Vaibhav Gupta

Alwar, Rajasthan | vaibhavgupta1148@gmail.com | +91-7424809555

linkedin.com/in/vaibhav-gupta-b929ba237

PROFESSIONAL SUMMARY

AI/ML Engineer with 2+ years of experience architecting Agentic AI systems and Generative AI pipelines. Expert in multi-agent frameworks (LangGraph), LLMs (Qwen, Llama 3), and RAG systems. Proven track record of delivering scalable AI solutions with measurable business impact.

EXPERIENCE

Machine Learning Engineer   *Delhi | Mar 2024 - Present*
Glarus Technology

- Architected autonomous multi-agent framework using Llama 3.2B & Groq API for NL-to-code generation, achieving 95% accuracy in HTML5/Tailwind output.
- Engineered parallel agent pipelines for UI generation, reducing content latency by 40% vs sequential baselines.
- Developed production virtual try-on system integrating SCHP, MODNet & PIFuHD for photorealistic garment rendering.

AI Engineer   *Delhi | Jun 2023 - Mar 2024*
Softage.ai

- Designed high-performance RAG pipelines with LangChain & Pinecone, improving response accuracy by 25% for chatbots.
- Built Large Action Models (LAMs) for RL-based autonomous decision-making in predictive modeling scenarios.
- Optimized ML inference pipelines across NLP/CV domains, reducing latency by 35% for recommendation systems.

AI/ML Research Intern   *Remote | Jan 2023 - May 2023*
Gilbert Research Center

- Transitioned 3 academic prototypes (Sentiment Analysis, Diabetes Prediction) to production demos for healthcare sectors.

KEY PROJECTS

Agentic Resume Optimization System
Tech: LangGraph, FastAPI, Groq (Qwen 32B), Tavily, Streamlit

- Built multi-agent system with 4 specialist agents for automated resume optimization, improving ATS match rates by 45%.
- Implemented real-time company research & JD gap analysis using Groq and Tavily API.

Enterprise Conversational AI
Tech: LangChain, GPT-4, Pinecone, FastAPI

- Developed autonomous chatbot handling multi-domain queries with sub-200ms latency using optimized RAG pipelines.

MCQ Generator (GenAI)
Tech: PyTorch, Hugging Face, LLMs

- Fine-tuned LLM for auto-generating exam-grade MCQs, deployed for ed-tech platforms with 92% quality rating.

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, Java, R, HTML5/CSS3
AI Frameworks: LangGraph, LangChain, PyTorch, TensorFlow, Hugging Face, FastAPI
GenAI: Multi-Agent Systems, RAG, Prompt Engineering, GPT-4, Llama 3, Qwen
Cloud/Tools: AWS, GCP, Docker, Git, Pinecone, Streamlit, MLflow

EDUCATION & CERTIFICATIONS

B.Tech in Computer Science   /
2019 - 2023

- Certifications: AI Engineer (Pro5.ai) | ML & Data Science (GeeksforGeeks) | Advanced Python (RKCL, Grade A)