

Harish V

Chennai, Tamilnadu | 📞 9080561537 | ✉️ eduhar0311@gmail.com | 🔗 [LinkedIn](#)

Summary

Accomplished Software Engineer with a strong background in developing AI-driven solutions and enhancing system performance through data-driven methodologies. At Accenture, I led a team of 5 in optimizing chatbot response accuracy by 20%, streamlining production releases for a 15% boost in efficiency, and reducing UI response times by 25% across multiple platforms. I have engineered advanced systems including a multi-source document retrieval and QA platform, an automated blog generator, and a classification model that improved accuracy by 32%—leveraging technologies such as Large Language Models (LLMs), NLP, Generative AI, LangChain, and cloud-based analytics. Certified in Microsoft Azure Administration and IBM Data Science, I excel in delivering scalable, high-impact solutions in both live production and analytics-driven environments.

Education

Bachelor of Commerce - Corporate Secretaryship | Chennai, Tamilnadu | **CGPA:** 8.5 | **Graduation:** Jul 2022

Experience

Accenture | Software Engineer (Sep 2022 - Mar 2025)

Leveraged data-driven methodologies to enhance chatbot performance in live production environments, achieving a 20% improvement in response accuracy and contextual relevance. Managed a team of 5 specialists, overseeing monthly production releases and driving a 15% increase in operational efficiency. Applied advanced analytics and statistical techniques to validate multi-option outputs, ensuring data integrity and optimal system performance. Conducted comprehensive UI and functional testing across Mac and Windows platforms, reducing response times by 25%. Analyzed extensive Power BI reports to extract actionable insights on project deals, employee productivity, and client interactions across 3+ countries. Regularly engaged with leadership to present data-driven insights and strategic recommendations, influencing key business decisions. Collaborated with cross-functional teams to integrate analytics insights into product enhancements, significantly improving user experience and overall system reliability.

Skills: Large Language Models (LLMs), Natural Language Processing (NLP), Generative AI & Deep Learning, Machine Learning & RAG, SQL & Python.

Projects

Advanced Multi-Source Document Retrieval & QA System

- Developed a multi-source ingestion system using LangChain to process web data and PDFs with advanced vector embeddings from Ollama and OpenAI.
- Engineered scalable retrieval pipelines using FAISS and Astra DB (Cassandra) to manage 50+ documents, ensuring fast and accurate search.
- Implemented chain-of-thought reasoning with ChatGroq and Llama3, reducing query response times by 30%.
- Built a dynamic Streamlit interface for seamless user interaction and detailed document similarity search.
- Achieved measurable improvements in system efficiency and accuracy, optimizing performance for high-level technical applications.

Automated Blog Generation with LLama 2 & LangChain

- Engineered an automated blog generator using LLama 2 via CTransformers, reducing content creation time by 40%.
- Integrated LangChain's PromptTemplate to dynamically tailor blog content for various professional profiles (e.g., Researchers, Data Scientists), ensuring precision and contextual relevance.
- Developed a responsive Streamlit interface for real-time input handling, including topic selection, word count, and target audience, optimizing user experience.
- Tuned model parameters to achieve high-quality, context-aware blog outputs that adhere to specific word count requirements.
- Demonstrated measurable improvements in content generation accuracy and speed, significantly enhancing overall productivity.

Advanced Classification Model Using LLM (Freelancing)

- Developed a classification model tailored for telecall data containing multiple intents.
- Enhanced multi-intent handling by leveraging Large Language Models (LLMs).
- Implemented LangChain and Ollama to run a quantized Llama 3.2 1B model efficiently on CPU, ensuring cost-effective and scalable processing.
- Achieved a 32% increase in classification accuracy, leading to more precise and reliable predictions.

Textual Insights Generation using Amazon GenAI

- Developed an automated text extraction solution using AWS Textract and Lambda.
- Implemented Generative AI-based prompt-driven retrieval for enhanced text insights.
- Utilized Python, AWS Bedrock, and LLMs for efficient text processing.

Certifications

- ✓ Microsoft Certified: Azure Administrator Associate
- ✓ IBM Data Science Professional Certificate Specialization