Rohith Ramakrishnan

Ann Arbor, MI | rrohith@umich.edu | (734) 476-5544 | LinkedIn | scholar.google.com

EDUCATION

University of Michigan Ann Arbor, MI

Master of Science, Data Science. *May 2028*

Relevant Coursework: Machine Learning, RDBMS, Deep Learning, Statistical Inference, Natural Language Processing, Computer Vision

Amrita Vishwa Vidyapeetham Coimbatore, India

Bachelor of Technology, Computer Science and Artificial Intelligence. GPA: 9.3/10. *May 2023*

EXPERIENCE

Data Scientist, Generative AI & Machine Learning May 2023 – Aug 2026

Rolls-Royce *Bengaluru, India*

- Designed a production agentic RAG system using specialized agents, OCR, domain-tuned embeddings, vLLM-served quantized models, and hybrid vector/BM25 retrieval, reducing engineering-analysis turnaround from 3-4 days to under 4 hours.
- Automated structured extraction from safety-engineering documents using LLM guardrails and evaluation pipelines, reducing review and approval time from 36 hours to 2 hours.
- Fine-tuned a dense embedding model via supervised and unsupervised techniques on curated aviation data, outperforming OpenAI text-embedding-ada-002 on retrieval tasks by 23% in an internal benchmark.
- Led a team of three data scientists developing aviation knowledge graphs and built a unified semantic and keyword search platform across 15 engineering data sources, reducing search time by 44%.
- Led migration of agentic deployment from Docker on Azure Cloud to Enterprise Databricks using LangGraph, reducing deployment time and improving system scalability.
- Developed a proof-of-concept for open LLMs on engineering data by quantizing models, benchmarking on performance, and deploying on local closed GPU hardware.

Data Science Intern Dec 2022 – May 2023

Rolls-Royce *Bengaluru, India*

- Trained and deployed a custom named-entity recognition model integrated with OCR post-processing, achieving an F1-score of 0.85 to extract critical terms from complex PDF layouts.
- Advanced a turbine-engine anomaly-detection system from proof of concept to MVP using time-series ML, reducing fault-detection latency from one week to seconds.

SKILLS

ML/AI: PyTorch, TensorFlow, JAX, scikit-learn, Hugging Face, LlamaIndex, LangChain, LangGraph

LLM Inference & Search: vLLM, Triton Inference Server, TensorRT-LLM, TGI, Apache Solr, Elasticsearch, BM25

Data and Infrastructure: PySpark, Dask, Databricks, Docker, Azure, CI/CD

Languages: Python, Java, MATLAB

PUBLICATIONS

- Ramakrishnan, R., et al. "Analysis of Text-Semantics via Efficient Word Embedding Using Variational Mode Decomposition." Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation, 2021.
- Ramakrishnan, R., et al. "CT Image Enhancement Using Variational Mode Decomposition for AI-Enabled COVID Classification." Artificial Intelligence on Medical Data, 2023.
- Ramakrishnan, R., et al. "Data-Driven Volatile Cryptocurrency Price Forecasting via Variational Mode Decomposition and BiLSTM." International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 1, 2023.

AWARDS & HONORS

- GBS Awards, Highly Commended, "Making a Difference", Rolls-Royce (2025)
- Contributed Triton streaming/chat inference and Sentence-BERT reranking support to LlamaIndex (2024)
- Silver Award, Generative AI & LLM Innovations, Rolls-Royce (2023)