

# Mithlesh Upadhyay

Data Scientist | Gen AI/ML Engineer

✉ mithleshupadhyay515@gmail.com ☎ +91-9311-3040-75 📍 Chitrakoot, MP

🌐 /in/mithlesh-upadhyay/ 🗨 /mithleshupadhyay

## Profile

- **Data Scientist** with **2.7 years of experience**.
- Hold **M.Tech** in **Artificial Intelligence** from Delhi Govt Engineering College.
- Developed projects utilizing **LLM, RAG, Embeddings, Python, Pandas**, Matplotlib, and NLP techniques.
- Successfully fine-tuned open source models like LLaMA 2 and PoC for **RAG for AI chatbot** building.
- **Managed teams** for data science and AI content and course preparation.
- Currently looking out for **Data Scientist | Generative AI/ML Engineer** roles.

## Skills

### Machine Learning & Deep Learning

TensorFlow, PyTorch, LLMs, RLHF, GPT, QLoRA,  
Model Deployment (SageMaker, Docker)

### Data Engineering & Pipelines

Python, Pandas, ETL, Data Cleaning

### Cloud Technology

AWS (SageMaker, EC2, S3), Docker, Kubernetes,  
GitHub

### Vector Databases

Chroma DB, FAISS, Pinecone

### Natural Language Processing (NLP)

NLTK, Spacy, BERT, BART, Prompt Engineering

### Data Visualization & EDA

Matplotlib, Seaborn, Plotly

### Frameworks

LangChain, LlamaIndex, FastAPI, Flask, Streamlit


### Vector Databases

SQL, MySQL, MongoDB

## Professional Experience

### SDE-I (AI Engineer), Propellyr.com

11/2023 – 08/2024

- **Fine-tuned** open source models like **LLaMA 2 using QLoRA** on proprietary datasets and conducted inference.
- Implemented RAG for AI chatbot building and conducted **proof of concept (POC)** for various AI initiatives.
- Extracted data from the Propellyr website and applied **embedding with Faiss DB** to build an interactive chatbot.
- Prepared a dataset to fine-tune Llama 2 for a **flight ticket booking chatbot**.
- Developed **scalable APIs**.
- Huggingface profile - <https://huggingface.co/mithlesh> 

Delhi, India

### Software Developer, ValueMatrix.ai

04/2023 – 09/2023

- Solved bugs and created new functionalities using JavaScript, Reactjs, MongoDB, HTML, and CSS.
- Participated in meetings and collaborated within teams.

Bengaluru, India

- ValueMatrix is an **AI-powered hiring platform** where I resolved bugs using ReactJS.

#### Python Developer, *FORESIET*

09/2022 – 12/2022

Bengaluru, India

- Developed functions for **APK crawlers**.
- Foresiet is a leading cybersecurity company where I worked as a **Python developer**.

#### Senior Technical Content Engineer, *GeeksforGeeks*

09/2017 – 12/2021

Noida, India

- Managed teams for data science and AI content and course preparation.
- Published articles, provided feedback, and managed content.
- **Published more than 7000 articles**, 3000 improvements, 9000 provided feedback.
- **Managed team of 15+ members** to deliver products before the deadlines.
- Mentored and guided new members.

### Education

---

#### M.Tech, Artificial Intelligence, *Delhi Technological University (Formerly DCE)*

07/2021 – 06/2023

Delhi, India

- Artificial Intelligence, Machine Learning, NLP, Artificial Neural Network, Deep Learning, Computer Vision

#### Bachelor of Engineering (CSE), *Vindhya Institute of Technology & Science*

09/2008 – 05/2012

Satna, MP, India

- Algorithms and Data Structure, DBMS, OS, CN

### Projects

---

#### Flight Agent Chatbot using RAG - LangChain

- Developed a chatbot for flight-related queries using the **RAG (Retrieval-Augmented Generation)** model.
- Used **OpenAI** Embeddings and **Chroma** as vectorstore, and **gpt-3.5-turbo** for chatbot.
- The chatbot, integrated with **LangChain**, provides users with **real-time assistance** and information regarding flights and related services.

#### Sales Manufacturing Prediction

- Analyzed and **compared the performance** of various **regression models** for predicting sales in manufacturing.
- Explored different regression techniques to find the most accurate model for **sales forecasting**.
- Applied **feature engineering techniques** to improve model performance.

#### mithlesh/LLaMA-2-ayurveda

- Uploaded to the Hugging Face Model Hub.
- This project involves an **ayurveda-specific** adaptation of the **LLaMA 2 language model**.
- Users can access and utilize the model for tasks related to ayurvedic **text processing and generation**.

#### Fine\_Tune\_LLaMA\_2\_using\_QLoRA

- **Fine-tuned the LLaMA 2** model on proprietary datasets.
- Conducted **inference** using the fine-tuned model.
- Fine-tuned Llama 2 using different datasets on **Amazon SageMaker**.
- Developed a **dataset for flight ticket booking** information in JSON format.

## Publications

---

### **A Comprehensive Analysis of Sentiment Analysis: Approaches, Applications, and Classifier Comparisons, *IEEE*** [↗](#)

16/12/2023

- Reviewed popular methods for **sentiment analysis** and how they are used in different applications.
- Analyzed how **text data** from social media and messaging apps can provide valuable feedback.
- Created a guide for **choosing the right** sentiment analysis models based on specific needs.
- IEEE Xplore DOI: 10.1109/ICAC3N60023.2023.10541746 [↗](#)

## Courses

---

### **GATE CS Score Card, *GATE Authority***

- Qualified GATE CS with good AIR multiple times, (AIR - **2226** in year 2022 out of 1 lac aspirants)

### **Generative AI with Large Language Models, *Coursera***

- Foundational knowledge, practical skills, and a functional understanding of how generative AI works

### **Learn LangChain: Build #22 LLM Apps using OpenAI & Llama 2, *Udemy***

- Build Real World LLM powered applications with LangChain, OpenAI, Llama2, Hugging Face.