

MOHIT GUPTA

Mathura, Uttar Pradesh, India – 281003

Mobile: +91 7300554546 | mohit.gupta.cr1@gmail.com | [LinkedIn](#)

Summary

Results-driven Data Scientist with 2.5 years of experience in NLP, LLM fine-tuning, and AI agent frameworks. Skilled in Python, Docker, and microservices, focusing on AI infrastructure and automation. Adept at aligning data solutions with business goals to improve efficiency and decision-making.

Technical Skills

- Programming:** Python (PyTorch, TensorFlow)
- ML & AI:** LLMs, Transformers, NLP, Multi-Agent Systems, RAG (Retrieval-Augmented Generation), LangChain, LangGraph
- Infrastructure:** Docker, Flask, FastAPI, Kubernetes, Elasticsearch, Nginx
- Cloud Platforms:** Google Cloud Platform (GCP), Microsoft Azure
- Databases:** MongoDB, SQL
- Tools:** Pandas, NumPy, Scikit-learn, Selenium, BeautifulSoup

Experience

Data Scientist

IntelleWings Pvt. Ltd. | Chandigarh, India | Mar 2022 – Sep 2024

- Met data requirements in alignment with business needs provided by clients, ensuring accuracy and efficiency in deliverables.
- Developed NLP pipelines processing 500K+ documents daily, reducing processing time by 60%.
- Built automated data extraction using custom LLMs (Llama, Mistral), achieving 92% accuracy.
- Engineered scraping infrastructure to extract and process over 500k+ data points daily with high reliability and uptime.

Projects

Multi-Agent News Intelligence

- Built a news intelligence dashboard showing trending India/Global news by genre, with search-based AI summaries and source citations.
- Implemented a nightly two-step pipeline on GCP (scraping → processing: cleaning, embeddings, indexing, summarization).
- Designed modular agents (SearchAgent, ReportingAgent, orchestrator) for semantic search, trend analysis, and reporting.
- Project Repository: [GitHub Link](#)

LLM-based Information Extraction

- Fine-tuned a quantized LLM (using BitsAndBytes & QLoRA) for efficient extraction of structured data from financial PDFs.
- Incorporated a Retrieval-Augmented Generation (RAG) setup to enhance context relevance during data extraction.
- Engineered custom prompt strategies to enhance information retrieval accuracy.
- Achieved 95% accuracy, drastically reducing manual data processing time and effort.

Financial Risk API Development

- Developed and maintained 100+ APIs using Flask for real-time financial risk data processing, ensuring 24/7 uptime.
- Designed a hybrid data storage solution using MongoDB for lightweight document storage and Elasticsearch for large-scale data retrieval.
- Implemented and managed web crawlers to continuously update financial data from raw sources, ensuring data freshness and accuracy.
- Optimized API performance and scalability to support high-volume financial risk assessments.
- Integrated real-time monitoring systems to proactively detect and resolve API performance issues.

Education

Bachelor of Technology (B.Tech) in Computer Science

GLA University | 2019–2023