PRAVEENA S

Generative AI Engineer | LLMs | RAG | AI Agents | LangChain | LangGraph | MCP | Python

Coimbatore, Tamil Nadu | +91 7695963156 | praveenasuku3@gmail.com | linkedin.com/in/praveena-suku | github.com/Praveenasuku

PROFESSIONAL SUMMARY

Generative AI Engineer with experience across data analytics and AI, currently building Retrieval-Augmented Generation (RAG) pipelines and LLM applications at Pumo Technovation using LangChain, ChromaDB, Pinecone, embeddings, and prompt engineering. Agentic AI portfolio for data centre operations includes a production RAG assistant with hybrid search and RAGAS evaluation, a LangGraph agent with tool calling and human-in-the-loop approval, and a Model Context Protocol (MCP) server with text-to-SQL guardrails, delivered through FastAPI, Docker, and AWS. Previously built GPT-4 chatbots, Flask-based AI apps, and ARIMA forecasting models. MBA in Business Analytics.

TECHNICAL SKILLS

- **Generative AI & LLMs:** Large Language Models (LLMs), OpenAI GPT-4, Google Gemini, Anthropic Claude, Prompt Engineering, Structured Outputs, NLP
- **Agentic AI:** LangChain, LangGraph, AI Agents, Agent Orchestration, Model Context Protocol (MCP), Tool Calling, Human-in-the-Loop
- **RAG & Vector Databases:** Retrieval-Augmented Generation (RAG), ChromaDB, Pinecone, Qdrant, Embeddings (Hugging Face), Semantic Search, Hybrid Search, Reranking
- **Evaluation & Observability:** LLM Evaluation, RAGAS, Langfuse, Guardrails
- **Programming & APIs:** Python, SQL, FastAPI, Flask, Streamlit, REST APIs
- **Databases & Deployment:** MySQL, PostgreSQL, Docker, AWS (EC2, S3), Git, GitHub, GitHub Actions (CI/CD), Jenkins
- **Data & Analytics:** Pandas, NumPy, Time Series Forecasting (ARIMA), Power BI, Tableau, Excel

PROFESSIONAL EXPERIENCE

Generative AI Engineer | Pumo Technovation India Pvt Ltd | Coimbatore | Jun 2026 – Present

- Built Retrieval-Augmented Generation (RAG) pipelines using LangChain to deliver accurate, context-grounded answers from internal enterprise documents in a data centre environment.
- Implemented vector search with ChromaDB and Pinecone, covering document ingestion, chunking, and embedding generation across internal document collections.
- Tuned chunking and retrieval parameters (chunk size, overlap, top-k) to improve answer relevance and retrieval accuracy.
- Applied prompt engineering to improve response relevance and reduce hallucinations in LLM-powered applications.

Data Analyst Trainer | Qtree Technologies | Coimbatore | Dec 2025 – Jun 2026

- Built hands-on data analytics, machine learning, and Generative AI examples in Python and SQL, including LLM-based applications.
- Worked across data analysis, ML model building, and LLM application development using modern AI tools.

Data Analyst | Arcadia (Urjanet Energy Solutions Pvt Ltd) | Chennai | May 2024 – Jul 2025

- Built AI-enhanced analytics solutions, including real-time chatbot improvements using GPT-4, generating customer engagement insights.
- Developed predictive and forecasting models (ARIMA, trend analysis) to support business planning.
- Transformed raw data into dashboards and reports using Power BI, Tableau, SQL, and Excel, enabling faster data-driven decisions.
- Automated repetitive data workflows with Jenkins, reducing manual effort and reporting errors.
- Collaborated with cross-functional and international teams to deliver insights for strategic initiatives.

Data Science Intern | Aapin Solution Limited | Coimbatore | Sep 2023 – Dec 2023

- Applied data analysis and machine learning techniques in Python to real-world business problems.
- Developed automated data-cleansing scripts, reducing processing time by 20% and improving data accuracy.
- Performed exploratory analysis and model building using Python, SQL, and Tableau as part of project teams.

AI Engineer | Vegas Project Management Service | Coimbatore | Mar 2023 – Sep 2023

- Enhanced chatbots using GPT-4 and NLP techniques, including a ChatGPT-style assistant grounded in a custom company knowledge base.
- Designed and built a Flask-based web app for AI-driven user interactions, increasing user engagement by 25%.
- Delivered sentiment analysis and predictive analytics projects.

PROJECTS

DC-Ops Copilot: Production RAG Assistant | LangChain, Qdrant, FastAPI, Streamlit, RAGAS, Langfuse, Docker, AWS

- RAG assistant that answers data centre operations questions from runbooks, SOPs, and incident reports, with source citations.
- Hybrid search (BM25 + vector), cross-encoder reranking, and metadata filtering, plus a refusal guardrail for out-of-scope questions.
- RAGAS evaluation on a 30-question golden set, comparing chunk size and top-k configurations.
- Semantic caching for repeated questions, plus Langfuse tracing for per-query latency and cost monitoring.
- Containerized with Docker and deployed on AWS EC2, with source documents stored in Amazon S3.

EnergyGuard: Energy Bill Intelligence Agent | LangGraph, Pydantic, ARIMA, FastAPI, Streamlit

- LangGraph agent that extracts structured fields from electricity bill PDFs, detects cost and usage anomalies, and forecasts spend with ARIMA.
- Human-in-the-loop approval for flagged bills, with checkpointing so interrupted runs resume safely.
- Field-level extraction accuracy measured against ground-truth data from synthetic bills.

DC-Ops MCP Server: AI Tools for Data Centre Analytics | MCP, PostgreSQL, Python, Docker Compose

- MCP server exposing a PostgreSQL operations database, the RAG assistant, and the bill-analysis agent as tools for any MCP client.
- Read-only SQL guardrails (read-only database role, SELECT-only validation, row limits), with text-to-SQL accuracy evaluated on a 40-question test set.
- RAG API, agent API, MCP server, and PostgreSQL run as containerized microservices with Docker Compose.

AI Search Engine (ChatGPT-style) | Python, LLM APIs, SQL

- Built an AI search engine that retrieves information and generates concise answer summaries using LLM APIs.
- Displayed relevant images alongside results and stored user queries in a SQL database for search history and analytics.

Portfolio Optimization using Machine Learning | Python, Machine Learning

- Applied Efficient Frontier analysis to optimize investment portfolios, maximizing returns while managing risk.
- Used Hierarchical Clustering and PCA to build data-driven allocation strategies that improved diversification.

EDUCATION

- **MBA, Business Analytics & Finance** | KV Institute of Management and Information Studies | Sep 2021 – Apr 2023
- **B.Com, Computer Applications** | SNMV College of Arts and Science | 2018 – 2021

CERTIFICATIONS

- Postgraduate Program in Data Analytics with Machine Learning | Imarticus Learning | 2023
- Generative AI & LLMs: Building ChatGPT with a Custom Knowledge Base | Vegas Project Management Service
- SQL in Data Science | Great Learning
- Data Visualization with Power BI | Great Learning