Sarthak Agrawal

AI Engineer, RippleLinks

Highly skilled AI Engineer with 1+ years of experience developing/deploying production-ready intelligent systems using modern LLM/VLM agent frameworks. Adept at building scalable FastAPI backends, containerized ML solutions, and training fine-tuned models. Passionate about applying research to real-world use cases through clean, efficient code.

sarthak.agrawal1311@gmail.com

+91 8871928567

Bengaluru, India

linkedin.com/in/sarthakagrawal11

github.com/IamME1311

WORK EXPERIENCE

AI Engineer
RippleLinks

06/2024 - Present *Bengaluru, Karnataka*

Achievements/Tasks

- Designed and deployed a custom AI Agent using open-source models (Flux, Stable Diffusion) to synthesize images from unstructured client briefs (PDFs, emails).
- Fine-tuned LoRA models to support ethnically/nationally underrepresented image generation scenarios.
- Built an async FastAPI backend with intelligent GPU memory management, enabling model execution on low-VRAM (1GB) systems.
- Containerized the solution with Docker and deployed across AWS and GCP.

Data Science Intern
Innomatics Research Labs

01/2024 - 04/2024

Achievements/Tasks

- Developed a subtitle-aware video search engine using NLP to improve relevance and accessibility
- Created a RAG pipeline with LangChain and Gemini 1.5 Pro, enabling real-time contextual response generation
- Built a sentiment analysis model using Flipkart review data (8,500+ samples) to derive product insights

EDUCATION

MCA in Machine Learning and AI
Amity University, Greater Noida

01/2022 - 01/2024 *Greater Noida, Uttar Pradesh*

CGPA

- 8.5 CGPA
- ML, Deep Learning, Cloud Architecture
- Transformers Architecture
- Reinforcement Learning

SKILLS

Python | FastAPI | Git | Docker | LlamaIndex | LangGraph | SmolAgents | AWS, GCP | CI/CD | Semantic Search | AI Agents | RAG | Embeddings | Collaboration | Research-to-production | Problem Solving

PROJECTS

ClauseGuard - Your Expert Legal Assistant (04/2025 - 07/2025)

- Built a contract intelligence agent using **LlamaIndex** and **LlamaParse** to parse and structure complex legal documents for intelligent querying and summarization.
- Integrated with **Llama Cloud** for scalable document storage and embedding indexing, enabling real-time semantic search across large corpora.
- Leveraged **Gemini 2.5 Flash** for fast and accurate clause-level analysis and retrieval, ensuring contextually relevant legal insights.
- Developed natural language query support for clause interpretation, risk identification, and legal Q&A workflows.
- Delivered a low-latency, scalable solution suitable for internal legal teams and enterprise contract workflows.

Semantic Subtitle Search Engine (02/2024 - 04/2024)

- Engineered a semantic search engine over **84,000 subtitle files** spanning movies, TV series, and anime to enable deep contextual search and retrieval.
- Utilized the **all-MiniLM-L6-v2** sentence transformer model to embed subtitle segments into vector space for efficient semantic comparison.
- Implemented **cosine similarity-based ranking** to accurately match user queries with relevant dialogue across diverse media content.
- Achieved significant improvements in relevance and recall over traditional keyword-based subtitle search.
- Enabled use cases such as quote search, theme-based discovery, and character-specific dialogue tracking.

INTERESTS

Hiking | Football | Camping