

# Dev Garg

+1(979)326-3907 | devgargd7@gmail.com | [in/devgargd7](#) | [Q/devgargd7](#)

## EDUCATION

### Texas A&M University

*Master of Science - Computer Science*

Jan 2024 - Present

*College Station, TX*

- GPA: **4.0/4.0**
- Courses: Software Engineering, Large Language Models, Deep Learning

### Indian Institute of Technology, BHU

*Bachelor of Technology - Electronics Engineering*

July 2017 - May 2021

*Varanasi, India*

- GPA: 8.87/10.0
- Courses: Natural Language Processing, Computer Vision, Probability and Statistics

## EXPERIENCE

### Societe Generale GSC

*Software Engineer*

June 2021 - Dec 2023

*Bengaluru, India*

- Played an integral role in every phase of the Agile software development lifecycle for credit risk exposure calculation applications employing **Big Data** technologies.
- Achieved an **50%** reduction in validation time by implementing automatic regression tests. Additionally, designed an autoscaling system slashing compute costs by approximately **€ XX,XXX** per quarter. Led the development of multiple alerting features and provided mentorship to new team members.
- Exposure: *Java, Spring Boot, Apache Spark, Kafka, Elasticsearch, Jenkins, SQL, Azure*

### Societe Generale GSC

*Software Engineer/Data Science Intern*

May 2020 - June 2020

*Remote*

- Developed an **ML-based** incident resolution recommendation system, reducing operational risks and improving response times by **38%**. Worked closely with cross-functional teams to design, develop, and test the MVP, meeting strict deadlines.
- Exposure: *Python, Scikit-learn, NLTK, Pandas, spaCy, Networkx*

## PROJECTS

### Agentic Self-Corrective Retrieval-Augmented Generation

- Designed and implemented an adaptive **RAG** system leveraging *LangChain* and *ChromaDB* vector store based on user-supplied documents and URLs. Leveraged *Google Custom Search API* to complement document-based retrieval reducing hallucinations and increasing the robustness of generated answers.
- Evaluated system performance with *RAGAS* metrics, resulting in a **20%** increase in answer relevance and a **5%** enhancement in faithfulness. Deployed on *AWS EC2* using *Docker Compose* with an intuitive *Streamlit* UI, and a 4-bit quantized *LLaMA3* model operational via an *Ollama* server for seamless and scalable operations.

### News Aggregation and Recommendation System

- Developed a web-based News Aggregator that periodically fetches RSS data from various news outlets, centralizing the latest stories. Implemented a **Hybrid Recommendation System** to suggest articles based on similarity, user behavior, news freshness, and personalized content curation according to user interests and reading history.
- Utilized *BERT* embeddings to capture semantic and contextual nuances and *pandas* for data manipulation. Deployed on *AWS Lambda* and *DynamoDB* for scalable serverless services, with an interface using *FAST API*.

### Stock Market Charting App

- Developed a **full-stack application** for analyzing and visualizing complex financial data through intuitive charts with *FusionCharts* and *Angular*. Implemented **RESTful** CRUD operations, Role-based Authorization, and JWT Authentication to ensure data security, reliability, and integrity using *Spring Boot* and *PostgreSQL*.

## TECHNICAL SKILLS

**Languages:** Python, Java, C++, Javascript, SQL

**Frameworks:** Tensorflow, Pytorch, Pandas, Spark, Spring Boot, React, CUDA

**Tools/Platforms:** Git, Azure Datalake, Docker, Kubernetes, AWS, GCP, ETL

**Interests and Experience:** Machine Learning, Big Data Systems, Web Development

**Certifications:** Google Cloud Professional MLE, Machine learning Spec., MLOps Spec., GenAI with LLMs