

AI Engineer (LLM & RAG Specialist)

Company: E2M

Experience Required: 2-3 years

Location: Ahmedabad, Gujarat, India

Job Type: Full-time

Department: Artificial Intelligence & Machine Learning

Job Summary

E2M is looking for a highly skilled AI Engineer with 2-3 years of experience specializing in Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), prompt engineering, and vector databases. The ideal candidate will design and implement intelligent conversational and knowledge-driven systems using state-of-the-art AI technologies.

Key Responsibilities

- Design, develop, and deploy AI applications leveraging LLMs and RAG-based architectures.
- Build and optimize prompt engineering pipelines to improve LLM response quality and contextual accuracy.
- Integrate and manage vector databases (e.g., Pinecone, ChromaDB, Qdrant, PGVector) for semantic search and context retrieval.
- Develop APIs and backend services that interface with LLMs for chatbot and automation use cases.
- Implement scalable pipelines for embedding generation, data indexing, and knowledge retrieval.
- Continuously fine-tune LLMs and optimize inference performance for production-level workloads.
- Collaborate with product and engineering teams to deliver end-to-end AI-driven solutions.

Required Skills

- Strong experience working with Large Language Models (e.g., OpenAI GPT, Gemini, Claude, Llama).
- Hands-on experience in RAG system design and implementation.
- Expertise in prompt engineering and prompt optimization techniques.
- Experience with vector databases such as ChromaDB, Pinecone, Qdrant, or PGVector.
- Proficiency in Python and frameworks like LangChain, LlamaIndex, or Haystack.
- Understanding of API development using FastAPI or Flask.
- Good knowledge of embeddings, semantic search, and document retrieval systems.
- Familiarity with cloud deployment (AWS, GCP, or Azure) and containerization (Docker, Kubernetes).

Education

Bachelor's or Master's degree in Computer Science, Artificial Intelligence, Data Science, or a related field.

Preferred Qualifications

- Experience in building AI chatbots, knowledge assistants, or automation systems using LLMs.
- Understanding of MLOps for LLM lifecycle management.

- Experience with fine-tuning or instruction-tuning of open-source LLMs.
- Knowledge of prompt evaluation and benchmarking frameworks.

Benefits

- Flexible working hours and hybrid work option
- Health insurance and performance-based incentives
- Continuous learning and certification support in GenAI and LLM domains
- Opportunity to work on cutting-edge AI projects and global client solutions

Application Instructions

Interested candidates can send their resume, project portfolio, or GitHub profile to careers@e2msolutions.com with the subject line 'AI Engineer (LLM & RAG) – Ahmedabad'.