

Pankaj

Data Scientist

Remote, India | pankajshakya627@gmail.com | pankajshakya627.github.io | linkedin.com/in/pankajshakya627 | github.com/pankajshakya627

PROFILE SUMMARY

Data Scientist with 4+ years of experience designing, deploying, and optimizing AI/ML solutions—specializing in generative and agentic AI frameworks. Expert in building scalable, cloud-based pipelines and leveraging Python, PyTorch, and AWS to deliver high-impact results. Adept at translating complex data insights into actionable strategies within Agile environments.

EXPERIENCE

ITC Infotech - Gurugram, Haryana, India

February 2024 – Present

Data Scientist (Associate IT Consultant)

- Developed a RAG system using OpenAI and LangChain, enhancing customer Q&A efficiency.
- Conducted data analysis to inform product strategy, communicating insights to stakeholders.
- Developed agentic AI workflows with LangGraph and Autogen to automate complex, multi-step data processes, cutting manual intervention by 60%.
- Collaborated with cross-functional teams to enhance AI/ML services.

Jasper Colin - Noida, U.P., India

September 2023 – February 2024

Executive

- Led fine-tuning initiatives for **large language models** (LLAMA 2) using PEFT techniques and integrated generative AI models enhanced NLP tasks.
- Deployed models on AWS EC2 Cloud instances, optimizing performance and ensuring high availability for end-users.
- Financial Data Analysis:** Performed comprehensive analyses on **financial news** and company data, providing actionable insights into market trends and portfolio performance.

LTIMindtree - Bengaluru, Karnataka, India

March 2021 - August 2023

Python Automation Developer

- Collaborated on the development of over 15 innovative features using Python and REST API.
- Developed Python documentation sites, reducing search time by 10%, demonstrating expertise in Git, GitHub, and various cloud tools.
- Automated over 300 features for Cisco storage servers, leveraging Python and QA expertise to drive testing excellence and product innovation.

EDUCATION

Rajkiya Engineering College – Ambedkar Nagar, U.P.

July 2016 - September 2020

Bachelor of Technology in Information Technology

SKILLS

- Programming Languages:** Python, R, DBMS, C++
- Data Analysis:** Pandas, NumPy, Scikit-learn, SciPy, statsmodel,
- Machine Learning:** TensorFlow, Keras, **PyTorch**, NLTK, NLP, Gensim, RNN, LSTM, **CNN**, OpenCV
- Generative AI (GenAI):** **Large Language Models (LLM)**, Langchain, Streamlit, Transformers (**Hugging Face**) LangGraph, Autogen, Pinecone, PEFT/LoRA
- Data Visualization:** Matplotlib, Seaborn, Plotly
- Databases:** MySQL, PostgreSQL, MongoDB
- Other Tools:** Git, AWS SageMaker, AWS Bedrock, **GCP**, **Docker**, Databricks, Kubernetes, Linux
- MLOps & Monitoring:** MLflow, Weights & Biases, Model drift detectionion

PROJECTS

Fine-Tuning a Large Language Model

- Applied Parameter-Efficient Fine-Tuning (PEFT) techniques to fine-tune the Llama-3.2-3B model.
- Resulting in improved task-specific performance with significantly reduced computational requirements, thereby enhancing sustainability and cost-effectiveness.

Agent Query Helper Chatbot

- Built a RAG-based chatbot for PDF queries using FAISS and Llama-3.1-8B, evaluated using the DeepEval framework
- Tools Used: Python, Pytorch, Huggingface, Streamlit, Langchain, Pinecone, Vector Embeddings







Invoice Recognition Streamlit App

- Built end-to-end invoice processing pipeline integrating Azure AI Vision, OpenAI, and Streamlit.
- Achieved 96% extraction accuracy and automated data ingestion for downstream analytics.

[Find Other Project Here](#) 

OTHERS

Certifications:

- | | | |
|---|---------------------------|---|
| • Generative AI with Large Language Models | - DeepLearning.AI and AWS | 




 |
| • Python for Data Science , AI & Development | - IBM | |
| • Applied Plotting, Charting & Data Representation in Python Data Science Methodology | - IBM | |
| • The Top 5 Machine Learning Libraries in Python | - Udemy | |
| • Complete Python programming from Basics to Advance level | - Udemy | |
| • National Workshop on "Image Processing and Computer Vision & Application" | | |

Awards:

- Strategic Contributor – LTIMindtree (2022)
- Rising Star Award – LTIMindtree (2021)