

ANANT SHARMA

AI/GenAI Engineer | Agentic AI, LLMs & Python | 1+ YOE

Bangalore, KA | 8851779323 | asindia23@gmail.com

linkedin.com/in/codelinechef | github.com/codelinechef | codelinechef.github.io/Portfolio-Website

SUMMARY

AI/GenAI engineer shipping production Python systems (FastAPI, Pydantic) across LLMs, Generative AI, and agentic AI. Builds agentic workflows and workflow-automation solutions with LangChain tool calling, external tool integrations, multi-step reasoning, MCP-aligned APIs that optimize business processes and reduce operational cost. Engineers REST API integrations tested with Pytest, LangSmith, and Postman, deployed on Docker + AWS with CI/CD (GitHub Actions).

WORK EXPERIENCE

AI Software Developer — CanopusGBS

Mar 2025 – Present | Bangalore, KA

- Cut enterprise document-review turnaround time by 30% across 500+ submission cycles by leading AI/ML solution development of a scalable production RAG pipeline in Python API integrations (OpenAI API, LangChain, ChromaDB), document chunking, embeddings, vector search, and reranking integrated into existing REST review workflows.
- Reduced sector-wise report generation from several hours to under 5 minutes and cut operational costs by 65% by building a business-process-automation workflow with feature engineering and AI-driven anomaly detection over structured datasets.
- Increased operational throughput by 40% and improved data reliability by developing a Django + SQL backend service with workflow automation, scheduled data-quality automation, cross-team alerting, and audit logging driving continuous process improvement.
- Led AI/ML solution development for three GenAI initiatives by conducting research and feasibility analysis for Generative AI use cases and presenting demos that aligned solutions with business goals, collaborating with leadership and cross-functional stakeholders across Agile/Scrum sprints to influence technical strategies.

PROJECTS

DeepWeave — Agentic AI Middleware & Orchestration Layer

Python, FastAPI, LangGraph, LangChain, MCP-aligned tool-calling, AsyncIO, multi-agent orchestration

- Cut token consumption by 35% and LLM API spend by 28% by building context-engineering middleware in FastAPI for a LangGraph/LangChain multi-agent orchestration framework orchestrating agentic workflows subagent coordination, tool calling and external tool integrations, and asynchronous context handoff across a ReAct-style multi-step reasoning pipeline while optimizing prompts and context for latency and token efficiency — an automation framework that reduced operational cost.
- Reviewed and enforced output quality and best practices by implementing QA evaluation (guardrail) agents that validate every AI output against business intent, blocking defective responses before production.

RemembrAI — LLM Memory & Retrieval Agent

Python, LangGraph, vector embeddings, conversational state & memory management, pytest, LangSmith

- Achieved 100% recall across retrieval benchmarks and eliminated ~94% of duplicate record entries by building a semantic vector-search memory agent with conflict-resolution logic.
- Gated every release through a CI-style validation layer by authoring automated pytest suites and LangSmith evaluation frameworks for regression gating.

MLOps Observability Pipeline

Python, Prometheus, Grafana, Galileo, OpenAI API, AWS

- Increased monitoring automation by 70% and shortened incident detection from manual review cycles to real-time alerts by deploying a live MLOps observability stack with anomaly detection and model-performance tracking.

SKILLS

AI & Machine Learning: Machine Learning, AI/ML solution development, supervised learning, feature engineering, anomaly detection, predictive analytics, EDA, data processing; PyTorch, TensorFlow, scikit-learn

Generative AI & Agentic: Generative AI, LLMs, prompt engineering & optimization (accuracy, latency, token efficiency), RAG (chunking, embeddings, vector search, reranking), embedding pipelines, AI agents, agentic workflows, multi-step reasoning pipelines, multi-agent & workflow orchestration, LangChain, LangGraph, LlamaIndex, CrewAI, Agents SDK, Model Context Protocol (MCP), tool calling & external tool integrations, context management, conversational state & memory handling, output validation / guardrails, NLP / semantic search, Promptfoo, OpenTelemetry

Backend & Languages: Python (AsyncIO, async/await, type hints, OOP), FastAPI, Pydantic, Django, REST & MCP-aligned APIs for AI service integration, Postman, SQL, TypeScript, JavaScript, C/C++, Postman,

LLM Providers: OpenAI, Anthropic Claude, Google Gemini, DeepSeek, Llama

Databases (Relational & NoSQL) & Vector Stores: PostgreSQL, MySQL, ChromaDB, Pinecone, Azure Cosmos DB; feature engineering, EDA, predictive analytics

Cloud, DevOps & Evals: AWS, Docker, CI/CD, GitHub Actions, Git, pytest, LangSmith, Galileo, Prometheus, Grafana

Delivery: workflow automation, business process automation (BPA), data-quality automation, continuous process improvement, automation framework design, stakeholder management, system design, Agile/Scrum, technical documentation

EDUCATION

B.Tech, Computer Science & Engineering (AI/ML Specialisation)

2021 – 2025

SRM Institute of Science and Technology, Chennai, Tamil Nadu

Certifications: Supervised Machine Learning — DeepLearning.AI & Stanford, 2026 • Google AI Essentials — Google, 2026