

Aman Kumar

+91 6206666035 | amankumarramakr@gmail.com | github.com/amankumarsbg21 | linkedin.com/aman-kumar

Professional Summary

AI Engineer with 2 years of hands-on experience architecting and deploying enterprise-grade, agentic AI systems. Specialized in building multi-agent workflows, memory-aware applications, and secure RAG pipelines using Python, LangChain, LangGraph, and Flask. Proven ability to integrate large language models (Gemini, GPT-4) with strict enterprise governance, NeMo Guardrails, and robust backend logging to deliver scalable, production-ready AI solutions.

Experience

Developer - TCS AI WisdomNext

Kolkata, India

Tata Consultancy Services

July 2024 - Present

- Architected secure, multi-agent AI workflows using LangGraph and ReAct prompting, automating complex enterprise processes and reducing manual effort by 35%.
- Designed and deployed LLM-based microservices via Flask REST APIs, establishing seamless integration points for internal state-managed Angular dashboards.
- Engineered Advanced RAG pipelines featuring hybrid search and cross-encoder reranking, drastically improving document retrieval accuracy and eliminating context hallucinations.
- Implemented enterprise-grade security layers using NeMo Guardrails to enforce strict input/output validation, prevent prompt injection, and automatically mask PII before processing.
- Built custom logging and state-management pipelines to trace complex agent trajectories, manage thread IDs for persistent memory, and monitor token consumption in production environments.

Summer Intern- DRDO

Balasore, Odisha, India

Defence Research and Development Organisation (DRDO) ITR Chandipur

June 2023 - July 2023

- Integrated Alpha-Beta filter algorithm for radar signal tracking, improving prediction stability by 15% in high-noise environments.
- Processed, scaled, and analyzed 1000+ radar datasets utilizing custom data transformation pipelines to enhance tracking accuracy.
- Collaborated with research engineers to validate algorithmic performance and reliability for defense-grade systems.

Projects

Autonomous Research & Task Orchestration Platform(Multi-Agent System)

- Engineered a multi-agent AI system (LangGraph) with 5 specialized agents, reducing task completion time by ~30%.
- Implemented long-term memory (Pinecone/ChromaDB + RAG), improving multi-turn accuracy by ~20%.
- Added security guardrails - prompt injection detection, sandboxed execution (Docker) — blocking 90%+ adversarial test cases.
- Designed a human-in-the-loop approval & role-based access across 8+ tool integrations, zero unauthorized actions in testing.
- Deployed custom logging/monitoring tracking latency & token usage across 200+ test runs.

Brain Tumor Detection Using CNN and SVM

- Built a custom CNN model trained on 3,000+ labeled brain MRI images using multiple convolutional layers for feature extraction and SVM for precise tumor categorization.
- Achieved 99.5% training accuracy and approximately 98% testing accuracy, demonstrating high reliability in detecting brain tumors from MRI scans.

Skills

AI & LLM Frameworks: AutoGen, CrewAi, MCP, Advanced RAG, Multi-Agent Systems, LangChain, LangGraph, Advanced RAG, Agentic Workflows, ReAct, Prompt Engineering, Agent-to-Agent Communication, AI Evaluation .

Backend & Frontend: Python, Flask, FastAPI, Microservices, Angular (State Management, Component Architecture), TypeScript.

Models: Gemini, GPT-4o, Llama 3, Claude 3.5.

Enterprise AI & Governance: NeMo Guardrails, PII Masking, Structured Outputs, AI Observability & Evaluation.

Data & Databases: PostgreSQL, Pinecone, Chroma, Data Transformation (ColumnTransformers, Scaling).

Cloud & Deployment: GCP, Git, CI/CD, Observability & Monitoring, Docker.

Education

C.V Raman Global University

Bhubaneswar, India

Bachelor of Technology in Electronics and telecommunication Engineering (CGPA: 8.41)

October 2021 - May 2024