

Vishnu Pratap Singh Rathore

✉ rathorevishnu26@outlook.com 📞 +91 77371 47617 🌐 LinkedIn 🐙 GitHub

Summary

AI/ML Engineer with 3+ years of experience building Python-based automation tools, LLM-powered applications, and data pipelines. Proven track record designing RAG systems, ML classification models, and SQL-driven data agents. Hands-on with Pandas, NumPy, FastAPI, PostgreSQL, Docker, and AWS. Experienced in delivering data-quality and intelligence tooling with high ownership and minimal supervision. Fully remote-ready.

Technical Skills

Languages & Data: Python (Pandas, NumPy, Scikit-learn), SQL (PostgreSQL, SQLite), Bash

AI / ML / LLMs: LangChain, LangGraph, HuggingFace Transformers, TensorFlow, Keras, OpenCV, Diffusion Models, ReAct Agents, Prompt Engineering

LLM Models: LLaMA 3, Gemma, GPT-4o, Code LLaMA, Groq API

Vector Stores & RAG: FAISS, Pinecone, AstraDB, Chroma

Backend & DevOps: FastAPI, Flask, Django, PostgreSQL, Redis, Docker, Microservices, Linux, Git

Cloud & Infra (exposure): AWS, ROS2

Interfaces: Streamlit, Gradio

Experience

Backend Developer

Sep 2025 – Present

ebizneeds, Jodhpur

- Building and maintaining production backend services using FastAPI, PostgreSQL, and Redis within a microservices architecture.
- Working with Docker for containerization and service orchestration across distributed backend systems.
- Collaborating with cross-functional teams to design scalable APIs and data pipelines for client-facing applications.

Robotics & AI Engineer

Jan 2025 – Sep 2025

ROBO AI Hub, Jodhpur

- Designed and deployed multi-modal AI pipelines integrating Vision, Language (LLaMA 3, Gemma, GPT-4o), and Audio models into production systems.
- Built MCP Client-Server infrastructure enabling structured tool-use for LLM agents, managing data flow between sensors and AI decision layers.
- Implemented ReAct Agent frameworks using LangChain and LangGraph for autonomous task planning, tool-calling, and self-correction loops over structured and unstructured data.
- Deployed open-source LLMs on edge hardware for low-latency offline inference; managed model configuration, versioning, and monitoring.
- Built Generative AI pipelines across modalities: text, image (Diffusion), audio, and video understanding.

AI/ML Developer (QA + AI)

Aug 2023 – Feb 2025

Square Bits, Jodhpur

- Built a production RAG system for Website Q&A using LangChain, FAISS, Groq, and LLaMA 3.2 – enabling semantic search and structured data retrieval across large unstructured datasets.
- Developed an SQL Database Natural Language Interface using LangChain's SQLAgent, translating plain-English queries into optimized SQL against PostgreSQL – directly analogous to LLM-driven data quality configuration over vendor datasets.
- Applied Pandas and NumPy extensively for data preprocessing, cleaning, and transformation across multimodal datasets (text, video, audio).
- Designed and executed automated test plans for AI-integrated software, collaborating with developers to identify and resolve data and model-quality issues.

Data Science & ML Intern

Mar 2021 – Sep 2021

GIP Technologies Pvt. Ltd., Jaipur

- Built NLP text classification and sentiment analysis models in Python using Scikit-learn, Pandas, and NumPy – including end-to-end feature engineering, training, and evaluation.
- Contributed to data visualization and analysis tooling for customer feedback datasets, providing actionable insights to business stakeholders.

Projects

SQL Database Natural Language Interface *LangChain SQLAgent · PostgreSQL ·*

FastAPI

- LLM-powered tool that translates natural-language queries into optimized SQL with automatic schema introspection – directly applicable to configuring data quality checks over vendor datasets.

RAG-Powered Website Q&A Chatbot *LangChain · FAISS · Groq · LLaMA 3.2 ·*

Streamlit

- Full RAG pipeline that ingests, indexes, and semantically retrieves content from large unstructured datasets with source-cited, confidence-aware responses.
- [Source Code \(GitHub\)](#)

LangGraph Research Helper Agent *LangGraph · OpenAI · Tavily · Python*

- Multi-step autonomous agent using LangGraph state machine – plans queries, validates intermediate results, and outputs structured reports; demonstrates ML-driven workflow automation.
- [Source Code \(GitHub\)](#)

Customer Churn Prediction Model *TensorFlow · Keras · Pandas · Streamlit*

- End-to-end ML pipeline: Pandas-based data ingestion and cleaning, feature engineering, deep learning model training, and Streamlit deployment with live inference.
- [Source Code \(GitHub\)](#)

Education

Bachelor of Computer Applications (BCA)

2021

University of Rajasthan, Jaipur

Led University Technical Club · Represented university at Global Entrepreneurship

Summit 2020

Certifications

- **Generative AI with LangChain & HuggingFace** ([Source Code \(GitHub\)](#)) – Udemy, April 2025 ([Certificate](#))
- **Computer Vision with OpenCV and Deep Learning** ([Source Code \(GitHub\)](#)) – Udemy, January 2025 ([Certificate](#))
- **AI and Machine Learning with Python** ([Source Code \(GitHub\)](#)) – Udemy, April 2024 ([Certificate](#))