

TARUN KUMAR RATHORE

+91 7999002506 | tarun422rathore@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

Generative AI Engineer Performance-driven Developer specializing in RAG architectures and multimodal systems utilizing LangChain, OpenAI, Gemini and Llama. Proficient in fine-tuning and model optimization via Groq, OpenRouter, and DeepSeek to enhance model reliability, streamline workflows, and deliver scalable AI solutions.

SKILLS

Languages: Python, HTML, CSS

AI/ML: Generative AI, LangChain, Hugging Face, NLP, Deep Learning, OpenCV, Statistics

Data Analytics & Visualization: Power BI, NumPy, Pandas, EDA, Matplotlib, Seaborn

Database Management: MySQL, PostgreSQL

Cloud & Deployment: Azure, Streamlit, Gradio

MLOps & Version Control: GitHub, MLOps, LLMOps, CI/CD

Project Experience

Fine-Tuned Gemma-2B LLM for Quote Generation | Python, PEFT, LoRA, Kaggle GPUs

- Fine-tuned Gemma-2B using LoRA and FP16 precision via Hugging Face PEFT to achieve high-quality, domain-specific text generation with minimal trainable parameters.
- Optimized training on dual Tesla T4 GPUs, successfully managing VRAM constraints, library dependencies, and secure API management via Kaggle Secrets.

Agentic AI Pipeline Design | LangGraph, Groq, Llama-3.3, LangSmith, Streamlit

- Engineered a stateful LangGraph pipeline using Llama-3.3 via Groq for modular, ultra-low latency AI decision-making.
- Integrated LangSmith observability and a Streamlit interface for transparent, real-time conversation state management.

Dynamic RAG Q&A System | LangChain, Groq, ChromaDB, Gradio, HuggingFace

- Built a RAG-based document Q&A system using LangChain, ChromaDB, and Sentence-Transformer embeddings.
- Enabled fast inference with Llama-3.3-70B (Groq) and a Gradio interface with source tracking.

Intelligent OCR & Reasoning System | DeepSeek-OCR, OpenRouter, Python

- Built an automated OCR pipeline using DeepSeek-OCR to convert images into structured Markdown.
- Developed a reasoning-aware engine via OpenRouter to preserve stateful logic with DeepSeek-V3.

CERTIFICATIONS

- Google Data Analytics Certificate - Coursera
- Oracle Data Science Professional - Oracle
- Deloitte Data Analytics Job Simulation - Deloitte
- Machine Learning with Python - IBM Cognitive Class
- AI Passport - EY & Microsoft

ACHIEVEMENTS

- Received the NSS "B" Certificate for national service and leadership.
- State-level achiever in Vedic Mathematics, demonstrating strong mental math skills.
- Earned 20+ Google Cloud Skill Badges in Data and AI.

EDUCATION

Chhattisgarh Swami Vivekanand Technical University (CSVTU) | 2021-2025

Bachelor of Technology, Computer Science - 79.72%

Saraswati Higher Secondary School, Korba , Chhattisgarh

12th - 95.4% 2021 | 10th - 89% 2019