Koushik M S

Mobile: +91 9514148569 | Email: koushikms21102000@gmail.com | geeksforgeeks

TECHNICAL SKILLS

Python | JavaScript | Java | Fast API | Node.js | Express.js | AI(GenAI) | Machine Learning | Supervised and Unsupervised Learning | Deep Learning | Mongo DB | Milvus | Mongoose | Microservices | Pandas | NumPy | Matplotlib | Seaborn | Scikit-learn | TensorFlow | Keras | NLTK | RAG | LangChain | Redis | Kafka | MySQL | Prompt Engineering | Fine-Tuning | Docker | K8s | RESTful API

AWS | S3 | Azure | Cosmos DB | App Services | Function App | Key vaults | Storage Account | AKS | API Management Service | GIT | GitHub | Splunk | Mongo Compass | Postman | MySQL Workbench | Application Insight | VS Code | Eclipse | Redis Insight | NoSQL Booster

WORK EXPERIENCE

IBM | GenAI Backend Developer

September 2022 – Current

As a **GenAI Backend Developer**, I focused on designing and building intelligent backend systems that leverage **LLMs, vector search, and RAG architectures** to solve complex enterprise problems. My work involved integrating GenAI pipelines with cloud platforms, optimizing data ingestion and retrieval workflows, handling large unstructured datasets, and implementing secure, scalable backend services. I collaborated closely with stakeholders to deliver AI-driven solutions that reduced manual effort, improved operational efficiency, and enabled faster, more accurate decision-making.

- **Database Optimization**: Migrated from a customer-specific database architecture to a multi-tenant model using 4 common databases. This reduced database costs by over **40%**.
- **Performance Enhancement**: Implemented **server-side pagination** for high-volume APIs, improving response time from **2 seconds to 250ms**.
- **Query Cost Reduction**: Optimized database queries by eliminating unnecessary aggregation lookups, reducing **Cosmos DB** request unit (RU) consumption by **10-20%**.
- **Caching Strategy**: Integrated **Redis in-memory caching** for frequently accessed configuration data, improving performance and reducing load on the database.
- **API Architecture Improvements**: Refactored heavy internal functions into standalone APIs and scheduled them with **Azure Logic Apps cron jobs**, improving response time and system reliability.
- **Automation of Customer Onboarding**: Designed and developed new APIs and a UI interface to automate customer and user onboarding processes, cutting manual onboarding time by **30-40 minutes per customer**.
- **Client Interaction & Documentation**: Regularly collaborated with clients to gather, analyse, and document technical and business requirements, ensuring successful and timely feature delivery.

GenAI-Based OCI (Oracle Cloud Infrastructure) Log Analysis and Recommendation System:

- Designed and implemented a **GenAI-based log analysis system** to automatically identify issues and generate remediation recommendations from OCI logs.
- Built an end-to-end **RAG (Retrieval Augmented Generation) pipeline** using LangChain, vector embeddings, and LLMs for intelligent log analysis.
- Automated ingestion of the **last 7 days of OCI logs**, including errors, warnings, audit events, and system failures across multiple services and regions.
- Implemented log preprocessing and overlapping chunking to remove noise and improve semantic understanding during similarity search.
- Generated high-quality vector embeddings using **gpt-text-embedding-3-large** and stored them in **Milvus** for scalable and fast vector search.
- Enabled **semantic similarity search** to retrieve the top relevant log chunks based on user queries or log IDs.

- Integrated **GPT-5.2** to perform root cause analysis and generate clear, step-by-step remediation recommendations.
- Reduced log investigation time and manual debugging effort while improving security by avoiding direct access to OCI services.

GenAI-Based OCI (Oracle Cloud Infrastructure) Log Analysis and Recommendation System:

- Designed and implemented a **GenAI-based solution** to analyze SAP BPMN process diagrams and automatically generate standardized process documentation, eliminating manual effort and SAP expert dependency.
- Built a pipeline to ingest **BPMN process diagrams in XML format**, leveraging the hierarchical structure of XML for accurate preprocessing and analysis.
- Implemented **structured XML preprocessing** to remove unnecessary tags and irrelevant data, improving input quality for downstream GenAI processing.
- Designed and implemented a **structured chunking mechanism** to handle large BPMN files, preserving semantic relationships between tags (lanes, tasks, events, gateways) while overcoming LLM token limitations.
- Optimized LLM usage by analysing user queries, identifying relevant BPMN sections, and sending **only the required XML chunks** to the model, reducing cost and improving response accuracy.
- Developed a **query-based BPMN analysis feature** that allows users to ask process-related questions and receive accurate, context-aware insights from BPMN data.
- Implemented **automated document generation** by processing BPMN sections with specialized prompts and generating standardized documentation using the **Python docx library**.
- Evaluated multiple LLMs and finalized **Granite-3.1-8B** for its strong understanding of structured XML data and support for large context windows (16K–32K tokens).
- Reduced manual BPMN analysis effort, improved scalability, and achieved cost savings through GenAI-driven automation.

EDUCATION

Anna University - Madras Institute of Technology Chennai, Tamil Nadu, India.
Bachelor of Engineering in Electronics and Communication CGPA: 8.03/10 *June 2018 - June 2022*

AWARDS AND CERTIFICATIONS

- Received a **$500 USD** cash award and eCard as recognition for driving client delight and contributing to IBM's Base Account Growth initiative by leveraging application expertise and client relationships to help expand business.
- IBM **Generative AI** Foundations(Developed an automated multilingual toxic comment detection system using **GPT-4o**, handling language detection, translation, toxicity analysis, and conditional MongoDB storage, **reducing manual moderation by 90%**)
- Microsoft Certified: Azure AI Fundamentals.