

Senior Software Engineer specializing in LLM-powered NLP systems, with experience designing and deploying AI solutions across legal, financial, and analytics domains. Passionate about building scalable, real-world applications using GenAI and ML.

Work Experience

Senior Software Engineer	Revenue	Sep 2022–Present
Data and AI team	Remote, India	
<ul style="list-style-type: none">Built and deployed a scalable Retrieval Augmented Generation (RAG) system for natural language interaction with 100+ contracts and legal documents to provide actionable insightsLeveraged multi-modal LLM tool-calling for custom NER-based data extraction from multiple domain-specific documents. Integrated a human-in-the-loop learning pipeline using PEFT (LoRA) for efficient and domain-adaptive fine-tuning.Embedded AWS QuickSight Q into the Revenue Platform, enabling real-time, natural language dashboards backed by custom database views helped convert 6+ PoC customers into full-time clients.Implemented and optimized a platform-wide natural language search system utilizing embeddings, vector databases, and LLMs.Built custom data transformation pipelines using Pandas, reducing customer onboarding time by 50% and efficiently handling datasets exceeding 100,000 records with 99.9% accuracy.Implemented extractive and abstractive text summarization techniques for legal documents and contracts, enhancing efficiency in information retrieval and decision-making.		
Software Engineer	Kaar Technology	2020–2022
Data Analytics team	Chennai, India	
<ul style="list-style-type: none">Integrated ML models with SAP Plant Maintenance workflows and deployed models to SAP Cloud Foundry.Built a virtual tour guide chatbot for a major tourist attraction in Saudi Arabia using SAP ConversationalAI to answer FAQs and provide navigation. Built a similar bot for staff to create incidents, quick data lookups, and maintenance.Developed machine learning models for predictive maintenance, achieving 92% accuracy in detecting potential equipment failures.Integrated the client's SAP system with Remote Function Calls (RFC), enabling real-time data synchronization.		

Education

- M.Tech in Artificial Intelligence**, Indian Institute of Technology Madras (IITM) 2025–Ongoing
- B.Tech in Electronics and Communication**, Velammal Engineering College 2017–2021

Technologies and Skills

- GenAI : OpenAI APIs, Transformers, Hugging Face, PEFT (LoRA, QLoRA, Adapters), LangChain, ChromaDB, FAISS
- ML & NLP : NumPy, Pandas, Scikit-learn, spaCy, NLTK, PyTorch, Keras
- Backend : Python, Streamlit, FastAPI, Django, REST APIs, SQLite, PostgreSQL, Celery, Docker
- Cloud : AWS (EC2, S3, Bedrock, Textract, Comprehend, QuickSight Q), Azure (AI Document Intelligence)

Projects

Natural Language to SQL (NL2SQL) Chatbot

<https://github.com/BalaMithran/NL2SQL>

- Built a full-stack **NL2SQL system** using OpenAI GPT-4, FastAPI, PostgreSQL and Streamlit to translate natural language queries into executable SQL over structured databases
- Implemented **entity value detection** and **query intent classification** to dynamically return KPIs, tables, or visualizations (e.g., bar/line charts), simulating a BI dashboard experience.
- Benchmarked SQL generation accuracy using random samples from **Spider** and **WikiSQL** datasets, achieving ~90% **execution accuracy** on single-turn queries with known schema