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 to present to senior management.
- Researched emerging trends in AI and collaborated with technical and non-technical specialists to identify and automate complex business processes. Quantified product success using custom KPIs, dashboards and presented to stakeholders.

**Data Scientist, Fidelity Investments, Boston, MA** *Jan – Jun 2021*

**PySpark, Machine Learning, PostgreSQL, Sagemaker, Airflow, SparkML, Python, Autoencoders, Tableau**

- Forecasted stock prices for efficient portfolio management using advanced multi-variate time-series models resulting in portfolio 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**.

**Graduate Teaching Assistant, Northeastern University, Boston, MA**

**Sep – Dec (2020 and 2021)**

**Python, Machine Learning, Data Analysis, General Programming**

- Mentored 300+ students from non-CS backgrounds in Python and Data Science for ‘Programming with Data’(DS2000). Held office hours, created and graded assignments.
- Managed Practicum classes, reviewed and graded project presentations

**Summer Research Fellow, IIT, Hyderabad, India**

**May – Aug 2017**

**Python, Machine Learning, Data Analysis, General Programming**

- Researched and developed novel electrical circuits to optimize and design a hybrid solar-wind power generator by collecting large streaming data from numerical relays using Deep Neural Networks. Boosted output voltage by 27%
- Collaborated with chemists and computer scientists on developing machine learning and vision methodologies to identify drug molecular structures with minimal side effects.
- Led the design of new drugs and conducted simulations on over 100K synthetic data samples, achieving a significant reduction in numerous side effects. Successfully presented research findings to the AI council at IIT, winning the best project award for innovative contributions.

**Research Assistant, SJCE, Mysuru, India**

**Aug 2017 – Apr 2018**

**Electric Motors, Transformers, Python, Machine Learning, Data Analysis, General Programming, Deep Learning**

- Studied the effects of signal variation on the performance and longevity of electrical components by conducting extensive experiments in carefully simulated environments. Systematically catalogued comprehensive data encompassing a wide array of critical features, ensuring detailed analysis and accuracy.
- Developed a machine learning model for fault detection of high-capacity voltmeters using isolation forest, achieving 0.93 F1 score thereby enhancing the safety of complex electrical circuits.
- Collaborated with industry partners to perform predictive analysis and pre-fault detection of induction motors for 6 types of electrical faults using Deep Neural Networks with 0.97 accuracy. Documented and presented findings to stakeholders.

## **TECHNICAL SKILLS**

---

**Languages/Databases:** Python, R, C, MySQL, SQL Server, PostgreSQL, XPath, Snowflake, RStudio

**Frameworks/Libraries:** TensorFlow, PyTorch, Keras, Sklearn, NumPy, Pandas, ChromaDB, Hugging Face

**Large Language Models:** Fine-Tuning, Prompt Engineering, VectorDatabases, Quantization, Agents

**Machine Learning:** Regression, Classification, Clustering, Anomaly Detection, Dim. Reduction

**Computer Vision:** CNN, GNN, Image Classification, Segmentation, Augmentation, Object Detection

**NLP:** OCR, 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

**Big data Processing:** MapReduce, Scala, AWS (EC2, EMR, S3), PySpark, Airflow, Koalas, Amazon Redshift

## **ACADEMIC PROJECTS**

---

**Predictive Analytics of Employees Turnover | Logistic Regression, Random Forest, SVM, ANN**

- Developed an application to retain customers and to optimize human resources by training and validating logistic regression, random forest, SVM and ANN classifiers on large data to achieve 0.92 F1 score.

**Customer Segmentation of NYC Bank Data | Sklearn, PCA, Autoencoders, K-Means, ggplot2, Tableau**

- Recommended strategies to enable efficient personalized customer service using K-Means clustering and PCA. Built an interactive tableau dashboard to present findings.

**Intelligent Movie Recommendation Using Hybrid Deep Learning | Python, PyTorch, CNN, Autoencoders**

- Built a User-Item based Collaborative Filtering Recommendation System on MovieLens data using a Hybrid Deep Learning Model with RBMs and Autoencoders with 81% accuracy, 0.73 personalization.

## **PUBLICATION and OTHER ACHIEVEMENTS**

---

**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 ICECCOT- 2018 and the paper is published in IEEE Xplore digital library ([link](#))

**Winner – Global Low Code Challenge at Fidelity Investments: Internal News Summarizer using GPT 3**

**Mar 2022**

**Generative AI Content Writer, Packt Publications | Hosts informative Data Science Workshops and Webinars**