TARUN KUMAR RATHORE

+91 7999002506 | ✉ tarun422rathore@gmail.com | in **LinkedIn** | ◯ **GitHub** | 🌐 **Portfolio**

SUMMARY

Generative AI Engineer with hands-on experience in building RAG systems, agentic workflows, and intelligent document pipelines. Experienced in LLM fine-tuning, optimization, and low-latency inference using LangChain, Groq, and DeepSeek to deliver scalable, reliable AI solutions.

SKILLS

- **Languages:** Python, HTML, CSS
- **AI/ML:** Generative AI, LangChain, LangGraph, Hugging Face, NLP, Deep Learning, OpenCV
- **Data Analytics & Visualization:** Power BI, Statistics, NumPy, Pandas, EDA, Matplotlib, Seaborn
- **Database Management:** MySQL, PostgreSQL, ChromaDB (Vector DB)
- **Cloud & Deployment:** Azure, Streamlit, Gradio
- **MLOps & Version Control:** GitHub, CI/CD, LLMOps

Project Experience

Cost-Efficient LLM Customization for Enterprise Content Generation | LoRA, Gemma-2B
- Built a resource-efficient LLM fine-tuning pipeline for domain-specific content generation with LoRA.
- Optimized training under GPU memory constraints using FP16 precision and gradient accumulation.
- Created an interactive Gradio interface to validate and demonstrate model outputs.

Scalable Agentic AI Workflow for Low-Latency Decision-Making | LangGraph, Groq
- Designed a stateful agentic AI workflow with LangGraph to manage multi-step decision-making in LLM applications.
- Integrated Llama-3.3 via Groq to enable ultra-low latency inference for real-time interactions.
- Added LangSmith-based observability and a Streamlit interface to support transparent workflow tracing and state visualization.

Document-Grounded AI Q&A System Using RAG | LangChain, ChromaDB
- Built a RAG-based document Q&A system to ground LLM answers and mitigate hallucinations.
- Implemented semantic retrieval using Sentence-Transformer embeddings with ChromaDB.
- Enabled fast inference via Groq and added source-aware Q&A using a Gradio interface.

Intelligent Document Understanding System Using OCR & LLM Reasoning | DeepSeek
- Built an OCR-based pipeline to convert scanned documents into structured Markdown.
- Integrated LLM-based reasoning to preserve context and logical flow across extracted content.
- Enabled automated document understanding by combining OCR extraction with reasoning in a unified pipeline.

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

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

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