

Praveen K. Varkala

pvarkala33@gmail.com | +1 (980) 307-4046 | Charlotte, NC

EXPERIENCE

Probe Practice Solutions LLC
AI Engineer

Jul 2024 – Present
Round Rock, TX

- Developed a claims explanation pipeline using fine-tuned GPT models and LangGraph, simplifying billing codes into patient-friendly summaries, reducing customer support queries by 40%.
- Built an LLM-based document processing system using Hugging Face Transformers and LangGraph’s advanced orchestration, extracting and summarizing key billing details to improve claim resolution time by 30%.
- Implemented a claims summarization tool powered by OpenAI APIs, Pinecone for vector retrieval, and LangGraph to orchestrate the end-to-end process, reducing processing time by 25% with CrewAI integration.

Hexagon Capability Center India
Software Developer

Aug 2022 – Dec 2022
Hyderabad, India

- Optimized Oracle SQL databases processing 10TB+ data with Hive, integrating Docker for scalable workflows.
- Built ETL pipelines in Snowflake and HDFS, processing 500GB+ daily, with AWS Lambda for automation.
- Applied regression models for SLA breach predictions with Datadog, identifying risk patterns and improving on-time delivery rates by 15%.

CMR Infotechs
Data Scientist

May 2019 – May 2022
Hyderabad, India

- Created a rule-based chatbot for a banking application using spaCy and NLTK, deployed via Docker on AWS Lambda to handle over 1M+ queries monthly.
- Built modern front-end integrations for banking workflows using Next.js and Tailwind CSS, connecting REST and GraphQL APIs to deliver responsive, interactive dashboards, reducing manual tasks by 20%.
- Secured financial transactions for 500K+ accounts using PostgreSQL on RDS, integrating Prisma to streamline database interactions and enhance performance.
- Migrated rule-based NLU to a GPT-2 system on AWS SageMaker, deploying with Docker for scalability and improved accuracy in natural language understanding.
- Engineered fraud detection APIs using Flask, integrated with PostgreSQL and optimized queries with Prisma to enhance transaction security and efficiency.
- Automated banking workflows for 100K+ accounts by implementing CI/CD pipelines using GitHub Actions, significantly reducing manual effort by 30%.

EDUCATION

University of North Carolina at Charlotte
Master of Science in Computer Science; Concentration in Data Science

May 2024
Charlotte, NC

ACADEMIC PROJECTS

End to End LLM Training and Optimization

- Built data pipeline for language model training, trained on 110M tokens with custom tokenizer and DeepSpeed.
- Enhanced performance through Supervised Fine-Tuning (SFT), DPO, and RLHF with PPO.

Collaborative Multi ReAct Agent System with Langgraph

- Developed multi-agent system with LangGraph and PythonREPL, reducing task completion by 40%
- Integrated Tavily Search and GPT-4 via LangChain for real-time economic data and automated reporting.

OTHER

- Languages and Libraries:** Python, SQL, JavaScript, TensorFlow, PyTorch, Langchain, Pandas, NumPy.
- Technologies:** Machine Learning, Feature Engineering, AWS, MLOps, Predictive Modeling, Statistical Analysis.