

# Vishwanath (Vish) Guruvayur

+1 (571) 602-2034 | [vishguru2704@gmail.com](mailto:vishguru2704@gmail.com) | [linkedin.com/in/vishgp](https://www.linkedin.com/in/vishgp) | [vishguru.com](https://vishguru.com)

## SUMMARY

Data Scientist with 3+ years of experience building ML systems and analytics pipelines across pharma, ops, and e-commerce. Strong focus on scalable deployment, experimentation, and LLM-based retrieval, with peer-reviewed research in multimodal RAG. Skilled in Python, SQL, and AWS, delivering production-grade, interpretable data products with model monitoring expertise. Passionate about bridging research and real-world impact through responsible, scalable AI.

## EXPERIENCE

### AI Researcher

Charlottesville, VA (Remote)

LMI Inc. and UVA School of Data Science

Nov 2024 – Present

- Engineered a modular multimodal RAG evaluation pipeline with **10,000+ vectors** using Sentence Transformers, CLIP, and Pinecone, enabling research on content retrieval and rapid experimentation across diverse embedding vector spaces.
- Co-authored **peer-reviewed paper [PDF]** accepted at AIED 2025, presenting innovative techniques for multimodal LLM retrieval integrating audio, visual, and textual data from digital textbooks and classroom sources.
- Recipient of “Most Compelling Data Visualizations Award” at Capstone Showcase for designing an interactive visual model that explained RAG vector retrieval dynamics and highlighted performance patterns across multimodal inputs.
- Currently leading research and development of an agentic LLM assistant that leverages structured retrieval over temporally-linked academic content to explore new methods of enhancing student support and informing instructional strategy.

### Senior Data Scientist (Team Lead)

Bethlehem, PA (Remote)

Piramal Pharma Ltd.

Jan 2022 – Jun 2024

- Established the Data Science Team for the Critical Care vertical, leading key projects and mentoring junior data scientists to integrate advanced analytics and data products into core business functions using best practices.
- Prevented **\$1.2M** in quarterly losses by implementing counterfactual analysis of inventory schedules, write-offs and sales time-series forecasting using Monte Carlo simulations and S-ARIMA, improving inventory and sales strategies.
- Partnered directly with the CFO to develop market segmentation models and intelligence dashboards that informed new product launch and acquisition strategies, shaping portfolio decisions and competitive positioning.
- Reduced ACoS **from 45% to 8% within 6 months** for high-priority e-commerce products via ad campaign optimizations by leveraging alternative data using NLP (VADER, SpaCy), market segmentation, and A/B testing.
- Received the “Piramal Excellence Award” for engineering **30+ analytics pipelines** with workflows monitoring both KPIs and ML model health, improving operational efficiency by 85% through proactive retraining aligned with business goals.

### Data Scientist (Intern)

Mumbai, India

Piramal Pharma Ltd.

Jul 2021 – Dec 2021

- Achieved **15% quarterly increase** in market penetration **across 400,000 retailers** by building the company’s first product recommendation engine using the Tree-based FP-Growth algorithm from scratch.
- Boosted high-performing sales reps **retention rates to 93%** by developing an attrition prediction model using XGBoost and Logistic Regression on historic sales performance and attrition data.

## EDUCATION

### University of Virginia

Charlottesville, VA

Master of Science in Data Science, GPA: 4.0/4.0

Jun 2024 – May 2025

Data Science Ambassador • Teaching Assistant - Statistical Learning • Range Resident

Certifications: *AWS Certified Solutions Architect – Associate*, *AWS Certified Machine Learning – Specialty*

### Birla Institute of Technology and Science (BITS Pilani)

India

Bachelor of Engineering in Mechanical Engineering, Minor in Data Science, GPA: 3.8/4.0

Aug 2018 – Dec 2021

## SKILLS

**Languages:** Python, R, SQL, Scala, JavaScript, Bash

**AI and ML:** Stat Modeling (Scikit-Learn, XGBoost), Deep Learning (TensorFlow), LLMs (LangChain), NLP (NLTK, SpaCy)

**Data Engineering:** Spark, Snowflake, Hadoop, MLflow, Docker, APIs, Web Scraping

**Cloud and Tools:** AWS (Certified), Azure, Databricks, Snowflake, Git, Conda, Flask

**Visualization:** Power BI, Plotly Dash, Seaborn, Matplotlib