

# Deepak Yadav

✉ deepakyadav54official@gmail.com ☎ 8919659441 📍 Bengaluru, Karnataka

🌐 linkedin.com/in/deepak-yadav540 🐙 github.com/D-Yadav

## PROFILE

Software Engineer with 4.5+ years of professional experience in Software Development and AI, specializing in development pipelines. Demonstrated proficiency in GenAI, Prompting, Machine Learning, and Backend development, and spearheaded the development of innovative and scalable AI infrastructure using LLaMA-based models with both cloud and on-premise GPU deployments.

## SKILLS

### Programming

Python (Flask, FastAPI, Scripting), Java, C++, SQL  
Databases: MySQL, PostgreSQL, ORACLE, HUE, MongoDB

### Machine Learning (Supervised | Unsupervised)

Regression, Classification, Ensembles Learning, clustering, Anomaly detection, DBSCAN  
Visualization Tool: Tableau

### NLP

Text Pre-processing, Vector Embedding, Transformers, Attention Models

### Tools (Development)

VS Code, Github Co-pilot, Databricks, Jupyter Notebook, AWS Sagemaker

### Gen AI Tools:

Claude.ai, Perplexity, ChatGPT, Github Co-pilot, Gemini

### Generative AI

LLMs (Models: Llama CPP, Mixtral), Fine Tuning, Prompting(CoT)  
Vector Database: Neo4J | Data Storage: Elasticsearch | Monitoring: Grafana, Prometheus  
Frameworks: Langchain, LlamaIndex, Huggingface, OpenCV,

### Deep Learning (Neural Networks)

CNN, ANN, RNN, LSTM, Transformer model.  
Frameworks: Keras, Tensorflow, Pytorch

### MLOps and Cloud

Docker & Kubernetes, AWS, GCP, Heroku, Udeploy  
Others : ADO, Jira, Confluence, Teamcity, Autosys, Kabana

### Bigdata (Pyspark, Hadoop Ecosystem)

Operating Systems: Windows, UNIX

## PROFESSIONAL EXPERIENCE

### AI Engineer, RiskInsight Consulting Pvt Ltd

**Client: Standard Chartered Bank | Project: MOSAIC**

- Architected scalable LLaMA-based infrastructure for real-time inference and embedding generation across cloud and on-premise deployments.
- Developed end-to-end MLOps pipelines with Bunicorn, JFrog, and secure model serving using NGINX/Keycloak, integrated with Grafana/Prometheus for observability.
- Built high-performance vector databases with Neo4j/Elasticsearch and contributed to SAS-to-Python conversion projects, leveraging prompt engineering on platforms like Claude AI.

07/2024 – present  
Bengaluru, India

### Senior Software Engineer, HCL Technologies Ltd

**Client: Citibank | Project : TTS-TADE [2021 - 2024]**

- Developed innovative solutions leveraging ML for large-scale data analysis.
- Designed a POC on trade data using Machine Learning, saving costs for the client. Applied Generative AI with LLMs at Citibank to enhance accuracy and efficiency in financial decision-making.
- Designed and implemented ETL using PySpark and Medilion architecture.

10/2019 – 06/2024  
Bangalore & Chennai, India

**Client: Citibank | Project: Olympus API [2019 - 2021]**

- Worked with a cross-functional team to gather requirements and developed web services with Java, SQL and RestAPI.
- SQL Query optimization, Shell Scripting, Logs monitoring, DevOps and logs cleanup.

## PROJECT (COMPANY)

---

### MOSAIC IAAS, RISKINSIGHT [Client: Standard Chartered Bank]

07/2024 – present

**Tech:** Generative ai, Prompting, Python, Fastapi, AWS, Postgre SQL, Docker

**Roles & responsibilities:**

- Orchestrated LLaMA model infrastructure deployment for cloud/on-premise GPUs, enabling real-time inference and enterprise embedding generation
- Developed MLOps pipeline using Bunicorn/JFrog with custom validation for secure versioned model deployment
- Implemented Neo4j/Elasticsearch vector database with optimized queries, integrated NGINX/Keycloak auth and Grafana/Prometheus monitoring.
- Spearheaded SAS to Python conversion project incorporating Prompt Engineering, Software testing, Data Validation and GenAI tools.

### TTS TRADE, HCL Technology Ltd [Client: Citibank]

06/2021 – 06/2024

**Tech:** ML Algorithm, Statistics, Decision Tree, LLMs, Prompting, SQL.

**Roles & responsibilities:**

- Orchestrated ML solutions for trade data analysis and LLM applications, driving financial decision improvements at Citibank
- Built optimized ETL pipelines using PySpark on Medilion architecture
- Applied advanced AI analytics to enhance financial operations accuracy and client value

### Olympus API, HCL Technology Ltd [Client: Citibank]

10/2019 – 05/2021

**Tech:** Java, SQL, Udeploy, Teamcity, Autosys, Postman

**Roles & responsibilities:**

- Developed and tested REST APIs with integrated database log monitoring for debugging.
- Optimized system performance through efficient queries, shell script automation, and DevOps best practices.
- Maintained system health through proactive log monitoring and regular maintenance.

## PROJECTS (PERSONAL)

---

### Text-to-Speech-Conversion

- Developed Text-to-Speech web application using gTTS API, enabling MP3 output with customizable audio preprocessing and tokenization, end-to-end deployment.

### Deep Face Authenticator, [Personal-Project] | End-to-End | SimpliLearn

- Computer vision-based authentication system by using state-of-art algorithms like FaceNet and MTCNN.
- CNN, pattern recognition, face detection, face-embedding, neural network integration, MLOps & Cloud.

## TRAINING | BOOTCAMP

---

### Data Scientist and Business Intelligence, SimpliLearn

04/2023 – 10/2023

- Acquired in-depth knowledge of ML, DL, NLP, CV and Tableau.
- Successfully designed and implemented end-to-end projects.

{Remote}, Bengaluru

## EDUCATION

---

### B.Tech [ECE] - SVCET Chittoor A.P [CGPA : 7.7],

Jawaharlal Nehru Technological University

09/2015 – 04/2019

Ananthapur, India

## LANGUAGES

---

### Language and Personal Skills:

**Language:** English[IELTS-7.0], Hindi, Nepali

**Personal Skills:** Adaptability & Flexibility Good  
Communication & Team Leader.